# Algal Blooms as Marine Ecosystem Risk: Forecasting Spread and Biogeochemical Stress

**A Dissertation**
**Submitted to the Graduate School of**
**Information Science and Technology**
**Hokkaido University**

**WANG HAOJIONG**

**September 5, 2023**

# Abstract

Algae, accounting for less than one percent of Earth's total photosynthetic biomass, are remarkable carbon drawdown contributors, fixing nearly half of the world's organic carbon, especially during algal blooms. However, escalating concerns surround algal blooms, their persistence, and distribution, serving as indicators of both global climate shifts and local anthropogenic pressures. These phenomena intertwine with coastal marine ecosystems worldwide, where tide disruptions and human-induced disturbances increasingly degrade water quality, fostering frequent algal blooms. The growing duration of these blooms poses a significant threat, impacting vital ecological processes and services, such as carbon cycling and sequestration. Yet, unraveling the complex interplay between ecological factors and environmental stressors, along with deciphering algal bloom patterns, remains a formidable challenge due to limited data and a lack of universally applicable analytical approaches.

An innovative predictive model merges transfer entropy network inference with a forecasting graph neural network to anticipate both blooming and non-blooming epidemic scenarios, along with their underlying environmental factors that elucidate bloom sources, causes and systemic risk. This model exhibits strong predictive capabilities, extracting crucial ecosystem features even in the absence of spatial dependencies. A novel 2D entropic ecosystem mandala is introduced, wherein the ecological impact, manifested through the distribution's Cyanobacteria-driven chlorophyll-a (CHL-a) randomness, correlates proportionally with systemic environmental stress, governed by erratic oceanic, climatic, and coastal nutrient factors. Originally, a spatial risk was defined based on CHL-a magnitude, persistence and shifts. Through a case study in Florida Bay (FL Bay), we unveil how algal bloom shifts endure in shallow regions with elevated dinoflagellate-to-diatom ratios, underscoring Cyanobacteria's pivotal role in phytoplankton dynamics and the influence of terrestrial discharge on marine microbiome equilibrium. This un-

folding scenario presents formidable challenges, notably the heightened potential for green-blue algal blooms (associated with river dominance) to trigger harmful red tides, with cascading socio-ecological impacts spanning carbon cycling disruptions and entrenched eutrophication in coastal ecosystems. A universal threshold on the top $20\%$ Pareto extremes of CHL-a, distinctly defines bloom and non-bloom phases, independent of endemic or epidemic categorization, driven by distinct eco-environmental interactions, where the paramount biogeochemical stress follows a scale-free structure with CHL-a acting as the central hub.

Predicting algal blooms in the short and long term is crucial for assessing the well-being of ecosystems, encompassing coastal-marine environments, species, and human populations. Furthermore, it offers insights into the effects on environmental processes like carbon sequestration. However, the escalating disruption in biogeochemical balance compromises our capacity to forecast algal blooms, barring during outbreaks when intervention becomes belated. This deficiency hampers the investigation and management of the underlying eco-environmental factors triggering undesirable algal bloom occurrences and propagation. And our ideas improved this difficulty to some extent, both in terms of causal inference and model prediction.

# Acknowledgments

I would like to thank my esteemed supervisor Prof. Matteo Convertino for his invaluable supervision and support during my PhD program. Additionally, I am also deeply grateful to Prof. Ohgane for his assistance and guidance during my final year, and Prof. Tsutsui for his advice and insights. I also thank Dr. Jie Li and Dr. Elroy Galbraith for valuable discussions on my research, and all my labmates, for their help and encouragement. My appreciation also goes out to my family and friends for their encouragement and support throughout my studies.

# Publications

## Articles

1. H. Wang, E. Galbraith, and M. Convertino, "Algal Bloom Ties: Spreading Network Inference and Extreme Eco-Environmental Feedback", *Entropy*, vol. 25, no. 4, 636, pp. 1–18, Apr. 2023.

2. H. Wang and M. Convertino, "Algal Bloom Ties: Systemic Biogeochemical Stress and Chlorophyll-a Shift Forecasting", *Ecological Indicators*, vol. 154, 110760, pp. 1–18, Oct. 2023.

## Conference Papers

1. H. Wang T. Tsutsui, and M. Convertino, "Classification of Rich-Classes but Scarce-Samples Images via Multi-modeling: the Humpback Whale Epitome",Proc. IEEE Global Conference on Life Sciences and Technologies (LifeTech), pp. 290–292, Mar. 2022.

2. H. Wang, E. Galbraith, and M. Convertino, "Risky Blooms: Space-Time Chlorophyll-a Analysis and Forecasting",Proc. International Symposium on Communications and Infor- mation Technologies (ISCIT) 2023, Oct. 2023.

## Book Chapter

1. M. Convertino and H. Wang, "Envirome Disorganization and Ecological Riskscapes: The Algal Bloom Epitome", In Risk Assessment for Environmental Health, CRC Press: Boca Raton, FL, USA, 2022; pp. 327–346.

# Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 A sensescape purview for ecosystem decision making under risk and possibilities.

Healthy ecosystems look, sound, and smell better than compromised ones. This is rather clear when we hear the symphony of a healthy vibrant forest vs. a cacophonic one. Environmental disorganization alters the balance of senses in species. Yet, we can use sensorial perceptions from ecosystem data – particularly audio-image ones to gather ecological disorganization – to quantify and judge the health of ecosystems in relation to systemic environmental pressure or noise. However, many questions arise from this evidence, such as: How senses of biodiversity are related to habitats (the environmental chambers) and other factors such as disturbances that alter the symphony (organization and function) of ecosystems? For how long these sense-base traits are manifested and when do they lead to irreversible collapse including cascading environmental extremes? What is the degree of irregularity, magnitude and persistence of environmental pressure leading to irreversible collapse? Can we use species somatic traits and species proportions to characterize habitat fitness more precisely than using macro-environmental features only? What are the most sensitive univariate and bivariate pattern indicators of ecosystem shifts? What is the degree of variability of ecosystem health for equivalent habitats worldwide or nearby habitats and their socio-ecological teleconnections?

Senses (images, sounds, tastes, smells, textures, and collective behavior such as velocity and trajectory of species) spreads as waves in the environment and cause symptoms of dysbiosis such as bleaching in coral reefs or plant stress in a

water-limited environment; stress that limits evolutionary processes leading to optimal function and organized ecological aggregation. Accurate sensors (including their location and sampling frequency) are yet needed to capture scents of senses and synthesize ecosystems' organization informing about current state, closeness to shifts, environmental determinants (diffused or localized, acute or chronic), and countering solutions.

An extra element is however needed beyond senses. Perception is when sensory information is selected, organized, and interpreted where models (as perceptions) are used to extract salient features that (i) capture dynamics of patterns informative of processes (yet evidencing basic mechanisms) and (ii) define early and long-term decisions, at higher levels of decision making, to counter ecological risk and adaptive response, or enhance its function based on the collective organization of species and habitats (i.e. biodiversity). Perceptions (i.e. artificial neurons in pattern-oriented machine-learning or any neuromorphic models) reflect species computation (with different level of cognition) that is the basis of collective behavior. Then, the question is how we can leverage this "ecological computing" through sensed information for improving ecosystems through precise interventions, plans and policy based on eco-history and feasible scenarios? What is the spatial and functional configuration of species decision trees (due to habitat/non-trophic and food-webs/tyrophic) and their perceived risks conditional to environmental stress? Are there external and beneficial environmental stimuli that we can manage (e.g. light and sound) to positively alter collective intelligence? How can we re-engineer collective sensing of species that has been compromised, including our co-mutual sensing with all species?

Species perceive patterns without thinking of processes and ecosystem pathologists/engineers should do that equivalently by minimizing undesired ecological dysbiosis (that in a computing sense can be about erroneous risk assessment). The focus should be on patterns because patterns reveal ecosystem health related to salient processed senses at multiple scale and the underpinning causal networks. Perceptions that must be optimally communicated – through visualization and other media – to strengthen collective behavior enhancing ecosystems. In this sense data are not just the dry input of models but the fundamental senses and vehicle of artistic visualization to deliver optimal information. This is an ambition rather than mere objectives, because it come with a cohesive and broader long-term vision of improving earth ecosystems through science and applications at local scale and high resolution.

## 1.2 Signals of ecosystem change

The Earth's ecosystems consist of intricate networks of interdependent species and processes, providing crucial services and supporting life as we know it. The most hardest challenge is to preserve the stability of these ecosystems, which does not imply a state of unchanging equilibrium, but rather a state of change that can be sustained with a certain degree of order [3]. These changes can be quantified into certain specific 'signals' that provide valuable indications about the state of the ecosystem. Signals of ecosystem change comprise a diverse array of indicators, such as changes in species abundance, alterations in trophic interactions, modifications in habitat structure, and shifts in ecosystem functions. However, some signals are early 'warning signs' given by ecosystems (i.e., 'non-equilibrium indicators' [4] that manifest as statistical properties of ecological data preceding a regime shift), which are valuable for studying ecosystem stability. The quantization of these signals include increased temporal autocorrelation [5], variance [6], skewness [7] [8], and rising spatial heterogeneity [9] [10]. Monitoring and interpreting these signals can provide 'advanced warning' of an approaching regime shift, enabling proactive management interventions. There are also signals that represent catastrophic shifts in the ecosystem, also known as regime shifts or ecological tipping points, represent sudden and profound changes in the structure, function, and dynamics of ecological systems. While there is a complex interplay between gradual environmental change and abrupt tipping points that can push ecosystems beyond their critical thresholds [11]. This catastrophic shift has the potential to trigger irreversible ecological damage. It can compromise biodiversity, cascading effects of trophic interactions and ecosystem functioning, and the potential for regime shifts to trigger further disturbances or create alternative steady states. Therefore, interpreting and monitoring these ecological signals allows us to gain a deeper understanding of the status of the ecosystem and take proactive measures to maintain its stability.

### 1.2.1 Senses of marine ecosystem change

Oceans are the basis of life in an ecological and evolutionary perspective [12] [3]. Marine ecosystems are vast, dynamic, and diverse systems that cover more than 70% of our planet's surface [13]. They encompass a complex web of interactions between living organisms and their surrounding environment, playing a vital role in sustaining life on Earth. The importance of marine ecosystems extends far beyond

the boundaries of the oceans. It is essential for maintaining biodiversity, producing oxygen, regulating the climate, and sequestering carbon dioxide [14]. However, 'signals' of marine ecosystem change are extremely important in the face of the delicate balance of them. For example, changes in water temperature, ocean currents, and availability of food can cause 'shifts in the distribution of marine species' [15] [16]. Some species may move to higher latitudes or deeper waters as the ocean temperature increases [17]. Shifts in ocean currents can affect nutrient availability, water temperature, and the distribution of marine organisms. El Niño and La Niña events, for instance, can alter oceanic conditions in the Pacific Ocean, impacting marine ecosystems across vast areas [18] [19]. Declines in fish populations may indicate overfishing or changes in ocean conditions, which is one indication of changes in abundance [20]. In addition, coral bleaching occurs when corals expel the symbiotic algae living in their tissues due to stress, such as increased water temperature or pollution [21]. This phenomenon is also a warning signal of environmental changes and can lead to coral reef degradation and loss of biodiversity. Increased levels of carbon dioxide in the atmosphere can lead to ocean acidification. This process involves the absorption of $CO_2$ by seawater, resulting in lower pH levels [22]. Acidification can have adverse effects on shell-forming organisms like corals, mollusks, and some plankton species. Alterations in predator-prey relationships or changes in the trophic structure of the ecosystem can indicate shifts in the marine environment [23] [24]. For example, a decline in top predators like sharks can disrupt the balance of the food chain and have cascading effects on the entire ecosystem. Reductions in species diversity and the loss of key species within an ecosystem are strong indicators of ecosystem change. A decrease in biodiversity can result from habitat destruction, pollution, overfishing, and other human-induced factors. It is crucial to recognize these signals and the interconnectedness of them, as multiple factors often contribute to changes in marine ecosystems. Monitoring and understanding these indicators can help scientists and policymakers take appropriate conservation measures to protect and restore marine environments.

### 1.2.2 Algal blooms: epitome of marine ecosystem change

Algae are primary producers that undergo photosynthesis, utilizing carbon dioxide ($CO_2$) from the atmosphere to synthesize organic carbon compounds. During algal blooms, certain species proliferate rapidly, leading to increased carbon fixation [25]. This process temporarily removes CO2 from the atmosphere and con-

tributes to the organic carbon pool. However, rapid and excessive growth of certain types of microscopic algae in aquatic ecosystems is an important manifestation of algal blooms [26]. This is a 'signal' that the marine ecosystem is under stress. Algal blooms always happen in both freshwater and marine environments, such as lakes, rivers, estuaries, and oceans. They are often caused by a combination of environmental factors, including warm water temperatures, high nutrient levels (such as nitrogen and phosphorus), calm water conditions, and increased sunlight. These conditions create an ideal environment for the rapid proliferation of certain algal species. Algal blooms can also have significant ecological impacts. When algal blooms occur, the excessive growth of algae often leads to their decay and death. As the algae decompose, the organic carbon they contain is released back into the environment [27]. This process can result in the production of carbon dioxide and other greenhouse gases, contributing to the atmospheric CO2 levels and potentially impacting global climate change. Some algal blooms, particularly those composed of certain species of dinoflagellates, can deplete oxygen levels in water bodies through a process called eutrophication [28]. Excessive algal growth consumes oxygen during decomposition, leading to hypoxic or anoxic conditions [29], which depletion of oxygen can create 'dead zones [30]' where marine life cannot survive. In such environments, the breakdown of organic matter is often incomplete, resulting in the production of methane, a potent greenhouse gas [31]. Methane can further contribute to climate change. Algal blooms can disrupt the normal carbon fluxes between different components of the ecosystem [32]. Additionally, the toxins produced by algae can harm or kill fish, shellfish, marine mammals, and other aquatic organisms [33], disrupting the balance of ecosystems and causing economic losses in fisheries and tourism industries.

## 1.3 Ecosystem dynamic change perception

The need for ecosystems can be quantified by assessing the value and importance of ecosystems for the planet's functioning. All ecological indicators are measurable variables that provide insights into the condition and functioning of an ecosystem. Examples include water quality parameters (e.g., pH, dissolved oxygen), nutrient levels, pollutant concentrations, and temperature. Monitoring these indicators can help detect changes in ecosystem dynamics. These are also the most sensitive variables for ecosystem transformation. In addition, biodiversity conservation are also

essential aspects of quantifying the need for ecosystems. Biodiversity is a fundamental aspect of ecosystems. Surveys can be conducted to quantify changes in species richness, abundance, and composition over time. These surveys may involve field observations, specimen collection, camera trapping, acoustic monitoring, DNA analysis and so on. Whereas the recognition of biodiversity has a crucial important relationship with habitat. Habitats with a higher number of species are generally perceived as more biodiverse. This perception is influenced by the visual presence and diversity of organisms within a habitat. Habitats such as rainforests or coral reefs, known for their high species richness, are often seen as highly biodiverse. Biodiversity perception is also influenced by the complexity and structural diversity of a habitat. Habitats with a variety of vegetation types, microhabitats, and niches are often perceived as more biodiverse. For example, a forest with multiple layers of vegetation, including the canopy, understory, and forest floor, may be seen as more biodiverse than a simple grassland. The presence of diverse ecological interactions within a habitat can influence biodiversity perception. Habitats that support intricate food webs, symbiotic relationships, and other ecological interactions are often seen as more biodiverse. For example, a coral reef with its complex network of species interactions and dependencies is recognized for its biodiversity. However, it's important to note that biodiversity perception may not always align with scientific assessments of biodiversity. Different factors, such as cultural backgrounds, personal experiences, and knowledge levels, can influence how individuals perceive and value biodiversity within a habitat.

There are many ways to quantify changes in ecosystem dynamics, however the combination of computer technology has revolutionized the way we study and understand ecosystems. Ecological modeling can be used to simulate ecosystem dynamics and predict how they may change under different scenarios. Models can incorporate various ecological processes, such as nutrient cycling, species interactions, and climate dynamics, to project future changes and assess the impacts of different factors on ecosystem dynamics. While ecological models can accurately capture the qualitative behavior of real ecosystems, they may not always accurately capture quantitative details [34]. The complexity of ecosystems makes it difficult to accurately model their behavior, and computational limitations may further affect our ability to understand and manage these systems. Apart from that, there are also remote sensing techniques to monitor changes in ecosystem characteristics. It can provide data on land cover, vegetation density, changes in landscape patterns, and track deforestation, urbanization, and habitat fragmentation, among other ecosys-

tem changes. Typical practices are the use of remote sensing data and machine learning technique for the detection and classification of algal blooms [35] [36] and the quantification of phytoplankton biomass [37]. The combination of remote sensing and computer technology allows a large amount of ecological data to be effectively analyzed and visualized.

### 1.3.1 Ecosystem information and status assessment

A complex network is a mathematical representation of a system with interconnected elements or nodes, capturing the patterns of interactions and dependencies between them. Complex networks provide a framework to study the structure, dynamics, and properties of complex systems in various domains. They reveal nontrivial network structures, connectivity patterns, and emergent properties, allowing for the analysis of information flow, resilience, synchronization, and other processes.Complex networks serve as a powerful tool for modeling, prediction, and understanding complex systems across disciplines, including social networks, biological systems, technological networks, and ecological networks. Therefore, complex networks have an significant contribution in describing ecosystem structure, dynamics, and function.

Mutualistic interactions (such as pollination or seed dispersal networks) and competitive interactions can be represented as networks. Complex networks provide a framework to study species interactions beyond predator-prey relationships. Analyzing these networks helps understand the dynamics of species interactions, the emergence of cooperation or competition, and the consequences of species loss or introduction. Meanwhile, complex networks can be used to model metapopulations, which are populations of species occupying interconnected habitat patches. Network analysis helps evaluate the connectivity and spatial structure of metapopulations, predict the spread of species or diseases, and assess the effectiveness of conservation strategies in maintaining population viability. So, networks can represent ecological connectivity patterns, such as landscape connectivity for species movement or dispersal. Network-based metrics and algorithms help identify corridors, stepping stones, or other critical connectivity features that support gene flow, colonization, and the maintenance of biodiversity across fragmented landscapes. Classical complex networks such as food webs, which depict the trophic interactions between species in an ecosystem, can be analyzed as complex networks. Network analysis helps identify key species (such as top predators or keystone species)

and understand the flow of energy and nutrients through the food web. It allows the quantification of various network metrics (e.g., connectivity, centrality, robustness) to assess the stability and resilience of ecological communities (such as the paper by J. Li et al. [38]). Complex network analysis provides tools to assess the resilience and robustness of ecosystems in the face of perturbations or species loss or extreme environmental stress. By modeling the network structure and simulating cascading effects, researchers can understand how the removal or disturbance of key species or links impacts the stability and functioning of ecosystems.

### 1.3.2 Multi-factor ecosystem risk prediction

GNNs combine the power of deep learning techniques with the ability to model and process structured data represented as graphs. They in ecology demonstrate their potential to unravel complex ecological systems, uncover hidden patterns and relationships, and make accurate predictions. GNNs can be utilized to model and predict the distribution of species across landscapes. By incorporating spatial information, environmental variables, and species interactions as a graph structure, GNNs can learn complex relationships and patterns to evaluate the effects of species loss or introduction, assess the stability and resilience of ecological communities, and generate more accurate predictions of species occurrence and abundance. GNNs have been applied to analyze the spread and dynamics of infectious diseases within ecological systems. By integrating contact networks, environmental factors, and disease attributes, GNNs can simulate and predict disease transmission, evaluate intervention strategies, and identify key nodes or areas for targeted control measures.

GNNs offer a powerful approach for multifactor ecosystem risk prediction by leveraging the strengths of graph representation, feature learning, and dynamic modeling. GNNs can integrate diverse data sources, including remote sensing data, ecological surveys, socio-economic data, and climate models, to provide a comprehensive understanding of ecosystem risks [39]. By assimilating and analyzing these multiple data types within the GNNs framework, a more accurate and holistic prediction of multifactor ecosystem risks can be achieved. Specifically, GNNs represent the ecosystem as a graph, where nodes represent different components (e.g., species, habitats, environmental variables) and edges represent the interactions or relationships between them. This graph representation enables the incorporation of multiple factors and their interactions into the risk prediction process. Features can be learned and extracted from the graph structure, including environmental vari-

ables, ecological attributes, and human-related factors. By propagating information through the graph and aggregating features from neighboring nodes, GNNs can capture the spatial, temporal, and relational aspects of the ecosystem, enhancing the prediction accuracy. GNNs also can model and predict ecosystem risks by considering multiple factors 'simultaneously'. By combining the learned features and incorporating various risk indicators. It can provide a comprehensive assessment of the multifactorial risks to ecosystem health and resilience. Furthermore, GNNs can capture the temporal dynamics of ecosystems by incorporating time-dependent factors and historical data. This enables the prediction of how risks may evolve and change over time, allowing for early detection and proactive management of ecosystem risks. The predictions generated by GNN-based risk models can support decision-making and inform conservation strategies. By identifying the key factors contributing to ecosystem risks and quantifying their relative importance, GNNs can guide the prioritization of interventions, adaptive management approaches, and the allocation of resources for risk mitigation.

# Chapter 2

# Algal Bloom Risk: Spatio-Temporal Inference and Prediction

## 2.1 Introduction

Blue carbon ecosystems play a vital role in the sequestration and storage of carbon [40] [41]. However, human activities have profoundly impacted coastal ecosystems, including mangroves, seagrass beds, and salt marshes, resulting in substantial depletion of blue carbon stocks throughout the 20th and early 21st centuries. Among the most severe challenges faced by these ecosystems are harmful algal blooms (HABs), which induce oxygen depletion, fish fatalities, and vegetation loss. Moreover, HABs discharge significant quantities of carbon dioxide and other greenhouse gases, contributing to the augmentation of climate change. Effectively managing and mitigating the repercussions of HABs on blue carbon ecosystems holds critical importance for both their preservation and climate change mitigation. An illustrative instance of a locale frequently experiencing intense HABs outbreaks is central Florida Bay. The predominantly enclosed nature of this region, characterized by constrained water exchange, results in high concentrations of cyanobacteria that can pose risks to aquatic life and human well-being [42]. Altered levels of inorganic nutrients, including those stemming from hurricanes and the vast reservoir of warm water in the Gulf of Mexico, collectively contribute to the onset of bloom outbreaks [43]. Analysis of ecosystem dynamics often extracts stress indicators from the observed network topology of ecological phenomena over time [44]. Effectively managing and predicting HABs necessitates the establishment of a compre-

hensive factor model grounded in topological networks, capable of accounting for robust causal relationships among diverse factors while accommodating interactions among multiple variables concurrently to facilitate accurate bloom predictions. This study introduces a novel approach that simultaneously addresses dynamic ecological processes and the interplay of multiple causal factors, thereby enhancing the precision of ecological dynamic observations and forecasting.

## 2.2 Methods and Materials

### 2.2.1 Data sets and data processing

The Southeast Environmental Research Center (SERC) at Florida International University established an extensive water quality monitoring network comprising 28 strategically positioned stations within Florida Bay. Each of these stations, akin to nodes within a network framework, gathers monthly data encompassing parameters such as Chlorophyll-a (CHL-a), total organic carbon, inorganic and organic nitrogen and phosphorus (TN&TP), turbidity (TURB), pH, salinity (SAL), water temperature (TEMP), and dissolved oxygen [45, 46]. Leveraging six sets of comprehensive water quality monitoring data from August 1992 to September 2008, collected across the 28 monitoring stations within the bay (depicted in Fig. 2.1 A), we conducted our analysis [45]. These data sets encompass monthly readings of total nitrogen (TN), total phosphorus (TP), salinity (SAL), chlorophyll a (CHL-a), water temperature (TEMP), and turbidity (TURB). CHL-a, a widely employed metric for algal biomass [47], serves as an indicator of ecosystem health. Elevated CHL-a values often signify heightened nutrient levels, fostering the proliferation of cyanobacteria and elevating the risk of Harmful Algal Blooms (HABs) [48]. Employing the Pareto Principle, we characterized an algal bloom state when CHL-a values resided in the top 20% range, whereas non-bloom states corresponded to the remaining 80% of CHL-a values.

### 2.2.2 Measuring variable influence

We measured the information content of a time series variable $X = x_1, x_2, \ldots, x_t$ of length $t$ using Shannon entropy,

$$H(X) = -\Sigma p(x) log_2 p(x), \tag{2.1}$$

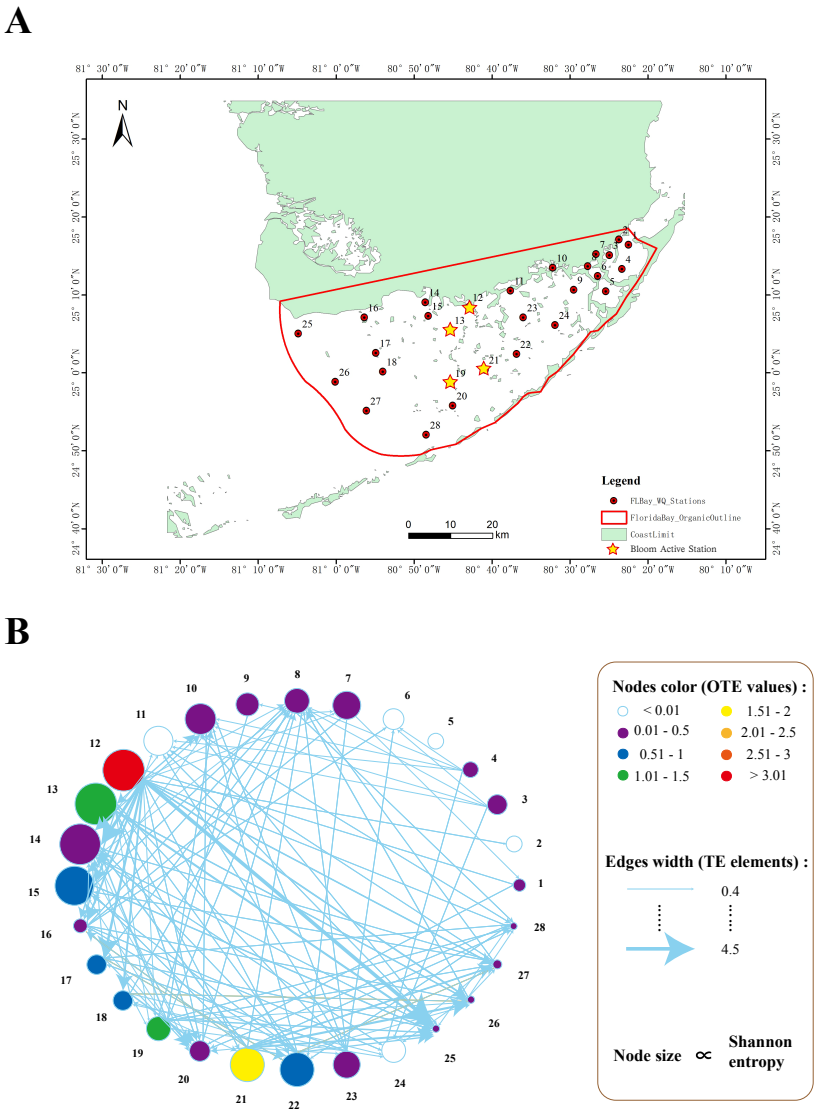


Figure 2.1: **Data monitoring stations map and HABs space inference network.** The numbers in (B) correspond to the station numbers in (A). (B) is based on bloom (CHL-a) data from August 1992 to September 2008. In part (B), the node size depends on the Shannon entropy of the CHL-a for the station, the node colour is proportional to the outgoing transfer entropy (OTE), and the edge size is proportional to the value of the corresponding transfer entropy.

where $p(x)$ is the probability of a value of $X$ being observed. The greater the information content, the higher the $H$, and the more unstable the variable. For multivariate, we considered the amount of information of $X_i$ ($i \in n$, $n$ is the number of variables) under the condition that the variable $X_j$ ($j \in n$) is known,

$$H(X_i|X_j) = -\Sigma p(x_{i,t}, x_{j,t}) log_2 p(x_{i,t}|x_{j,t}), \tag{2.2}$$

The transfer entropy ($TE$) from one time series variable $X_j$ to another $X_i$,

$$TE_{X_j \to X_i} = \sum_{X_j, X_i} p(x_{i,t+1}, x_{i,t}, x_{j,t}) log \left( \frac{p(x_{i,t+1}|x_{i,t}, x_{j_t})}{p(x_{i,t+1}|x_{i,t})} \right), \tag{2.3}$$

where $x_i$ and $x_j$ are the value of $X_i$ and $X_j$ at times $t$ or $t+1$. $TE$ quantifies the extent to which one variable influences another: higher values indicate greater influence.

In the case of ecosystem dynamics, there is a two-way effect between any two variables [49]. So, $TE$ can be calculated for both directions, i.e., $X_i \to X_j$ and $X_i \leftarrow X_j$. But the following way facilitates the identification of stronger impact variables; we call it strong causality.

$$TE_{diff(X_i,X_j)} = \max_{20\%}(TE_{X_i \to X_j} - TE_{X_i \leftarrow X_j}). \tag{2.4}$$

When $TE_{diff(X_i,X_j)} > 0$, $X_i$ is said to cause $X_j$; when $TE_{diff(X_i,X_j)} < 0$, it will be set to 0; and when $TE_{diff(X_i,X_j)} = 0$, there is no connection between the two variables. To retain only the most efficient relationship in a network, the Pareto Principle was utilized to preserve the top 20% values of $TE_{X_i \to X_j} - TE_{X_i \leftarrow X_j}$ solely.

We quantified the influence of a variable $X_i$ as its total outgoing transfer entropy ($OTE$) by using

$$OTE_{X_i} = \sum_{X_j} TE_{diff(X_i,X_j)}. \tag{2.5}$$

A large value of $OTE_{X_i}$ indicates that $X_i$ has a large average impact on its network, proportional to its activity level.

### 2.2.3 Spatial and ecological information network

To monitor the spatial pattern (trends in spread and influence) of algal blooms (indicated by large values of CHL-a), we constructed an information network among

the stations [41]. In this network (the structure follows the Fig. 2.2 B), nodes represent the stations, and links among them are the $TE_{diff}(X_i, X_j)$ (Eq. 2.4), where $X_i$and $X_j$ are the time series of CHL-a at the different stations. Using $TE_{diff}(X_i, X_j)$ means this network was directional, displaying the strong causality between the two interacting stations; no link between nodes indicates negligible interaction strengths.

The spatial model of blooms is an inferred model based on information from CHL-a data only. However, understanding the interactions between water quality factors is also vital to bloom management. In order to explore the interaction between CHL-a and other ecologically relevant variables, a separate information network was constructed for water quality factors in each station (the structure follows Fig. 2.2 B). In this network, nodes are the set of environmental variables (i.e., CHL-a, TN, TP, SAL, TEMP, TURB). Similar to the spatial network idea, the links among them are $TE_{diff}(X_i, X_j)$ (Eq. 2.4). Herein, $X_i$ & $X_j$ are the time series of measurements for the different water quality variables.

### 2.2.4 Network-based inferred bloom prediction

Eq. 2.4 serves as a connection among several variables, enabling the prediction of a specific variable while taking into account multiple adjacency variables. The structural framework follows Fig. 2.2 A.

#### Node feature matrix

In general, the extraction of time series features is done by convolutional operations. A convolution kernel is a feature extractor. Inspired by Transfer Entropy Graph Neural Network (TEGNN) [50], we used Convolution Neural Networks (CNN) filters with kernel sizes of $k = 3, 5, 7$ to extract temporal scale features since kernels of different sizes can extract time series features with different significance for a variable. For each node 3-time series features ($f_{X_i} = [{f^{(1)}}_{X_i}, {f^{(2)}}_{X_i}, {f^{(3)}}_{X_i}], i \in n$) can be obtained. So, the feature matrix is $F \in \mathbb{R}^{(n \times d)}$ for all nodes, where $n$ represents the number of nodes, and $d$ is the number of features ($d = 3$).

#### Node embedding

The determination of node adjacency in the graph structure is based on the transfer entropy difference matrix (Eq. 2.4). This matrix can be utilized in combination

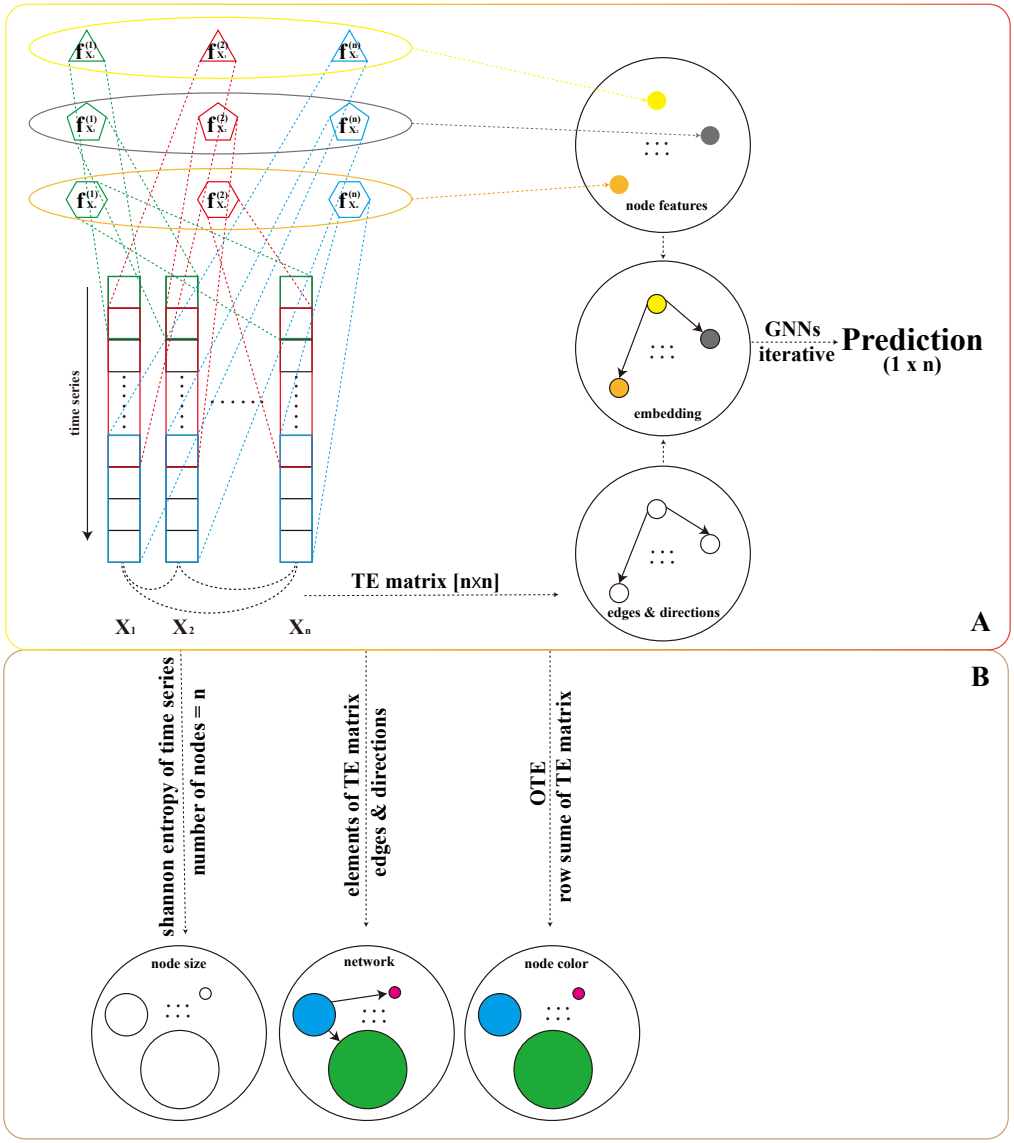


Figure 2.2: **Information inference network and prediction modeling framework.** (A) is a causality-based graph neural network framework with transfer entropy as a priori information. (B) indicates the size of the nodes in the causal inference network, the network composition (including the size and direction of the edges), and the node colours, respectively.

with the node feature matrix ($F \in \mathbb{R}^{(n \times d)}$) to obtain node embedding and create a multivariate time series prediction model. The hidden state of a node $X_i$ is represented by a $v$-dimensional vector $h_{X_i}$ that contains information about the node's neighborhood.

$$h_{X_i} = \mu(f_{X_i}, te_{diff[X_i]}, f_{ne[X_i]}, h_{ne[X_i]}), \tag{2.6}$$

where $f_{X_i}$ indicates the features of node $X_i$. $te_{diff[X_i]}$ denotes the features of edges connected to $X_i$ (Eq. 2.4) in the hidden layer. Features of the nodes adjacent to node $X_i$ is denoted as $f_{ne[X_i]}$. The hidden state representation of the nodes connected to this node $X_i$. The update of the hidden status follows

$$h_{X_i}^{(m+1)} = \mu(h_{X_i}^{(m)} W_{\alpha}^{(m)} + \sum_{X_j \in N(X_i)} h_{X_j}^{(m)} W_{b}^{(m)}). \tag{2.7}$$

Function $\mu$ is a transition function that maps inputs to $v$-dimensional space. $h_{X_i}^{(m)}$ denotes the state of the node $X_i$ in the hidden layer $m$, $W_a^{(m)}$ and $W_b^{(m)}$ are parameter matrices and $N(X_i)$ represents the neighbors of the node $X_i$. $X_j$ can be arbitrary nodes connected to $X_i$. The node features $f_{X_i}$ (in the Eq. 2.6) are iterated many times in different hidden layers of the neural network. Therefore, the forward propagation update of a certain node $X_i$ can be performed as Eq. 2.7. However, the output of the network is obtained by passing the state $h_{X_i}^{(m+1)}$ and the feature $f_{X_i}$ to the output function (Eq. 2.8). Meanwhile, the output dimension of the last graph neural network layer is 1, which is used as the prediction result.

$$O_i = g(f_{X_i}, h_{X_i}^{(m+1)}). \tag{2.8}$$

**Loss function**

To optimize the parameters of the above model, we used the $L_1$ loss function,

$$L_1(\hat{O}_t, O_t) = \min_{\phi} \sum_{t \in T} \sum_{i=1}^{q} ||\hat{O_{i,t}} - O_{i,t}||, \tag{2.9}$$

where $\hat{O}_t$ is the prediction result of the model, $O_t$ is the actual result; $q$ denotes the number of variables; $T$ is the set of time stamps used for training and $\phi$ represents all training parameters, including $W_{\alpha}^{(m)}$ and $W_{\beta}^{(m)}$ (Eq. 2.7). The training parameters are updated to minimize the loss function.

## 2.3 Results and Discussion

### 2.3.1 Analysis of spatial dynamics of algal blooms

Using time series of CHL-a from 28 spatially distributed water monitoring stations (see Fig. 2.1 A), we constructed a strong causal information network (see Fig. 2.1 B) to explore spatial dynamics of algal blooms between 1992 and 2008. First, the bloom information content of a station (the time series of CHL-a) was quantified by Shannon entropy (see Eq. 2.1. This is represented as the node size in Fig. 2.1 B). Stations with unstable bloom information content have high Shannon entropy. The bloom surrounding stations located in the center of the bay (12, 13, 14, and 15) were the most unstable, with those at the extreme ends – in the East and West – were relatively stable (see node sizes in Fig. 2.1 B). This is due in large part to the geographical location of the central region, which limits the exchange of water between the bay and the open ocean, resulting in the accumulation of various nutrients in the water column. This is a reason why the central part of the bay is the hardest hit by bloom, and it does not dissipate for a long time.

We considered the spatial transfer of the bloom as the spatial interaction between any two stations, which was quantified as the difference in transfer entropy ($TE_{diff}$, see Eq. 2.4) between their respective CHL-a time series. We considered strong unidirectional causality, capturing the extent to which the bloom surrounding one station causes uncertainty blooms elsewhere. The inferred network confirmed that the blooms in central Florida bay were most likely the source of bloom activity elsewhere in the bay. As seen in Fig.2.1 B, the directions of interaction were largest from these information hotspots to other stations in the region. This further illustrates how a complex information network can simulate the spatial state and dynamics during an algal bloom. This further demonstrates the ability of complex information networks to effectively model the spatial state and dynamics of algal blooms.

We further exploited this information network by analyzing the total effective outgoing impact of a station, calculated as its total outgoing transfer entropy ($OTE$, see Eq. 2.5). In this scenario, a high OTE reflects a high rate of activity (information dissemination) from that station to others in the network. We grouped the OTE into eight levels according to its magnitude. As can be seen in Fig.2.1 B., Stations 12, 21, 13, and 19 are the most active stations during the bloom in Florida bay. They were more likely to be the source of any bloom activity but were less likely to be

affected by the activity around other stations. In contrast, Stations 2, 5, 6, 11, and 24 were the least active stations, having little impact on the rest of the network. This suggests that any observed bloom activity is likely to originate in the central bay and radiate outwards. And to a large extent, the eastern bay is significantly impacted by the spread of blooms from other regions, and these blooms are generally not indigenous to this area.

### 2.3.2 Inferred network warning for algal blooms

We utilized the strong causality network as a basis to study the interaction between blooms and water quality and the abruptness of bloom occurrence by observing blooms at the most active station (each station is independent). The region with the highest bloom activity was found to be Station 12 in central Florida bay (Fig. 2.1 B), where blooms exploded continuously for 18 months from February 1993 to July 1994 (Fig. 2.3). In addition to observing the bloom indicator CHL-a among stations, we also analyzed the interaction between blooms and five other water quality factors (TN, TP, SAL, TEMP, and TURB). At Station 12, where the bloom persisted for an extended period, TN was identified as another significant driver of its continued occurrence, along with CHL-a (see color nodes in Fig. 2.3 B). In addition, all the observed factors except TP have higher unstable water quality information (see the size of nodes in Fig. 2.3 B). A prolonged bloom was observed to cause significant changes in water salinity in the vicinity of Station 12, while the other water quality factors acted as both controllers and regulators. The direct effects of a prolonged bloom on water quality are extensive, and almost all observed water quality factors are altered after a long bloom period (like the CHL-a impact on others in Fig. 2.3 B). Such effects play significant roles in the occurrence of future blooms. Although blooms frequently occurred at Station 12 over the next decade, they were less severe and short-lived (see Fig. 2.3 A). However, a sudden outbreak of intense blooms in September 2006 prompted us to examine the network of water quality relationships for the preceding four months in order to identify the factors contributing to the sudden bloom (Fig. 2.3 C). Our analysis revealed that TN, CHL-a, SAL, and TURB were all in active states prior to the peak bloom, while the water temperature was passively affected (see the color of nodes in Fig. 2.3 C). Although the information content on unstable water quality factors is lower than that on persistent blooms (node sizes in Fig. 2.3 B & C), it can be shown that the sudden outbreak of blooms is the result of water quality imbalance caused by multiple

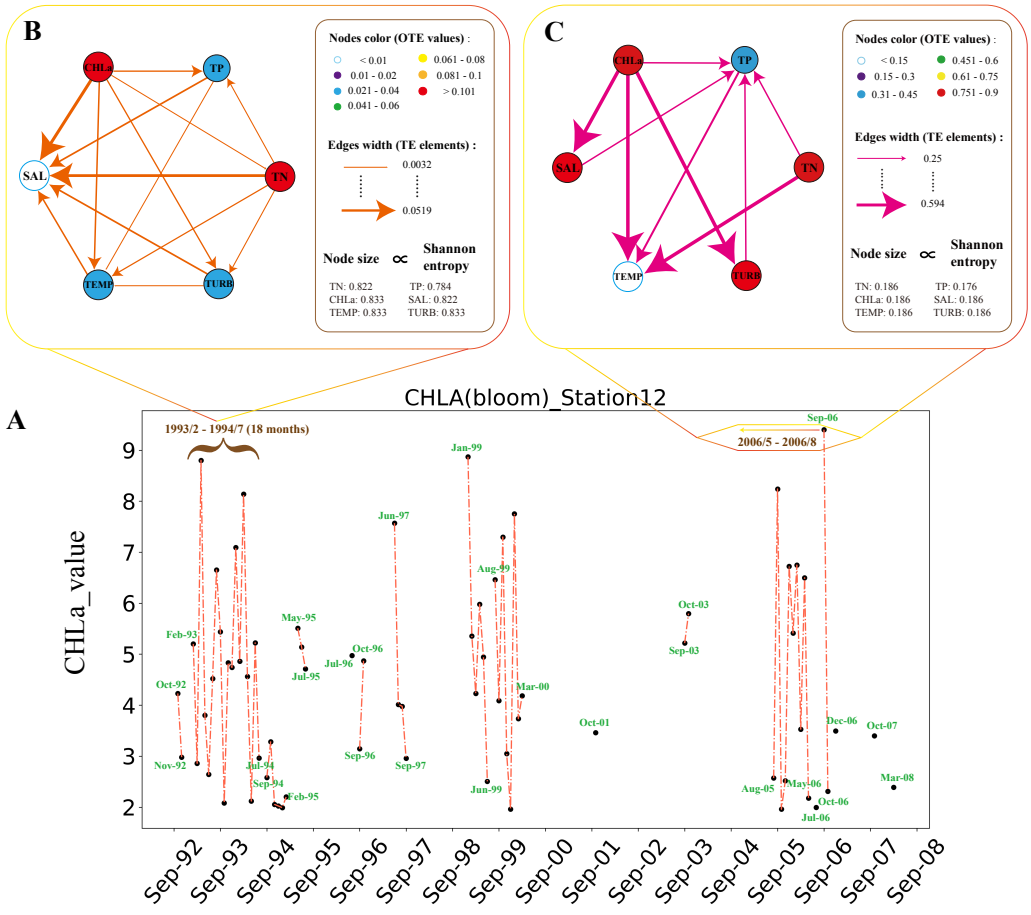


Figure 2.3: **Prolonged and sudden blooms at a typical station (Station 12).** (A) indicates the moment of onset and persistence of bloom at station 12. (B) demonstrates the causal effects of water quality factors on prolonged blooms. (C) displays the interactions of water quality factors before the sudden bloom peak.

factors interacting in depth.

### 2.3.3 Bloom prediction and analysis

We used four active stations as examples to analyze the bloom prediction ability of our network-based inference method from an ecological perspective. In the line charts (Fig. 2.4 a, d, g & j), all parts above zero imply the occurrence of blooms. So, on the temporal scale, we can conclude that the predicted and actual values are almost synchronized, and there is no severe delay. Our method achieved good results in predicting the temporal scale of the blooms outbreak. On a spatial scale, we used the peak moments of bloom at each active station as a reference to compare the actual and predicted values of blooms across the whole Florida bay (see Fig. 2.4 maps). As can be seen from the maps, the real areas of bloom occurrence are extremely similar to our predicted bloom conditions. At the peak bloom at Station 12, the most severe bloom should be concentrated around Station 14 and northeast regions (Fig. 2.4 b). The predictive map shows a similar situation (Fig. 2.4 c). In August 2007, the bloom at Station 21 reached the maximum value; however, the area of severe bloom at this time remains in the north-central and northeastern regions for the whole Florida bay (Fig. 2.4 e). The predictions at this time are biased, with the severe bloom stations occurring in the northwest and northeast areas (Fig. 2.4 f). But the predicted location deviations are not significant. This is because there are fewer bloom data at Station 21, leading to spatial prediction bias. The peak bloom at Station 13 was in October 1999. But the most serious areas of bloom across Florida bay at this time were around Stations 14 and 15 in the northcentral region (Fig. 2.4 h), and the predictions are identical to this (Fig. 2.4 i). December 2005 was the peak of blooms at Station 19. But the bloom was rampant around Stations 14, 15, and 16 in addition to 19, while significant bloom was also observed around Stations 4 and 5 in the northeast (Fig. 2.4 k). The forecast map also shows water blooms in the northwest, north-central, and northeast and near Station 19, which are consistent with the real situation (Fig. 2.4 l). Based on our analysis, we can confidently state that our method exhibits strong performance in terms of temporal accuracy and spatial reliability. This indicates the efficacy of our approach in capturing patterns and making predictions at various scales.

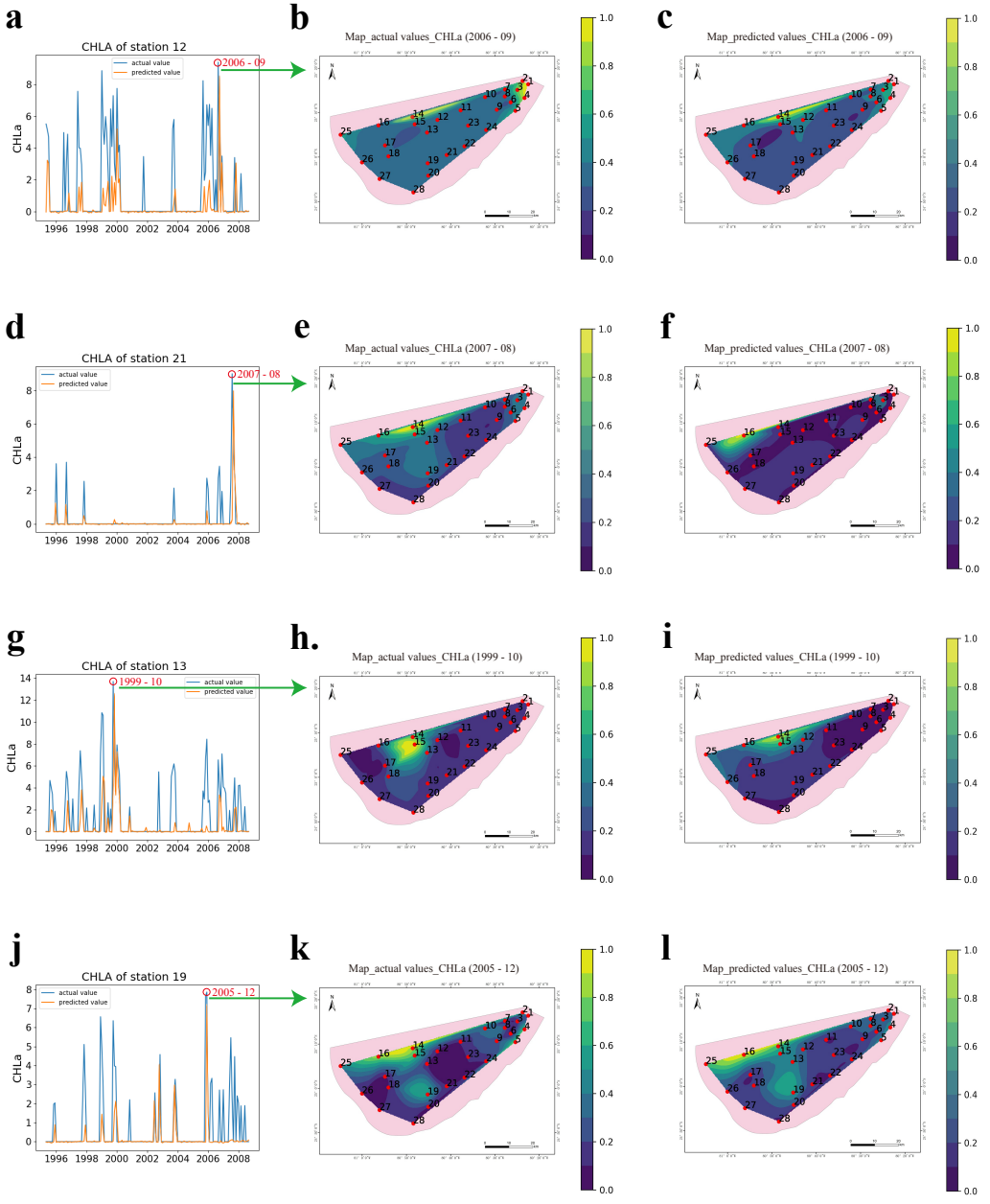


Figure 2.4: **Temporal and spatial scales analysis and prediction of active stations (Stations 12, 21, 13 and 19).** b, e, h and k are the actual distributions of blooms at the corresponding peak moments. while c, f, i, and l are the corresponding predicted distributions of blooms.

## 2.4 Conclusions

Harmful algal blooms (HABs) are a phenomenon where certain species of algae grow rapidly and produce toxins that can be harmful to humans and marine life, marine ecosystems, and even the economy. Some of the potential hazards of HABs also include fish and shellfish poisoning [49], respiratory problems, skin irritation [51], and even death. It is crucial to observe the temporal as well as spatial scales of HABs as they can have a significant impact on the frequency and duration of HABs. HABs can occur sporadically or frequently, depending on various environmental factors such as nutrient levels, water temperature, and weather patterns. Therefore, monitoring HABs over a long period can help identify trends and changes in their occurrence and severity. In addition, HABs can occur in specific areas or spread over large regions, making it important to understand their spatial distribution. This information can help identify areas at risk of HABs and support the development of management strategies to mitigate their impact. Prolonged blooms can have significant ecological impacts, such as depleting oxygen levels in the water, which can lead to the death of fish and other aquatic life. The decay of large amounts of algae can also lead to the production of methane and other greenhouse gases that contribute to global climate change [52]. In order to avoid sudden outbreaks of blooms, it is important to monitor nutrient levels and water quality, reduce nutrient pollution from sources such as agricultural runoff and wastewater treatment plants, and implement management strategies such as controlling water flow and introducing natural enemies of harmful algae. And to prevent sudden outbreaks of blooms, it is important to monitor nutrient levels and water quality, reduce nutrient pollution from sources such as agricultural runoff and wastewater treatment plants, and implement management strategies such as controlling water flow and introducing natural enemies of harmful algae.

Our current work has been analyzed and evaluated from an ecological perspective. The results show that our method is very effective in terms of multi-variable and simultaneous consideration of multi-factor interactions. This differs from using traditional model analysis techniques but rather assesses the effectiveness of the method from a more intuitive perspective. This approach is valuable for studying and solving real-world problems such as ecology and environment, the internet, social networks, and so on. In addition, methods like ours are currently in high demand in the pharmaceutical industry. They are used to simulate protein production and assist researchers in discovering potential new drugs. However, our future

work still requires model refinement to assess the accuracy of the model from the computer technology perspective.

# Chapter 3

# Spreading Network Inference and Extreme Eco-Environmental Feedback

## 3.1 Introduction

### 3.1.1 Health and complexity of marine ecosystems

Algal blooms manifest as abnormal changes in phytoplankton communities within aquatic ecosystems, including estuaries and lakes [53] [54]. Despite discussions about the global increase in algal blooms attributed to intensified monitoring and emerging impacts [55], satellite images by Dai et al. [55] highlight a worldwide increase in these blooms, raising concerns for local ecology and global climate. These blooms, as observed through satellite imagery [55], pose various ecological and climate-related concerns. They are highly destructive and persistent [56] [57], leading to ecological catastrophes such as reduced vegetated communities, widespread sponge mortality, and loss of marine habitat geomorphological structure [58] due to habitat calcification.

Despite the extensive damage to aquatic ecosystems, the mutual influence between blooms and the environment has received limited attention from scientists and policymakers. Algal blooms result from changes in nitrogen (N) and phosphorus (P) but can also alter the N/P balance [59] and temperature [60], affecting carbon sequestration in blue carbon ecosystems like seagrass [61]). These effects can be exacerbated by local and global climate change [62].

While literature on the effects of blooms on the environment is limited, some studies have explored the relationship between phytoplankton and water quality during bloom conditions [63], climatic and regional variations in phytoplankton as bloom features [64] [46], and habitat-specific effects related to local planktonic biogeochemical stress [44]. Fewer studies have examined the spatial spread of blooms as complex networks and predicted blooms based on spatially explicit biogeochemical factors.

A comprehensive biocomplexity study, similar to the one we propose, would be essential to define micro-macro feedback loops crucial for risk assessment, management, and policies aimed at minimizing eco-environmental imbalances resulting from blooms. Blooms epitomize marine ecosystem health, as their emergence is closely linked to altered ecohydrological factors on a basin-scale, stemming from both land and ocean sources. This leads to rapid and persistent phytoplankton increases with short-term [65] and long-lasting systemic effects on ecological function and the environment. This extends beyond a single species or humans, contributing to the gradual degeneration of ecosystem function from its optimal or baseline state in relation to initial or desired conditions.

Marine microbial food webs consist of heterotrophic protists, phytoplankton, prokaryotes and viruses (i.e., the ocean microbiome). Together, they are responsible for a large part of the production, respiration and nutrient transfer in oceans; they affect, for instance, the carbon cycle both in blue carbon habitats and in the ocean via the carbon pump. As marine ecosystems are increasingly affected by anthropogenic disturbances both from land and ocean, predicting ecosystem responses above critical environmental pressure relies on a better understanding community dynamics, including their composition, spatial/temporal distribution and interactions. Long-term observations are especially useful for this, and both Galbraith and Convertino [44] and Galbraith et.al. [66] provided clear ecological patterns to use as indicators of ecosystem health in relation to ocean microbiome variability intended as a complex network. Chlorophyll-a (CHLa) seems to be the best indicator of community health; however, currently there is the need to quantify how much CHLa variability implies changes in ecological effects (e.g., blooms) and long-term effects, such as on the environment and ecosystem function (e.g., carbon cycle). Coastal and marine ecosystems that experienced marine heatwaves, which were particularly significant in 2014-2015 worldwide, provide a unique opportunity to study how warming affects community dynamics (namely, microbiome interactions) and how imbalance of the latter affects the environment back in the long term.

The presented tool for ecosystemic risk assessment and the results from FL Bay are the main innovations of this paper. The topological network structure is an effective and intuitive way to describe the dynamical dependencies among diverse of analogous units of an ecosystem, or ecological communities composed of hundreds or thousands of populations of species [67] [68]. This is particularly important for marine ecosystems where both structural networks (such as coastal and marine habitat connections and flows) and functional networks (such as prokaryotic and eukaryotic interactions) are not directly visible or known. Yet, causal network discovery and inference models (e.g., see Li and Convertino [69] for an articulated discussion about ecosystems) are particularly important for mapping the ecological baseline on which current ecosystem assessment and future predictions of ecosystem patterns (tangibly liked to ecosystem services) can be made. Complex networks have great potential to help in solving contemporary real-world problems in a wide range of fields [70] [71] [72] [73] [38] [74]. Complex networks have been used to analyze the dynamics of pseudo-periodic time series [75] and the functional dynamics of complex systems [76] [77] [78] [79]. Furthermore, networks have become an excellent method for the study of functional and structural dependencies among very complex units with different temporal dynamics [80] [81] [82] [83]. However, most of the considered networks in the literature and particularly those inferred in ecosystems, typically represent relationships based on known or assumed affiliations [84] [85] or fixed connections [86]. This makes it difficult to represent the independent local properties of each node and, more importantly, the unique dependencies between different nodes. This issue is particularly relevant for algal blooms where the biogeochemical networks are hypothesized to vary dramatically over time and space. This has been verified by recent studies on prokaryotic networks whose topological variability was strongly related to systemic ecological stress [44] [66]. Nonetheless, no analyses have been made so far on bloom spreading networks, and this research presents a novel template for characterizing and predicting algal blooms based on chlorophyll-a and associated water quality factors.

### 3.1.2 Ecological patterns of spreading networks of algal blooms indicator

Species, encompassing microscale eukaryotes, function within dynamic ecosystems, placing immense importance on their capacity to react to shifts in environmental dynamics. The capability for an effective collective response, influencing

the restoration of optimal ecosystem conditions, hinges on the exchange of pertinent information among species. Consequently, ecosystems rely profoundly on networks of eco-environmental interactions. This fundamental aspect underlies the progression of ecosystems toward states of low entropy, which is exemplified by the scale-free distribution of CHLa, as explored in this study [87]. Furthermore, this process is instrumental in enabling adaptation to novel environmental stress states [88], some of which may be undesirable, including persistent and substantial blooms. In the present study, we emphasize the pivotal role of information transfer as a prominent characteristic driving collective eco-environmental dynamics that lead to the occurrence of algal blooms. Connectomics is broadly defined as the study of structural and functional networks (the connectome), which are maps of a system (such as the nervous system), mainly in the brain; however, this concept has been extended to ecosystems (see Convertino and Valverde Jr. [89]) to characterize both functional species interaction networks, their stimuli with the environment or the envirome itself as set of interdependent environmental processes [44] and habitat networks [69]. The connectome enables understanding of how spreading information is processed (coded, stored, transmitted and decoded in an information sense, which can be any ecological information) at and among different scales of the system (e.g., one node and the whole system, while also considering cross-scale dependencies). As the connectome serves as the fundamental information framework of ecosystems, possessing knowledge of it empowers the enhancement of predictive capabilities, both in the short and long term, for representing ecosystem dynamics. Addressing the aforementioned requirements, specifically the identification of spreading bloom trajectories and their potential environmental ramifications, we showcase the efficacy of an information-theoretic approach in deducing bloom networks and biogeochemical feedback. The optimal information flow model was initially devised for deducing species interaction networks within any ecosystem based on abundance data [69], subsequently extended to predict fish biodiversity patterns [90] and eco-environmental interactions within the ocean microbiome [44]. The ecological time series underpinning ecological dynamics are particularly important for assessing ecological states and early warning signals of shifts [91] before the inference of ecological networks. The proposed model applies transfer entropy (TE) differences (to target the salient directed interactions) to infer a spatial network strategy that can identify the sources and sinks of bloom outbreak as well as foretell changes probabilistically in water quality factors (in average and fluctuations) when blooms happen. Through the model, we specifically infer and analyze

the spatial ecological corridors determining bloom spread and direct interactions between CHLa and environmental factors to quantify the environmental effects of ecological dysbiosis; previous efforts (see Wang and Convertino [41]) focused on the whole set of biogeochemical interactions useful for forecasting outbreaks, except for bloom spreading networks. Historically, CHLa has commonly served as an indicator of blooms, owing to its sensitivity to environmental shifts, straightforward monitoring, and proficiency in reflecting phytoplankton biomass [47]. However, its status as a comprehensive indicator of ecosystem health in relation to ecosystem function remains unverified. We present a discussion on the outcomes of applying this model to instances of algal blooms observed in Florida Bay, located within the Florida Everglades National Park, during the period spanning 2005 to 2006, marked by a recurrence of substantial blooms. Given its distinctive lagoon configuration and climatic conditions, Florida Bay (Fig. 3.1) experiences recurrent algal blooms [42], aligning with the frequency observed in numerous other aquatic ecosystems situated in subtropical and tropical climates. As a result, a robust and dynamic predictive model is imperative for algal blooms, with the aim of facilitating decision-making processes and the prevention of blooms.

## 3.2 Methods and Materials

To represent the Fig. 2.2(B) in detail, the proposed TE network inference model that can be used for variable interaction discovery at multiple scales is explained. Its structure is graphically shown in Fig. 3.2.

### 3.2.1 Data preprocessing

We used a threshold-based quantile regression method (analogous to Nelson et al. [46] ) to establish an average threshold of $\geq 2\mu gL^{-1}$ on CHLa, universally applied to all stations, to distinguish bloom from non-bloom states across all stations. Initially, the dataset for this study spanned 2004 to 2006, corresponding to before, during and after a severe bloom outbreak in Florida Bay in 2005 [92] in terms of a CHLa extreme. In 1999, several blooms were observed in the same area but with lower CHLa extremes [46]. Then, the dataset (comprising all 2004, 2005 and 2006 CHLa monthly data) was filtered to include only those months and stations with CHLa values exceeding the critical blooming threshold, i.e., those months and stations indicating sustained bloom conditions. As a result, the final dataset contained

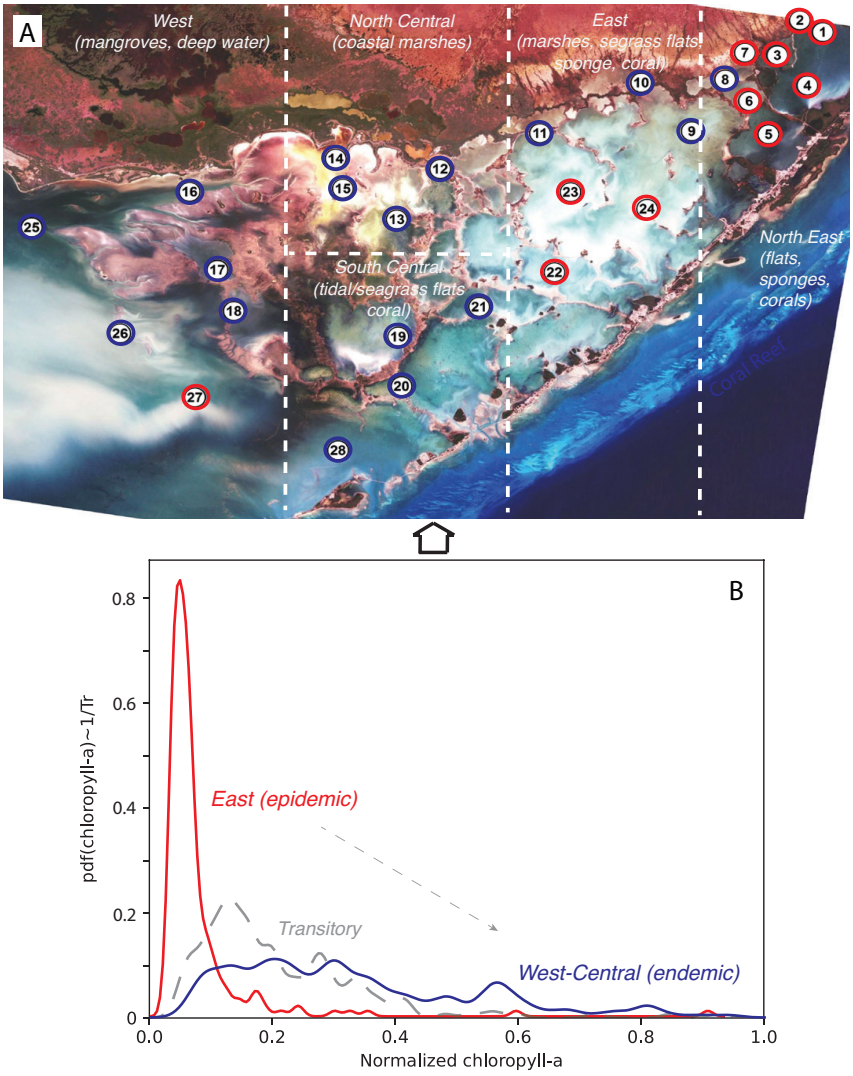


Figure 3.1: **Florida Bay and area classification based on CHLa dynamics.** The red–blue classification in plot (A) is related to the probabilistic structure of CHLa as highlighted in plot (B). Plot (A) also highlights the main habitats and species present in FL Bay.

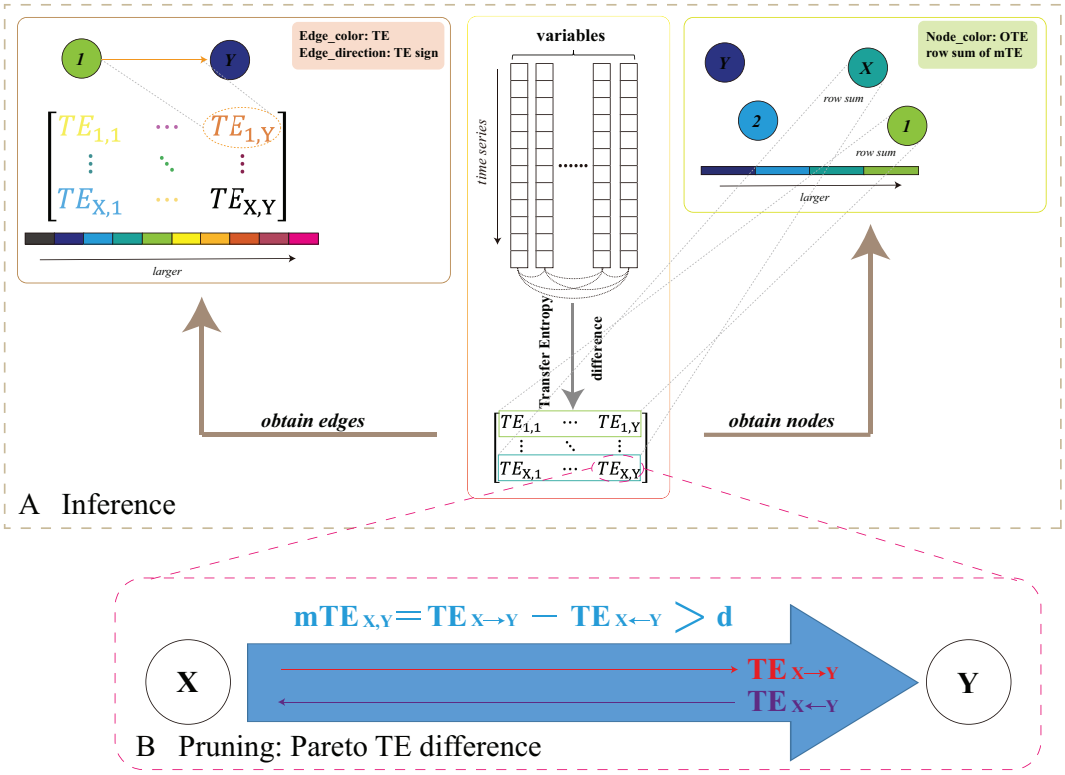


Figure 3.2: **Ecological corridor inference model.** The structure of the TE inference model. Here variables are annotated as $X$ and $Y$ generically. $X$ can be thought as $CHLa$ and $Y$ all other environmental variables. The first step of the proposed model is to infer variable pairwise interaction as TE, and node collective influence (OTE), determined via Eqs. 2.5 and 2.4, respectively. The second step is to prune the network considering only salient Pareto interaction via thresholding TE differences with a threshold $d$ of causal significance that is set to consider the top 20% TEs (Eq. 2.4) necessary and sufficient predict bloom spreading.

18, 63 and 136 rows of measurements (i.e., months) for the 2004, 2005 and 2006 bloom periods (pre-, peri- and post-bloom), respectively. More generically, 2005 can be considered as epitomic of bloom outbreaks, while 2004 and 2006 are representative of early and post-bloom periods.

### 3.2.2 Ecosystem organization and connectome

The entropy of the ecosystem, manifesting ecological disorganization in relation to CHLa variability, is dependent on the probability distribution functions (pdfs) that affect TE calculated on the pdf divergence and asynchronicity. The TE variability of an area, or the whole system, can be decomposed into eco-environmental interactions (considering CHLa and environmental factors acting as determinants or effect of ecological imbalance) and the ecological areal interactions underpinning bloom spread. This variability affects the organization propagation of CHLa (i.e., how randomly distributed CHLa is), and in an information-balance equation, can be written as the spatio-temporal convolution of the aforementioned components composing the ecosystem connectome,

$$\overbrace{H(\text{CHLa})}^{\text{eco-function}} = \sum_{m,n} \int_0^t \overbrace{\underbrace{(1 - TE(X_m, \text{CHLa}_m))}_{\text{eco-env feedback}} * \underbrace{(1 - TE(\text{CHLa}_{m,n}))}_{\text{eco-corridors}}}^{\text{eco-connectome}} \, \mathrm{d}\tau \ , \quad (3.1)$$

where, $X$ stands for all other environmental variables except for CHLa, and $m, n$ stands for the location of each area being monitored over the period $t$. The specific TE chose in Eq. 3.1 is related to TE analytics and posed objectives, later to be specified. It should be noted that the time delay $\tau$ between eco-env factors in Eq. 3.1 has been set to one due to the sub-monthly variability of CHLa and the resolution of the data. Equation 3.1 is focused CHLa patterns where networks are the backbone determinants of the ecological "weave" (CHLa interconnected patterns) that can be potentially controlled. Space and time are the dimensions along which CHLa is considered, plus other dimensions along gradients of environmental features on which stress-response patterns and related features (e.g. early-warning signals and risk thresholds) can be derived. Networks that define sources, sinks, pathways and determinants to guide monitoring and environmental control for bloom prevention. In this paper, we specifically analyze the spatial ecological corridors determining bloom spreading and direct interactions between CHLa and environmental factors (second and first term in Eq. 3.1, where for the latter only $CHLa_m \rightarrow X_m$ interactions are considered) to quantify environmental effects of ecological dysbiosis;

Wang and Convertino [41] focused instead on the whole set of biogeochemical interactions useful for forecasting except for bloom spreading networks.

### 3.2.3 Eco-environmental network inference

Transfer entropy ($TE$), constructed from information entropy [93], measures the causal relationship between two asynchronous and divergent variables (expressed as time series) $X$ and $Y$ (in the bivariate form, yet not accounting for second-order indirect interactions) by quantifying the predictive information flow between variable pairs [94]. Previously the TE-based model, called "Optimal Information Flow" (in relation to the maximization of the systemic entropy to gather the largest information), was used to discover causal relationships in human and aquatic ecosystems, e.g., for bacteria [38, 44, 66] and fish interactions [69] and assess ecosystem health. The information flow, and thus the predictive relationship between variables, is bi-directional. In this paper we took the form of bivariate TE (yet skipping interactions higher than the third-order that is our first modeling assumption considering the weakly third-order interactions between environmental factors [3]) and calculated the difference between the pairwise information flows to identify the strongest causal factor. It is expressed as Eq. 2.4.

In this study we considered only positive $TE_{diff(X_i,X_j)}$ where $X_i = CHLa$ and $X_j$ are all other environmental factors for eco-environmental feedback in Eq. 2.4, and all $TE_{diff(X_i,X_j)}$ where $X_i$ and $X_j$ are both CHLa in two different stations. Additionally, in the TE calculation we did not investigate the optimal time-delay between $X_i$ and $X_j$, nor the optimal set of factors that are predictive of CHLa, due to the fact that: (i) bloom eco-env feedback occur at scale lower than one month (at which data are available), and (ii) our interest into the whole systemic dynamics. This first part of all TE inference is considering all pairs of variables (Fig. 3.2A). The unbounded causality matrix, or more precisely the predictive causality matrix **TE** unconstrained to any prediction of biodiversity patterns, based on calculated TEs without predictive environmental factors of ecological patterns in an Optimal Information Flow perspective, can be constructed as follows:

$$\mathbf{TE} = \begin{bmatrix} TE_{diff(1,1)} & \cdots & TE_{diff(1,X_j)} \\ \vdots & \ddots & \vdots \\ TE_{diff(X_i,1)} & \cdots & TE_{diff(X_i,X_j)} \end{bmatrix}. \tag{3.2}$$

where TE in indeed a difference of transfer entropies as in the Transfer Entropy

Graph Neural Network model (TEGNN) (originally developed by Duan et al. [95] and applied to algal blooms by Convertino, M. et al. [3]) in contrast to the Optimal Information Flow model (OIF) originally developed by Li and Convertino [69]. For each year, two networks were constructed, each defined by an underlying matrix of transfer entropy differences **TE**. One inferred matrix was a spatial network in which the 28 stations were nodes and the causal influence among them were the edges. The time series used to calculate the transfer entropy differences (Eq. 2.4) in this network were the time series of CHLa measurements at each station. The second inferred network was a water quality network, in which the nodes were the water quality factors (CHLa, TN, TP, SAL, TEMP, and TURB) and the edges were the causal influence among them.

In this study, however, the causality matrix underlying the water quality network was further filtered to focus only on the effect of CHLa on other water quality factors in relation to the objective to quantify these eco-environmental feedback; the reverse effect of water quality factors on CHLa and the interactions among water quality factors were not considered in this task but in Convertino and Wang [3] and Wang and Convertino [41]. Additionally, the values in the matrices for each of the three years were normalized to enable better comparison of the inferred interactions. This second part of salient TE selection is about the network pruning (Fig. 3.2B).

## 3.3 Results and Discussion

### 3.3.1 Spatio-temporal spreading and fluctuations

To infer and characterize the spreading networks of blooms, underpinning ecological risk, we considered Florida Bay blooms in between 2004 and 2006. We novelly inferred a spatial influence network, underpinning bloom spread, among a set of spatially distributed water monitoring stations. This was achieved by deriving a TE matrix from spatio-temporal patterns of CHLa derived from monitored stations (see Section 3.2.2). The TE matrix for 2004 suggested that the study site was free of severe blooms, except for a few stations in the northwest: specifically, stations 16, 14, 25, and 26 (Fig. 3.1) at least in 2004 where the resurgence of blooms was observed after the big bloom in 1999 [3, 46].

Ecological spreading corridors are defined by the most divergent and asynchronous CHLa among nodes, yet defining the most likely interdependent area, at least in a predictive causality sense (causality considering all other feasible con-

nections, that are all other nodes in this case). Divergence and asynchronicity, as highlighted by Li and Convertino [69], are related to the difference in pdfs of CHLa (in two nodes) at different or equivalent time periods, respectively. Spreading can certainly related to marine currents, however, in this study the purpose is not to define the precise mechanisms underpinning ecological patterns but rather to define patterns' backbone networks that are the salient spreading networks. This does also identify the potential coastal areas of influence of biogeochemical loads in Florida Bay, yet, the maximum extent of blooms; something that is really poorly quantified but necessary for bloom prevention. In analogy, runoff in terrestrial basins is predicted equivalently to CHLa, where the amount (and distribution) of water in different locations changes in an asynchronous way and dependent on river network spreading defining timing and divergent volume. True causality, leaving aside the feasibility of its assessment, must be done considering all areas where CHLa can spread, that in a bay is virtually everywhere; however, this is challenged by the data-limitation that is constrained only to the used stations in this case. The properties of network edges, representing eco-environmental interactions, depend on $TE_{diff(X_i,X_j)}$ (Eq.2.4). The edge directions of the spreading network (Fig. 3.3) suggested that an algal bloom would have initiated around station 16 and then move west with a preferential direction toward northwest.The edge colors (proportional to $TE_{diff(X_i,X_j)}$) suggested that the bloom was moderately strong but localized in 2004 (Fig. 3.3A) with a high probability to continue growing in the Bay (due to $TE_{diff(X_i,X_j)}$ directions). The spatial spreading $TE_{diff(X_i,X_j)}$ matrix for 2005 revealed large and widespread bloom outbreaks, concentrated in the western and central areas of the Bay (Fig. 3.3B). In that year, the spatial influence was the strongest near stations 25, 16, 14, and 12 in the northwest and station 28 in the south region. The edge colors indicate that the bloom at all stations was moderately strong and also very likely to continue in the NE direction. Station 28 seems largely affected by many other stations in the bloom spreading, and yet likely a sink node with potentially strong ecological effects also considering its proximity to the FL coral reef. The matrix for 2006 (Fig. 3.3C), showed the most extreme area interactions as well as a reversal in the spreading of blooms, i.e. moving from NE to central areas. The edge colors implied that bloom activity was extremely high, covering a wide area of Florida Bay. Nonetheless, the resulting graph suggested that after the largest outbreak the bloom moved from the easternmost into the north-central area, while the bloom in the west region had dissipated.

We showed how by analyzing information flow among spatially distributed nodes,

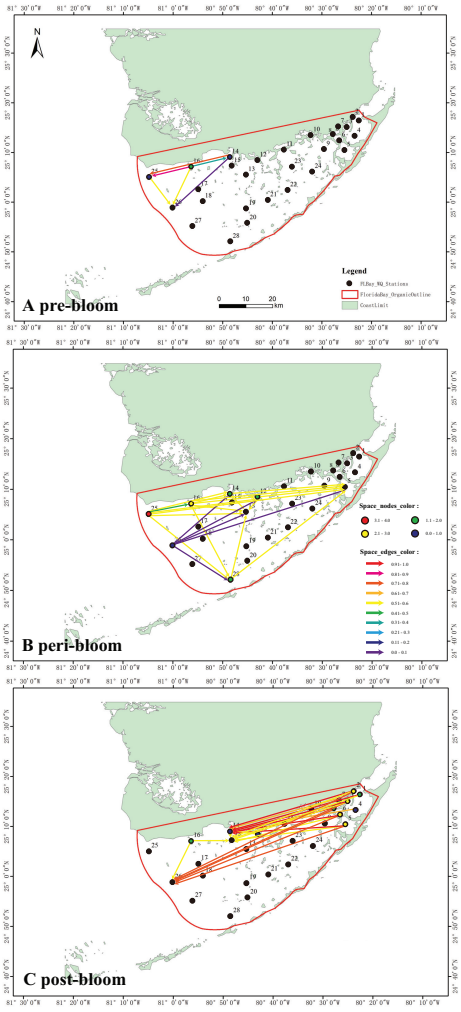


Figure 3.3: **Inferred spatial CHLa for the 2004, 2005 and 2006 pre-, peri- and post-bloom periods in Florida Bay.** Link and node color (from blue to red) is proportional to mTE based on CHLa interdependence between node pairs, and OTE considering only $TE_{CHLa->Env}$ where Env stands for all other environmental factors. East to West node and link dynamic increase is observed from 2004 to 2006 as well as a spreading network transition from regular/Small-World to Scale-Free and Regular (or uniform) with long-range connections for 2004, '05 and '06. Each year corresponds to different bloom precursor area and environmental factors (Central and North-West more affected by nutrients), widespread and extremely localized outbreak (North-East more affected by temperature and turbidity and sequential effects of spreading.

it is possible to model the spatial spread of a phenomenon, like algal blooms. In addition, this approach is able to detect sources, sinks, directions and salient pathways of bloom spreading. Due to various unaccounted factors such as wind intensity and direction, current direction, and bathymetry to name a few, there is a certain dynamic spatial change of blooms that is not attributed to the aforementioned factors. However, the model can take into account any environmental factor if available and yet it can attribute the degree of variability of CHLa. In a complex network sense, the bloom spatial network in 2004 is small in scale and regular in topology but has an obvious active station (station 16) that is an actively connected hub for bloom spreading. Therefore it is much easier to take measures against blooms at this time (whether possible) or to prevent triggers by controlling environmental determinants. This area is well-know to be heavily influenced by nutrient efflux from the Everglades [46]. Particularly with the outbreaks of blooms in 2005, and later in 2006 the network has many more area that are very active and affected, yet blooms management becomes more difficult. Over time a spreading network transition is observed from a regular/Small-World in 2004, to Scale-Free in 2005, and Regular (or uniform) topology in 2006 with long-range connections.

### 3.3.2 Impacts of blooms on water quality

The investigation of the impact of CHLa extremes (magnitude, duration and frequency) on ecosystem health is a poorly covered topic in science. To explore how algal blooms impacted water quality in Florida Bay, we analyzed how CHLa impact other water quality variables using TE (see Section 3.2.2). We focused our analysis on how CHLa implicated potential changes in water quality – in terms of predictive causality – for stations where extreme blooms were most likely. At the most active station in 2004 (i.e., station 16 characterized by coastal marshes that is likely the source of blooms; see Fig. 3.3A blooms did not affect TN, TP, SAL, and TURB, except for a slight effect on water temperature (see Fig. 3.4A). Rather, TN, TP, SAL, and TURB, likely driven by riverine efflux in the Bay, triggered CHLa changes leading to blooms as highlighted in Wang and Convertino [41] and Convertino and Wang [3]. In 2005, the impact of blooms on other water quality factors was mostly evident at station 25 that is a deep-water mangrove habitat, where the blooms were the most intense (see Fig. 3.3B). CHLa induced not only water temperature changes but also variations into total nitrogen and salinity (TN and SAL) with higher impact

for the latter (Fig. 3.4B). In 2006 (see Fig.3.3C) where blooms are the most extreme but localized (NW area), the effect of blooms on water quality peaked at station 3, followed by stations 5 and 2, and then station 6 in terms of magnitude (Fig. 3.4C). Stations 2, 3, experienced blooms throughout the year, while station 6 had a relatively short blooms (7 months as reported in data). At stations 3 and 6 (characterized more by tidal flats) blooms induced changes in water temperature, salinity, total phosphorus, and turbidity, while at stations 2 and 5 (characterized more by submerged marshes) blooms led to substantial fluctuations in total nitrogen, total phosphorus, salinity, and turbidity. Information flow patterns (TE patterns) suggest that blooms are first strongly causing water temperature alterations, then enhancing salinity and nitrogen, and later impacting other nutrients (phosphorous) and turbidity. This is aligned to the understanding of underlying microbiological processes [3]. In the vicinity of station 2, blooms can cause a large change in salinity, while the effects on TN,TP, and TURB are less significant. As blooms are a manifestation of eutrophication in water bodies, large amounts of phytoplankton cause dramatic changes in the total phosphorus and turbidity such as near station 3, with minor influence on temperature and salinity. Around station 5, the bloom had a strong influence on turbidity and salinity, with minor impact on TN and TP due to the deeper water in this area. Despite bloom near station 6 is relatively short, it still caused elevated changes on both salinity and turbidity and in a minor way on water temperature and total phosphorus. In general, the occurrence of blooms had a serious effect on total phosphorus, salinity, and turbidity in the eastern zone of Florida Bay; a worrisome condition because of the highly valuable biodiversity in that area comprising a wide set of sponge, fish and coral species. Our results reveal that algal bloom severity also cause environmental degradation a posteriori beyond the direct causal effect of environmental change (particularly from temperature in the ocean, and nutrients form estuarine efflux) in triggering blooms a priori. Certainly, the primary causal pathway is about temperature leading to CHLa changes; however, the inferred networks is also manifesting the feedback of CHLa change on temperature, and while this can be minor with respect to the first mechanism, it is also possible in relation to algal overgrowth and local temperature increase. This substantiates environmental changes due to ecological imbalance [60], such as the "oceanic positive feedback mechanis" that can lead to further increases in phytoplankton growth, chlorophyll a concentration, and temperature looms are ecological processes that consume energy and yet increase local temperature; precisely algal blooms absorb light from the sun and carbon from the atmosphere which increases the temperature

of surface water. Whether this can be captured by our data or other data is an open question, but what is certainly true is that the bidirectional CHLa-temperature feedback is inferred, as well as CHLa-salinity. Rising temperature, also related to local eutrophication, implies more evaporation from waterbodies and yet higher salinity in case the hydrology is not changed. Of course, if algae are too much (in term of biomass) too much oxygen is depleted when they die and this creates hypoxia and cascading risks such as the death of species and the emergence of toxins. This can also lead to an exceeded capacity of zooplankton to sink carbon to the bottom of the ocean; thus, increase in size and frequency in blooms is not good, also for the generated temperature that is a co-occurring risk factor.

### 3.3.3 Bloom intensity and regional dependence

We explored the interaction dynamics of blooms by analyzing the annual probability distribution, or pdf, of Outgoing Transfer Entropy (OTE, see Fig.3.5) pre-, while- and post-bloom. OTE quantified the extent to which blooms around one area can predict CHLa dynamics (in terms of value and distribution) in other areas: higher OTE values indicate higher area interactions, yet higher spreading and predictability. In 2004, the OTE ranged between 0 and 1.7; most of values with a non-zero probability were between 0 and 0.6. It can be seen that most of the stations have no bloom, resulting in a low probability of large values of OTE, but a high probability of low OTE. The pdf is bimodal with a leptokurtic character. In an ecological sense, the dynamics is characterized by highly localized blooms, and few traces of bloom emergence in other areas. Thus, the bloom spatial network system was relatively contained in 2004 and corresponding to a regular/small-world topology (Fig.3.3). This also corresponded to a simple low-TE dynamics of eco-environmental interactions (Fig. 3.4). In 2005, the OTE range increased to a maximum of 4.0, with most of OTE having higher probability than in 2004. In 2006, the range of OTE increased even further to a maximum of 13, with all OTE values having higher probability. This also corresponded to a shift in the pdf to a more platykurtic; yet, highlighting more widespread and common bloom dynamics. From the perspective of complex networks, the number of nodes with large OTE values increased over time. This indicates that an explosive spread of blooms across FL Bay. Therefore the initial energy dissipation is higher over time. In 2005 the system is in an active and complex state, that makes the management of blooms extremely challenging. The 2006 pdf has higher entropy because more Poisson than all previous years.

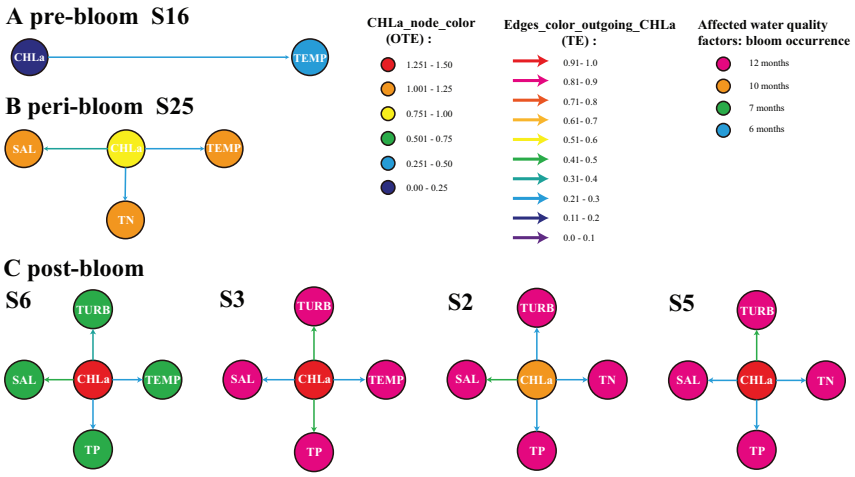


Figure 3.4: **Inferred biogeochemical networks for 2004, 2005 and 2006 pre-, peri- and post-bloom periods in Florida Bay**. The purpose was to quantify local eco-environmental impacts for bloom sources. Yet, only four nodes in 2006, and one node for 2004 and 2005 were considered because those are the most active in terms of CHLa's OTE. However, blooms are spreading phenomena and even other nodes are involved. Stations 16 and 25 are characterized by mangrove habitats in the West region, while stations 2, 3, 5 and 6 (displayed proportionally to a gradient of potential impact of CHLa on the environment) are characterized by coastal marshes and marine flat habitats in the East region of Florida Bay. The color of directed edges is proportional to ranges of mTE for $TE_{CHLa->Env}$ only. The node color for CHLa is proportional to OTE while for other water quality factors depend on the frequency of local blooms during that year (yet, manifesting the potential impact of CHLa on the environment): specifically, blue, green, orange and pink are for 6, 7, 10, and 12 months of bloom occurrence.

The pdf of OTE proves that OTE reflects the probabilistic state of ecosystems with particular reference to algal blooms in this case. The higher the entropy the larger the effect of blooms and the higher the ecological effects; interestingly for FL Bay we notice the the highest the entropy the more scale-free the bloom spreading network although a time-delay may exists between ecological effects (CHLa that is more random like in 2006) and the largest spreading network (that is in 2005) signifying potential long-term effects. By flipping the pdf it is possible to get information on the ecosystem potential landscape informing about energy dissipation, likelihood of shifts and relative stability of bloom conditions (Fig.3.5). The energy dissipation of the system, that is the potential amount of energy consumed by ecological processes, is visualized and $\propto max[p(OTE)] - p(OTE)$, and scales with $\sim 1/p(OTE)$, therefore the higher the leptokurtic character of the pdf the lower the energy dissipation (such as in 2004). The energy potential also gives the number of ecological states (metastable states are identified by the point where the pre- and post-curvature of the energy landscape diverges in sign; those are represented by the balls in Fig. 3.5), the probability of a configuration to be stable (the lower the energy potential with respect to all other states the higher the stability) and the likely shifts among states (proportional to the slope of the energy potential), all of which define the "resilience" of the ecosystem that is rapidity to bounce back to initial states. Higher entropy corresponds to higher energy dissipation in relation to larger and more random OTEs. This implies lower probability of CHLa stable states which are much closer to each other and increasing in number, yet implying higher likelihood of shifts, with larger ecological impacts. For Florida Bay, the energy dissipation is also increasing in average value for pre-, while-, and post-bloom periods indicating a diminishing resilience and loss of complexity of the system; this also highlights the persistent effects of blooms despite their relative short duration.

## 3.4 Conclusions

The study uniquely proposed a model based on optimal Transfer Entropy (TE), as TE differences, to infer bloom spatial dependencies used to pinpoint risk areas and pathways to target monitoring and controls. Blooms show non-trivial spreading patterns manifested by network transitions with different stability that determine their persistence and potential ecological effects. For Florida Bay, we predominantly highlighted the spatial trends and the neglected impacts n water quality of algal

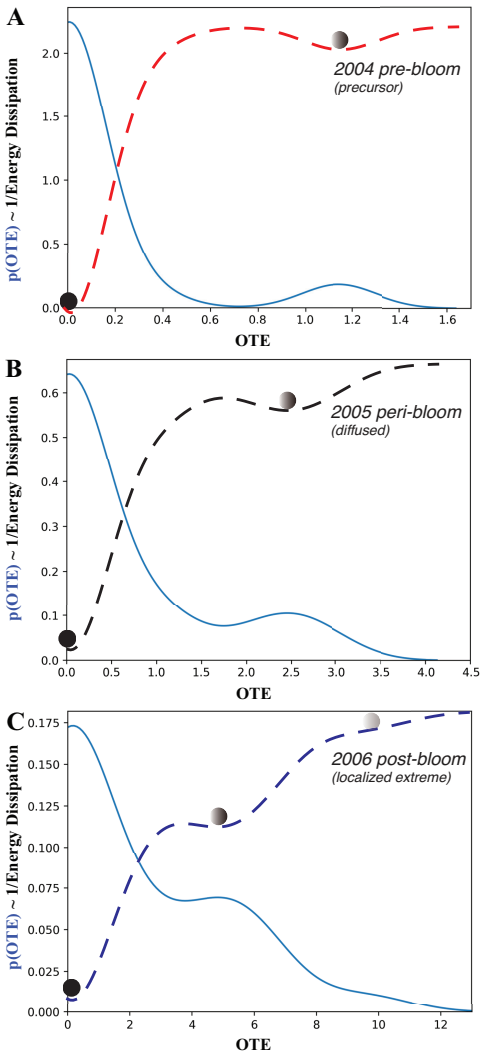


Figure 3.5: **Probability distribution of CHLa's collective influence and ecosystem potential.** A, B, and C are for 2004, 2005 and 2006 pre-, peri- and post-bloom periods in Florida Bay. CHLa's collective influence is assessed based on OTE range and distribution, where the latter defines energy potential (in dashed red, black and blue for the 2004, 2005 and 2006 aligned to the distinct epidemic, transitory and endemic dynamics as in Fig. 3.1B), stability of ecosystem states and transition probabilities from one to another.

blooms. The following specific results are worth mentioning.

- We showed how CHLa patterns carry information of underpinning ecohydrological networks (and associated spreading determinants such as nutrients) that support ecosystem function and services. Salient Pareto interactions are defined via thresholding TE differences with a threshold of causal significance that is set to consider the top 20% TEs (related to the tail of scale-free CHLa probability distribution function), i.e. necessary and sufficient interaction to predict the risk of bloom spreading. More generally, the discovery and inference of the "ecosystem connectome" (as biogeochemical determinant and spreading networks) allows for the assessment of ecosystem health (quantified by the proximity to an optimal condition such as non-bloom state), investigation of causal determinants and their sources, proximity to ecosystem shifts, and targeted ecohydrological controls.

- Through spatial analysis of the bloom spreading networks, we showed how regions not previously involved in blooms (i.e. the highly biodiverse NE tidal-flat habitats with corals and sponge) were caused by large imbalances of CHLa in the western and central blooms that were causally involved. The latter regions are characterized by CHLa that is more randomly distributed and higher probability of CHLa extremes. This probabilistic structure, reflecting the spatial distribution of CHLa, is likely tipping eastern regions to similar bloom endemics. From the perspective of complex networks, this bloom event (2004-2006) evolved from a spatial network with a localized trigger area and a small-world topology to a random topology with long-range spatial diffusion. In 2005, where most stations were blooming, the spatial spreading network was scale-free (theoretically optimal in a purely topological and predictive sense [4, 96]) with a more random biogeochemical network including CHLa (topologically suboptimal), that underpins the dichotomy between structural and functional networks for ecological risks;

- In terms of temporal dynamics, subsequently the first bloom outbreak, persistent and recurring blooms were observed for several NE areas with long-lasting environmental impacts on turbidity and salinity aggravated by temperature increase. Bloom sources were related to central coastal marshes and to a lower extent to mangrove habitats. We further showed that blooms are a recurring and persistent phenomenon over a long period of time with continuous

outbreaks in interdependent regions. This leads to higher energy dissipation and larger instability dictated by the more random distribution of CHLa that is associated to a more uniform network with long-range connectivity regardless of habitats, likely leading to loss of ecological heterogeneity;

- The analysis of biogeochemical factors affecting water quality showed that the occurrence of blooms could only affect small fluctuations of temperature at the beginning of the blooms; however, repeated bloom outbreaks largely affect other biogeochemical factors (such as salinity, turbidity, and CHLa triggering hysteresis or memory effects) that are hardly controllable systemically due to the loss of vegetation and other keynote species. The concentration of CHLa can be influenced by temperature and salinity, and changes in CHLa concentration can, in turn, have indirect effects on water temperature through various ecological processes. In some regions, facilitated by shallow-water habitats, the water temperature increase can stimulate phytoplankton growth and increase the concentration of CHLa. The increased CHLa can, in turn, absorb more sunlight, that can lead to local warming of the water in some cases. In long-term the persistence of blooms, i.e., high CHLa, may also alter nutrient cycling as highlighted also by other studies with the term "oceanic positive feedback mechanism" [62], and our model is able to infer this secondary causal pathway together with the primary one, where temperature change leads to CHLa change and blooms. This underscores that blooms management should start from the source, otherwise blooms environmental impacts will gradually expand and become uncontrollable, affecting also ecosystems stability and resilience, yet settling them into undesired ecological states.

Even though the extent, duration, and spatial arrangement of blooms are influenced by a multitude of factors, CHLa variability (regardless of any triggering factor) still exhibits a considerable level of predictability and control from an ecosystem perspective, encompassing both predictive and ecological engineering models. We have introduced a data-driven inferential model designed for ecological insight, aimed at discerning risk patterns (including sources and pathways), trajectories, and causative factors. Our proposed model, focused on spatial and biogeochemical network inference, furnishes valuable insights for the prediction and management of blooms. For instance, it aids in identifying areas for monitoring and implementing nature-based solutions, such as targeting coastal blue-carbon habitats, to curb the progression of eco-environmental imbalances and their consequences.

Future endeavors will delve into precisely quantifying critical thresholds—whether habitat-specific, climate-specific, or universal—as early warning indicators of environmental factors (including controls) that lead to persistent blooms. This will also encompass accounting for systemic stress, representing the condition of habitats stemming from their ecological history.

# Chapter 4

# 2D Entropic Ecosystem Mandala: Shifts in Chlorophyll-a Dynamics Under Systemic Biogeochemical Stress

## 4.1 Introduction

### 4.1.1 Land-ocean function and phytoplankton as health ecoindicator

Encompassing more than two-thirds of Earth's surface, the ocean functions as a vital carbon sink by absorbing roughly 31% of the atmospheric $CO_2$ emissions. Phytoplankton thriving in the ocean's surface contributes to half of Earth's oxygen through photosynthesis. The ocean has absorbed 90% of the heat generated by global warming. The influence of the ocean resonates across every corner of the planet, with a particularly profound impact on coastal regions harboring the greatest biodiversity and marine primary productivity. In the context of the ocean, productivity primarily pertains to the organic matter generated by phytoplankton—microscopic plants suspended in the ocean. These organisms, most of which are single-celled, play a pivotal role in carbon and oxygen processes. Chlorophyll-a (CHL-a) stands as the quintessential systemic indicator of ecosystem function. This is due to its capacity to infer net primary productivity (NPP), which quantifies the annual car-

bon fixation through photosynthesis per unit area of the Earth's surface. Satellite-derived chlorophyll-a concentration, expressed in milligrams of carbon per square meter per day (mgC/m2/day), underpins the determination of NPP. However, imbalances within CHL-a concentrations can yield inefficiencies in carbon sequestration by the global ocean, carrying long-term climate risks. Conversely, natural and controlled blooms, if viable, play a role in regulating the carbon cycle and ecosystems on a comprehensive scale. Furthermore, the relationship between changes in NPP derived from CHL-a and fishery yield has been extensively demonstrated. Temporal variability in NPP exhibits a significant and positive correlation with increased yields, particularly for fish of higher trophic levels [97].

CHLa also holds significant relevance within coastal habitats, which are ecotones undergoing transitions yet exhibiting high fragility. These habitats play a pivotal role in furnishing numerous ecosystem services. They contribute to functions such as nutrient filtration from land to the deep ocean, carbon cycling, support for fisheries, recreational opportunities, and a diverse array of ecosystem services that collectively contribute to our"ecological security," aligning with various sustainable development goals. In the realm of coastal ecosystems, estuaries assume exceptional importance as transitional habitats situated between land and ocean. The imbalances within these ecosystems yield far-reaching societal consequences, particularly considering that approximately 40% of the global population resides along coastlines. Estuarine ecosystems are characterized by their high dynamism and complexity, attributed to the coexistence of a multitude of diverse habitats within a confined spatial expanse. These habitats are subject to the influence of a plethora of factors, including localized riverine dynamics and diffused oceanic influences. Geographic heterogeneities—such as climate variations and land development—further contribute to the intricate landscape of estuarine ecosystems [46]. The intricate interplay between ecological dynamics and environmental factors is compounded by the effects of climate change, notably alterations in precipitation and temperature, decreases in pH, and rising sea levels. Anthropogenic pressures, including coastal land reclamation and intensive aquaculture, further exacerbate these complex interdependencies.

For these habitats, phytoplankton biomass is an effective indicator to characterized estuarine ecosystem's health, and its variations can reflect the dynamic processes of this ecosystem type [46, 92]. Fluctuations in phytoplankton dynamics are influenced by several factors and influence a variety of important ecological communities such as seagrass, coral, and fish communities [64, 98] that contribute

to ecosystem function, including carbon cycling. Anthropogenic causes of phytoplankton variability is mainly related to ecohydrological adn biogeochemical factors affecting phytoplankton biomass directly and causing blooms [46, 92, 99, 100]. It should be paid attention that erroneous nutrient management may lead unexpected phytoplankton changes. For instance, unbalanced nutrient reduction led the shift of dominant species of red tides (in oligotrophic seas) from diatoms to dinoflagellates, despite coastal nutrient controls for de-eutrophication, related to combined offshore habitat alteration changing the hydrodynamics. Thus, efforts controlling blue-green tides can generated red tides. This was the case of Mirs bay in Southern China [101]. Therefore, while systemic ecosystem management is necessary, it should be kept in mind that it is really hard to revert certain abrupt ecosystem shifts once they happen; however, some ecological shifts may happen gradually so monitoring and countering actions are necessary to avoid persistent blooms before whole ecological communities are compromised.

Furthermore, in conjunction with the occurrence of gradual and persistent blooms, instances of rapid and sudden blooms are increasingly prevalent on a global scale. These rapid-onset blooms pose challenges to the efficacy of bloom monitoring and forecasting capabilities. Yet, there exists no consensus regarding whether these abrupt blooms are emerging as the new norm, given that the prevalence of gradual blooms may have also risen. The emergence of these sudden blooms could potentially be linked to significant hydrological alterations that reduce freshwater inflow within coastal ecosystems. This highlights the pressing need to adapt to the accelerated onset of blooms, especially within the context of a warmer future. The impacts of blooms extend beyond local ecosystems. They play a pivotal role in shaping substantial shifts within the global climate, as alterations in the ocean contribute to large-scale hydrological changes, including the escalation of droughts [102]..

### 4.1.2 Florida bay epitomises complex algal anomalies and ecosystem risks

Florida Bay holds the distinction of being the largest estuarine ecosystem within the state of Florida, encompassing a vast expanse measuring 2,200 $km^2$ [103]. Situated in a subtropical region, it is characterized as a multi-habitat inner shelf lagoon, positioned at the convergence point of the Gulf of Mexico and the Atlantic Ocean [46]. A significant portion of the bay's territory falls under the jurisdiction of the Everglades National Park, and it bears direct influence from the Ever-

glades ecosystem. Partial influence extends to encompass the city of Miami and the Florida Keys. Florida Bay's configuration can be envisioned as a succession of basins demarcated by shallow mud banks. These basins are adorned with seagrass and sponges, interspersed with vegetated islands—including islands abundant with mangroves—painting a diverse ecological tapestry [104]. Figure 4.1 shows a satellite image of the bay. The primary point-sources of terrestrial freshwater to the bay, beyond diffuse precipitation, are Taylor Slough and C-111 canals [103], and saline flux sources are the Gulf of Mexico and Atlantic Ocean [105]. Considering the unique configuration of bay's habitats and anthropogenic alterations of the Everglades in the last 50 years (structurally and functionally as flows), as well as increased evapotranspiration due to warming climate, the ecosystem has experienced high freshwater scarcity with amplification of ocean pressure into phytoplankton. There have been a number of historical analyses of environmental factors in Florida Bay. Some of the best known are, for example, studies on regional clustering of the bay based on water quality characteristics [99], seagrass distribution features [106], and phytoplankton community features [107]. Water management practices within the northern Everglades wetlands during the twentieth century, reduced freshwater flows in Florida Bay by 59% [108]. This exacerbated the ecological impact of salinity due to the severe scarcity of freshwater in the Bay. As a result, from 1987, phytoplankton increased with the death of large seagrass community in Florida Bay [109]. In the 1990s, seagrass communities in central and western Florida Bay declined substantially, due to the widespread and persistent occurrence of microalgal blooms [107]. In such poor water conditions, turbidity in in West and Central Bay also became a critical factor. Diatoms grow faster than cyanobacteria under conditions of greater turbidity [110], although these may not necessarily trigger bloom. However, West Bay, that was dominated by faster-growing diatoms at that time [107], it experienced more severe blooms than Central Bay. Nitrogen and phosphorus are criitical nutrients determining phytoplankton growth [111, 112]. The organic and inorganic fractions of nitrogen and phosphorus have a strong influence on the occurrence of blooms. Nitrogen enhancement in West Bay and phosphorus enhancement in East Bay lead to changes in phytoplankton community composition [113], [46, 100]. Normally, water temperature in Florida Bay ranges from 26-27°C, but over time it continued to rise coupled with more frequent onsets of blooms. Studies have clearly shown a positive causal relationship between water temperature and phytoplankton biomass growth under bloom conditions [46]. Complex interactions and proportions among all these environmental

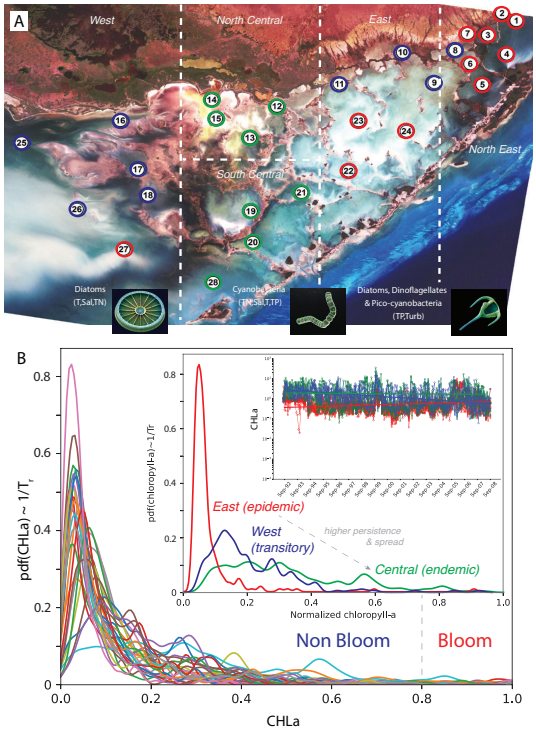


Figure 4.1: **Florida Bay and Chlorophyl-a Distribution**. (A) Satellite imagery of Florida Bay (from NASA Landsat 7, October 2000) and FIU Water Quality monitoring stations (`http://serc.fiu.edu/wqmnetwork/SFWMD-CD/index.htm`). Green and red stations are characterized as endemic and epidemic bloom dynamics with Poisson/uniform and power-law distribution of CHL-a, as extreme pdfs; blue stations are characterized by exponential or Poisson or gamma distributions and are classified as transitory dynamics. Erratic behavior of CHL-a may result into carbon emission and biodiversity loss, thus particularly for endemic areas. For epidemic dynamics any generalized extreme value distribution can be possible. Epitomic stations are 6 and 7 for East, 14 and 20-21 for Central, and 25-26 for West. (B) Probability distribution function (pdf) of normalized CHL-a separating top 20% values (identifying blooms) from the rest (corresponding to the 80-20 Pareto principle that considers the top 20% CHL-a events in magnitude; note that makes-up less than 20% of blooms). Pdf(CHL-a) is inversely proportional to CHL-a extreme return period ($T_r$) that corresponds to extreme events with low exceedance probability; however, the magnitude of CHL-a is very different for different stations. Images of diatoms, cyanobacteria and dinoflagellates are used under license from `Shutterstock.com`.

determinants exist to trigger blooms. However, previous studies between blooms and water quality have mainly involved simple principal component [99, 100] and empirical orthogonal function analyses [63]. In recent years, quantile regression methods [46, 114] have been widely used to analyze spatial and temporal trends in phytoplankton biomass, but not for detecting differences in the relative importance of water quality factors for bloom and non-bloom states. These issues apply to all estuaries worldwide, so they represent a grand-challenge for intensifying blooms and more broadly for ecological security affecting local ecosystems and global climate via teleconnections on eco-hydro-carbon cycles.

### 4.1.3 Impacts of multiscale blooms and risk forecasting model

Around 20% of the net primary production of phytoplankton, amounting to 5–10 Gt C, is channeled into the deep ocean via the biological pump. The nature and scale of carbon transported to the deep ocean are influenced by the size structure of phytoplankton communities, a factor potentially impacted by changing blooms. A study by Irion et al. [115] unveiled intriguing scaling relationships between daily carbon uptake rates of individual cells and their volumes, particularly in small diatoms and nonsilicified cells. These findings can be extrapolated to the scale of algal blooms, offering insights into the volume of carbon export within the ocean. This scaling holds substantial value for elevating micro-level phytoplankton information to macro-level functions. This study adopts a similar approach, albeit leveraging nonlinear biogeochemical networks, to extend local CHL-a dynamics to the ecosystem scale. This methodology anticipates the spatial diffusion of blooms without the inclusion of spreading corridors, diverging from the approach employed by Wang et al. [41]. The collective efforts outlined are instrumental in quantifying the systemic impact of ”flesh blooms” (sudden onset or intensification of blooms) across micro to macro scales. This comprehensive approach is crucial for assessing ecosystem health—a convergence toward optimal function—and establishing a baseline for ongoing tracking.

The primary innovation of this study lies in the investigation of causal relationships between water quality variables—namely TN, TP, SAL, TEMP, TURB—and chlorophyll a (CHL-a), which functions as a pivotal ecological indicator of phytoplankton alterations. These relationships collectively form biogeochemical networks and are harnessed for the dual purposes of forecasting and formulating sys-

temic risks. The overarching objective was to establish robust and salient ecological indicators of shifts within ecosystems and to comprehend how these indicators can offer insights into forthcoming trajectories and eco-environmental determinants, some of which can be managed. Extensive literature attests to CHL-a's efficacy as an indicator of cyanobacteria-driven blue-green algal blooms, and to some extent, dinoflagellate blooms responsible for red or brown blooms. Cyanobacteria are primarily associated with freshwater inflows, while dinoflagellates thrive in marine environments. Despite being less prevalent in diatoms, CHL-a was found to serve as a reliable predictor of diatom concentration in water, which can precede the occurrence of blooms [116]. Preliminary evidence hints at a connection between changes in cyanobacteria and alterations in the diatom/dinoflagellate ratio. This ratio, in turn, can be adopted as a holistic indicator of ecosystem health and the potential correlation of blooms with either freshwater or marine-specific stressors. Specifically, exceptionally low values in the diatom/dinoflagellate ratio signal a concerning ecosystem state, indicating heightened ocean acidity. This is alarming since dinoflagellates have the capacity to produce neurotoxins that can detrimentally impact both species and human health. Moreover, diatom species exhibit lower carbon-to-CHL-a ratios and higher photosynthetic rates per unit of carbon compared to dinoflagellates [115, 117]. While diatoms significantly contribute to marine food webs, global oxygen production, and carbon sequestration, it's crucial to maintain a balanced composition of species, unlike an excessive abundance of dinoflagellates.

In light of the above issues, the specific objective of this study was to design and quantify an ecosystemic risk function considering the dynamics of CHL-a information in terms of magnitude, persistence and extreme shifts. Specific tasks were to: (1) objectively identify a systemic threshold that separates bloom from non-bloom conditions; (2) construct a causal biogeochemical network from an inference and forecasting model that reveal the relative importance of water quality factors for local bloom emergence and their duration as forecasts based on the non-linear biogeochemical relationships; and (3), assess CHL-a shifts and persistence of blooms, and cluster habitat- and region-specific bloom risk. Eco-environmental feedback and their relative dynamic force was predicted for blooms and non-blooms conditions with a novel causal network-inference model coupled to a machine-learning model (TEGNN); TEGNN was applied in diverse ecological regions of Florida Bay to test habitat-specificity of forecasting ability. This pattern-oriented model differs from traditional complex networks, that is representative of simple probabilistic

graphical links which usually requires the assumption of the specific form of relationships between variables; whereas our approach is based on entropy and the difference in transfer entropy [41], which more intuitively predicts the dynamics of the complex system without the need to assume variable relationships, added to a forecasting model and a risk model. Despite the application to blooms, TEGNN is broadly promoted for its use in ecosystem science and engineering for inference, forecasting and strategic management of spreading phenomena.

## 4.2 Methods and Materials

### 4.2.1 Eco-environmental data

The dataset is still provided by the Florida International University Southeast Environmental Research Center (FIU SERC), i.e. six sets of water quality monthly data monitoring by 28 water quality stations from August 1992 to September 2008 [45, 46]. CHL-a has often been used as an indicator of blooms given its sensitivity to environmental changes, ease of monitoring, and ability to reflect phytoplankton biomass effectively [47], but has not yet been verified as a systemic indicator of ecosystem health related to ecosystem function. We used 80/20 principle to establish an average threshold of $\geq 1.841 \mu g L^{-1}$ on CHL-a, universally applied to all stations, to distinguish bloom from non-bloom states across all stations. A timeline of CHL-a bloom and non-bloom periods was used to define the corresponding states for the remaining water quality factors (i.e. TN, TP, SAL, TEMP, and TURB).

As for reference, Fig. 4.2 shows the composite seven-day mean of CHLa concentration (in mg/m$^3$) in the surface ocean layer of Florida Bay, extracted from NASA MODIS data. Fig. S1 shows the same for the larger South-West Florida region and the Gulf of Mexico to emphasize how algal blooms are much larger phenomena and FL Bay blooms may have some influence from Norther Everglades regions in the Gulf, such as areas nearby Tampa Bay. Northern. blooms may spread south to mangrove habitats of Florida bay and alter western phytoplankton. In this study, in conjunction with these analyses of spatial environmental factors, the bay was dissected into three macro-habitat regions, namely the east, central and west bay (Fig.4.1) reflecting mangrove, salt-marsh and seagrass flat habitats. This was for a better habitat-specific characterization of the ecosystem risk in terms of forecasting and attribution to environmental determinants. The main phytoplank-

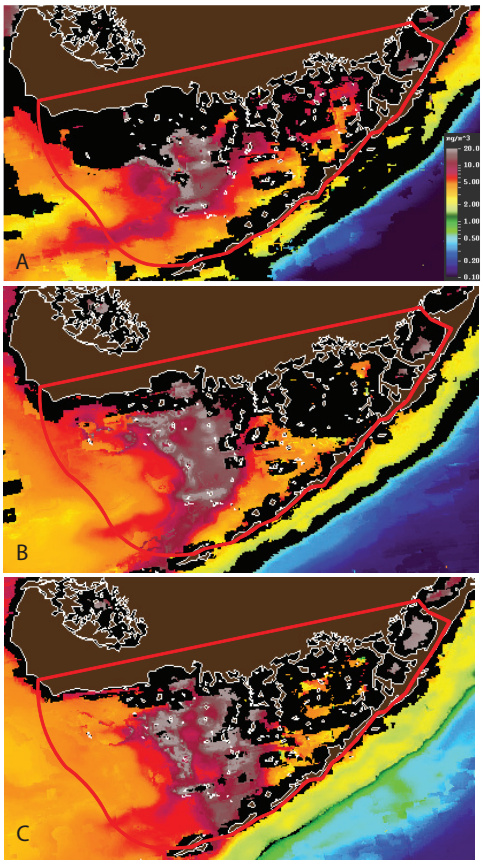


Figure 4.2: **Composite seven-day mean of CHL-a concentration (in mg/m$^3$) in the surface ocean layer of Florida Bay.** CHL-a, for 7-day mean of October-November-December 2005 (A, B, C), is estimated by the Optical Oceanography Laboratory at USF (`https://optics.marine.usf.edu/cgi-bin/optics_data?roi=FLKEYS&Date=10/15/2005#C20052822005288.QKM.FLKEYS.7DAY.L3D.SST.png`). Estimates, in $\mu g/L = mg/m^3$ are done with the most updated calibration and algorithms in the SeaDAS processing software (SeaDAS is a comprehensive software package for the processing, display, analysis, and quality control of ocean color data, `https://seadas.gsfc.nasa.gov/`) based on multiple reflectance indicators for deep [1] and shallow waters [2]. Over clear, shallow waters (<30 m) or over very turbid coastal waters or river plumes, it is often overestimated as other components (colored dissolved organic matter, suspended sediments, ocean bottom) interfere with the algorithm. See Fig. S1 for CHL-a of the whole South-West FL region.

ton community in the western bay is composed by diatoms, the central area of the bay is dominated by a unicellular picoplanktonic cyanobacterium (*Synechococcus spp.*), and in the eastern region by a mixed community of dinoflagellates, cyanobacteria and diatoms. *Synechococcus spp*, has a significant impact in the North Central Bay, including causing changes in salinity [118–120], nitrogen and phosphorus [111, 112], and other water quality factors [121]. For this reason and characterize even more ecological effects, the central region of the bay was further divided into South-central and North-central regions. Due to the complexity of community composition throughout the whole eastern region, also considering the C-111 canal in that area, the eastern and north-eastern regions (with low and high environmental stress, respectively) were also divided.

### 4.2.2 Thresholds for phytoplankton blooms

The transition from "non-bloom" to bloom dynamics can be assessed from the probability distribution function or its exceedance distribution function of the CHL-a. Matteo C. et al. [3] showed how at the transition point there is the largest divergence between probabilities defining the critical CHL-a threshold at which blooms start to spread. [41] focused on how spatial spreading networks are associated with different bloom states, while in this paper we focused on of the whole set of biogeochemical interactions useful for forecasting bloom emergence and how systemic environmental triggers are changing across habitat types. The first task, however, is to classify what blooms are from CHL-a data; in other words, what CHL-a values constitute a bloom. Previous methods [47], considered as threshold for the bloom emergence the CHL-a value corresponding to the $75^th$ CHL-a percentile of a reference system designated as the level at which CHL-a values departed from baseline condition; this was applied to the whole Florida Bay as one threshold. Natalie G Nelson et al. [46] instead considered the empirical cumulative distribution function (ECDF) to quantify the $95\%$ confidence interval of the critical CHL-a threshold; this value was set as station-specific. Natalie G Nelson et al. [46] assumed that the threshold was consistent with the initiation of blooming, whereby the majority of threshold values fell below $2\mu gL^{-1}$. However, the method of Natalie G Nelson et al. [46] is not suitable for evaluating data collected in stations with significantly different skewed distributions, and the thresholds they obtained are still biased estimates in a probabilistic sense. Therefore, taking into account the above considera-

tions, for the design of a systemic ecological criteria of water blooms we considered the normalized probability distribution function for CHL-a and its maximum divergence as *probabilistic tipping point* (Figs. 4.1 and Fig. 4.3) manifesting collective (distribution) changes in CHL-a, where CHL-a is the salient eco-indicator. This criteria largely corresponds to the generic Pareto principle characterizing bursting phenomena where outbreaks (corresponding to the largest 20% CHL-a values) describe 80% of the CHL-a variability. Thus, with this criteria the largest 20% of CHL-a values characterize bloom dynamics for the whole Florida Bay (regardless of the station considered). This Pareto criteria also identifies salient Pareto links in terms of eco-environmental feedback and spatial interactions [41].

### 4.2.3 Informational risk profiling

Risk of algal blooms (unconditional to any impact) can be defined as the product of bloom magnitude M (as CHL-a range, i.e. maximum minus average CHL-a), persistence P (i.e., H(CHL-a), or the frequency of occurrence in a deterministic sense), and shifts S (defined by the Kullback–Leibler divergence $KL(CHL-a_t, CHL-a_{t+\Delta t})$ between the pdfs of CHL-a in two subsequent periods, i.e. the time up to a point and the new distribution with new events. Thus, the impact-independent risk is analytically defined as:

$$R(CHL-a) = M \times P \times S = \\ (CHL-a_{max} - \langle CHL-a \rangle) \times H(CHL-a) \times TE(CHL-a_t, CHL-a_{t+\Delta t}) \tag{4.1}$$

The shift S is an increasing variable between a non-random and random distribution, manifesting an increasing risk due to loss of predictability (or increased loss of information or permutation entropy [122, 123], complexity of a time series intended as chaotic or functional networks) that increases RMSE and H(CHL-a). Alternatively, to quantify S, TE between two CHL-a of two time-periods can be used when accounting for time delays. The risk can also be constructed considering spatial connections C to risky areas; however, spatial connection were not considered in this paper but in [41]. The ranking of community risk can identify risk sources and sinks, clusters of community at risks (here divided into three classes), and feature/value thresholds for M, P, and S accounting for all facets of ecological

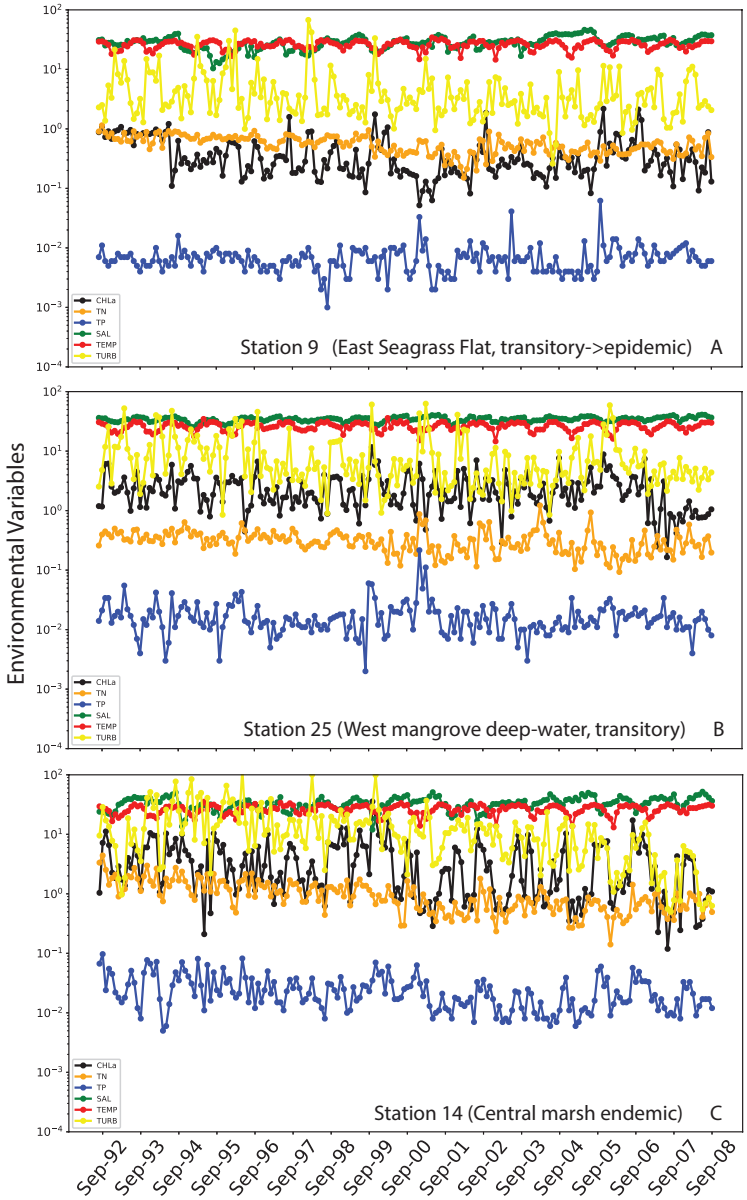


Figure 4.3: **Biogeochemical Time Series.** Biogeochemical variables are shown as normalized values over time for representative habitats with different CHL-a dynamics. The log-value was chose to appreciate the mutual variability together with CHL-a.

imbalance (i.e., intensity of change, uncertainty and uncertainty structure, leaving aside connections to other areas). Eq. 4.1 is then used to classify three classes of risk (low, medium and high) based on the predictability, ecological organization and magnitude of CHL-a. CHL-a was forecasted by TEGNN as described below.

### 4.2.4 Ecosystem causality inference model

Causality networks help to synthesize the structure of complex ecosystems, including the nexus between ecosystem structure, function, and services where the latter are based on desirable threshold over functions. This section provides a detailed analysis of a complex network model that takes infer causal relationships between variables. A schematic of the causality network model is illustrated in Fig. 4.4. The entropy of ecosystems, manifesting the ecological disorganization in relation to CHL-a variability, is dependent on the probability distribution functions (pdfs) that affect TE calculated on variable pdfs' divergence and asynchronicity. TE variability of an area, or the whole system, can be decomposed into eco-environmental interactions (considering CHL-a and environmental factors acting as determinants or effect of ecological imbalance) and ecological areal interactions underpinning bloom spread; the latter was studied by [41] for FL Bay. This variability affects the organization propagation of CHL-a (i.e., how randomly distributed CHL-a is) and in a information-balance equation can be written as the spatio-temporal convolution of the aforementioned components composing the ecosystem connectome, i.e.: Eq. 3.1. Equation 3.1 is focused CHL-a patterns where networks are the backbone determinants of the ecological "weave" (CHL-a interconnected patterns) that can be potentially controlled. Space and time are the dimensions along which CHL-a is considered, plus other dimensions along gradients of environmental features on which stress-response patterns (environment-bloom change) and related features (e.g. early-warning signals and risk thresholds) can be derived. Networks help to define sources, sinks, pathways and environmental determinants to guide monitoring and control for bloom prevention. In this section, we specifically analyze the functional ecological connections determining bloom initiation and forecasting used as ecological reservoir computing (first term in Eq. 3.1, where for the second term was addressed in Wang H. et al. [41]). It should be noted that in Eq. 3.1 we refer to all TEs, yet $H(CHL - a)$ is dependent on the distribution of all interactions and not just the average value. According to the Schef-

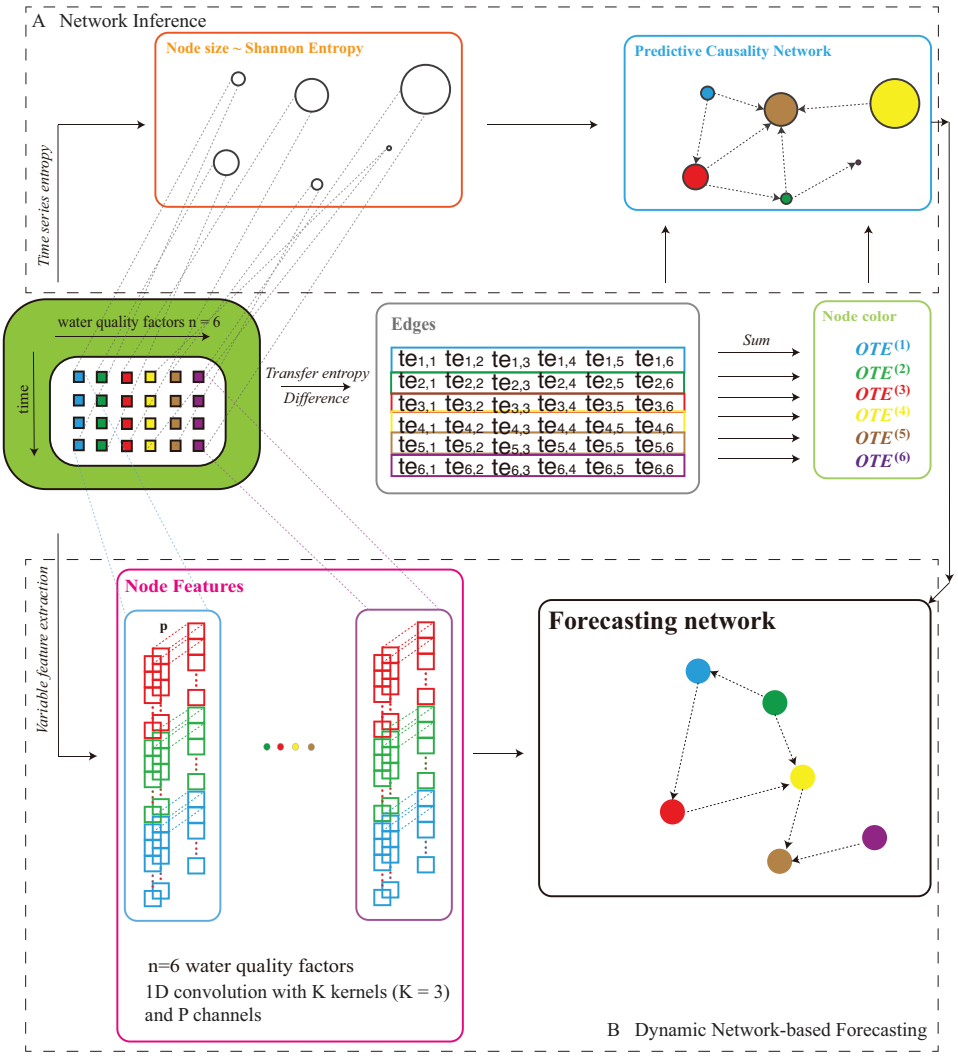


Figure 4.4: **Network TE Inference and CNN Forecasting Model**. A and B, at the TE network inference and CNN forecasting of TEGNN given time-series information of biogeochemical variable. TE allows for network discovery since relationships between micro phytoplankton and macro environmental features are not widely known. The dotted arrows in both networks indicate that they are directed networks, where the specific direction depends on TEs as in Eq. 2.4. Despite these TE relationships are bidirectional we focus, via TE differences, on predominant direct interactions for ecological predictability that is important to evaluate ecosystems (such as bloom risk profiling) as well as for optimizing monitoring.

fer et al. [124] and Hirota et al. [125] the energy potential defining ecosystems states and energy dissipation can be derived from the probability of CHL-a, i.e. $U = -\sigma^2/2$ pdf$(CHL-a)$ log pdf$(CHL-a) = \sigma^2/2\ H(CHL-a)$; this establishes a clear connection between increasing entropy, energy dissipation of a system with potential instability, and lower predictability (proportional to the average value of TE).

To predict CHL-a values over time in Florida Bay, based on the TE-inferred biogeochemical network we used TEGNN [50]. We used time series features extracted by Convolutional Neural Networks (CNN) instead of all variables as a whole, as node expressions to achieve the forecasting accuracy maximization objective of CHL-a that identifies bloom regimes in Florida Bay (Fig.4.4). The first component of TEGNN is related to the inference of the optimal TE network (with characterization of node uncertainty and uncertainty propagation based on TE, where the latter is a form of statistical complexity related to uncertainty structure; Fig. 4.4A), and the second component is about the feature-based extraction of variables' feature for the CNN forecasting using the TE backbone network structure. The algorithms for node feature extraction, node embedding and model training are the same as Section 2.2.4.

### 4.2.5 Assessment of the predictive capacity of the model

The Root Mean Squared Error (RMSE) was used to quantify the accuracy of the forecast for the whole period considered, like:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(O_i - \hat{O}_i)^2}\ , \tag{4.2}$$

where $\hat{O}_i$ denotes forecasts and $O_i$ is the observations, and $m$ is the number of observations. The model was calibrated and validated in a way to minimize the systemic RMSE across the whole forecasted period.

## 4.3 Results and Discussion

### 4.3.1 Chlorophyll-a as an ecosystem indicator

CHL-a was confirmed as the key ecosystem health indicator, because: (i) it is an indicator of systemic phytoplankton community biomass, and its concentration reflects the comprehensive impact of systemic biogeochemical stress determining water quality, thus underpinning the fine balance between ecology and the environment with cascading risks for species at multiple scales; (ii) it is highly sensitive of ecological spatial habitat features and ecological history defining phytoplankton spread and persistence (yet, compounding risks); (3) it is easy to monitor both considering discrete water sampling and inference from remote sensing (i.e., risk sensing). Therefore, CHL-a is the leading and most suitable ecological indicator for of aquatic ecosystems like Florida Bay, and in general it works really well for characterizing phytoplankton imbalance, potentially leading to ecological catastrophes, in coastal and marine ecosystems. Fig. 4.1 shows its probabilistic characterization and how this defines distinct blooming dynamics with diverse spreading behavior and environmental determinants' organization.

### 4.3.2 Bloom and non-bloom regime threshold

In Fig. 4.4 we identified the normalized probability distribution of all CHL-a throughout the whole Florida Bay as a power-law distribution (i.e., "sharp peak and long tail"). Thus, according to the Pareto rule (80% of CHL-a values depends on the largest 20% CHL-a), we selected the largest 20% CHL-a value as the blooming condition for the whole Florida Bay. The application of the 80/20 rule identifies only the extreme peaks that for sure are blooming; therefore, disregarding peaks that are only relevant locally without any bloom spreading. With this Pareto criteria we obtained the threshold $1.84 \mu g L^{-1}$ for dividing bloom and non-bloom conditions throughout Florida Bay. This criteria is much more systemic than the one of Nelson et al. [46] that provides a site-dependent threshold ranging from 0.86 to 7.70 $\mu g L^{-1}$ without considering how much one area is the byproduct of other spreading blooms assessable via normalized pdfs' divergence. Blue-green-red colors for nodes in Fig. 4.1 are proportional to the Shannon Entropy of each station reflecting the disorganization of CHL-a. Small-World and Scale-Free organization of Non-Bloom and Bloom CHL-a networks underpin habitat networks: both environmental and eco-

geomorphological gradients (water flow and quality as well as vegetation-habitat configuration) determine distribution and intensity of CHL-a over space and time. Non-Blooms and Bloom networks are determined via time-delayed CHL-a probabilistic changes (values, probability and their divergence) as Transfer Entropy via the Optimal Information Flow model [69].

### 4.3.3 Biogeochemical fluctuations in algal blooms

Fig. 4.3 shows the fluctuations of biogeochemical factors over time for epitomic habitats. For East seagrass habitats (A) CHL-a is increasing with amplified fluctuations as a sign of shifts, mostly related to increase in TP, temperature and salinity. For West mangrove habitats (B) CHL-a is mildly decreasing, however with large fluctuations associated to TN and turbidity increase in magnitude and fluctuations. Central marsh habitats (C) are showing a decreasing CHL-a associated with a decreased TP, TN and turbidity; however, their distribution is pseudo-uniform with high level of blood endemism in contrast to Eastern habitats that are endemic. Across all habitats, it is noticeable the large seasonality of temperature and salinity (defining CHL-a seasonal component) and the peak-dynamics (defining sharp CHL-a peaks/blooms) of Turbidity, TN and TP (in order of fluctuation magnitude) where TN peaks are the smallest and largest for Eastern and Central habitats. Yet, Central marsh habitats are characterized by the most randomly organized biogeochemical load affecting CHL-a in multiple and less predictable ways. Considering epidemic and endemic bloom classification it is possible to define how environmental features changed irrespectively of the habitat. Epidemic areas were shown to experience a large increase in TP, temperature and salinity; endemic areas experienced a decrease in nutrients and turbidity but net increase in temperature and salinity; instead, transitory deep-water areas were found fairly constant in their environmental features. All habitats are on average experiencing increasing variance in all eco-environmental indicators, that may suggest a global instability of the bay ecosystem. For small blooms (mostly epidemic) the most salient factors were salinity, phosphorous and turbidity, while for large blooms (mostly endemic) the top factors were turbidity, nitrogen and temperature; yet large-blooms vs. flash blooms are much more affected by land efflux. This result emphasizes how systemic factor variability, played out by hydrodynamical, biogeochemical and global ocean change factors (reflected by turbidity, nitrogen and temperature, respectively) de-

termine large blooms as previously shown by published literature independently of the geographical area considered.

### 4.3.4 Multi-factor predicting for bloom dynamics

The left plots of Fig. 4.5 show data and TEGNN forecasts of CHL-a that is the best bioindicator of algal bloom emergence and spread. Bloom dynamics is shown for three different magnitude classes (from small to large) based on average and maximum CHL-a (defining the magnitude M) as well as changing considering CHL-a pdf (i.e., $\sim$ power-law, gamma/Poisson or exponential, and uniform for small, transitory and large blooms). Right plots show the ranking of environmental predictors (based on Outgoing Transfer Entropy) for the CHL-a collective dynamics of the envirome represented by the functional networks (Fig. 4.6). Functional networks can be thought as the ecosystem sensing networks related to the spread of CHL-a and its sensitivity on environmental factors. Fig. S2, and S4 shows temporal forecasts and underpinning biogeochemical networks for key West, Central and East stations with different degree of endemicity and risk. Leaving aside the memory/self-predictability of historical CHL-a (i.e., one month earlier in this study) for small blooms the most salient factors are nitrogen, salinity, phosphorous and turbidity (station 9 in shallow reef habitats); while for large blooms are turbidity, nitrogen and temperature (station 14 that is a coastal habitat nearby extended drainage efflux) in order of importance, despite turbidity and TN are decrease in value but increasing in fluctuations (yet manifesting "critical slowing down" phenomena with potential ecological shifts). Intermediate bloom dynamics (transitory in deep-water habitats epitomized by station 25) is characterized by a combination of land and ocean factors but less about temperature fluctuations. These results emphasize how systemic factor variability, that are hydrodynamical, biogeochemical and global ocean change factors (reflected by turbidity, nitrogen and temperature as well as salinity, respectively) determine large blooms. In particular small-bloom eastern dynamics is regulated by ocean factors (salinity-dominated, phosphorous time-point loads, and lower ecological memory that means higher dependency on previous CHL-a) while large blooms are regulated by estuarine efflux factors (slow turbidity- and nitrogen-dominated with higher ecological memory) as observed for blooms worldwide. It should be noted, however, how size and fluctuations of CHL-a peaks are increasing for small-bloom dynamics habitats possibly associated to TP increase and shifts of adjacent central habitat blooms (Fig. 4.5A). Ocean temper-

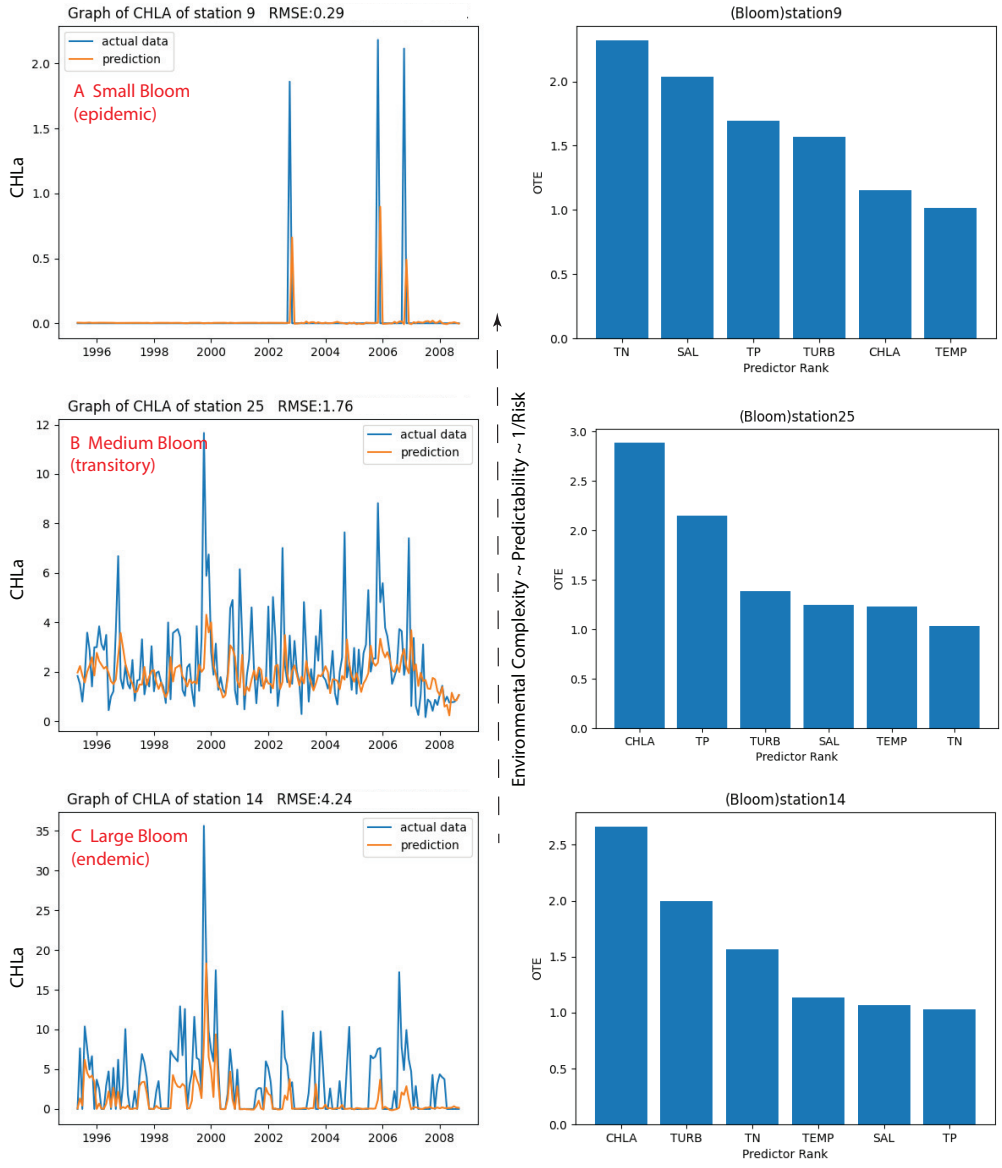


Figure 4.5: **Bloom Forecasting and Causal Attribution to Eco-environmental Factors.** CHL-a predictability decreases with envirome randomness that is the opposite of organized complexity (from small to large bloom dynamics where CHL-a is more affected by multiple factors directly) where the former is evaluated on each single CHL-a peak via RMSE (left plots). The top eco-environmental predictors are identified by the Outgoing Transfer Entropy (OTE) Jie2019 that is quantifying how much one environmental predictor affect all others (through the envirome as a network defined by TE and CNN, yet considering dynamical features).

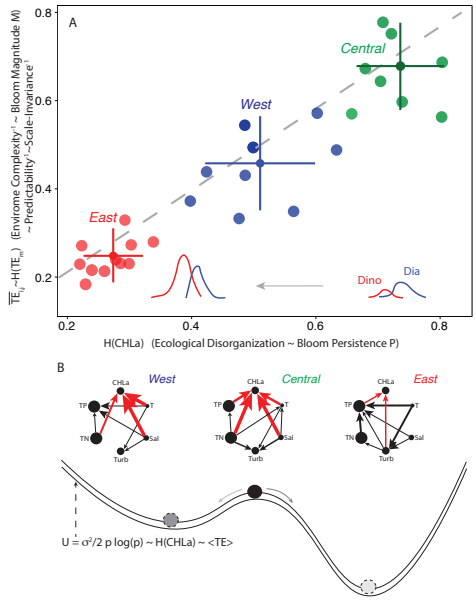


Figure 4.6: **Envirome Entropy, Chlorophyll Randomness, and Bloom Instability.** Entropic mandala to define potential healthy and diseased bloom conditions considering structural complexity of the envirome and bloom randomness. A. The ecosystemic bloom risk is proportional to the envirome (black network) or the biogeochemical network (black and red link network) randomness convoluted to ecological effects (i.e. CHL-a in this case), where CHL-a is affected by many environmental factors directly with a certain time delay (the latter are decreasing while approaching the peak of blooms). Colored points in A are for each station in each area. B. High energy (dissipation) potential, based on TE distribution, is associated to more random envirome (or pseudo scale-free biogeochemical networks where CHL-a is the affected hub) and high CHL-a entropy (loss of power-law distribution of CHL-a with larger and more persistent extremes determining blooms). The position of balls in the energy dissipation landscape shows the magnitude of the dissipation, proportional to H(CHL-a), and the darker the color, the higher the instability of the ecological bloom state; the color of arrows in the energy landscape is proportional to the likelihood of shifts defining risks (the shift risk is higher for the steepest gradients, yet more vulnerable areas are East areas).

ature seems to affect blooms in terms of large scale temporal trends considering temperature regular seasonality. The predicted value of these environmental factors (based on the network in which all factors are predicted simultaneously) is shown in Fig. 4.7 over Florida Bay (Fig. S3 for the second largest bloom in 2005); these collective dependencies affecting predictions define the potential energy dissipation of the system (Fig. 4.6) in term of phytoplankton biomass with cascading ecological consequence. The randomness of ecological processes (CHL-a) has deep impact on ecological computation intended both as computational forecasting, likely species cross-sensing and species-environment sensing. From top to bottom in Fig. 4.5, the decrease in statistical complexity (from a time series that is clearly associated to a scale-invariant distribution, very far from a uniform distribution where the modes of variability are many more despite being "simpler" statistically speaking) lead to a decrease in both predictability and emergence of scale-free blooms. Blooms disrupt the positive eco-environmental feedback created by the interplay of habitat features and microorganisms; a finding that is evident from the iferred network changes. Analogous studies showed how information-dissipation occurs much faster while epidemic outbreaks are approached (and yet the information-dissipation length increases [126] with an associated decrease in the characteristic time delay between events and a decrease in forecasting accuracy, that is a sign of "critical slowing down"; this coincides with increased interactions between ecosystems elements (environmental factors or areas) that precede critical transitions. This rapid information decay implies that previous ecosystem states are less and less informative the more algal bloom outbreaks are approached and become persistent. In general, we show how the consideration of functional eco-environmental networks increases bloom forecasting and this network-based ecological reservoir [127] is more tangible during non-bloom periods and habitats where networks are more organized.

As evident in Fig. 4.7 (and Fig. S3 for the second largest bloom in 2005), non-bloom CHL-a is much more homogeneous across areas. Bloom CHL-a is more heterogeneous and spreading from station 14 predominantly (mudbank habitat) to elsewhere (e.g., to station 9 that is a hard-bottom seagrass habitat) with preferential pathways as reported by Wang H. et al. [41]. The spatio-temporal dynamics relates to to the predictive importance, i.e. likely causality, of each factor for explaining CHL-a dynamics even without the inclusion of time delays. It is visible how 1999 is the largest bloom for the North Central Bay, and 2005 for the North-East Bay (Fig. S3). Synchrony among environmental factors and CHL-a is more important

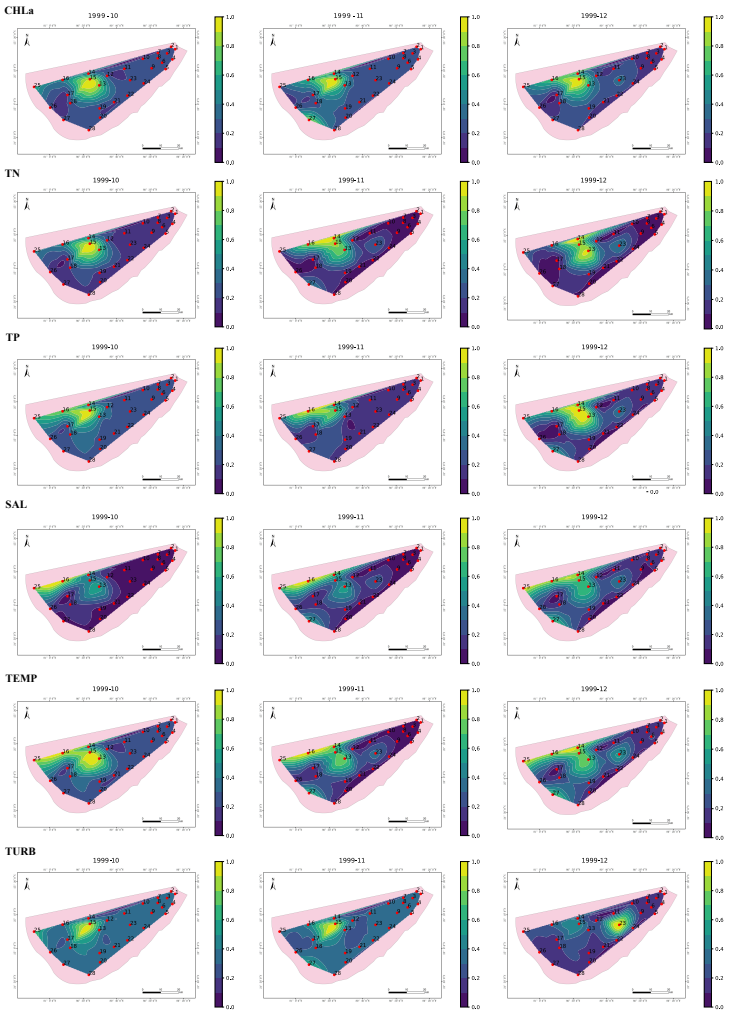


Figure 4.7: **Spatial Biogeochemical Patterns for the Largest FL-Bay Bloom.** Patterns for the 1999 bloom that is the largest and most widespread bloom up to date. Turbidity and phosphorous are much more spiky and coinciding with high CHL-a areas, whereas nitrogen have more contained fluctuations except for station 25 (deep-water sandy-bottom habitat proximal to mangrove coast). Salinity and temperature are much more seasonal and distributed homogeneously across habitat types, and divergent from CHL-a spatial peaks. This characterization holds both for spatial and temporal variability of environmental factors. Synchrony and asynchrony of eco-env factors over space-time (space overlap is approximatively the same of the overlap in time series) is a major component defining interactions of stations and factors leading to emergent blooms.

than divergence in predicting CHL-a extremes, as much as when considering station spatial networks [41]. This is because synchrony is more important for effective interaction (TE) much more than divergence of pdfs (Eq. 3.2).

### 4.3.5 Entropy-complexity mandala: bloom magnitude-persistence synthesis

Fig. 4.6 shows the mandala of potential risk, or vulnerability, accounting for systemic environmental pressure disorganization as entropy of biogeochemical TEs, and the ecological imbalance as entropy of CHL-a. It is rather clear the correspondence between H(TE) and H(CHL-a) which underpins a strong causality (considering distributions and their co-evolution, beyond values) between biogeochemical stress and CHL-a response. H(TE) is for TE among environmental determinants, and H(CHL-a) represent the entropy of the biogeochemical interactions and CHL-a which are proportional to the average bloom emergence (or magnitude $\propto CHL - a_{max} - \angle CHL - a\rangle$) and persistence P of algal blooms (see Fig. 4.8 for the characterization of risk). The higher H(CHL-a) the higher the persistence with larger average CHL-a value but not necessarily larger extremes, and lower predictability of fluctuations (related to the distribution of TEs, associated to low emergence of scale-invariant patterns, and yet low environmental and statistical complexity due to proximity to uniform distributions, or departure from non-linear distributions). H(TE) is then a measure of statistical complexity (where complexity is lower for more uniform distribution of TE) defining the structure of networks leading to the emergence of blooms in an area. The scale-free geometry (solely based on connections) is associated to the increase in direct feedback between environmental factors and CHL-a, leading to high and more uniformly distributed TEs that cause higher and more homogeneously distributed CHL-a. Energy dissipation is the highest for the eco-environmental network that is structurally more scale-free (CHL-a becomes the hub), but functionally (in terms of TE) more random, particularly the envirome defined as the set of interdependent water quality variables. CHL-a in the East Bay is more power-law distributed with the smallest average and large extremes; in the Central Bay, CHL-a is more uniform with the highest average and variance (extremes with much higher probability); while in the West Bay, CHL-a (exponentially distributed) has intermediate average and variance between East and West. Fig. 4.6 however, does not account for shifts in CHL-a to define the

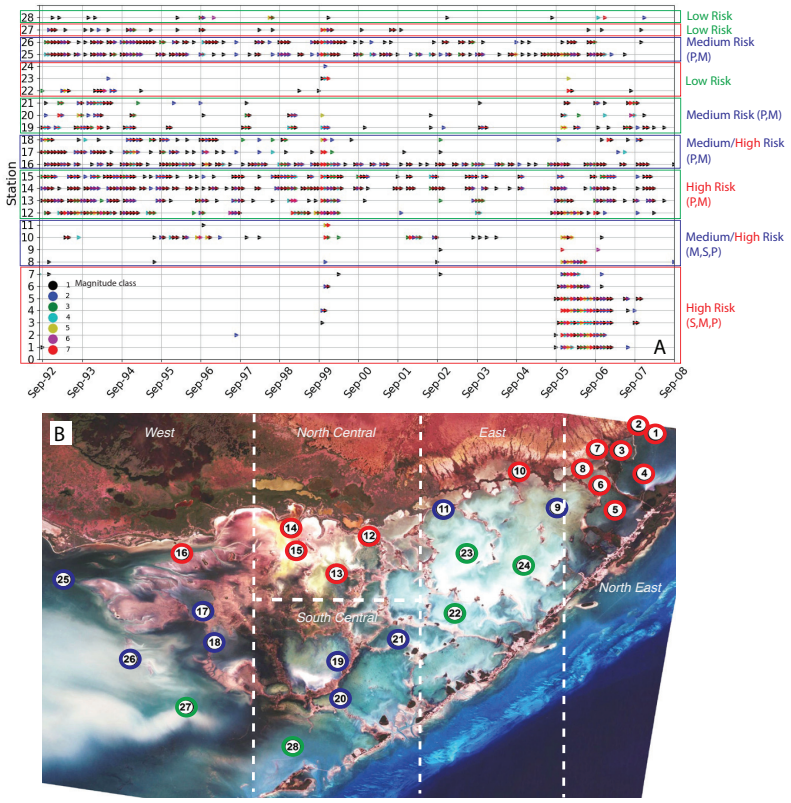


Figure 4.8: **Riskgram: risk profile of bay communities based on algal bloom Persistence, Extreme Magnitude and Shift.** Risk of algal blooms (unconditional to any impact) is defined as the product of bloom magnitude M (considering extreme value and not only average CHL-a), persistence P, and shifts S (see Eq. 4.1). Triangles indicate that a bloom is occurring at the considered station for the month on the x-axis. The triangle color is proportional to CHL-a, the warmer the color (from black to red), the more severe the bloom in terms of CHL-a magnitude. Bloom classes are defined for each station by an increasing $\Delta(CHL-a) = (CHL-a_{max} - CHL-a_{min})/7\ \mu g/L$, that gives a detailed information on shifts risk profiling. Red lines between two triangles indicate that blooms happen continuously between adjacent months (this defines bloom duration). Three classes of risk (low, medium and high, identified by green, blue and red colored text) are based on grouping values of Risk as in Eq. 4.1, and nodes are shown in plot B based on these risk classes.

systemic risk by also considering temporal changes in distributions as in Eq. 4.1 that can shift risk classes of ecological communities (Fig. 4.8). Blooms are highly dissipative, disorganized and unstable phenomena, with potential cascading effects across the whole bay. Functional networks in Fig. 4.6B show arrows whose width is proportional to TE and topology is more complex for more scale-free envirome or biogeochemical networks (e.g., for East Bay) conferring ecological stability and higher predictability; links in red are for direct interactions of environmental factors with CHL-a that in non-bloom conditions are less directed toward CHL-a; yet CHL-a is less subjected to strong variability in temperature and TP. Solely structural analyses of ecological networks, without considering the TE distribution, may lead to a mislead classification of networks such as a scale-free network for Central Bay due to the centrality of CHL-a. Environmental triggers may lead to ecological states that are more persistent and with higher bloom intensity (smaller recurrence time and higher CHL-a), which are also much less predictable due to the more equivalent probability of very different CHL-a ranges. We argue that increase in entropy of CHL-a may contributes to shifts from diatom- to dinoflagellate-dominated ecosystems with increased dia-dino niche divergence (related to unbalanced competition-cooperation between these phytoplankton species), implying ingrained harmful algal blooms. This hypothesis is made considering the information about diatoms and dinoflagellate in FL Bay (Fig 4.1) and the space-time evolution of algal blooms. Vice versa, overlap of niches underpin ecological stability, in principle. The energy potential in Fig. 4.6B can be considered as the ecological fitness landscape of possibilities. The position of balls in the energy dissipation landscape of Fig. 4.6B shows the magnitude of the dissipation, proportional to H(CHL-a), and the darker the color, the higher the instability of the ecological bloom state. The average and max value of CHL-a is related to randomness of the envirome and this apparent simplicity/homogeneity is not good for the environment and the ecology. This also affects predictability where computational complexity, defined as the number of layers of the GNN, is larger. that impacts the forecasting accuracy. Therefore, we observe a duality between ecological and computational complexity, where beneficial increase in the former (scale-free) correspond to a detrimental decrease in the latter (more features that are more disconnected and changing over time, with lower cross-autocorrelation and less predictability of cause-effects). Complexity is inversely proportional to randomness; the higher the randomness the lower the complexity and also the network function efficiency, including the computational one; therefore, more blooms also mean lower predictability.

### 4.3.6 Non-linear bloom risk

Bloom is a systemic manifestation of water eutrophication related to both oceanic and land stress. The frequent occurrence of bloom has negatively affected the marine ecosystem of Florida Bay. According to the bloom threshold (CHL-a $1.841 \mu g L^{-1}$) in Florida Bay, the persistence of bloom and their timing an be calculated, to characterize their patterns with the goal to prevent the occurrence of blooms relay in following seasons. Through observations throughout the monitoring period, blooms have seasonality, where in FL Bay the autumn is the peak with long-lasting periods of bloom after outbreaks. Based on the analysis of the two largest blooms in 1999 and 2005, we observed a tendency for the September-December blooms (post tropical cyclone season) to shift from the west and central bay to the east over time (Figs 4.2,4.8 and Fig. S1). Fig. 4.7 shows the linear interpolation (linear geokriging) of values of CHL-a and environmental factors across monitored stations that are stations where the eco-environmental feedback are inferred. In Fig. 4.8A, we visualized the blooming dynamics ("ecogram") at each monitoring station and corresponding risk classification ("riskgram") based on magnitude, persistence and shifts (Eq. 4.1). So that, we can observe the duration of bloom and the time points of each outbreak, as well as monitoring stations where the bloom occurs frequently. Stations in Fig. 4.8B are re-ranked according to the risk class. Yet, Fig. 4.8 is also a riskgram, that is a risk profile of bay communities based on magnitude, persistence and shifts of CHL-a above the blooming threshold. Shift increases for an increasing divergence between non-random and random distributions manifesting higher risk due to loss of predictability (equivalently, increased loss of information, or decreased complexity of a time series intended as departure from scale-free distributions (Fig 4.5C for station 14) or functional networks of environmental determinants (Fig. 4.6B for the Central lagoon), that increases RMSE). Triangles in Fig. 4.8 indicate that a bloom is occurring at the considered station for the month on the x-axis. The triangle color is proportional to CHL-a, the warmer the color (from black to red), the more severe the bloom in terms of CHL-a magnitude. Bloom classes are defined for each station by an increasing $\Delta(CHL-a) = (CHL-a_{max} - CHL-a_{min})/7\ \mu g/L$, that gives a detailed information on shifts (proportional to permutation entropy) risk profiling compared to classifying blooms only on one threshold as in Nelson et al. [46]. For instance,

$\Delta(CHL-a) = 30-8 = 22$ for station 14 (higher max and average) for station 1, that are more uniformly and scale-free distributed but belong to the same risk class because of (P,M) and S for the former and the latter. The top $20\%$ blooms belong to the 7th class. In Fig. 4.8 red lines between two triangles indicate that blooms happen continuously between adjacent months (this defines bloom duration), yet the longer and more frequent the lines the more persistent the blooms. It is noticeable that from 2005 onward, blooms expand to the North-East of FL Bay and have an average increase of 20% in CHL-a (that corresponds to a reclassification of blooms to class 6 and 7 particularly in the Central and North-East areas). Three classes of risk (low, medium and high, identified by green, blue and red colored text) are based on grouping values of Risk as in Eq. 4.1, and nodes are shown in plot B based on these risk classes. It is noticeable the impact of M and S in shifting node risk classes based on P solely.

### 4.3.7 Missed causality, relative anomaly, inevitability and controllability

Here we try to address aspects that are larger in scope than details of the paper. For instance, answering questions about "missed" causal factors, whether bloom are truly anomalous, and what is controllable or inevitable due to natural variability.

Are blooms truly locally-driven by nutrient efflux or more related to global climate? Are blooms truly "man-caused" or natural? CHL-a outbreaks and subsequent blooms, e.g. the massive booms in 1999 and 2005, are largely coinciding with the strongest tropical cyclone seasons in South Florida. For instance, hurricanes like Katrina, Rita, and Wilma in 2005 (Fig. S7), likely caused high salinity, sea-surface temperature, nutrients and phytoplankton redistribution toward the center of the bay and all along coastal habitats. In 1999 category 2 Hurricane Irene cut across FL Bay, category 4 Hurricane Floyd passed closer to East Bay habitats, and Tropical Storm Harvey passed closer to the northern coastal areas of FL Bay. In 2005 category 4 Hurricane Dennis passed very close to West FL Bay, category 5 Hurricane Katrina cut across FL Bay, category 5 Hurricane Rita passed south of FL Bay, category 5 Hurricane Wilma passed in the northern coastal areas of FL Bay (Fig. S7). The 1999 Atlantic hurricane season had five Category 4 hurricanes (Fig. S7) – the highest number recorded in a single season in the Atlantic basin, previously tied in 1961, and later tied in 2005. With 28 storms (27 named storms and one unnamed), the 2005 Atlantic hurricane season set a new single-year record for most

storms (that also implied severe economic losses [128]), surpassing the total of 20 from 1933. By considering blooms, their number was higher in 2005 (considering Eastern areas that bloomed much more, particularly stations 3, 4 and 5, potentially because of Floyd) with higher magnitude around West FL Bay, but 1999 outbreaks (CHLa maximum, mostly as single peaks around central and western areas; station 14 and 27 predominantly because of Irene likely) was larger despite fewer and less intense hydroclimatic extremes. Station 13, 14 and 15 are always fluctuating in a persistent endemic regime regardless of extreme tropical cyclones. While our findings do not expand the predictions of previous studies that emphasized the role of storm-induced nutrient fluxes and freshwater supply as primary drivers of productivity gradients in bays, we pinpoint how *biogeochemical organization* is driving blooms. Further investigations are needed to determine how much of that is truly an anomaly in contrast to natural ecological cycles where mortality of species is needed for instance to absorb carbon. Hurricanes and other tropical cyclones may be the primary triggers that alter the biogeochemical distribution, where the latter can be more harmful if nutrients loads (from land) are higher than established threshold. However, not considering these extreme phenomena (cyclones and loads) is fine for forecasting purposes (and not attribution to environmental causes, i.e., "etiognosis") because their effect as biogeochemical imbalance – with CHL-a is the salient ecosystem indicator – is already incorporated.

Relatively surprisingly, in the non-bloom regime for most stations, CHL-a is the most active factor and it has the largest influence on other factors. This confirms the regulatory function of phytoplankton for all aquatic habitats and how this function is altered when other biogeochemical factors become predominant; yet, this highlights the transition from a functionally scale-free to a disorganized network in non-bloom and bloom regimes where CHL-a is the leading and effect hub, respectively. Therefore, the balance of ecosystem productivity, dictated by interactions, is important rather than single-threshold factors. Additionally, these results provide evidence on the power of *invariant predictive causality* (as TE or affine information-theoretic measures that quantify *disorder, its structure and shifts*) in representing salient ecological processes where gradients of uncertainty reduction correspond to the relative importance of underlying factors that by virtue of their divergence influence each other. Based on our analyses, it is clear that blooms and water quality factors are interdependent bidirectionally, where the occurrence of blooms can lead to long-term water-quality changes and fluctuations of those baseline the new bloom emergence and the persistence of old ones. This underpins *bidirectional eco-*

*environmental feedback* at multiple scales – from phytoplankton community structure to teleconnected climate – that are often neglected. While independent changes in biogeochemical factors are important, we showed how their non-linear interactions is much more concerning because those can constitute systemic alterations into ecosystem functions. How much are these interactions controllable and are bloom truly increasing [55] or it is just a byproduct of increased monitoring [129]? It should always be kept in mind that any analysis has *spatio-temporal reference scales* defining initial and boundary conditions over which an anomaly is classified. Therefore, care should be placed when making conclusions. In FL Bay, blooms are shifting in some locations that never experienced peaks for our monitoring period and area considered, yet no certainty can be placed on whether that is absolutely anomalous.

The above arguments raise questions about the *controllability* of blooms: how much are truly controllable and should we control them? We can control efflux but we cannot control cyclones; however, we may be able to influence global climate change via local climate adaptation, where climate change adaptation is the process of adjusting to current or expected effects of climate change via climate eco-engineering (acting on *ecological and environmental flows*). In the context of bays this may be possible via preservation and restoration of habitats and reduction of their anthropogenic pressure. In FL Bay, the central part of the bay is attracting most of the of biogeochemical changes in other areas, with higher oceanic influence from the Western part and terrestrial efflux influence from the Easter part. Yet there is a certain inevitability of the central bay to be a sink for blooms as well as a source after its ecological imbalance when that happens. *Inevitability* also in relation of the fact that ecological imbalance causes environmental decline and emergence of subsequent blooms with potential hysteresis, and ecological memory due to long-lasting effects of extreme hazards and eutrophication. To avoid this, a proactive consequentialist ecology (ecosystem engineering) is needed vs. purely descriptive analyses. Beyond restoration and protection, *rewiring ecological corridors* (by re-shaping the lost geomorphological heterogeneity and necessary fresh- and salt-water flows in an applied ecohydrology sense) is necessary as well as *integrated monitoring* that considers models as sensors that integrate information for optimal and adaptive decision making. Therefore, models for investigating causes, ranking malfunction, and delineating solutions. Models, can be improved for instance by coupling extreme climate forecasts with ecological forecasts, incorporating geomorphological information of habitats like bathymetry, and including

multiscale ecoinformation such as phenotypical features of species, to (i) provide high-resolution of malfunction and disservices, and (ii) pinpoint where we should improve response capacity. In this perspective monitoring is essential where models as *Digital Information Models* are like processing sensors that assimilate data non-linearly to produce information maps (risk, pathways and perception divergence e.g., about known and objective ecological risk) for stakeholder action.

Lastly, an *opportunistic ecosystem purview*, more than consequentialist on impacts, is arguably needed where blooms are seen as an opportunity rather than a curse, and action can be done via ecohydrological foundations and eco-inspired technology such as "Daphnia biofiltrators" [130]. This is also in light of bloom natural role as guardians of the carbon cycle [3]. The same can be said about tropical cyclones as regulators of climate. Opportunities can come from asking how to *enhance ecosystem services* we can get from blooms, for example by (i) enhancing the response capacity to extreme hazards such as tropical cyclones; and (ii) enlarging the carbon sequestration capacity that can help phytoplankton balance. At large, cyclones and blooms may also be opportunities as energy sources. In this view, the ecology is seen as a *engineerable asset* where mitigation, adaptation, and development exist together and ecological change is the basis of any economic change [131]. This opportunist view seems necessary due do the already passed "ecological tipping point" where declines are leading to tangible compounding hydroclimatic extremes [132].

## 4.4 Conclusions

Oceanic Chlorophyll-a concentration (CHL-a) is considered as the universal proxy for phytoplankton biomass, and yet its alteration are arguably related to ocean dysbiosis (i.e., imbalance not necessarily related to impacted populations). Because of the key role of phytoplankton in the global cycle of earth's elements (production and export of nutrients), mapping and understanding the spatio-temporal distribution and variability in CHL-a is of primary importance. This is also particularly important for formulating risk and decision priorities in imbalanced coastal and marine ecosystems, especially those at the verge of ecological shifts. Global and local pressure such as related to human activities and climate warming have altered the ecohydrology of many inland and coastal wetlands worldwide, such as for the Everglades in South Florida (USA) that is an epitome considering its size and

complexity. Water management plans in the Everglades have severely impacted the coastal and marine ecosystems of Florida Bay, which contributed to the increase in frequency and magnitude of algal blooms. Beyond local ecological impacts, this has likely implications for climate regulation as well since carbon fixation (CO2-fixation) by marine phytoplankton accounts for about half the Earth's primary production. The central objective of this study was to explore the causal relationships between biogeochemical factors and CHL-a, altogether considered as a biogeochemical network whose topology alteration leads to algal bloom emergence and persistency. The following points are worthwhile mentioning.

- **Ecological network organization and forecasting**. Predictive causality of flash blooms (sensu [69] as Transfer Entropy organization of ecological networks), emerging from species collective sensing modulated by the environment, was shown to enhance forecasting accuracy. It is noticeable how the envirome is more structurally organized (and not functionally for ecological flows) as a small-world network for small blooms (higher indirect connections for CHL-a with lower magnitude), while more scale-free (higher direct connections with higher magnitude) for large blooms; this results occur when small and large bloom interdependencies over time are not considered. In light of the apparent scale-freeness of bloom-dynamics, predictability increases with complexity (from large to small blooms if considered separately), likely because less and more organized factors (and their less sustained interactions including time-delays) are important for the variability of CHL-a, including higher habitat heterogeneity that has a beneficial effect. This may also relate to a lower degree in spatial connections between regions that are balanced and largely independent from each other in terms of CHL-a; connectivity that is enhanced during blooms and increases predictability if considered [41]. Increasing blooms may disrupt the positive ecological connections facilitated by both habitats and microorganisms; an hypothesis to verify in further studies but that is certainly evident from inferred networks resulting more disorganized over time. We highlighted how for epidemic dynamics, baselined by the highest environmental complexity, the highest forecasting skills are achieved considering the coincidence of magnitude, timing and duration of observed and predicted CHLa; this also speculates on how the optimal Pareto dynamics of CHL-a has implications for optimal sensing, predictability and eco-engineering.

- **Critical Chlorophyll-a threshold for bloom emergence**. A systemic ecological criteria of water blooms was validated through the normalized probability distribution function of CHL-a and its maximum divergence as *probabilistic tipping point* manifesting collective (distribution) changes in blooms. It should be noted that these collective change may have different persistence levels depending on a sequence of CHL-a values over time defining the CHL-a pdf. This criteria largely corresponds to the generic Pareto principle where blooms (corresponding to the largest 20% CHL-a values) describe 80% of the CHL-a variability. Risk warning may be loose for station-dependent threshold whose blooming may initiate earlier (due to distributional vs. value changes) in relation to spatial spreading. Yet, it is important to quantify the change in probability distribution vs. a single value approach on statistical quantiles that may not reflect the change in dynamics: for instance the quartile may be very similar for very diverse probability distribution function manifesting the persistency of a process. Endemic conditions are persistent for blooms only if CHL-a is large, and yet CHL-a value-threshold criteria are meaningful only for non scale-free distributions. The selection of a station-dependent critical threshold like in Nelson et al. [46] is also suboptimal because that does not consider blooms in their systemic non-linear interdependencies (probabilities informs about the node collective response to changing environmental conditions) where small values, such as for power-law distributed CHL-a, may be very meaningful vs. small values for normally distributed CHL-a. The former case underpins an incipient change toward a bloom outbreak (bloom onset). Further studies would be needed to verify the specificity of CHL-a distributions in different habitats worldwide and their blooming threshold stability.

- **Non-linear risk characterization**. Risk was originally defined in in terms of probability distribution (pdf) and their change vs. risk classes defined only on set threshold on CHL-a, because CHL-a values are meaningful in different ways depending on the distribution they belong to. Each distribution defines persistence or antipersistence with respect to all other variables realized over space and time, as well as their expectation, being more or less random. That is the reason for which CHL-a dynamics must be characterized as endemic or epidemic (based on its pdf), while the CHL-a blooming threshold identifies only single outbreaks with no assessment of their persistence (the higher

the more uniform the pdf is). Therefore, a decision analytical framework (where analytics come from a proper analytical characterization of the pdfs and its shifts) must place higher risk on randomly distributed areas (endemic blooms) for the same extreme values of CHL-a, shifts of these distribution and fluctuations, and persistence of extreme values. For FL Bay, fluctuations and persistence of large CHL-a are increasing for endemic areas, and magnitude and distribution are increasing for epidemic areas, which constitute a great risk on ecosystems. Increasing bloom persistence is reported in areas that have lower entropy, and areas with higher disorganization with higher free-energy (energy dissipation is lower for lower TEs); persistence in new areas that that is across months and not only in summer/post-summer periods. Our approach, compared to Nelson et al. [46] and similar papers, is (1) informational (in terms of pdf characterization/dynamics, that is about pdf shape and predictability of bloom patterns, beyond absolute probability value; blooms driven by local ecology and ecological corridors, where predictability is always inversely proportional to the bloom risk); (2) systemic (considering phytoplankton CHL-a population normalization across ecological communities over space, because CHL-a is always dependent on other communities); and, (3) and temporal (considering how pdfs change over time to characterize ecological history). Thus, bloom risk (or vulnerability more precisely, since risk is always associated to an impact) should be assessed on the three above elements and not just on average magnitude or extreme outbreaks to incorporate persistent and shifting degrees of bloom.

- **Biogeochemical disorganization and critical slowing-down**. Eutrophic ecosystems are demonstrated to experience more persistent and/or erratic blooms manifested by disorganized biogeochemical network topologies. Small and repeated stress should not be overlooked because they can have large effects in the long run via critical slowing down [90, 133]. This is for instance the case of NE habitats in FL Bay that are experiencing erratic and more persistent blooms with slow but critical ecological collapse driven by excess nutrient loads and ocean temperature fluctuations. Dinoflagellate, for example were increasing due to temperature anomalies and nutrient efflux. We emphasized the critical slowing down (CSD a' la Scheffer [124]) of blooms that became smaller but more erratic and widely distributed in space. This temporal trend is evident from 2005, after a hiatus in in between 2000 and

2005 approximately, that is a classical signature of critical transition from one long-term dynamics to another: variance and average of indicators (in this case top 20% CHL-a as indicator of blooms) diminish before a change in ecosystem regime followed up by an increase in variance. It is also interesting to observe that after 2005, algal blooms started to expand to NE areas of Florida Bay (stations 1-10 previously not impacted by blooms) closer to South Florida agricultural and urban areas; yet, manifesting a likely increasing pressure from these areas. Unfortunately that sector of Florida Bay hosts many unique species, such as marine sponges, and due to the shallow-water habitats, ecological communities are very sensitive to environmental change which poses a very high risk for many rare and native species. This regime shift also likely favors the emergence of invasive species and pathogens, as well as the potential release of toxins affecting human health via direct exposure and sea-food contamination [134–137]. These elements cumulate and increase the systemic eco-environmental impact of the whole ecosystem.

# Chapter 5

# Conclusions

The aim of this dissertation is to offer insights into how causal network reasoning and graph neural networks can be utilized to generate innovative ideas for unraveling a complex ecological problem, namely algal blooms. A causal inference model is established based on transfer entropy difference to quantify the one-directional information flow between all pairs of variables, thus simulating the dynamics of algal blooms. Furthermore, one-directional transfer entropy differences are combined with graph neural networks to achieve time-series predictions of algal blooms. By analyzing the information dynamics, structure, and functional features of the causal inference network, it enables the identification of the dynamic evolution of the response of algal blooms to environmental changes, or even the ability to track it without spatial dependencies. With the assistance of macroecological analyses, these results can be utilized to comprehend how algal blooms respond to fluctuations in environmental factors and changes in geography. Additionally, causal relationships (i.e., transfer entropy differences) are employed as a priori information to construct graph neural networks. This design serves to emphasize the connections between variables, effectively avoiding the redundancy of information stemming from high dimensions. Through the incorporation of prior causal information acquired through Transfer Entropy difference, the model eliminates the necessity to independently identify pivotal variables for accurate forecasting. Thus, this study provides insights into complex ecosystem problems, contributes to the proposition of targeted preventive approaches and measures, and achieves an effective prediction of complex ecosystems with multiple environmental factors.

Chapter 2 outlines the development of both causal inference models and causal neural network models, primarily leveraging them for the dynamic analysis and

prediction of the algal blooms indicator, CHLa. Our application of the causal inference model to the complete algal blooms observation period, spanning August 1992 to September 2008, yielded a causal network involving 28 observation stations for CHLA. This network effectively pinpointed the most influential and active stations pertaining to algal blooms, providing a framework for subsequent analyses regarding the interrelationships of water quality factors within these stations. However, it is imperative to address the heightened concern surrounding prolonged and abrupt blooms, particularly in the stations identified as most influential. Abrupt blooms may result from sudden fluctuations in water quality factors or prolonged accumulation over time. Conversely, prolonged blooms are consistently accompanied by significant passive shifts in specific water quality factors. By scrutinizing the causal relationships amidst environmental factors prior to abrupt blooms and establishing the causality between water quality factors during prolonged bloom conditions, we meticulously dissected the specific triggers and clear risk indicators for these two distinctive bloom phases at individual stations. In addition, we demonstrate more intuitively the prediction results of the causality-based graph neural network for CHLa from an ecological perspective. Comparison of the actual and predicted conditions of peak CHLa periods at specified station using maps clearly shows the geographical deviation between the actual and predicted situations. Thus, our causal graph neural network model achieves accurate geographic prediction of CHLa (i.e., algal blooms) without based on spatial dependence.

Chapter 3 places its emphasis on the practical application of causal inference modeling to the algal blooms event of 2005. The most effective manner to validate an inferential model resides in its comparison against the actual occurrence of an event. Given the prominent status of the 2005 Florida Bay bloom event, comprehensive records detailing its circumstances are accessible. By contrasting our causal inference findings prior to and following the 2005 bloom outbreak, a congruence emerges with the authentic spatial trend of the bloom – namely, the emergence of blooms in the western and central sectors of Florida Bay, which subsequently spread toward the eastern regions over time. This particular bloom event also incited considerable fluctuations in water quality on a large scale. Consequently, we undertook the quantification of the bloom's repercussions on water quality factors at diverse stations, proposing targeted strategies for control in response. Furthermore, we undertook an evaluation of the ecosystem's stability prior to and subsequent to the event, grounded in the probability distribution of CHLa's collective influence and ecosystem potential. This assessment showcased the transformation in the ecosys-

tem's state before and after the bloom event, concurrently highlighting the robust capacity of the ecosystem for self-regulation.

Chapter 4 presents a comprehensive analysis and evaluation of our proposed causal inference method and causality-based graph neural networks. In addition to the main bloom indicator CHL-a, reasoning and predictive analyses of other relevant environmental factors, as well as ecosystem risk evaluation, were incorporated. Consequently, based on the causal environmental factors behind CHL-a's dynamic variability, constructing a predictive biogeochemical network that underlies algal bloom emergence, persistence, shifts, and ecosystemic risks. Noteworthy findings include the enhancement of predictive accuracy through predictive causality of flash blooms organized by Transfer Entropy, and the identification of critical CHL-a thresholds for bloom outbreaks. Risk characterization involves systemic ecological criteria of water blooms, and the evaluation of risk is non-linear due to probability distribution dynamics. Furthermore, the study highlights the influence of biogeochemical disorganization and critical slowing-down in eutrophic ecosystems, where persistent and erratic blooms signal impending ecological collapse. Temporal shifts in bloom patterns underscore the ecosystem's vulnerability, impacting marine life and human health. This investigation contributes to understanding the multifaceted interplay between CHL-a dynamics, ecological shifts, and the systemic risks associated with changing environmental conditions.

The results from all these studies evidence the efficiency of the developed causal inference model and causality-based graph neural network. Notably, when applied to real-world ecological challenges, these tools exhibit robust performance and distinct advantages, unraveling intricate system dynamics, information flow, and system stability. Moreover, their prowess in multivariate time series prediction within complex systems becomes evident. Beyond the scope of the present study, these concepts stand adept at extrapolating and anticipating system behaviors and information dynamics across a spectrum of intricate scenarios, extending to domains like medicine, social networks, and brain sciences, where conventional methodologies may falter.

# Bibliography


[1] C. Hu, Z. Lee, and B. Franz, "Chlorophyll aalgorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference," *Journal of Geophysical Research: Oceans*, vol. 117, no. C1, 2012.

[2] J. E. O'Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain, "Ocean color chlorophyll algorithms for seawifs," *Journal of Geophysical Research: Oceans*, vol. 103, no. C11, pp. 24 937–24 953, 1998.

[3] M. Convertino and H. Wang, "Envirome disorganization and ecological riskscapes: The algal bloom epitome," in *Risk Assessment for Environmental Health*. CRC Press, 2022, pp. 327–346.

[4] L. Xu, D. Patterson, S. A. Levin, and J. Wang, "Non-equilibrium early-warning signals for critical transitions in ecological systems," *Proceedings of the National Academy of Sciences*, vol. 120, no. 5, p. e2218663120, 2023.

[5] T. J. Hefley, K. M. Broms, B. M. Brost, F. E. Buderman, S. L. Kay, H. R. Scharf, J. R. Tipton, P. J. Williams, and M. B. Hooten, "The basis function approach for modeling autocorrelation in ecological data," *Ecology*, vol. 98, no. 3, pp. 632–646, 2017.

[6] R. W. Day and G. P. Quinn, "Comparisons of treatments after an analysis of variance in ecology," *Ecological monographs*, vol. 59, no. 4, pp. 433–463, 1989.

[7] V. Guttal and C. Jayaprakash, "Changing skewness: an early warning signal of regime shifts in ecosystems," *Ecology letters*, vol. 11, no. 5, pp. 450–460, 2008.

[8] ——, "Spatial variance and spatial skewness: leading indicators of regime shifts in spatial ecological systems," *Theoretical Ecology*, vol. 2, pp. 3–12, 2009.

[9] L. H. Loke and R. A. Chisholm, "Measuring habitat complexity and spatial heterogeneity in ecology," *Ecology Letters*, vol. 25, no. 10, pp. 2269–2288, 2022.

[10] S. Dobbert, R. Pape, and J. Löffler, "How does spatial heterogeneity affect inter-and intraspecific growth patterns in tundra shrubs?" *Journal of Ecology*, vol. 109, no. 12, pp. 4115–4131, 2021.

[11] M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, and B. Walker, "Catastrophic shifts in ecosystems," *Nature*, vol. 413, no. 6856, pp. 591–596, 2001.

[12] G. B. Bonan and S. C. Doney, "Climate, ecosystems, and planetary futures: The challenge to predict life in earth system models," *Science*, vol. 359, no. 6375, p. eaam8328, 2018.

[13] M. J. Kaiser, S. Jennings, D. N. Thomas, D. K. Barnes *et al.*, *Marine ecology: processes, systems, and impacts*. Oxford University Press, 2011.

[14] M. A. St. John, A. Borja, G. Chust, M. Heath, I. Grigorov, P. Mariani, A. P. Martin, and R. S. Santos, "A dark hole in our understanding of marine ecosystems and their services: perspectives from the mesopelagic community," *Frontiers in Marine Science*, vol. 3, p. 31, 2016.

[15] D. E. Stevenson and R. R. Lauth, "Bottom trawl surveys in the northern bering sea indicate recent shifts in the distribution of marine species," *Polar Biology*, vol. 42, no. 2, pp. 407–421, 2019.

[16] K. M. Kleisner, M. J. Fogarty, S. McGee, J. A. Hare, S. Moret, C. T. Perretti, and V. S. Saba, "Marine species distribution shifts on the us northeast continental shelf under continued ocean warming," *Progress in Oceanography*, vol. 153, pp. 24–36, 2017.

[17] A. R. Spence and M. W. Tingley, "The challenge of novel abiotic conditions for species undergoing climate-induced range shifts," *Ecography*, vol. 43, no. 11, pp. 1571–1590, 2020.

[18] M. H. Glantz *et al.*, *Currents of change: impacts of El Niño and La Niña on climate and society*. Cambridge University Press, 2001.

[19] M. J. McPhaden, "El niño and la niña: causes and global consequences," *Encyclopedia of global environmental change*, vol. 1, pp. 353–370, 2002.

[20] M. Vetemaa, R. Eschbaum, A. Albert, L. Saks, A. Verliin, K. Jürgens, M. Kesler, K. Hubel, R. Hannesson, and T. Saat, "Changes in fish stocks in an estonian estuary: overfishing by cormorants?" *ICES Journal of Marine Science*, vol. 67, no. 9, pp. 1972–1979, 2010.

[21] J. R. Zaneveld, D. E. Burkepile, A. A. Shantz, C. E. Pritchard, R. McMinds, J. P. Payet, R. Welsh, A. M. Correa, N. P. Lemoine, S. Rosales *et al.*, "Overfishing and nutrient pollution interact with temperature to disrupt coral reefs down to microbial scales," *Nature communications*, vol. 7, no. 1, p. 11833, 2016.

[22] K. C. Hester, E. T. Peltzer, W. J. Kirkwood, and P. G. Brewer, "Unanticipated consequences of ocean acidification: A noisier ocean at lower ph," *Geophysical research letters*, vol. 35, no. 19, 2008.

[23] J. K. Baum and B. Worm, "Cascading top-down effects of changing oceanic predator abundances," *Journal of animal ecology*, vol. 78, no. 4, pp. 699–714, 2009.

[24] M. V. Abrahams and M. G. Kattenfeld, "The role of turbidity as a constraint on predator-prey interactions in aquatic environments," *Behavioral Ecology and Sociobiology*, vol. 40, pp. 169–174, 1997.

[25] L. Jiang, J. Eriksson, S. Lage, S. Jonasson, S. Shams, M. Mehine, L. L. Ilag, and U. Rasmussen, "Diatoms: a novel source for the neurotoxin bmaa in aquatic environments," *PloS one*, vol. 9, no. 1, p. e84578, 2014.

[26] K. G. Sellner, G. J. Doucette, and G. J. Kirkpatrick, "Harmful algal blooms: causes, impacts and detection," *Journal of Industrial Microbiology and Biotechnology*, vol. 30, pp. 383–406, 2003.

[27] B. D. Eyre and A. J. Ferguson, "Comparison of carbon production and decomposition, benthic nutrient fluxes and denitrification in

seagrass, phytoplankton, benthic microalgae-and macroalgae-dominated warm-temperate australian lagoons," *Marine Ecology Progress Series*, vol. 229, pp. 43–59, 2002.

[28] B. Dale, "Eutrophication signals in the sedimentary record of dinoflagellate cysts in coastal waters," *Journal of Sea Research*, vol. 61, no. 1-2, pp. 103–113, 2009.

[29] A. Altieri, "Dead zones: Low oxygen in coastal waters," in *Encyclopedia of Ecology*. Elsevier, 2019, vol. 1, pp. 22–34.

[30] E. H. Buck, "Marine dead zones: Understanding the problem." Congressional Research Service, Library of Congress, 2007.

[31] K. E. Limburg, D. Breitburg, D. P. Swaney, and G. Jacinto, "Ocean deoxygenation: A primer," *One Earth*, vol. 2, no. 1, pp. 24–29, 2020.

[32] G. M. Hallegraeff, "Ocean climate change, phytoplankton community responses, and harmful algal blooms: a formidable predictive challenge 1," *Journal of phycology*, vol. 46, no. 2, pp. 220–235, 2010.

[33] F. M. V. Dolah, D. Roelke, and R. M. Greene, "Health and ecological impacts of harmful algal blooms: risk assessment needs," *Human and Ecological Risk Assessment: An International Journal*, vol. 7, no. 5, pp. 1329–1345, 2001.

[34] M. Ushio, K. Watanabe, Y. Fukuda, Y. Tokudome, and K. Nakajima, "Computational capability of ecological dynamics," *bioRxiv*, pp. 2021–09, 2021.

[35] P. R. Hill, A. Kumar, M. Temimi, and D. R. Bull, "Habnet: Machine learning, remote sensing based detection and prediction of harmful algal blooms," *arXiv preprint arXiv:1912.02305*, 2019.

[36] ——, "Habnet: Machine learning, remote sensing-based detection of harmful algal blooms," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3229–3239, 2020.

[37] E. Montes, A. Djurhuus, F. E. Muller-Karger, D. Otis, C. R. Kelble, and M. T. Kavanaugh, "Dynamic satellite seascapes as a biogeographic framework for understanding phytoplankton assemblages in the florida keys national marine sanctuary, united states," *Frontiers in Marine Science*, p. 575, 2020.

[38] J. Li and M. Convertino, "Optimal microbiome networks: macroecology and criticality," *Entropy*, vol. 21, no. 5, p. 506, 2019.

[39] J. Yu, Y. Tian, X. Wang, and C. Zheng, "Using machine learning to reveal spatiotemporal complexity and driving forces of water quality changes in hong kong marine water," *Journal of Hydrology*, vol. 603, p. 126841, 2021.

[40] E. Mcleod, G. L. Chmura, S. Bouillon, R. Salm, M. Björk, C. M. Duarte, C. E. Lovelock, W. H. Schlesinger, and B. R. Silliman, "A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering co2," *Frontiers in Ecology and the Environment*, vol. 9, no. 10, pp. 552–560, 2011.

[41] H. Wang, E. Galbraith, and M. Convertino, "Algal bloom ties: Spreading network inference and extreme eco-environmental feedback," *Entropy*, vol. 25, no. 4, p. 636, 2023.

[42] E. J. Phlips, S. Badylak, and T. C. Lynch, "Blooms of the picoplanktonic cyanobacterium synechococcus in florida bay, a subtropical inner-shelf lagoon," *Limnology and Oceanography*, vol. 44, no. 4, pp. 1166–1175, 1999.

[43] A. Wachnicka, J. Browder, T. Jackson, W. Louda, C. Kelble, O. Abdelrahman, E. Stabenau, and C. Avila, "Hurricane irma's impact on water quality and phytoplankton communities in biscayne bay (florida, usa)," *Estuaries and Coasts*, vol. 43, pp. 1217–1234, 2020.

[44] E. Galbraith and M. Convertino, "The eco-evo mandala: Simplifying bacterioplankton complexity into ecohealth signatures," *Entropy*, vol. 23, no. 11, p. 1471, 2021.

[45] J. N. Boyer and H. O. Briceño, "South florida coastal water quality monitoring network," *FY2006 Cumulative Report South Florida Water Management District, Southeast Environmental Research Center, Florida International University (http://serc. fiu. edu/wqmnetwork/)*, 2007.

[46] N. G. Nelson, R. Munoz-Carpena, and E. J. Phlips, "A novel quantile method reveals spatiotemporal shifts in phytoplankton biomass descriptors between bloom and non-bloom conditions in a subtropical estuary," *Marine Ecology Progress Series*, vol. 567, pp. 57–78, 2017.

[47] J. N. Boyer, C. R. Kelble, P. B. Ortner, and D. T. Rudnick, "Phytoplankton bloom status: Chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of florida, usa," *Ecological indicators*, vol. 9, no. 6, pp. S56–S67, 2009.

[48] H. W. Paerl and T. G. Otten, "Harmful cyanobacterial blooms: causes, consequences, and controls," *Microbial ecology*, vol. 65, pp. 995–1010, 2013.

[49] L. M. Grattan, S. Holobaugh, and J. G. Morris Jr, "Harmful algal blooms and public health," *Harmful algae*, vol. 57, pp. 2–8, 2016.

[50] H. Xu, Y. Huang, Z. Duan, X. Wang, J. Feng, and P. Song, "Multivariate time series forecasting with transfer entropy graph," *arXiv preprint arXiv:2005.01185*, 2020.

[51] O. Pulido, "Phycotoxins by harmful algal blooms (habs) and human poisoning: an overview," *Int. Clin. Pathol. J*, vol. 2, no. 6, pp. 145–152, 2016.

[52] H. Xu, Y. Liu, Z. Tang, H. Li, G. Li, and Q. He, "Methane production in harmful algal blooms collapsed water: The contribution of non-toxic microcystis aeruginosa outweighs that of the toxic variety," *Journal of Cleaner Production*, vol. 276, p. 124280, 2020.

[53] T. J. Smayda, "What is a bloom? a commentary," *Limnology and Oceanography*, vol. 42, no. 5part2, pp. 1132–1136, 1997.

[54] H. W. Paerl, R. S. Fulton, P. H. Moisander, and J. Dyble, "Harmful freshwater algal blooms, with an emphasis on cyanobacteria," *TheScientificWorldJournal*, vol. 1, pp. 76–113, 2001.

[55] Y. Dai, S. Yang, D. Zhao, C. Hu, W. Xu, D. M. Anderson, Y. Li, X.-P. Song, D. G. Boyce, L. Gibson *et al.*, "Coastal phytoplankton blooms expand and intensify in the 21st century," *Nature*, pp. 1–5, 2023.

[56] H. K. Lotze, B. Worm, and U. Sommer, "Propagule banks, herbivory and nutrient supply control population development and dominance patterns in macroalgal blooms," *Oikos*, vol. 89, no. 1, pp. 46–58, 2000.

[57] E. M. Cosper, W. C. Dennison, E. J. Carpenter, V. M. Bricelj, J. G. Mitchell, S. H. Kuenstner, D. Colflesh, and M. Dewey, "Recurrent and persistent brown tide blooms perturb coastal marine ecosystem," *Estuaries*, vol. 10, no. 4, pp. 284–290, 1987.

[58] M. J. Butler IV, J. H. Hunt, W. F. Herrnkind, M. J. Childress, R. Bertelsen, W. Sharp, T. Matthews, J. M. Field, and H. G. Marshall, "Cascading disturbances in florida bay, usa: cyanobacteria blooms, sponge mortality, and implications for juvenile spiny lobsters panulirus argus," *Marine Ecology Progress Series*, vol. 129, pp. 119–125, 1995.

[59] K. Inomura, C. Deutsch, O. Jahn, S. Dutkiewicz, and M. J. Follows, "Global patterns in marine organic matter stoichiometry driven by phytoplankton ecophysiology," *Nature Geoscience*, vol. 15, no. 12, pp. 1034–1040, 2022.

[60] P. K. Dunstan, S. D. Foster, E. King, J. Risbey, T. J. O'Kane, D. Monselesan, A. J. Hobday, J. R. Hartog, and P. A. Thompson, "Global patterns of change and variation in sea surface temperature and chlorophyll a," *Scientific reports*, vol. 8, no. 1, p. 14624, 2018.

[61] A. C. Spivak, J. Sanderman, J. L. Bowen, E. A. Canuel, and C. S. Hopkinson, "Global-change controls on soil-carbon accumulation and loss in coastal vegetated ecosystems," *Nature Geoscience*, vol. 12, no. 9, pp. 685–692, 2019.

[62] C. J. Gobler, "Climate change and harmful algal blooms: insights and perspective," *Harmful algae*, vol. 91, p. 101731, 2020.

[63] A. B. Burd and G. A. Jackson, "An analysis of water column distributions in florida bay," *Estuaries*, vol. 25, no. 4, pp. 570–585, 2002.

[64] H. O. Briceño and J. N. Boyer, "Climatic controls on phytoplankton biomass in a sub-tropical estuary, florida bay, usa," *Estuaries and Coasts*, vol. 33, no. 2, pp. 541–553, 2010.

[65] T. C. Malone and A. Newton, "The globalization of cultural eutrophication in the coastal ocean: causes and consequences," *Frontiers in Marine Science*, p. 670, 2020.

[66] E. Galbraith, P. Frade, and M. Convertino, "Metabolic shifts of oceans: Summoning bacterial interactions," *Ecological Indicators*, vol. 138, p. 108871, 2022.

[67] X. F. Wang, "Complex networks: topology, dynamics and synchronization," *International journal of bifurcation and chaos*, vol. 12, no. 05, pp. 885–916, 2002.

[68] A. Ilany, A. S. Booms, and K. E. Holekamp, "Topological effects of network structure on long-term social network dynamics in a wild mammal," *Ecology letters*, vol. 18, no. 7, pp. 687–695, 2015.

[69] J. Li and M. Convertino, "Inferring ecosystem networks as information flows," *Scientific reports*, vol. 11, no. 1, pp. 1–22, 2021.

[70] J. C. Reijneveld, S. C. Ponten, H. W. Berendse, and C. J. Stam, "The application of graph theoretical analysis to complex networks in the brain," *Clinical neurophysiology*, vol. 118, no. 11, pp. 2317–2331, 2007.

[71] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "Complex networks in climate dynamics," *The European Physical Journal Special Topics*, vol. 174, no. 1, pp. 157–179, 2009.

[72] M. Coscia, F. Giannotti, and D. Pedreschi, "A classification for community discovery methods in complex networks," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 4, no. 5, pp. 512–546, 2011.

[73] J. H. Feldhoff, S. Lange, J. Volkholz, J. F. Donges, J. Kurths, and F.-W. Gerstengarbe, "Complex networks for climate model evaluation with application to statistical versus dynamical modeling of south american climate," *Climate dynamics*, vol. 44, no. 5, pp. 1567–1581, 2015.

[74] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths, "Complex network approaches to nonlinear time series analysis," *Physics Reports*, vol. 787, pp. 1–97, 2019.

[75] J. Zhang and M. Small, "Complex network from pseudoperiodic time series: Topology versus dynamics," *Physical review letters*, vol. 96, no. 23, p. 238701, 2006.

[76] D. Gfeller, P. De Los Rios, A. Caflisch, and F. Rao, "Complex network analysis of free-energy landscapes," *Proceedings of the National Academy of Sciences*, vol. 104, no. 6, pp. 1817–1822, 2007.

[77] J. Lin and Y. Ban, "Complex network topology of transportation systems," *Transport reviews*, vol. 33, no. 6, pp. 658–685, 2013.

[78] C. Zhou, L. Zemanová, G. Zamora, C. C. Hilgetag, and J. Kurths, "Hierarchical organization unveiled by functional connectivity in complex brain networks," *Physical review letters*, vol. 97, no. 23, p. 238103, 2006.

[79] C. Zhou, L. Zemanová, G. Zamora-Lopez, C. C. Hilgetag, and J. Kurths, "Structure–function relationship in complex brain networks expressed by hierarchical synchronization," *New Journal of Physics*, vol. 9, no. 6, p. 178, 2007.

[80] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[81] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[82] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, "Characterization of complex networks: A survey of measurements," *Advances in physics*, vol. 56, no. 1, pp. 167–242, 2007.

[83] M. E. Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.

[84] J. S. Andrade Jr, H. J. Herrmann, R. F. Andrade, and L. R. Da Silva, "Apollonian networks: Simultaneously scale-free, small world, euclidean, space filling, and with matching graphs," *Physical review letters*, vol. 94, no. 1, p. 018702, 2005.

[85] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "Pseudofractal scale-free web," *Physical review E*, vol. 65, no. 6, p. 066122, 2002.

[86] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, "Stochastic models for the web graph," in *Proceedings 41st Annual Symposium on Foundations of Computer Science*. IEEE, 2000, pp. 57–65.

[87] J. Hidalgo, J. Grilli, S. Suweis, M. A. Munoz, J. R. Banavar, and A. Maritan, "Information-based fitness and the emergence of criticality in living systems," *Proceedings of the National Academy of Sciences*, vol. 111, no. 28, pp. 10 095–10 100, 2014.

[88] E. Ramírez-Carrillo, O. López-Corona, J. C. Toledo-Roy, J. C. Lovett, F. de León-González, L. Osorio-Olvera, J. Equihua, E. Robredo, A. Frank, R. Dirzo *et al.*, "Assessing sustainability in north america's ecosystems using criticality and information theory," *PloS one*, vol. 13, no. 7, p. e0200382, 2018.

[89] M. Convertino and L. J. Valverde Jr, "Toward a pluralistic conception of resilience," *Ecological Indicators*, vol. 107, p. 105510, 2019.

[90] J. Li and M. Convertino, "Temperature increase drives critical slowing down of fish ecosystems," *PLoS One*, vol. 16, no. 10, p. e0246222, 2021.

[91] M. Convertino, A. Reddy, Y. Liu, and C. Munoz-Zanzi, "Eco-epidemiological scaling of leptospirosis: Vulnerability mapping and early warning forecasts," *Science of The Total Environment*, vol. 799, p. 149102, 2021.

[92] J. E. Cloern and A. D. Jassby, "Patterns and scales of phytoplankton variability in estuarine–coastal ecosystems," *Estuaries and coasts*, vol. 33, no. 2, pp. 230–241, 2010.

[93] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[94] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000.

[95] Z. Duan, H. Xu, Y. Huang, J. Feng, and Y. Wang, "Multivariate time series forecasting with transfer entropy graph," *Tsinghua Science and Technology*, vol. 28, no. 1, pp. 141–149, 2022.

[96] M. Martinello, J. Hidalgo, A. Maritan, S. Di Santo, D. Plenz, and M. A. Muñoz, "Neutral theory and scale-free neural dynamics," *Physical Review X*, vol. 7, no. 4, p. 041071, 2017.

[97] L. Conti and M. Scardi, "Fisheries yield and primary productivity in large marine ecosystems," *Marine Ecology Progress Series*, vol. 410, pp. 233–244, 2010.

[98] C. J. Madden, D. T. Rudnick, A. A. McDonald, K. M. Cunniff, and J. W. Fourqurean, "Ecological indicators for assessing and communicating seagrass status and trends in florida bay," *Ecological Indicators*, vol. 9, no. 6, pp. S68–S82, 2009.

[99] J. N. Boyer, J. W. Fourqurean, and R. D. Jones, "Spatial characterization of water quality in florida bay and whitewater bay by multivariate analyses: zones of similar influence," *Estuaries*, vol. 20, pp. 743–758, 1997.

[100] ——, "Seasonal and long-term trends in the water quality of florida bay (1989–1997)," *Estuaries*, vol. 22, pp. 417–430, 1999.

[101] S. Dai, Y. Zhou, N. Li, and X.-z. Mao, "Why do red tides occur frequently in some oligotrophic waters? analysis of red tide evolution history in mirs bay, china and its implications," *Science of The Total Environment*, vol. 844, p. 157112, 2022.

[102] M. Hoerling and A. Kumar, "The perfect ocean for drought," *Science*, vol. 299, no. 5607, pp. 691–694, 2003.

[103] D. T. Rudnick, Z. Chen, D. Childers, and T. Fontaine, "Phosphorus and nitrogen inputs to florida bay: the importance of the everglades watershed," *Estuaries*, vol. 22, pp. 398–416, 1999.

[104] D. F. Carlson, L. A. Yarbro, S. Scolaro, M. Poniatowski, V. McGee-Absten, and P. R. Carlson Jr, "Sea surface temperatures and seagrass mortality in florida bay: spatial and temporal patterns discerned from modis and avhrr data," *Remote Sensing of Environment*, vol. 208, pp. 171–188, 2018.

[105] L. Oey, T. Ezer, and H. Lee, "Loop current, rings and related circulation in the gulf of mexico: A review of numerical models and future challenges," *Geophysical Monograph-American Geophysical Union*, vol. 161, p. 31, 2005.

[106] J. Zieman, J. W. Fourqurean, and R. L. Iverson, "Distribution, abundance and productivity of seagrasses and macroalgae in florida bay," *Bulletin of marine science*, vol. 44, no. 1, pp. 292–311, 1989.

[107] E. Phlips, T. Lynch, and S. Badylak, "Chlorophyll a, tripton, color, and light availability in a shallow tropical inner-shelf lagoon, florida bay, usa," *Marine Ecology Progress Series*, vol. 127, pp. 223–234, 1995.

[108] T. J. Smith III, H. J. Hudson, M. B. Robblee, G. N. Powell, and P. J. Isdale, "Freshwater flow from the everglades to florida bay: a historical reconstruction based on fluorescent banding in the coral solenastrea bournoni," *Bulletin of Marine Science*, vol. 44, no. 1, pp. 274–282, 1989.

[109] J. W. Fourqurean and M. B. Robblee, "Florida bay: a history of recent ecological changes," *Estuaries*, vol. 22, pp. 345–357, 1999.

[110] B. Richardson, "Physiological characteristics and competitive strategies of bloom-forming cyanobacteria and diatoms in florida bay," *Bull. Mar. Sci*, vol. 38, pp. 19–36, 2009.

[111] P. Glibert, C. A. Heil, D. Hollander, M. Revilla, A. Hoare, J. Alexander, and S. Murasko, "Evidence for dissolved organic nitrogen and phosphorus uptake during a cyanobacterial bloom in florida bay," *Marine Ecology Progress Series*, vol. 280, pp. 73–83, 2004.

[112] J. N. Boyer, S. K. Dailey, P. J. Gibson, M. T. Rogers, and D. Mir-Gonzalez, "The role of dissolved organic matter bioavailability in promoting phytoplankton blooms in florida bay," *Hydrobiologia*, vol. 569, pp. 71–85, 2006.

[113] P. J. Lavrentyev, H. A. Bootsma, T. H. Johengen, J. F. Cavaletto, and W. S. Gardner, "Microbial plankton response to resource limitation: insights from the community structure and seston stoichiometry in florida bay, usa," *Marine Ecology Progress Series*, vol. 165, pp. 45–57, 1998.

[114] B. S. Cade and B. R. Noon, "A gentle introduction to quantile regression for ecologists," *Frontiers in Ecology and the Environment*, vol. 1, no. 8, pp. 412–420, 2003.

[115] S. Irion, U. Christaki, H. Berthelot, S. L'helguen, and L. Jardillier, "Small phytoplankton contribute greatly to co2-fixation after the diatom bloom in the southern ocean," *The ISME Journal*, vol. 15, no. 9, pp. 2509–2522, 2021.

[116] T. Hirata, N. Hardman-Mountford, R. Brewin, J. Aiken, R. Barlow, K. Suzuki, T. Isada, E. Howell, T. Hashioka, M. Noguchi-Aita *et al.*, "Synoptic relationships quantified between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types." *Biogeosciences Discussions*, vol. 7, no. 5, 2010.

[117] A. T. Chan, "Comparative physiological study of marine diatoms and dinoflagellates in relation to irradiance and cell size. ii. relationship between photosynthesis, growth, and carbon/chlorophyll a ratio 1, 2," *Journal of phycology*, vol. 16, no. 3, pp. 428–432, 1980.

[118] W. K. Nuttle, J. W. Fourqurean, B. J. Cosby, J. C. Zieman, and M. B. Robblee, "Influence of net freshwater supply on salinity in florida bay," *Water Resources Research*, vol. 36, no. 7, pp. 1805–1822, 2000.

[119] T. N. Lee, E. Johns, N. Melo, R. H. Smith, P. Ortner, and D. Smith, "On florida bay hypersalinity and water exchange," *Bulletin of Marine Science*, vol. 79, no. 2, pp. 301–327, 2006.

[120] C. R. Kelble, E. M. Johns, W. K. Nuttle, T. N. Lee, R. H. Smith, and P. B. Ortner, "Salinity patterns of florida bay," *Estuarine, Coastal and Shelf Science*, vol. 71, no. 1-2, pp. 318–334, 2007.

[121] J. A. Goleski, F. Koch, M. A. Marcoval, C. C. Wall, F. J. Jochem, B. J. Peterson, and C. J. Gobler, "The role of zooplankton grazing and nutrient loading in the occurrence of harmful cyanobacterial blooms in florida bay, usa," *Estuaries and Coasts*, vol. 33, pp. 1202–1215, 2010.

[122] F. Pennekamp, A. C. Iles, J. Garland, G. Brennan, U. Brose, U. Gaedke, U. Jacob, P. Kratina, B. Matthews, S. Munch *et al.*, "The intrinsic predictability of ecological time series and its potential to guide forecasting," *Ecological Monographs*, vol. 89, no. 2, p. e01359, 2019.

[123] S. V. Scarpino and G. Petri, "On the predictability of infectious disease outbreaks," *Nature communications*, vol. 10, no. 1, p. 898, 2019.

[124] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara, "Early-warning signals for critical transitions," *Nature*, vol. 461, no. 7260, pp. 53–59, 2009.

[125] M. Hirota, M. Holmgren, E. H. Van Nes, and M. Scheffer, "Global resilience of tropical forest and savanna to critical transitions," *Science*, vol. 334, no. 6053, pp. 232–235, 2011.

[126] R. Quax, D. Kandhai, and P. Sloot, "Information dissipation as an early-warning signal for the lehman brothers collapse in financial time series," *Scientific reports*, vol. 3, no. 1, pp. 1–7, 2013.

[127] M. Ushio, K. Watanabe, Y. Fukuda, Y. Tokudome, and K. Nakajima, "Computational capability of ecological dynamics," *Royal Society Open Science*, vol. 10, no. 4, p. 221614, 2023.

[128] L. J. Valverde Jr and M. Convertino, "Insurer resilience in an era of climate change and extreme weather: an econometric analysis," *Climate*, vol. 7, no. 4, p. 55, 2019.

[129] G. M. Hallegraeff, D. M. Anderson, C. Belin, M.-Y. D. Bottein, E. Bresnan, M. Chinain, H. Enevoldsen, M. Iwataki, B. Karlson, C. H. McKenzie *et al.*, "Perceived global increase in algal blooms is attributable to intensified monitoring and emerging bloom impacts," *Communications Earth & Environment*, vol. 2, no. 1, p. 117, 2021.

[130] P. Urrutia-Cordero, M. K. Ekvall, and L.-A. Hansson, "Controlling harmful cyanobacteria: taxa-specific responses of cyanobacteria to grazing by large-bodied daphnia in a biomanipulation scenario," *PLoS One*, vol. 11, no. 4, p. e0153032, 2016.

[131] S. P. Dasgupta, "The economics of biodiversity the dasgupta review abridged version," 2021.

[132] A. Bastos, S. Sippel, D. Frank, M. D. Mahecha, S. Zaehle, J. Zscheischler, and M. Reichstein, "A joint framework for studying compound ecoclimatic events," *Nature Reviews Earth & Environment*, pp. 1–18, 2023.

[133] R. D. Batt, T. Eason, and A. Garmestani, "Time scale of resilience loss: implications for managing critical transitions in water quality," *Plos one*, vol. 14, no. 10, p. e0223366, 2019.

[134] S. E. Fire, L. J. Flewelling, J. Naar, M. J. Twiner, M. S. Henry, R. H. Pierce, D. P. Gannon, Z. Wang, L. Davidson, and R. S. Wells, "Prevalence of brevetoxins in prey fish of bottlenose dolphins in sarasota bay, florida," *Marine Ecology Progress Series*, vol. 368, pp. 283–294, 2008.

[135] D. P. Gannon, E. J. B. McCabe, S. A. Camilleri, J. G. Gannon, M. K. Brueggen, A. A. Barleycorn, V. I. Palubok, G. J. Kirkpatrick, and R. S. Wells, "Effects of karenia brevis harmful algal blooms on nearshore fish communities in southwest florida," *Marine Ecology Progress Series*, vol. 378, pp. 171–186, 2009.

[136] C. A. Heil and A. L. Muni-Morgan, "Florida's harmful algal bloom (hab) problem: Escalating risks to human, environmental and economic health with climate change," *Frontiers in Ecology and Evolution*, vol. 9, p. 646080, 2021.

[137] M. Jenkins, S. Ahmed, and A. N. Barnes, "A systematic review of waterborne and water-related disease in animal populations of florida from 1999–2019," *Plos one*, vol. 16, no. 7, p. e0255025, 2021.

# Chapter 6

# Appendix

# Appendix A

# Supplement for Chapter 4

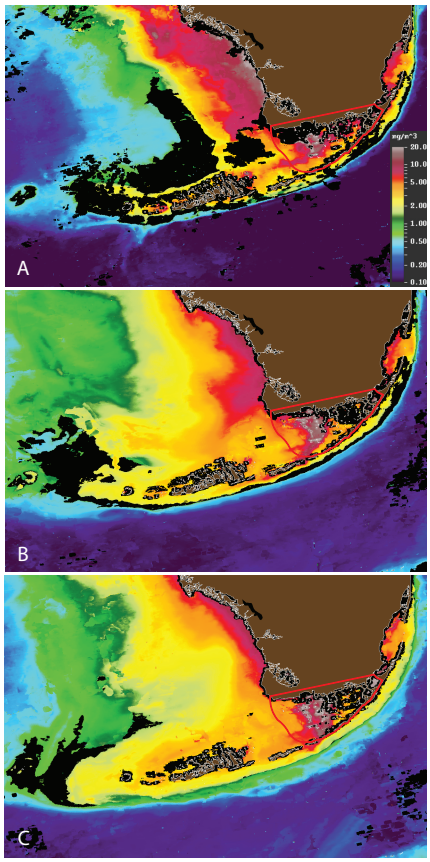


Figure S1: **Composite seven-day mean of CHL-a concentration (in $mg/m^3$) in 168 the surface ocean layer of South-West Florida.** CHL-a, for 7-day mean of October-November-December 2005 (A, B, C), is estimated by the Optical Oceanography Laboratory at USF (`https://optics.marine.usf.edu/cgi-bin/optics_data?roi=FLKEYS&Date=10/15/2005#C20052822005288.QKM.FLKEYS.7DAY.L3D.SST.png`). Estimates, in $mug/L = mg/m^3$ are done with the most updated calibration and algorithms in the SeaDAS processing software (SeaDAS is a comprehensive software package for the processing, display, analysis, and quality control of ocean color data, `https://seadas.gsfc.nasa.gov/`) based on multiple reflectance indicators for deep [1] and shallow waters [2] Over clear, shallow waters ($<30m$) or over very turbid coastal waters or river plumes, it is often overestimated as other components (colored dissolved organic matter, suspended sediments, ocean bottom) interfere with the algorithm.

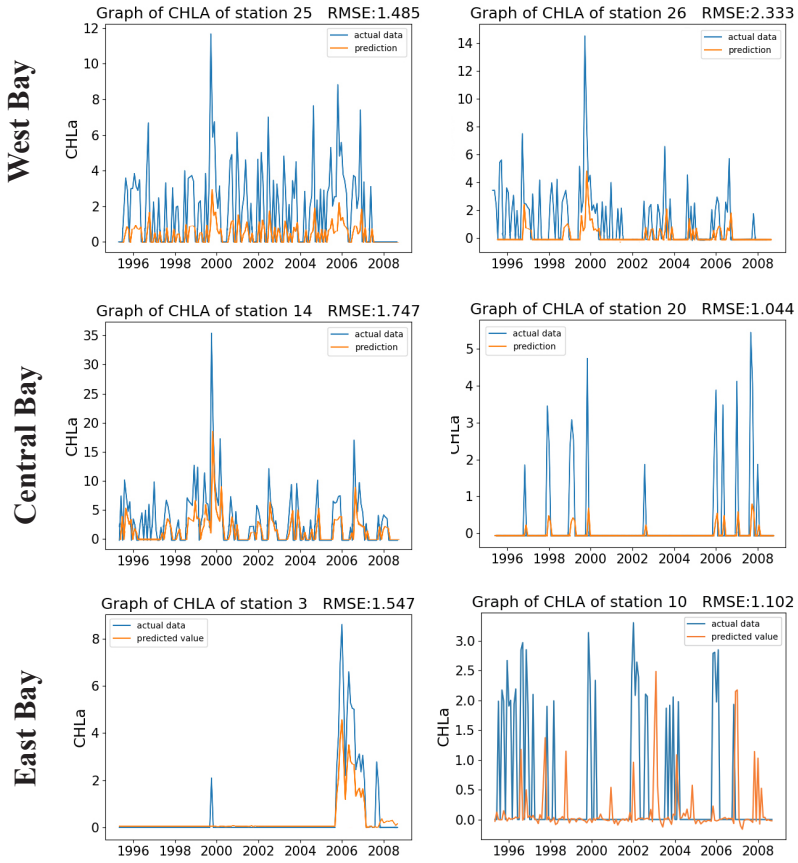


Figure S2: **Bloom Forecasting for representative West, Central and East stations.** In the West stations 20, 25, and 26 as low risk (despite being endemic for blooms) that are all very far from the coast with respect to high risk habitats, manifesting the lower ocean-pressure and the risk-sink character of these sites. In East and Central stations 3, 14 and 10 as high risk (where 3 is epidemic and the last two are endemic) that are all very coastal with respect to low risk habitats, manifesting the critical land-pressure effect. Thus, bloom risk should be assessed also on bloom persistence and shift and not just on average magnitude or extreme outbreaks to incorporate dynamical features of blooms. It is noticeable how for epidemic dynamics is baselined by the highest environmental complexity and implies the highest forecasting skills considering the coincidence of magnitude, timing and duration of observed and predicted CHL-a (from top to bottom plots).

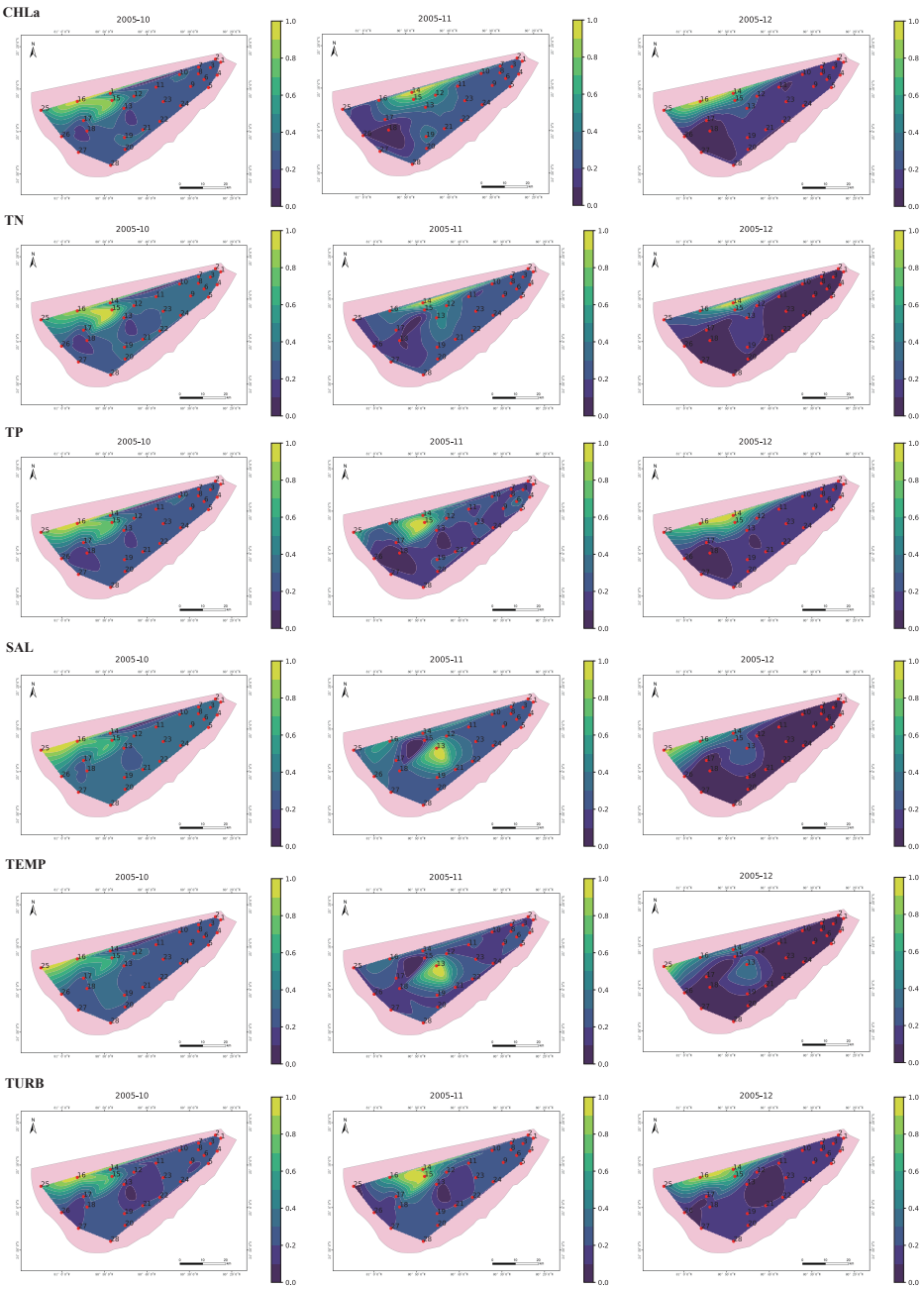


Figure S3: **Spatial Biogeochemical Patterns for the Largest FL-Bay Bloom.** Patterns for the 2005 bloom that is the second largest and most widespread bloom up to date. Patterns are similar to those of the 1999 bloom (Fig. 4.7).

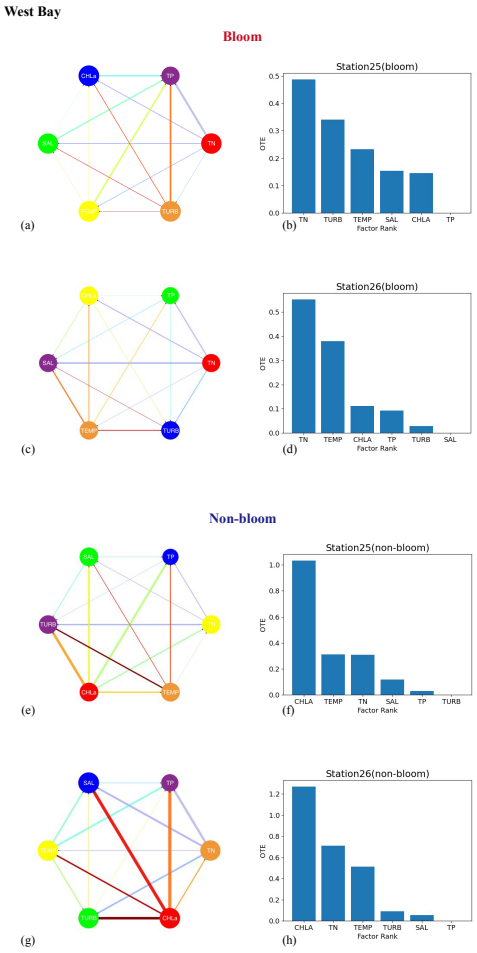


Figure S4: **Biogeochemical network and variable influence for representative mangrove habitats**. Causality network and OTE of biogeochemical variables for representative stations 25 and 26 in the western region under bloom and non-bloom regime. The size of each node is proportional to the Shannon Entropy of each variables, while the color is proportional to OTE. The higher OTE, the warmer the color. The width of each edge is proportional to $te_{i,j}(X)$, while the color is just to distinguish links. The direction is related to the dominant causality of factors $X_i$ and $X_j$, i.e. $TE_{X_i,X_j}$ (thresholded TE difference). The distance is related to the node configuration in a circle, made through the Python package "networkx layout". For visualization clarity all nodes in bloom condition are magnified 1000 times, and the nodes in non-bloom condition are magnified 100 times. Width edges are magnified 30 times.

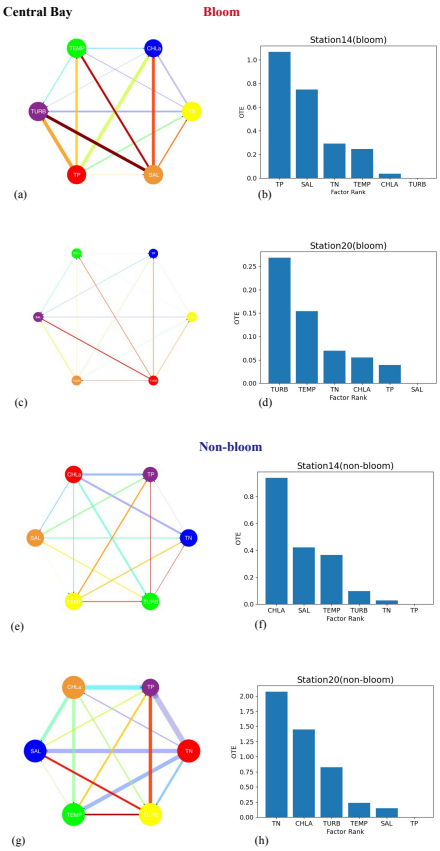


Figure S5: **Biogeochemical network and variable influence for representative marsh habitats**. Causality network and OTE of biogeochemical variables for representative stations 14 and 20 in the central region under bloom and non-bloom regime. The size of each node is proportional to the Shannon Entropy of each variables, while the color is proportional to OTE. The higher OTE, the warmer the color. The width of each edge is proportional to $te_{i,j}(X)$, while the color is just to distinguish links. The direction is related to the dominant causality of factors $X_i$ and $X_j$, i.e. $TE_{X_i,X_j}$ (thresholded TE difference). The distance is related to the node configuration in a circle, made through the Python package "networkx layout". For visualization clarity all nodes in bloom condition are magnified 1000 times, and the nodes in non-bloom condition are magnified 100 times. Width edges are magnified 30 times.

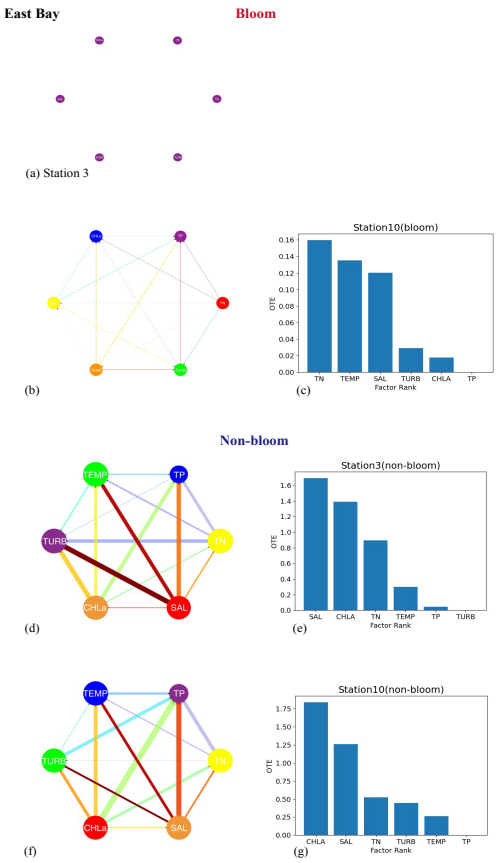


Figure S6: **Biogeochemical network and variable influence for representative coastal seagrass habitats**. Causality network and OTE of biogeochemical variables for representative stations 3 and 10 in the central and eastern region under bloom and non-bloom regime. The size of each node is proportional to the Shannon Entropy of each variables, while the color is proportional to OTE. The higher OTE, the warmer the color. The width of each edge is proportional to $te_{i,j}(X)$, while the color is just to distinguish links. The direction is related to the dominant causality of factors $X_i$ and $X_j$, i.e. $TE_{X_i,X_j}$ (thresholded TE difference). The distance is related to the node configuration in a circle, made through the Python package "networkx layout". For visualization clarity all nodes in bloom condition are magnified 1000 times, and the nodes in non-bloom condition are magnified 100 times. Width edges are magnified 30 times.

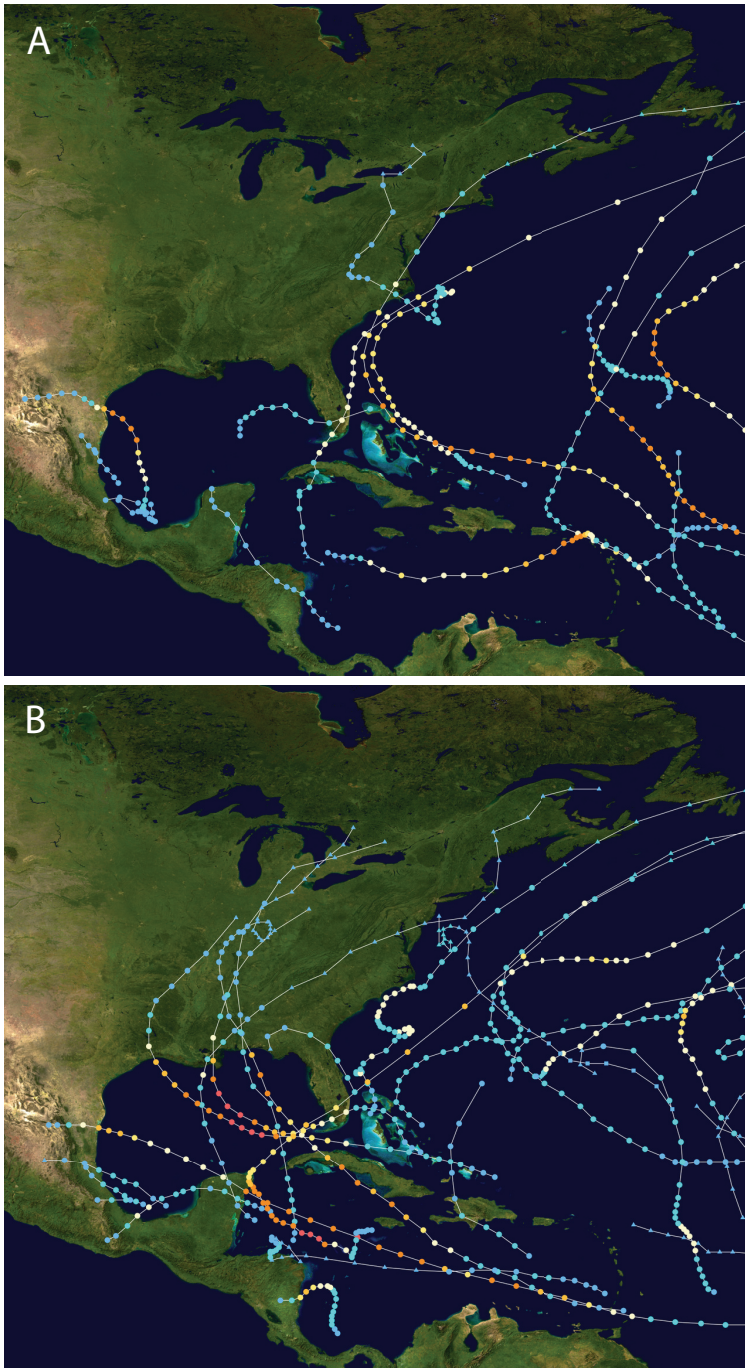


Figure S7: **North Atlantic basin tropical cyclones in 1999 and 2005**. All tropical cyclone tracks are shown for 1999 (A) and 2005 (B). Colors are proportional to cyclone's intensity at peak in each node, from category 1 to 5 (from light blue to red).