More than co-occurrence: what amplicon time series data can tell us

Ellen Oldenburg, Nima P. Saadat, Sofie Thielen, Ovidiu Popa

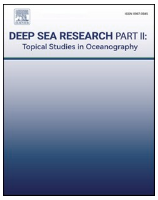
Suggested Citation:
Oldenburg, E., Saadat, N., Thielen, S., & Popa, O. (2026). More than co-occurrence: what amplicon time series data can tell us. Deep-Sea Research 2, 226, Article 105599.
https://doi.org/10.1016/j.dsr2.2026.105599

Deep–Sea Research II 226 (2026) 105599

Contents lists available at ScienceDirect

Deep-Sea Research Part II

journal homepage: www.elsevier.com/locate/dsr2



More than co-occurrence: what amplicon time series data can tell us

Ellen Oldenburg [a,b,c], Nima P. Saadat [a], Sofie Thielen [a], Ovidiu Popa [a,*]

[a] *Institute of Quantitative and Theoretical Biology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany*
[b] *Cluster of Excellence on Plant Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany*
[c] *Deep-Sea Ecology and Technology, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany*

ABSTRACT

The Arctic Ocean is undergoing unprecedented transformations due to climate change, with rising temperatures, shrinking sea ice, and shifting oceanographic conditions reshaping its ecosystem. At the center of these changes lie the Arctic's microbial communities that drive biogeochemical cycles, sustain primary production, and maintain ecosystem stability. Long-term ecological research is essential for understanding microbial community dynamics and their role in biogeochemical cycles, particularly in polar ecosystems. The HAUSGARTEN observatory, established in 1999, has provided unparalleled insights into Arctic marine microbial ecology. This review synthesizes 25 years of microbial re-search at HAUSGARTEN, revealing how advanced methodologies—such as amplicon sequencing, which has revolutionized microbial ecology by enabling the taxonomic characterization of complex communities through targeted marker genes, Fourier decomposition, Convergent Cross Mapping (CCM), and Energy Landscape Analysis (ELA)—have revolutionized our understanding of Arctic microbial ecology. By integrating time-series data with network-based approaches, we move beyond static snapshots to uncover the hidden rhythms of microbial life, from seasonal successions to long-term trends. We explore the interplay between environmental drivers and microbial community structure, emphasizing seasonal succession, functional adaptations, and the impact of Atlantification. Environmental conditions are constantly changing; therefore, there is a need for predictive models. By combining machine learning, deterministic modeling, and ecological theory, we are now poised to forecast how microbial communities will respond to future climate scenarios. From Graph Neural Networks (GNNs) to ARIMA forecasting, this review showcases the power of amplicon data in a time series frame work together with interdisciplinary approaches to tackle one of the most pressing challenges of our time. Arctic microbial communities are the key to understanding and mitigating the impacts of climate change, and this review is a guide to unlocking their secrets.

1. Introduction

The HAUSGARTEN observatory in the Fram Strait provides one of the longest time-series records of Arctic marine ecosystem dynamics (Soltwedel et al., 2016). Established in 1999, this Long- Term Ecological Research (LTER) site has facilitated the study of microbial communities in response to environmental shifts (Soltwedel et al., 2016). While amplicon sequencing has become a cornerstone of microbial ecology (Luisa and Anders, 2017; Knight et al., 2018), traditional snapshot analyses fail to capture the temporal dynamics necessary for understanding ecosystem stability and climate-driven changes (Weinreb et al., 2018; Schmid et al., 2011).

This review, titled ’More than Co-Occurrence: What Amplicon Time-Series Data Can Tell Us’, summarizes the state-of-the-art methodologies used to investigate 25 years of microbial time-series research at HAUSGARTEN. Additionally, it highlights the systematic application of interdisciplinary mathematical approaches to this type of time series data, with the aim of maximising the extraction and integration of ecological information. Rather than focusing on individual techniques, we emphasize how the combination of Co-Occurrence Networks, Convergent Cross Mapping (CCM), and other quantitative methods enhance the interpretability of microbial time-series data. Using these integrated approaches, we provide a framework for moving beyond descriptive analyses toward predictive modeling of Arctic microbial communities and their environmental drivers.

Arctic marine ecosystems are undergoing rapid transformations due to climate change, with rising temperatures, decreasing ice cover, and shifting oceanographic conditions altering the composition and function of microbial communities (Priest et al., 2023; Oldenburg et al., 2024c). These microorganisms play a fundamental role in biogeochemical cycling, primary production, and ecosystem stability (Priest et al., 2023; Oldenburg et al., 2024c). Understanding the temporal patterns and ecological interactions of Arctic microbial communities is therefore crucial for predicting how these systems will respond to ongoing environmental changes (Oldenburg et al., 2024a; Schmidt et al., 2017).

Traditional microbial ecology studies often rely on point-in-time

---

This article is part of a special issue entitled: LTER HAUSGARTEN published in Deep-Sea Research Part II.

* Corresponding author. Universitätstraße 1, 40225, Düsseldorf, Germany.
E-mail address: ovidiu.popa@hhu.de (O. Popa).

https://doi.org/10.1016/j.dsr2.2026.105599
Received 28 February 2025; Received in revised form 7 January 2026; Accepted 12 January 2026

analyses, which pro-vide valuable insights into community composition but lack the temporal resolution needed to track ecological successions, resilience, and long-term trends (Oldenburg et al., 2024b). The implementation of advanced time-series methodologies at HAUSGARTEN allows researchers to move beyond descriptive analyses and towards predictive modeling of microbial responses to environmental shifts (Oldenburg et al., 2024a; Banerjee et al., 2018). By applying long-term ecological research principles, we can identify key environmental drivers, establish causative links between microbial taxa, define keystone species and assess the functional stability of Arctic microbial networks over time (Oldenburg et al., 2024a).

Understanding microbial time-series data requires both advanced analytical approaches and a deep consideration of environmental drivers. In this review we outline the key environmental factors shaping microbial diversity in the Arctic, followed by an in-depth discussion of the analytical tools employed to interpret long-term datasets. Finally, we assess the implications of these findings for understanding microbial ecosystem resilience and offer recommendations for future research directions. By integrating time-series analyses with network-based approaches, we aim to provide a comprehensive synthesis of microbial community dynamics in one of the most rapidly changing environments on Earth. The Arctic Ocean represents a unique and highly sensitive ecosystem characterized by its semi-enclosed basin, geographic isolation, and strong seasonality. It receives inputs from both the Atlantic and Pacific Oceans, yet is markedly influenced by freshwater influx from surrounding continents, which contributes to stratification and oligotrophic conditions (Timmermans and Marshall, 2020). Seasonal phenomena such as prolonged darkness during the polar night, extensive sea-ice formation and melt, and nutrient limitation shape microbial productivity, succession, and food web dynamics (Kellogg et al., 2019). Within this dynamic system, the Fram Strait serves as the principal deep-water gateway between the North Atlantic and the Central Arctic Ocean. It

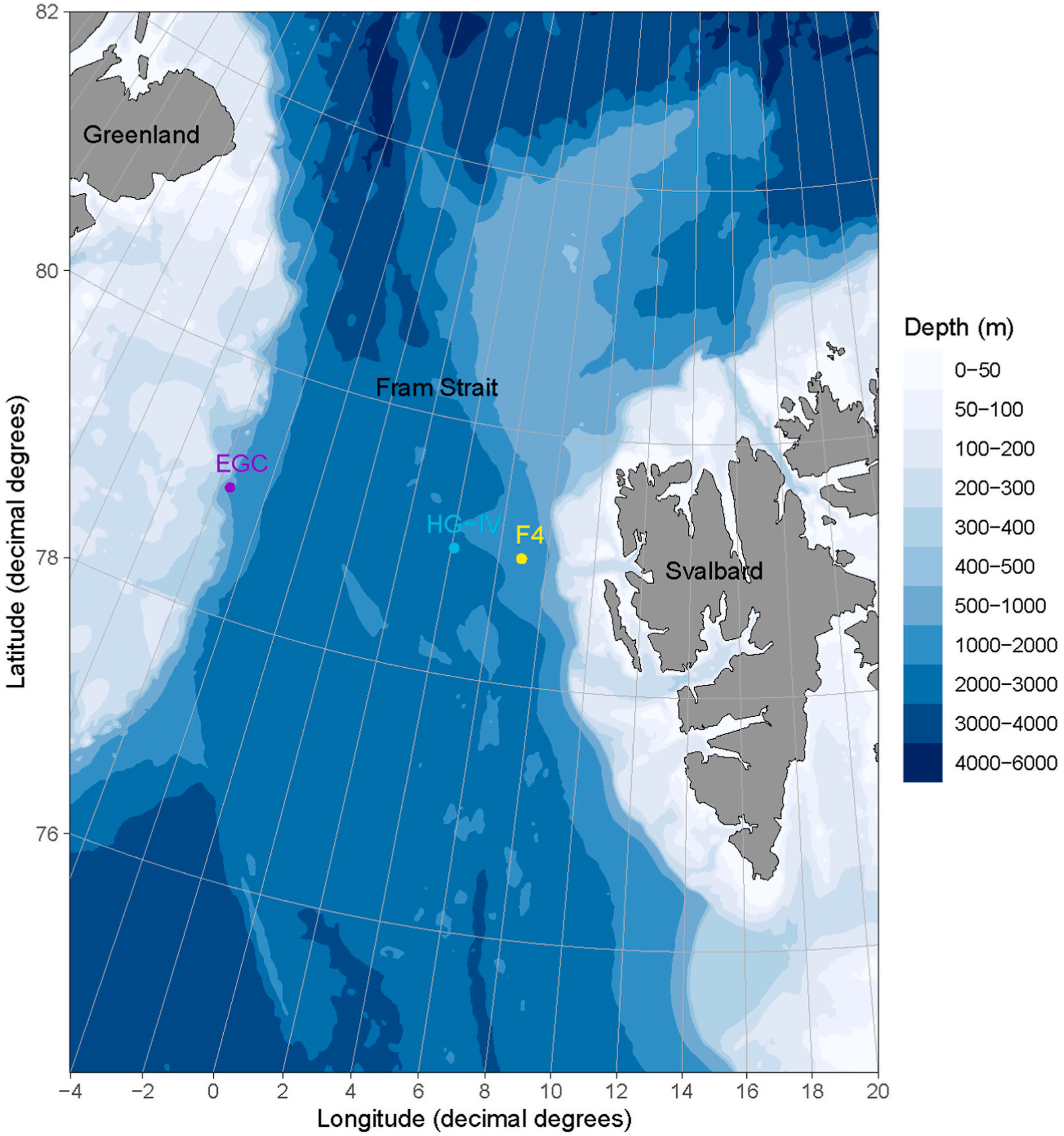


Fig. 1. Bathymetric map of the Fram Strait showing the location of the HAUSGARTEN long-term ecological research (LTER) observatory, including stations HG-IV, F4, and EGC. The map illustrates the regional setting of the Arctic-Atlantic interface and the depth gradient along the transect.

exhibits strong Atlantic inflow via the West Spitsbergen Current and polar outflow via the East Greenland Current, making it a key region to study microbial responses to Atlantification and climate-driven environmental change. The HAUSGARTEN LTER observatory, located in the Fram Strait, offers one of the most comprehensive long-term ecological records of Arctic deep-sea microbial communities. Its position at the interface of Atlantic and Arctic water masses makes it particularly suitable for assessing how microbial ecosystems respond to physical and biogeochemical shifts. Fig. 1 illustrates the geographical position of HAUSGARTEN and selected sampling stations within the Fram Strait, alongside regional bathymetry (Soltwedel et al., 2016).

2. Amplicon sequencing: balancing innovation and limitations in microbial ecosystems

Amplicon sequencing has revolutionized microbial ecology by enabling taxonomic characterization of complex microbial communities through targeted marker genes (e.g., 16S rRNA for bacteria and archaea, 18S rRNA/ITS for eukaryotes, g23 major capsid protein for phages) (Tian, 2015). This technology has provided unprecedented insights into microbial diversity, allowing researchers to uncover previously hidden aspects of community composition (Elena P' et al., 2020). By applying approaches such as principal component analysis (PCA), non-metric multidimensional scaling (NMDS), alpha and beta diversity measures, and static co-occurrence networks, it has become possible to capture snapshots of microbial community structures that were previously inaccessible. These methods have significantly advanced our understanding of microbial ecosystems by revealing key taxa, their relative abundances, and potential interactions, paving the way for further exploration of microbial dynamics over time (Paliy and Shankar, 2016). Additionally, while alpha diversity metrics describe within community richness and evenness (e.g., Shannon index, Simpson index), beta diversity assesses differentiation between samples using measures such as Bray-Curtis dissimilarity, Jaccard or UniFrac distance, providing insights into how microbial communities vary across environmental gradients or temporal scales (Anderson, 2001; Lozupone and Knight, 2005). These methods have significantly advanced our understanding of microbial ecosystems by revealing key taxa, their relative abundances, and potential interactions, paving the way for further exploration of microbial dynamics over time (Lozupone and Knight, 2008). However, amplicon sequencing data come with inherent limitations, particularly regarding the interpretation of abundance data. The sequencing output does not reflect absolute cell counts but rather the relative abundance of marker genes within a sample. This is problematic because different taxa may have varying numbers of rRNA gene copies, leading to biases in inferred microbial abundances (Větrovsky', Baldrian, 2013; Kembel et al., 2012; Klappenbach et al., 2000). Consequently, relative abundance measures must be carefully considered in statistical analyses, as they introduce compositionality constraints—Meaning that the rel-ative abundance of each taxon depends on the proportions of all others, which can distort ecological interpretations by creating artificial correlations or obscuring true abundance patterns (Weiss et al., 2017).

To address these issues, several statistical approaches and bioinformatics tools have been developed to correct for compositional biases and improve the robustness of microbial community analyses. Methods such as SparCC (Sparse Correlations for Compositional Data), Log-Ratio Transformations (e.g., Centered Log-Ratio, lgclr), and Bayesian based differential abundance models like Aldex2 allow for more accurate interpretation of relative abundance data in ecological studies (Tsilimigras and Fodor, 2016; Fernandes et al., 2014). These tools mitigate spurious correlations and enhance the reliability of inferred microbial interactions (Tsilimigras and Fodor, 2016), making them essential for modern amplicon-based analyses. Another critical challenge of amplicon sequencing is its inherent lack of functional information. While there are efforts to develop functional profiling tools (e.g., PICRUSt, Tax4Fun) that allow researchers to infer metabolic potential and ecosystem functions from abundance data, these tools are mostly limited or biased to particular environments. This limitation can be attributed to the bias inherent in the present sequencing data, which is mostly human-associated microbiome data (The Human Microbiome Project Consortium, 2012; Turnbaugh et al., 2007; Aβhauer et al., 2015). Gene databases (such as SILVA, Greengenes, and the Ribosomal Database Project) contain therefore a disproportionately high number of sequences from human-associated environments. This bias stems from the extensive research and clinical interest in human health and disease, which drives more frequent sampling and sequencing of these niches compared to many environmental habitats. Further biases are also due to incomplete reference databases, lack of strain-specific metabolic variability, and the assumption that phylogeny accurately reflects functional potential (Pilar Cabezas et al., 2024). While taxonomic identification is valuable, it hardly infers metabolic roles or ecological functions (Louca et al., 2016). This limitation complicates the interpretation of microbial interactions, as co-occurrence alone does not confirm ecological relevance (Bin et al., 2016). For example, detecting the simultaneous presence of two taxa in a dataset does not indicate a functional relationship, such as metabolic cross-feeding or competitive exclusion. This gap underscores the need for integrative approaches that move beyond taxonomic profiling toward a mechanistic understanding of microbial interactions (Faust and Raes, 2012).

In Arctic microbial studies, the choice of genetic markers significantly influences ob-served community composition. Amplicon-based techniques using the 18S rRNA v4 or v9 regions are prone to amplification bias and copy number variation, leading to divergent taxonomic profiles (Romero et al., 2025; Jamy et al., 2025). Shotgun metagenomics and the reconstruction of metagenome-assembled genomes (MAGs) offer PCR-free alternatives that enable both taxonomic and functional resolution. Although more resource-intensive, MAGs have recently uncovered novel Arctic lineages and metabolic traits (Royo-Llonch et al., 2021). These complementary approaches support the findings from the amplicon data and underscore the need for multi-method frameworks to fully capture microbial diversity and function in polar environments (Romero et al., 2025; Royo-Llonch et al., 2021; Jamy et al., 2025). Integrating amplicon data into a time-series framework markedly enhances its informational value by capturing dynamic temporal patterns. This approach transcends static snapshots, enabling the application of advanced analytical methods—such as Convergent Cross Mapping (CCM), Fourier transformation, Energy Landscape Algorithms (ELA), and ecological network analyses—to elucidate microbial dynamics, infer causal relationships, assess community stability, and forecast responses to environmental perturbations. All these methodological steps are visualized in Fig. 2. By revealing time-dependent microbial interactions, these techniques provide a more nuanced understanding of the ecological drivers and long-term trends governing microbial communities. Ultimately, the transition from static, abundance-based analyses to dynamic ecological modeling represents a significant advancement, offering a high resolution perspective on ecosystem processes as derived from amplicon data.

To support readers unfamiliar with advanced computational techniques, we provide a comparative summary of the core methods used in this review. Table 1 outlines their conceptual basis, strengths, limitations, and typical applications in Arctic microbial ecology.

While amplicon data provide powerful insights into the structure and dynamics of microbial communities, they often fall short in capturing the functional roles these communities perform in the ecosystem. To address this gap, recent approaches have focused on the integration of paired measurements, where amplicon sequencing is combined with functional or biogeochemical indicators such as chlorophyll concentration, bacterial pro-duction, oxygen fluxes, or toxin levels. Moreover, the transformation of relative into absolute abundances through pairing with microscopy or flow cytometric cell counts enables more quantitative ecological interpretation. Although such data are often difficult to obtain over long timescales, they have been successfully applied in

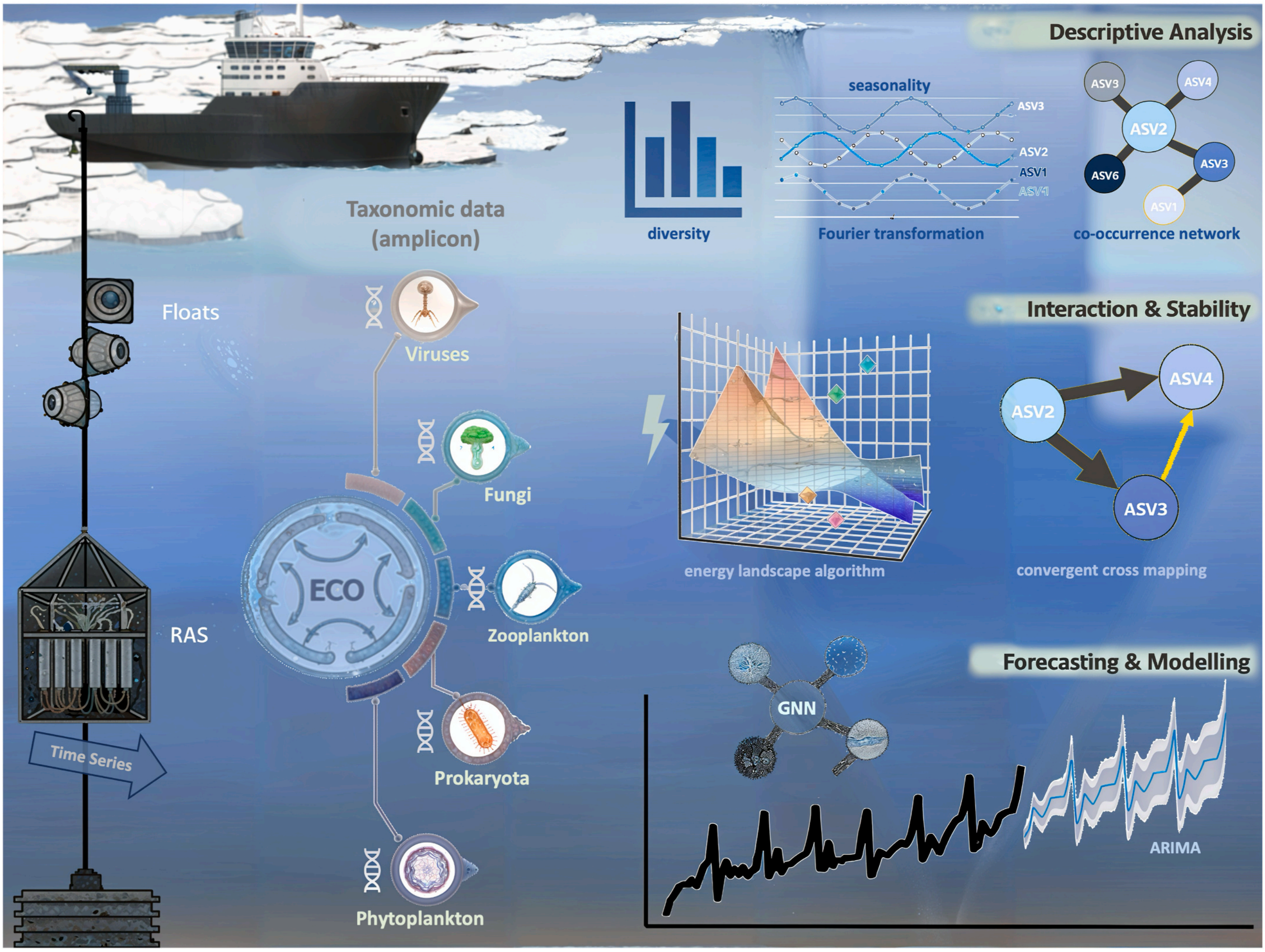


Fig. 2. Overview of analytical pathways The left panel illustrates long-term monitoring through a mooring system deployed from a research vessel, continuously collecting data over time. The middle panel highlights the diversity of amplicon-derived biological data, spanning viruses, fungi, zooplankton, prokaryotes, and phytoplankton. The right panel presents the stepwise analytical workflow, where data can be processed through descriptive analyses (e.g., alpha and beta diversity) or co-occurrence networks as an initial step. These outputs can then be used either independently or sequentially for interaction and stability assessments (e.g., Energy Landscape Analysis (ELA) and Convergent Cross Mapping (CCM)). Finally, the results from any of these analyses can be integrated or used separately for forecasting and modeling approaches, including Ordinary Differential Equations (ODE), Autoregressive Integrated Moving Average (ARIMA), and Graph Neural Networks (GNNs).

shorter campaigns and offer valuable calibration points for interpreting relative sequence data. Importantly, machine learning frameworks can be employed to build models linking taxonomic profiles with ecosystem-level processes. As highlighted by Ghannam and Techtmann (Ryan et al., 2021), even relatively small paired datasets can be used to train models that identify taxa-function associations, which can then be projected onto longer amplicon-based time series. This strategy allows for inference of microbial ecosystem functions beyond the limits of direct observation and enables prediction of critical rates or ecological states over extended periods. Such hybrid modeling approaches—when paired with rich contextual data—have the potential to reveal key drivers of microbial ecosystem services in the Arctic, even in data-limited environments such as the HAUSGARTEN observatory.

2.1. Advancing temporal analysis in microbial ecology

Time-series analysis offers a significant improvement over traditional approaches by incorporating a temporal dimension into microbial community studies. By tracking abundance fluctuations over time, researchers can monitor a particular ecological environment by identify periodicity, detect community shifts, and infer potential environmental drivers of microbial dynamics (Ducklow et al., 2009). In Arctic marine ecosystems for example, long-term monitoring has revealed microbial responses to seasonal cycles (Oldenburg et al., 2024a, 2024b), sea ice retreat (Oldenburg et al., 2024c), and changing oceanographic conditions (Oldenburg et al., 2024c), highlighting the value of temporal datasets (Oldenburg et al., 2024a, 2024b, 2024c).

Fourier transformation techniques facilitate the decomposition of microbial time-series data into distinct frequency components, thereby enabling the identification of dominant cyclic patterns within community structures. This approach is particularly effective for distinguishing recurrent seasonal dynamics from stochastic fluctuations, ultimately enhancing our capacity to predict microbial behavior under variable environmental conditions (Taylor et al., 2025). The integration of co-occurrence network analyses with causal inference methods, such as Convergent Cross Mapping, permits a transition beyond traditional correlation-based assessments. While conventional co-occurrence networks merely indicate taxa that tend to occur together, CCM reconstructs the potential influence of one species on another by evaluating whether the historical behavior of one species can reliably

Table 1
Summary of key analytical methods for microbial time-series data.

Method	Core Concept	Strengths	Limitations	Best Use
Fourier Decomposition	Breaks time-series into sinusoidal (seasonal) components	Identifies periodic patterns (e.g., seasonality), separates signal from noise	Assumes station-arity; limited to capturing cyclical trends	Detecting seasonal dynamics and temporal similarity be-tween taxa
Co-Occurrence Networks (CON)	Correlates taxa that vary similarly over time	Reveals co-fluctuating taxa, easy to visualize	No directionality; may reflect shared response rather than true interac-tion	Initial detection of associations and community structure
Convergent Cross Mapping (CCM)	Tests if past values of one taxon can predict another → infers causality	Provides directionality, distinguishes true interactions from shared drivers	Requires long, dense time-series; sensitive to noise	Inferring causal in-teractions between taxa or with environmental variables
Energy Land-scape Analysis (ELA)	Maps community states as energy minima/maxima to assess stability	Identifies keystone taxa, predicts regime shifts, reveals ecosystem stability	Computationally intensive; abstract concept for newcomers	Assessing ecosys-tem resilience, finding key stabilizers
Ordinary Differential Equations (ODE)	Mechanistic models to simulate taxon dynamics under changing condi-tions	Describes func-tional relationships, models long-term trends	Requires prior knowledge of system dynamics and strong assumptions	Simulating ecosys-tem responses to environmental changes
Graph Neu-ral Networks (GNN)	Learns patterns in complex interaction networks via deep learning	Integrates multiple data layers, predicts new relation-ships, scalable	Black-box model; difficult to interpret; requires large datasets	Predictive mod-eling of network changes under climate scenarios
Boruta Feature Selection	Identifies key envi-ronmental variables using Random Forest & statistical testing	Robust against overfitting, detects subtle drivers, works with com-plex data	Does not model in-teractions between variables	Selecting important drivers of community change
ARIMA (Forecasting)	Time-series fore-casting using past trends (autoregression & moving average)	Good for short-term predictions, interpretable	Limited with non-linear dynamics or sudden shifts	Forecasting abun-dance trends for individual taxa
DeepLOKI (CNN-based Imaging)	Classifies high-res images of plankton with Convolutional Neural Networks	Provides indepen-dent validation, detects rare morpho-types	Dependent on image quality and training data	Cross-validation of sequencing data and identification of rare taxa

predict the state of another. By differentiating direct effects from indirect associations, this methodology refines our understanding of microbial interactions—a critical enhancement for studying Arctic microbial networks, where biogeochemical cycling and food web interactions are closely linked to environmental variability (Oldenburg et al., 2024b). Moreover, recent methodological innovations such as DeepLOKI—a back-checking mechanism for time-series data based on image recognition—offer an additional layer of quality control by validating network-based inferences (Oldenburg et al., 2023). When combined with machine learning–driven approaches like graph neural networks, these techniques provide novel strategies for analyzing and interpreting microbial datasets at scale (Marre et al., 2020). Over the past 25 years, time-series analyses conducted at HAUSGARTEN have elucidated key trends in Arctic microbial ecology, capturing shifts in community composition associated with climatic perturbations (Soltwedel et al., 2016; Oldenburg et al., 2024c; Priest et al., 2023) and biogeochemical changes (Priest et al., 2023). The incorporation of advanced time-series methodologies has not only enhanced our understanding of microbial succession patterns but also established a robust foundation for predictive modeling of Arctic ecosystem responses to future climate scenarios. The subsequent sections will examine these advancements in detail, focusing on their applications in reconstructing microbial interactions, assessing ecosystem stability, and informing long-term ecological research strategies (Faust et al., 2015; Oldenburg et al., 2024a).

3. Synergies between quantitative methods for amplicon data and network analysis

Advances in bioinformatics and computational biology provide powerful tools to enhance the quantitative interpretation of amplicon data, mitigating their inherent limitations and expanding their applicability in ecological studies (Liu et al., 2020; Logares et al., 2012). By integrating multiple analytical techniques, researchers can extract deeper insights into microbial interactions (Weiss et al., 2016), co-occurrence patterns (Runge et al., 2019; Ushio et al., 2017), and the structure of Arctic ecological networks (Faust and Raes, 2012). While individual methods capture only specific aspects of these networks (Adam et al., 2015; Ye et al., 2015; Banerjee et al., 2016), their synergistic combination enhances our ability to infer species interactions (Weiss et al., 2016), predict ecological roles (Faust and Raes, 2012; Berry and Widder, 2014), and assess the stability of microbial communities under changing environmental conditions (Fujita et al., 2023). This section explores key methodologies, their individual strengths, and the insights gained by their combined application in advancing the study of Arctic microbial communities.

3.1. Fourier decomposition of amplicon time-series data

For ecological network construction, amplicon time-series data must be functionally comparable. Unlike absolute abundance values, relative abundance data alone cannot reliably detect behavioral patterns or interactions between taxa (Kurtz et al., 2015). Fourier decomposition provides a solution by extracting periodic components from time-series signals, allowing the identification of seasonal trends and underlying cyclic dynamics (Dokumentov and Hyndman, 2021). This approach is particularly useful in Arctic microbial studies, where environmental fluctuations—such as sea ice cover, nutrient availability (Priest et al., 2023), and light cycles (Oldenburg et al., 2024b; Zhang et al., 2010)—strongly impact microbial activity (Comeau et al., 2011; Priest et al., 2023).

Fourier decomposition has long been employed to analyze seasonality in ecological datasets (Scharlemann et al., 2008; Schmid et al., 2011). By decomposing signals into their constituent sine and cosine components, this method facilitates the quantification of seasonal fluctuations in microbial communities (Singh, 2018). Specifically, the seasonality of a given time series is quantified as:

$$s(t) = \frac{|f_2(t)|}{|f_0(t)|} \tag{1}$$

where t represents the t-th Fourier component (Bracewell, 2005). Through this transformation, the frequency, amplitude, and phase of each amplicon sequence variant (ASV) time series can be extracted (Singh et al., 2017), thereby providing robust temporal profiles for subsequent analyses. Similarities in these temporal profiles indicate analogous behavior over time, which enhances the detection of co-occurrence patterns and refines our understanding of microbial responses to both seasonal and long-term environmental shifts (Amin et al., 2023). By combining Fourier decomposition with subsequent network analyses, we can move beyond merely identifying microbial taxa that fluctuate in abundance to examining how different taxa respond to environmental drivers at distinct frequencies and phases (Danczak et al., 2022).

3.2. Constructing an ecological graph using Co-occurrence networks

To elucidate microbial interactions, it is essential to embed ASV time-series data within an ecological framework that accounts for seasonal patterns (Gilbert et al., 2012; Fuhrman et al., 2015). Co-occurrence network analyses facilitate the detection of taxa exhibiting analogous seasonal trends, thereby identifying potential ecological relationships (Faust and Raes, 2012; Berry and Widder, 2014). In these networks, pairs of taxa that display robust covariation in their temporal profiles are connected by undirected edges, forming the basis of an ecological graph (Weiss et al., 2016; Friedman and Alm, 2012; Widder et al., 2016). While co-occurrence networks provide a valuable initial approach for visualizing and describing microbial relationships, they have several inherent limitations (Lima-Mendez et al., 2015; Steele et al., 2011; Banerjee et al., 2016; Hirano and Takemoto, 2019). A fundamental issue is that correlation does not imply causation; taxa that co-vary over time may not be biologically interacting but could instead be independently responding to a shared environmental driver, such as seasonal fluctuations in temperature or nutrient availability (Jacob et al., 2015; Shade et al., 2012). Furthermore, co-occurrence networks do not distinguish between direct and indirect relationships (Kurtz et al., 2015). A taxon that appears strongly correlated with another may, in reality, be linked only indirectly through a third taxon, introducing noise and potentially leading to misinterpretations (Chaffron et al., 2010; Freilich et al., 2010; Barbera'n et al., 2012). Co-occurrence networks are also constrained by the use of static correlation thresholds, which can oversimplify ecological dynamics (Mandakovic et al., 2018; Lisa et al., 2018). Threshold-based approaches may inadvertently include weak or spurious correlations while excluding biologically meaningful but lower-strength associations (Weiss et al., 2016). This rigidity limits the ability to capture the complexity of microbial networks, where interactions may fluctuate dynamically over time (Widder et al., 2016). Additionally, because co-occurrence networks lack directional information, they cannot distinguish whether one taxon influences another or if their association is merely coincidental (Adam et al., 2015; Ye et al., 2015).

To address these limitations, additional methodological layers are required (Fisher and Mehta, 2014; Valencia-Agami et al., 2024). Integrating causal inference techniques such as Convergent Cross Mapping (CCM) represents a crucial advancement, enabling researchers to move beyond correlation-based analyses and infer true ecological interactions (Runge et al., 2019; Ushio et al., 2017). Unlike conventional co-occurrence approaches, CCM identifies directional causal links allowing for the differentiation between direct biological interactions and indirect associations (George et al., 2012; Deyle et al., 2016). This transition from undirected to directed networks provides a more accurate representation of microbial ecosystem structures and their responses to environmental fluctuations (Kurtz et al., 2015; Jacob et al., 2015; Shade et al., 2012). The following section explores how CCM can infer causality in microbial networks, enhancing the interpretability of co-occurrence-based ecological graphs and improving our understanding of microbial community dynamics.

3.3. Integrating directionality with Convergent Cross Mapping

The transition from undirected to directed ecological networks can be realized by employing Convergent Cross Mapping (CCM) to infer causal relationships between taxa (Runge et al., 2019; Chang et al., 2017). CCM reconstructs the state-space dynamics for each species from their time-series data by leveraging the concept of state-space embedding, where past observations of a variable are used to reconstruct its underlying dynamics in a multidimensional space (George et al., 2012; Takens, 2006). By testing whether the reconstructed states of one species can reliably predict those of another, CCM determines the strength and directionality of causal influence (Hsieh et al., 2005). A significant cross-mapping skill suggests that past values of one taxon contain information about the future states of another, indicating a potential causal relationship (Deyle and George, 2011; Ushio et al., 2018). This method is particularly useful for disentangling direct biological interactions from indirect associations driven by external environmental factors, allowing for a more refined understanding of microbial community dynamics and ecosystem stability (Cenci and Saavedra, 2019).

For example, suppose two microbial taxa (A and B) are strongly correlated in a co-occurrence network (Barbera'n et al., 2012). While this could indicate an ecological interaction, it could also mean that both taxa are responding to a common environmental factor, such as seasonal nutrient upwelling (Faust and Raes, 2012; Berry and Widder, 2014; Gilbert et al., 2012). By applying CCM, we can test whether time-lagged values of taxon A can predict changes in taxon B better than expected by chance (Wessling et al., 2022). If taxon A reliably predicts taxon B's abundance, we infer that A is influencing B rather than both simply co-varying with an unmeasured environmental variable. In this way, CCM helps distinguish between direct ecological interactions and indirect associations (Bonotto et al., 2022; Liu and Gaines, 2022).

However, CCM is inherently limited to pairwise comparisons and does not account for higher-order or multivariate dependencies. In complex microbial ecosystems where interactions often involve multiple taxa simultaneously, this can restrict interpretability. While some recent approaches have sought to extend CCM to multivariate settings, they remain computationally intensive and require longer, high-resolution datasets to be effective. Additionally, integrating biogeochemical or physiological parameters—such as bacterial production, nutrient concentrations, or oxygen uptake—into CCM analysis could help link microbial interactions to functional ecosystem outcomes. As demonstrated by Ghannam and Techtmann (Ryan et al., 2021), combining taxonomic and functional data allows for training models that capture ecosystem-level processes from time-series datasets. Co-occurrence networks provide a valuable starting point for understanding microbial associations but require causal inference methods like CCM to uncover true ecological dependencies (Zhu et al., 2019b). By identifying taxa that exhibit strong correlations in their seasonal patterns, co-occurrence networks provide a filtered set of taxa pairs that are likely to interact (Xue et al., 2018). This pre-selection significantly reduces the computational burden of CCM, allowing it to focus only on the most relevant relationships rather than testing every possible pairwise interaction (Barbera'n et al., 2012). Additionally, co-occurrence analysis helps to reveal broader community structures, which can then be refined using CCM to identify key regulatory species within microbial networks (Luo et al., 2024). However, CCM itself has limitations. It requires long, high-frequency time series with relatively low noise to reliably infer causality (Cobey and Baskerville, 2016). If sampling intervals are too sparse, or if the dataset contains excessive random fluctuations, the method may fail to detect true interactions (Eric Yuan and Shou, 2022; Plis et al., 2014). Additionally, CCM only establishes predictive relationships and does not directly provide mechanistic insights into the nature of the interaction (e.g., competition, facilitation, or trophic dependence) (Lytle and Tonkin, 2023). Furthermore, it assumes that causal relationships remain consistent over time, which may not always be the case in dynamic microbial ecosystems (Adam et al., 2015). Combining CCM with environmental metadata allows researchers to determine whether observed microbial interactions are primarily driven by biological dependencies or shared responses to external factors. These enhancements ensure that microbial network reconstructions are both statistically rigorous and ecologically meaningful, providing a clearer picture of the complex interactions shaping Arctic microbial ecosystems (Freire-Zapata et al., 2024).

3.4. Identification of key taxa using energy landscape analysis

Once a directed ecological network has been established, key microbial taxa can be identified using Energy Landscape Analysis (ELA) (Zhang and Nakaoka, 2024). This method quantifies the stability and resilience of microbial communities by mapping potential energy landscapes, where lower energy states indicate more stable ecological configurations (Suzuki et al., 2021). In Arctic microbial ecosystems, where environmental variability is extreme, identifying key taxa that maintain network stability is crucial (Oldenburg et al., 2024b; Libralato et al., 2006). It is important to note that the accuracy of these complex analyses is highly dependent on the quality of the initial data collection. Therefore, the necessity of recording accurate and comprehensive metadata cannot be overstated, as many projects are still designed with minimal consideration for this crucial aspect (Crandall et al., 2023; Gilbert et al., 2014; Faust and Raes, 2012).

ELA is particularly valuable for pinpointing taxa that function as keystone species—those that play disproportionately important roles in ecosystem stability. These taxa may act as core drivers of community structure, influencing biogeochemical cycles or facilitating symbiotic relationships within microbial networks. By analyzing energy landscapes, re-searchers can detect stable equilibrium states, where microbial communities persist under normal conditions, as well as unstable configurations that may indicate vulnerability to external disturbances (Oldenburg et al., 2024b; Banerjee and Schlaeppi, 2018).

A major strength of ELA is its ability to identify critical thresholds and potential regime shifts within microbial networks (Suzuki et al., 2021). When Arctic ecosystems experience perturbations—such as temperature anomalies, changes in nutrient flux, or loss of sea ice—certain microbial communities may transition between alternative stable states (Fujita et al., 2023). Detecting these transitions allows researchers to assess the resilience of microbial networks and predict whether they can return to their original state after disturbance or shift into a fundamentally altered community structure (Bardgett and Caruso, 2020).

The integration of ELA with methods such as Fourier decomposition and CCM further enhances our ability to assess microbial network dynamics (Oldenburg et al., 2024b). While Fourier transformed time-series data highlight periodic fluctuations in microbial abundance, ELA determines whether these fluctuations represent stable recurring patterns or shifts toward new ecological states (Suzuki et al., 2021). Similarly, combining ELA with CCM allows for the identification of not only key taxa but also their functional influence within the network, providing deeper insights into ecosystem drivers and long-term stability (Freire-Zapata et al., 2024; Søndergaard et al., 2011).

By leveraging energy landscape analysis in conjunction with other network-based methodologies, researchers can refine their understanding of Arctic microbial communities, offering predictive insights into how these ecosystems will respond to ongoing environmental change.

3.5. Quantitative analysis of taxa using deterministic mathematical models

Deterministic mathematical models, particularly those based on ordinary differential equations (ODEs), have long been used to describe ecological interactions (Alexander et al., 2020). The Lotka-Volterra system, one of the most widely used ODE frameworks, models predator-prey relationships and competitive interactions (Peter, 1978; Dedrick et al., 2023). These models provide a mechanistic understanding of how microbial taxa influence each other (Dedrick et al., 2023) and how environmental changes impact ecosystem stability (Gou et al., 2025).

Modern adaptations of these models incorporate additional environmental variables such as temperature, salinity, and nutrient fluxes, thereby improving their predictive capabilities (Joseph et al., 2016). In Arctic microbial networks, where external conditions fluctuate dramatically (Von Appen et al., 2021), integrating these factors allows for a more nuanced simulation of microbial dynamics under varying climate scenarios.

By integrating taxa identified by ELA as being in an energetic stable or unstable state into mechanistic ODE models, researchers can simulate the effects of environmental variability on Arctic microbial communities, providing valuable insights into their long-term stability and adaptation strategies (Suzuki et al., 2021). These models allow us to test hypotheses about the resilience of Arctic microbial networks under different climate change scenarios, identifying tipping points and thresholds for ecosystem transitions.

In addition to taxon-specific abundance data, these models can be extended to incorporate proxies of microbial ecosystem function—such as carbon flux, bacterial respiration, or nitrogen turnover—thereby enabling simulation of functional dynamics under varying conditions. Such integrative modeling can offer mechanistic insights into how microbial communities contribute to biogeochemical cycling and how these contributions may shift under future environmental scenarios.

Furthermore, deterministic models can be combined with time-series analyses and network-based approaches to refine predictions (Deyle et al., 2016). For example, Fourier-transformed time-series data can be incorporated into ODE models to capture periodic fluctuations in microbial activity, while CCM-derived causal relationships can improve model accuracy by informing taxon-specific interaction parameters. The combination of these techniques enhances the ability to predict ecosystem responses (Strogatz, 2001), making these models powerful tools for assessing microbial community dynamics in a rapidly changing Arctic environment.

3.6. Supplementing amplicon count data with convolutional deep neural networks

Despite the wealth of information provided by ASV time-series data, amplicon sequencing has inherent limitations that necessitate independent validation (Poretsky et al., 2014). The Lightframe On- Sight Keyspecies Investigation (LOKI) system, which captures high-resolution images of plankton and microbial aggregates (Schulz et al., 2010), offers an alternative means of quantifying microbial diversity (Daims and Wagner, 2007).

Deep-learning models, such as DeepLOKI, utilize convolutional neural networks (CNNs) to automate the classification of LOKI-generated images (Jogin et al., 2018; Oldenburg et al., 2023). By cross-referencing amplicon sequencing data with image-based classifications (Friederike Weiβ et al., 2024; Herna'ndez Medina et al., 2022), CNNs provide an additional quality control layer, ensuring the reliability of microbial abundance estimates. This integrative approach strengthens confidence in ecological inferences drawn from ASV time-series analyses and allows for real-time assessment of microbial biodiversity during expeditions (Marre et al., 2020).

Moreover, the combination of amplicon sequencing and CNN-based image classification provides a more robust assessment of microbial community composition by addressing potential biases in sequencing techniques (Marre et al., 2020; Friederike Weiβ et al., 2024). Amplicon sequencing can suffer from PCR amplification biases, primer mismatches, and differential sequencing efficiencies (Větrovsky', Baldrian, 2013; Kembel et al., 2012; Klappenbach et al., 2000), leading to inaccuracies in taxonomic abundance estimates (Kembel et al., 2012). By integrating independent image-based validation, researchers can cross-check the presence and relative abundances of key taxa (Friederike Weiβ et al., 2024), refining ecological models and improving the interpretation of microbial network structures (Zhu et al., 2019a; Daims and Wagner, 2007).

Additionally, this approach allows for the identification of rare but ecologically significant taxa that may be underrepresented in sequencing data due to primer inefficiencies (Herna'ndez Medina et al., 2022). However, it is important to note that many ASVs detected through sequencing lack corresponding morphological information, which limits the potential to link molecular and morphological observations. Moreover, especially in bacteria, morphological diversity is relatively low, and unrelated taxa can often appear morphologically similar, thereby reducing the taxonomic resolution of morphological observations. These limitations may be less pronounced in eukaryotic microorganisms, which often exhibit more distinct and diverse morphological features.

CNNs trained on LOKI images can detect morphological characteristics of specific organisms that may be overlooked in molecular analyses (Oldenburg et al., 2023), offering complementary insights into microbial diversity and function (Daims and Wagner, 2007; Oldenburg et al., 2023).

By merging high-throughput image classification with amplicon-based microbial monitoring (Zhu et al., 2019a; Daims and Wagner, 2007), researchers can improve the accuracy and depth of Arctic microbial network reconstructions (Zhu et al., 2019a), ensuring that ecological conclusions are built upon a solid, multi-method foundation (Marre et al., 2020).

3.7. *Ecosystem modeling: advanced machine learning techniques for precision environmental feature selection*

A fundamental challenge in studying Arctic marine microbial ecosystems is determining which environmental variables drive microbial community shifts and how these influences change over time. Deterministic models, such as ordinary differential equations (ODEs), rely on well-defined input parameters to describe microbial interactions mathematically. However, identifying the most relevant environmental drivers is often difficult due to the complex, nonlinear relationships between microbes and their surroundings. To address this, machine learning-based feature selection methods, such as Boruta (Kursa and Rudnicki, 2010), provide a robust way to determine the key environmental predictors shaping microbial communities. Boruta is an extension of the Random Forest algorithm designed to perform all relevant feature selection, rather than relying on traditional statistical significance measures. The core idea behind Boruta is to generate shadow variables, which are random copies of the original dataset's features. The algorithm then evaluates whether a real variable provides more predictive power than these shadow variables. If an environmental factor (e.g., sea ice cover, nutrient concentrations, or water temperature) consistently ranks higher than its randomized counterpart, it is considered an essential predictor.

This method helps avoid false negatives (incorrectly discarding important variables) and ensures that all ecologically significant drivers are accounted for (Kursa and Rudnicki, 2010).

Beyond identifying key environmental factors, Boruta can also highlight which microbial taxa are most influenced by specific conditions, providing a more detailed understanding of how environmental shifts drive community composition changes. This is particularly relevant in Arctic ecosystems, where seasonal shifts and climatic trends can have cascading effects on microbial assemblages (Banerjee et al., 2018).

3.7.1. *Limitations of Boruta and overcoming them with advanced ecological models*

While Boruta is a powerful feature selection method, it has certain limitations. One key drawback is that it does not explicitly model interactions between variables—it only ranks them based on importance. This means that while we may know that temperature and nutrient availability are key drivers, Boruta does not inherently reveal how these factors influence each other or how microbial interactions mediate these effects. Furthermore, because it relies on a Random Forest-based approach, it may struggle with datasets that have a high degree of multicollinearity (strongly correlated variables), potentially overestimating the importance of redundant features (Degenhardt et al., 2019).

To compensate for these limitations, integrating Boruta with ecosystem modeling frameworks—such as Energy Landscape Analysis (ELA), Convergent Cross Mapping (CCM), and ODE-based modeling—can provide deeper insights.

ELA can reveal stable and unstable ecosystem states by assessing how different environmental variables contribute to long-term microbial shifts (Suzuki et al., 2021). CCM allows us to infer causal interactions, helping to clarify whether a given environmental factor directly influences a microbial population or is simply correlated with it (Ye et al., 2015). ODEs can incorporate the selected Boruta-identified environmental factors into mechanistic models, simulating how microbial dynamics evolve under varying conditions. This multi-method approach enables not only the identification of key environmental drivers but also a deeper understanding of their ecological interactions, leading to more robust predictions.

3.8. *Node and edge predictions using graph neural networks*

With increasing network complexity, traditional analytical methods become computationally expensive (Justus et al., 2018). Graph Neural Networks (GNNs) offer a machine learning-driven solution by enabling scalable prediction of taxa relationships within Arctic microbial networks (Wu et al., 2020; Franco et al., 2008).

By training GNNs on existing ASV time-series networks (Shao et al., 2022; Jin et al., 2024), researchers can predict novel taxa interactions (Shao et al., 2022) and uncover previously unrecognized ecological dependencies (Wang, 2024). This approach enhances our ability to model microbial network evolution, providing a powerful tool for forecasting Arctic ecosystem dynamics under future climate scenarios (Spadon et al., 2021). Additionally, GNNs allow for the integration of multiple data modalities—including time-series data, co-occurrence networks, and environmental metadata—further improving predictive models.

A key advantage of GNNs is their ability to generalize complex relationships beyond the existing dataset, allowing for predictions about previously unobserved interactions (Feng et al., 2023). Unlike traditional methods that rely on predefined correlation thresholds, GNNs leverage deep learning to extract hidden patterns in ecological networks, identifying relationships that might be obscured by noise or sparse sampling (Przymus et al., 2025). This makes GNNs particularly valuable for Arctic research, where the limited availability of long-term data can make traditional statistical approaches less reliable.

Furthermore, integrating GNNs with other computational techniques—such as Fourier decomposition, CCM, and energy landscape analysis —enhances their predictive power (Thomas et al., 2016; George et al., 2012). By using GNNs to incorporate both biological interactions and environmental drivers, researchers can build more comprehensive models of Arctic microbial ecosystems, enabling improved forecasting of ecosystem responses to climate change.

By combining GNNs with the previously discussed methodologies, we can create an advanced, data-driven framework for understanding microbial network dynamics. This integration provides a holistic view of microbial interactions, helping to refine ecosystem models and ensure that predictions about Arctic microbial communities are both robust and biologically meaningful.

3.9. *Why forecasting matters in Arctic microbial research*

The Arctic Ocean is experiencing rapid environmental shifts, with warming temperatures, sea ice retreat, and changes in nutrient availability altering microbial communities. These shifts can have cascading effects on biogeochemical cycles, primary production, and ecosystem stability.

While observational time-series studies provide crucial insights into past and present microbial community dynamics, they do not inherently allow predictions about future ecosystem states. Forecasting models enable researchers to project microbial composition and abundance under different climate scenarios, identify early-warning signals of ecosystem instability, and evaluate microbial responses to environmental drivers such as temperature, salinity, and nutrient availability. However, to develop robust forecasting models, it is essential to combine statistical, machine learning, and mechanistic modeling approaches, ensuring that both data-driven predictions and ecological mechanisms are considered.

3.9.1. Classical time-series forecasting methods

A widely used statistical model, ARIMA (Box et al., 2015) is based on autoregressive components (AR), moving averages (MA), and differencing (I) to remove trends (Siami-Namini et al., 2018). It is particularly effective for analyzing microbial time series, as it allows researchers to identify seasonal fluctuations and predict short-term trends. One of ARIMA's main strengths lies in its interpretability, making it useful for detecting seasonal and long-term patterns in microbial abundance. However, it struggles with nonlinear ecological interactions and assumes that microbial dynamics will continue to follow historical patterns, which may not hold under climate-driven disruptions. To improve its effectiveness, ARIMA can be integrated with CCM and Co-Occurrence Networks to incorporate species interactions and causal drivers rather than relying purely on past trends. Additionally, Energy Landscape Analysis (ELA) can help validate stability predictions derived from ARIMA models, ensuring that the predicted ecosystem states align with observed resilience patterns.

Error, Trend, Seasonality models use exponential smoothing techniques to forecast microbial dynamics by capturing long-term trends and seasonality (Wang et al., 2020; Hyndman et al., 2002). Unlike ARIMA, ETS does not require stationarity assumptions, making it well-suited for microbial communities with clear seasonal successions, such as Arctic phytoplankton blooms. Despite these advantages, ETS models are limited in their ability to predict abrupt ecological shifts, such as sudden regime changes caused by extreme environmental events (Dakos et al., 2012). They also do not inherently account for species interactions or external drivers. To enhance their predictive power, ETS models can be integrated with Boruta for feature selection, identifying key environmental factors that drive microbial seasonality. Furthermore, coupling ETS with ODE models allows researchers to validate seasonality-based predictions using ecological theory, making it possible to distinguish between true seasonal patterns and externally driven fluctuations.

A powerful Bayesian machine learning method, Gaussian Process Regression (GPR) is particularly useful for ecological time-series analysis due to its ability to model uncertainty in microbial forecasts (Tonner et al., 2017; Wang, 2023). This makes it especially valuable in Arctic research, where microbial data is often sparse or unevenly sampled (Kupilik et al., 2018). GPR provides probabilistic forecasts with confidence intervals, ensuring that predictions include estimates of uncertainty (Tonner et al., 2017). However, it is computationally expensive, which can be problematic for long-term, high-resolution microbial time series. Moreover, GPR does not inherently model ecological interactions, making it less suitable for studying microbial network dynamics (Gibbs, 1998). To improve its application, GPR can be combined with CCM, helping to identify causal relationships between microbial taxa and environmental drivers. Additionally, Boruta feature selection can be used to pre-select relevant environmental variables, reducing computational complexity and improving model accuracy.

Long Short-Term Memorys (LSTM), a class of recurrent neural networks, are specifically designed for handling long-term dependencies in time-series data (Graves and Graves, 2012; Hochreiter and Schmidhuber, 1997). These models are particularly effective for microbial ecosystem research, as they can capture complex, nonlinear relationships between microbial taxa and environmental variables (Qu et al., 2019). Unlike ARIMA and ETS, LSTMs do not rely on stationarity assumptions and can handle irregular sampling and missing data better than traditional statistical models. However, a major drawback of LSTMs is their lack of interpretability, as they function as a ”black box” without providing insights into the underlying ecological processes (Che et al., 2018). Additionally, training LSTMs requires large datasets, which can be challenging in Arctic microbial research. To address these limitations, Boruta feature selection can be used to filter relevant environmental variables before training, and LSTMs can be integrated with ELA to validate that forecasts align with stable ecological states.

HMMs (hidden markov moddels) are valuable for identifying hidden ecological states in microbial time series, making them particularly useful for detecting regime shifts such as the Atlantification of Arctic waters. These models can infer transitions between stable and unstable microbial states, providing insights into early-warning signals of ecosystem tipping points (Scheffer et al., 2009). However, HMMs alone cannot determine causality and may misclassify states if the number of hidden regimes is not well-defined. To improve their reliability, HMMs can be combined with CCM, ensuring that inferred state transitions are biologically meaningful (Runge et al., 2019). Additionally, incorporating ELA alongside HMMs allows for better validation of stable and transitional ecological phases, ensuring that microbial regime shifts are accurately detected and contextualized.

To facilitate comparison between forecasting approaches, Table 2 summarizes key features, strengths, limitations, and application contexts of classical and machine-learning based forecasting methods.

Towards Predictive Ecosystem Modeling: Combining Feature Selection with Time-Series Forecasting.

Once key environmental variables and microbial interactions have been identified using Boruta (Kursa and Rudnicki, 2010; Wright and Ziegler, 2017) and validated through CCM, ELA, and ODEs (Moore et al., 2005), the next logical step is predicting future ecosystem states. This can be achieved using forecasting models, such as AutoRegressive Integrated Moving Average (ARIMA), which excels at predicting time-dependent trends in ecological data. ARIMA models can be trained

Table 2
Overview of forecasting methods used in microbial time-series analysis, including traditional statistical models (ARIMA, ETS), probabilistic frameworks (GPR), and machine-learning-based approaches (LSTM, HMM).

Method	Core Principle	Strengths	Limitations	Best Used For
ARIMA	Autoregressive Integrated Moving Average	Simple, interpretable, suitable for short-term trends	Limited to linear trends, no handling of sudden shifts	Short-term pre-dictions with well-behaved time series
ETS	Exponential Smoothing (Trend + Season-ality)	Handles trend and seasonality, easy to tune	No inclusion of external variables, sensitive to anoma-lies	Forecasting sea-sonal trends
GPR	Gaussian Process Regression	Flexible, proba-bilistic outputs, good for small datasets	Computationally expensive, requires kernel design	Uncertainty es-timation and nonlinear dynamics
LSTM	Long Short-Term Memory Neural Networks	Captures long-range dependencies, nonlinear patterns	Requires large training data, hard to interpret	Complex multivariate forecasting
HMM	Hidden Markov Models	Models regime shifts and state transitions	Difficult to scale, assumes Markov property	Detecting regime changes or discrete ecological states

on long-term time-series datasets from Arctic observatories (e.g., HAUSGARTEN) to forecast how microbial communities will shift under changing environmental conditions (Shumway et al., 2000). By iteratively adjusting environmental parameters within these models, researchers can explore potential future scenarios, such as the effects of warming temperatures or declining sea ice on microbial stability (Vincent and Laybourn-Parry, 2008).

By integrating feature selection (Boruta), interaction modeling (CCM, ELA, ODEs), and predictive time-series forecasting (ARIMA), we move towards a comprehensive ecological modeling framework capable of both explaining and predicting Arctic microbial dynamics. This combination of machine learning and deterministic modeling represents a powerful strategy for understanding how marine microbial communities respond to environmental shifts, ultimately enhancing our ability to monitor and mitigate the impacts of climate change on polar ecosystems (Scheffer et al., 2001).

3.9.2. *Hybrid approaches and future directions*

No single forecasting method is sufficient to predict microbial ecosystem dynamics in the rapidly changing Arctic environment. Instead, a hybrid approach, integrating multiple techniques, is needed to capture both short-term microbial fluctuations and long-term ecological trends. Classical methods such as ARIMA and ETS are useful for short-term seasonal forecasts, but they must be validated using ODE stability analysis and ELA to ensure their ecological relevance. Machine learning models, such as GPR and LSTMs, can capture nonlinear dynamics, but their predictive power is significantly enhanced when paired with Boruta feature selection and CCM for causal inference. Similarly, regime-shift models such as HMMs can be valuable for detecting ecosystem tipping points, but their accuracy can be improved by integrating them with ELA and network-based approaches. By leveraging a combination of data-driven forecasting and ecological modeling, researchers can develop more robust predictions about microbial ecosystem resilience under future climate scenarios. The integration of machine learning, statistical modeling, and ecological theory enhances both accuracy and interpretability, paving the way for a more comprehensive understanding of Arctic microbial dynamics.

3.10. *Key insights and future directions*

This review, ”More than Co-Occurrence: What Amplicon Time-Series Data Can Tell Us”, encapsulates 25 years of pioneering microbial research at the HAUSGARTEN observatory. Not only does it highlight the evolution of our understanding of Arctic marine microbial ecology in a rapidly changing climate, but it also demonstrates the transformative power of long-term ecological monitoring and quantitative analytical techniques. For instance, we observed a marked increase in SAR11 clade abundance in late-summer surface waters over the past decade, likely linked to fresher and warmer conditions. Similarly, shifts in picoeukaryotic community structure suggest changing nutrient dynamics associated with sea ice retreat (Hartmann et al., 2016).

This review highlights a number of key developments that are reshaping our understanding of Arctic microbial ecosystems and their response to environmental change. Long-term ecological time series from observatories such as HAUSGARTEN, operating since 1999, are fundamental to deciphering microbial community dynamics in the context of climate-driven shifts including sea-ice loss, Atlantification, and warming. These sustained observations allow the assessment of ecosystem stability and resilience, offering a temporal backbone for detecting ecological transitions. Arctic microbial communities play central roles in biogeochemical cycling and ecosystem functioning. Seasonal dynamics, resource fluctuations, and light availability shape their composition and activity, while keystone taxa contribute disproportionately to community stability. Understanding these dynamics requires both descriptive and predictive tools. Recent methodological advancements—ranging from amplicon sequencing and Fourier decomposition to co-occurrence networks, Convergent Cross Mapping (CCM), and Energy Landscape Analysis (ELA)—have enabled a more mechanistic view of microbial interactions across time. Moreover, the integration of deterministic models with machine learning tools such as Graph Neural Networks (GNNs) and ARIMA-based forecasts is now facilitating the prediction of microbial behavior under future climate conditions. These approaches also underscore the importance of environmental coupling. Interactions between microbial communities and environmental drivers such as temperature, nutrient availability, and mixing regimes are not static, but evolve through feedbacks that can now be partially captured by integrative models. Moving forward, progress in Arctic microbial ecology will depend on refining these models, increasing the spatial and temporal resolution of the underlying data, and including functional measurements wherever possible. In conclusion, the convergence of high-resolution ecological observations with modern quantitative approaches marks a paradigm shift in how microbial time-series data are analyzed and interpreted. Transitioning from static descriptions to dynamic, predictive, and functionally grounded models is essential for anticipating microbial ecosystem responses in the Arctic. As climate change continues to accelerate, these integrated approaches will be crucial for guiding research and informing conservation strategies in polar marine systems.

CRediT authorship contribution statement

Ellen Oldenburg: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Conceptualization. **Nima P. Saadat:** Writing – original draft, Methodology. **Sofie Thielen:** Resources, Data curation. **Ovidiu Popa:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used DeepL and other AI-based language tools for linguistic refinements of this manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Funding

This work was further supported under Germany's Excellence Strategy - EXC-2048/1 - project ID 390686111 (CEPLAS)(EO).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adam, Thomas Clark, et al., 2015. Spatial convergent cross mapping to detect causal relationships from short time series. Ecology 96.5, 1174–1181.

Alexander, P Browning, et al., 2020. Identifiability analysis for stochastic differential equation models in systems biology. J. R. Soc., Interface 17 (173), 20200652.

Amin, Kassem, et al., 2023. Applications of Fourier transform-infrared spectroscopy in microbial cell biology and environmental microbiology: advances, challenges, and future perspectives. Front. Microbiol. 14, 1304081.

Anderson, Marti J., 2001. “A new method for non-parametric multivariate analysis of variance”. en. Austral Ecol. 26 (1), 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x issn: 1442-9985, 1442-9993. https://onlinelibrary.wiley.com/doi/10.1111/j.1442-9993.2001.01070.pp.x (visited on 02/12/2025).

Aßhauer, Kathrin P., et al., 2015. “Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data”. en. Bioinformatics 31.17, 2882–2884. https://doi.org/10.1093/bioinformatics/btv287 issn: 1367-4811, 1367-4803. https://academic.oup.com/bioinformatics/article/31/17/2882/183768 (visited on 02/24/2025).

Banerjee, Samiran, Schlaeppi, Klaus, 2018. Marcel GA Van Der Heijden. “Keystone taxa as drivers of microbiome structure and functioning”. Nat. Rev. Microbiol. 16.9, 567–576.

Banerjee, Samiran, et al., 2016. Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. Soil Biol. Biochem. 97, 188–198.

Banerjee, Samiran, Schlaeppi, Klaus, Van Der Heijden, Marcel G.A., 2018. “Keystone taxa as drivers of microbiome structure and functioning”. en. Nat. Rev. Microbiol. 16 (9), 567–576. https://doi.org/10.1038/s41579-018-0024-1 issn: 1740-1526, 1740-1534. https://www.nature.com/articles/s41579-018-0024-1 (visited on 02/14/2025).

Barberán, Albert, et al., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J. 6 (2), 343–351.

Bardgett, Richard D., Caruso, Tancredi, 2020. Soil microbial community responses to climate extremes: resistance, resilience and transitions to alternative states. Philos. Trans. R. Soc. B 375.1794, 20190112.

Berry, David, Widder, Stefanie, 2014. Deciphering microbial interactions and detect- ing keystone species with co-occurrence networks. Front. Microbiol. 5, 219.

Bin, Ma, et al., 2016. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. ISME J. 10 (8), 1891–1901.

Bonotto, Giancarlo, et al., 2022. Identifying causal interactions between groundwater and streamflow using convergent cross-mapping. Water Resour. Res. 58.8, e2021.

Box, George EP., et al., 2015. Time Series Analysis: Forecasting and Control. John Wiley & Sons.

Bracewell, Ronald N., 2005. Fourier and other mathematical transforms. Mathe- matical Tools for Physicists 83–107.

Cenci, Simone, Saavedra, Serguei, 2019. Non-parametric estimation of the structural stability of non-equilibrium community dynamics. Nat. Ecol. Evol. 3.6, 912–918.

Chaffron, Samuel, et al., 2010. A global network of coexisting microbes from environmen- tal and whole-genome sequence data. Genome Res. 20 (7), 947–959.

Chang, Chun-Wei, Ushio, Masayuki, Hsieh, Chih-hao, 2017. Empirical dynamic modeling for beginners. Ecol. Res. 32, 785–796.

Che, Zhengping, et al., 2018. Recurrent neural networks for multivariate time series with missing values. Sci. Rep. 8.1, 6085.

Cobey, Sarah, Baskerville, Edward B., 2016. Limits to causal inference with state- space reconstruction for infectious disease. PLoS One 11 (12), e0169050.

Comeau, André M., et al., 2011. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PLoS One 6.11, e27492.

Crandall, Eric D., et al., 2023. Importance of timely metadata curation to the global surveillance of genetic diversity. Conserv. Biol. 37.4, e14061.

Daims, Holger, Wagner, Michael, 2007. Quantification of uncultured microorganisms by fluorescence microscopy and digital image analysis. Appl. Microbiol. Biotechnol. 75, 237–248.

Dakos, Vasilis, et al., 2012. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. PLoS One 7 (7), e41010.

Danczak, Robert E., et al., 2022. Inferring the contribution of microbial taxa and organic matter molecular formulas to ecological assembly. Front. Microbiol. 13, 803420.

Dedrick, Sandra, et al., 2023. When does a Lotka-Volterra model represent microbial interactions? Insights from in vitro nasal bacterial communities. mSystems 8.3, e00757, 22.

Degenhardt, Frauke, Seifert, Stephan, Szymczak, Silke, 2019. Evaluation of variable selection methods for random forests and omics data sets. Briefings Bioinf. 20.2, 492–503.

Deyle, Ethan R., George, Sugihara, 2011. Generalized theorems for nonlinear state space reconstruction. PLoS One 6 (3), e18295.

Deyle, Ethan R., et al., 2016. Tracking and forecasting ecosystem interactions in real time. Proc. Biol. Sci. 283.1822, 20152258.

Dokumentov, Alexander, Hyndman, Rob J., 2021. STR: seasonal-trend Decompo- sition Using Regression. en. arXiv:2009.05894. https://doi.org/10.48550/arXiv.2009.05894 [stat]. http://arxiv.org/abs/2009.05894 (visited on 02/26/2025).

Ducklow, Hugh W., Doney, Scott C., Steinberg, Deborah K., 2009. “Contributions of Long-Term Research and Time-Series Observations to Marine Ecology and Biogeochemistry”. en. Ann. Rev. Mar. Sci 1 (1), 279–302. https://doi.org/10.1146/annurev.marine.010908.163801 issn: 1941-1405, 1941-0611. https://www.annualreviews.org/doi/10.1146/annurev.marine.010908.163801 (visited on 02/14/2025).

Elena Pérez-Cobas, Ana, Gomez-Valero, Laura, Buchrieser, Carmen, 2020. Metage- nomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. Microb. Genom. 8 (6), e000409.

Eric Yuan, Alex, Shou, Wenying, 2022. Data-driven causal analysis of observational biological time series. eLife 11, e72518.

Faust, Karoline, Raes, Jeroen, 2012. Microbial interactions: from networks to mod- els. Nat. Rev. Microbiol. 10 (8), 538–550.

Faust, Karoline, et al., 2015. “Metagenomics meets time series analysis: unraveling microbial community dynamics”. en. Curr. Opin. Microbiol. 25, 56–66. https://doi.org/10.1016/j.mib.2015.04.004 issn: 13695274. https://linkinghub.elsevier.com/retrieve/pii/S1369527415000478 (vis- ited on 02/14/2025).

Feng, Junning, Liang, Yong, Yu, Tianwei, 2023. MM-GANN-DDI: multimodal graph-agnostic neural networks for predicting drug–drug interaction events. Comput. Biol. Med. 166, 107492.

Fernandes, Andrew D., et al., 2014. “Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis”. en. Microbiome 2.1, 15. https://doi.org/10.1186/2049-2618-2-15 issn: 2049-2618. https://microbiomejournal.biomedcentral.com/articles/10.1186/2049-2618-2-15 (visited on 02/24/2025).

Fisher, Charles K., Mehta, Pankaj, 2014. “The transition between the niche and neutral regimes in ecology”. In: Proceedings of the National Academy of Sciences, 111.36, pp. 13111–13116.

Franco, Scarselli, et al., 2008. The graph neural network model. IEEE Trans. Neural Network. 20 (1), 61–80.

Freilich, Shiri, et al., 2010. The large-scale organization of the bacterial network of ecolog- ical co-occurrence interactions. Nucleic Acids Res. 38.12, 3857–3868.

Freire-Zapata, Viviana, et al., 2024. Microbiome–metabolite linkages drive greenhouse gas dynamics over a permafrost thaw gradient. Nat. Microbiol. 9.11, 2892–2908.

Friederike Weiß, Josefine, et al., 2024. Unprecedented insights into extents of biologi- cal responses to physical forcing in an Arctic sub-mesoscale filament by combin- ing high-resolution measurement approaches. Sci. Rep. 14 (1), 8192.

Friedman, Jonathan, Alm, Eric J., 2012. Inferring Correlation Networks from Genomic Survey Data.

Fuhrman, Jed A., Cram, Jacob A., Needham, David M., 2015. Marine microbial community dynamics and their ecological interpretation. Nat. Rev. Microbiol. 13.3, 133–146.

Fujita, Hiroaki, et al., 2023. Alternative stable states, nonlinear behavior, and predictabil- ity of microbiome dynamics. Microbiome 11.1, 63.

George, Sugihara, et al., 2012. Detecting causality in complex ecosystems. science 338.6106, 496–500.

Gibbs, Mark N., 1998. Bayesian Gaussian Processes for Regression and Classification. PhD thesis, Citeseer.

Gilbert, Jack A., et al., 2012. Defining seasonal marine microbial community dynamics. ISME J. 6 (2), 298–308.

Gilbert, Jack A., Jansson, Janet K., Knight, Rob, 2014. The Earth microbiome project: successes and aspirations. BMC Biol. 12.1, 69.

Gou, Wenshuo, Xue, Haowen, Chang, Jiayi, 2025. Application of Lotka-Volterra algorithm model in ecosystem assessment. In: Smart Infrastructures in the Iot Era. Springer, pp. 437–451.

Graves, Alex, Graves, Alex, 2012. Long short-term memory. In: Supervised Sequence Labelling with Recurrent Neural Networks, pp. 37–45.

Hartmann, Manuela, et al., 2016. Resilience of SAR11 bacteria to rapid acidification in the high-latitude open ocean. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol. 92 (2), fiv161.

Hernández Medina, Ricardo, et al., 2022. Machine learning and deep learning applications in microbiome research. ISME communications 2.1, 98.

Hirano, Hokuto, Takemoto, Kazuhiro, 2019. Difficulty in inferring microbial community structure based on co-occurrence network approaches. BMC Bioinf. 20, 1–14.

Hochreiter, Sepp, Schmidhuber, Jürgen, 1997. Long short-term memory. Neural Comput. 9.8, 1735–1780.

Hsieh, Chih-hao, et al., 2005. Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. Nature 435.7040, 336–340.

Hyndman, Rob J., et al., 2002. A state space framework for automatic forecasting using exponential smoothing methods. Int. J. Forecast. 18 (3), 439–454.

Jacob, A Cram, et al., 2015. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. ISME J. 9 (3), 563–580.

Jamy, Mahwash, et al., 2025. New deep-branching environmental plastid genomes on the algal tree of life. bioRxiv, 2025–01.

Jin, Ming, et al., 2024. A survey on graph neural networks for time series: Forecast- ing, classification, imputation, and anomaly detection. In: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Jogin, Manjunath, et al., 2018. Feature extraction using convolution neural networks (CNN) and deep learning. In: 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). IEEE, pp. 2319–2323.

Joseph, J Vallino, Algar, Christopher K., 2016. The thermodynamics of marine biogeochemical cycles: Lotka revisited. Ann. Rev. Mar. Sci 8.1, 333–356.

Justus, Daniel, et al., 2018. Predicting the computational cost of deep learning mod- els. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE, pp. 3873–3882.

Kellogg, Colleen TE., et al., 2019. Strong seasonality in Arctic estuarine microbial food webs. Front. Microbiol. 10, 2628.

Kembel, Steven W., et al., 2012. “Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance”. en. PLoS Comput. Biol. 8, 10. https://doi.org/10.1371/journal.pcbi.1002743. Ed. by Christian Von Mering, e1002743. issn: 1553-7358. https://dx.plos.org/10.1371/journal.pcbi.1002743 (visited on 02/12/2025).

Klappenbach, Joel A., Dunbar, John M., Schmidt, Thomas M., 2000. “rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria”. en. Appl. Environ. Microbiol. 66 (4), 1328–1333. https://doi.org/10.1128/AEM.66.4.1328-1333.2000 issn: 0099-2240, 1098- 5336. https://journals.asm.org/doi/10.1128/AEM.66.4.1328-1333.2000 (visited on 02/12/2025).

Knight, Rob, et al., 2018. “Best practices for analysing microbiomes”. en. Nat. Rev. Microbiol. 16 (7), 410–422. https://doi.org/10.1038/s41579-018-0029-9 issn: 1740-1526, 1740-1534. https://www.nature.com/articles/s41579-018-0029-9 (visited on 02/12/2025).

Kupilik, Matthew, et al., 2018. Gaussian process regression for Arctic coastal erosion fore- casting. IEEE Trans. Geosci. Rem. Sens. 57 (3), 1256–1264.

Kursa, Miron B., Rudnicki, Witold R., 2010. Feature selection with the Boruta pack- age. J. Stat. Software 36, 1–13.

Kurtz, Zachary D., et al., 2015. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput. Biol. 11 (5), e1004226.

Libralato, Simone, Christensen, Villy, Pauly, Daniel, 2006. A method for identifying keystone species in food web models. Ecol. Model. 195 (3–4), 153–171.

Lima-Mendez, Gipsi, et al., 2015. Determinants of community structure in the global plankton interactome. Science 348.6237, 1262073.

Lisa, Röttjers, Faust, Karoline, 2018. From hairballs to hypotheses–biological insights from microbial networks. FEMS Microbiol. Rev. 42 (6), 761–780.

Liu, Owen R., Gaines, Steven D., 2022. Environmental context dependency in species interactions. Proc. Natl. Acad. Sci. 119 (36), e2118539119.

Liu, Yong-Xin, et al., 2020. A practical guide to amplicon and metagenomic analysis of microbiome data. Protein Cell 12 (5), 315–330. https://doi.org/10.1007/s13238-020-00724-8. https://academic.oup.com/protei.pdf/12/5/315/47844648/132382020article724.pdf. issn: 1674-800X. doi: 10.1007/s13238-020-00724-8.

Logares, Ramiro, et al., 2012. “Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches”. en. J. Microbiol. Methods 91 (1), 106–113. https://doi.org/10.1016/j.mimet.2012.07.017 issn: 01677012. https://linkinghub.elsevier.com/retrieve/pii/S0167701212002527 (visited on 02/25/2025).

Louca, Stilianos, Parfrey, Laura Wegener, Doebeli, Michael, 2016. Decoupling func- tion and taxonomy in the global ocean microbiome. Science 353.6305, 1272–1277. https://doi.org/10.1126/science.aaf4507 eprint: https://www.science.org/doi/pdf/10.1126/science.aaf4507 https://www.science.org/doi/abs/10.1126/science.aaf4507.

Lozupone, Catherine, Knight, Rob, 2005. “UniFrac: a New Phylogenetic Method for Comparing Microbial Communities”. en. Appl. Environ. Microbiol. 71 (12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005 issn: 0099-2240, 1098-5336. https://journals.asm.org/doi/10.1128/AEM.71.12.8228-8235.2005 (visited on 02/12/2025).

Lozupone, Catherine A., Knight, Rob, 2008. “Species divergence and the measure- ment of microbial diversity”. en. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Rev. 32 (4), 557–578. https://doi.org/10.1111/j.1574-6976.2008.00111.x issn: 1574-6976. https://academic.oup.com/femsre/article-lookup/doi/10.1111/j.1574-6976.2008.00111.x (visited on 02/18/2025).

Luisa, W. Hugerth, Anders, F. Andersson, 2017. “Analysing Microbial Community Composition through Amplicon Sequencing: from Sampling to Hypothesis Test- ing”. en. Front. Microbiol. 8, 1561. https://doi.org/10.3389/fmicb.2017.01561 issn: 1664-302X. http://journal.frontiersin.org/article/10.3389/fmicb.2017.01561/full (visited on 02/12/2025).

Luo, Meng, et al., 2024. Progress on network modeling and analysis of gut microecology: a review. Appl. Environ. Microbiol. 90.3, e00092, 24.

Lytle, David A., Tonkin, Jonathan D., 2023. Matrix community models for ecology and evolution. npj Biodiversity 2.1, 26.

Mandakovic, Dinka, et al., 2018. Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. Sci. Rep. 8.1, 5875.

Marre, Guilhem, et al., 2020. Deep convolutional neural networks to monitor coralligenous reefs: operationalizing biodiversity and ecological assessment. Ecol. Inform. 59, 101110.

Moore, John C., McCann, Kevin, de Ruiter, Peter C., 2005. Modeling trophic path- ways, nutrient cycling, and dynamic stability in soils. Pedobiologia 49.6, 499–510.

Oldenburg, Ellen, et al., 2023. DeepLOKI-a deep learning based approach to identify zooplankton taxa on high-resolution images from the optical plankton recorder LOKI. Front. Mar. Sci. 10, 1280510.

Oldenburg, Ellen, et al., 2024a. Beyond blooms: a novel time series analysis framework predicts seasonal keystone species and sheds light on Arctic pelagic ecosystem stability. bioRxiv, 2024–03.

Oldenburg, Ellen, et al., 2024b. “Beyond blooms: the winter ecosystem reset determines microeukaryotic community dynamics in the Fram Strait”. en. Commun. Earth Environ. 5 (1), 643. https://doi.org/10.1038/s43247-024-01782-0 issn: 2662-4435. https://www.nature.com/articles/s43247-024-01782-0 (visited on 02/13/2025).

Oldenburg, Ellen, et al., 2024c. Sea-ice melt determines seasonal phytoplankton dynamics and delimits the habitat of temperate Atlantic taxa as the Arctic Ocean atlanti- fies. ISME communications 4.1, ycae027.

Paliy, Oleg, Shankar, Vijay, 2016. Application of multivariate statistical techniques in microbial ecology. Mol. Ecol. 25.5, 1032–1057.

Peter, J Wangersky, 1978. Lotka-Volterra population models. Annu. Rev. Ecol. Systemat. 9, 189–218.

Pilar Cabezas, M., Fonseca, Nuno A., Muñoz-Mérida, Antonio, 2024. MIMt: a curated 16S rRNA reference database with less redundancy and higher accuracy at species- level identification. Environmental Microbiome 19.1, 88.

Plis, Sergey M., et al., 2014. High-order interactions observed in multi-task intrinsic networks are dominant indicators of aberrant brain function in schizophrenia. Neuroimage 102, 35–48.

Poretsky, Rachel, et al., 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 9 (4), e93827.

Priest, Taylor, et al., 2023. Atlantic water influx and sea-ice cover drive taxonomic and functional shifts in Arctic marine bacterial communities. ISME J. 17 (10), 1612–1625.

Przymus, Piotr, et al., 2025. Deep learning in microbiome analysis: a comprehensive review of neural network models. Front. Microbiol. 15, 1516667.

Qu, Kaiyang, et al., 2019. Application of machine learning in microbiology. Front. Microbiol. 10, 827.

Romero, Miguel F., et al., 2025. Global Metagenomics Reveals Hidden Protist Diversity.

Royo-Llonch, Marta, et al., 2021. Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. Nat. Microbiol. 6 (12), 1561–1574.

Runge, Jakob, et al., 2019. Inferring causation from time series in Earth system sciences. Nat. Commun. 10 (1), 2553.

Ryan, B Ghannam, Techtmann, Stephen M., 2021. Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. Comput. Struct. Biotechnol. J. 19, 1092–1107.

Scharlemann, Jörn PW., et al., 2008. Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. PLoS One 3 (1), e1408.

Scheffer, Marten, et al., 2001. Catastrophic shifts in ecosystems. Nature 413.6856, 591–596.

Scheffer, Marten, et al., 2009. Early-warning signals for critical transitions. Nature 461.7260, 53–59.

Schmid, P.J., et al., 2011. “Applications of the dynamic mode decomposition”. en. Theor. Comput. Fluid Dynam. 25 (1–4), 249–259. https://doi.org/10.1007/s00162-010-0203-9 issn: 0935-4964, 1432-2250. http://link.springer.com/10.1007/s00162-010-0203-9 (visited on 02/12/2025).

Schmidt, Niels M., et al., 2017. “Interaction webs in arctic ecosystems: determinants of arctic change?” en. Ambio 46.S1, 12–25. https://doi.org/10.1007/s13280-016-0862-x issn: 0044-7447, 1654- 7209. http://link.springer.com/10.1007/s13280-016-0862-x (visited on 02/13/2025).

Schulz, Jan, et al., 2010. Imaging of plankton specimens with the lightframe on-sight keyspecies investigation (LOKI) system. Journal of the European optical society-rapid publications 5, 10017s.

Shade, Ashley, et al., 2012. Fundamentals of microbial community resistance and resilience. Front. Microbiol. 3, 417.

Shao, Zezhi, et al., 2022. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1567–1577.

Shumway, Robert H., Stoffer, David S., Stoffer, David S., 2000. Time Series Analysis and its Applications, vol. 3. Springer.

Siami-Namini, Sima, Tavakoli, Neda, Namin, Akbar Siami, 2018. A comparison of ARIMA and LSTM in forecasting time series. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). Ieee, pp. 1394–1401.

Singh, Pushpendra, 2018. Novel Fourier quadrature transforms and analytic signal representations for nonlinear and non-stationary time-series analysis. R. Soc. Open Sci. 5 (11), 181131.

Singh, Pushpendra, et al., 2017. The Fourier decomposition method for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Math. Phys. Eng. Sci. 473 (2199), 20160871.

Soltwedel, Thomas, et al., 2016. “Natural variability or anthropogenically-induced variation? Insights from 15 years of multidisciplinary observations at the arctic marine LTER site HAUSGARTEN”. en. Ecol. Indic. 65, 89–102. https://doi.org/10.1016/j.ecolind.2015.10.001 issn: 1470160X. https://linkinghub.elsevier.com/retrieve/pii/S1470160X15005361 (visited on 02/04/2025).

Søndergaard, Martin, et al., 2011. Using chlorophyll a and Cyanobacteria in the ecological classification of lakes. Ecol. Indic. 11 (5), 1403–1412.

Spadon, Gabriel, et al., 2021. Pay attention to evolution: time series forecasting with deep graph-evolution learning. IEEE Trans. Pattern Anal. Mach. Intell. 44 (9), 5368–5384.

Steele, Joshua A., et al., 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J. 5 (9), 1414–1425.

Strogatz, S.H., 2001. Exploring complex networks. nature 410 (6825), 268–276. https://doi.org/10.1038/35065725.

Suzuki, Kenta, et al., 2021. Energy landscape analysis elucidates the multistability of ecological communities across environmental gradients. Ecol. Monogr. 91 (3), e01469.

Takens, Floris, 2006. Detecting strange attractors in turbulence. In: Dynamical Sys- Tems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the Uni- Versity of Warwick 1979/80. Springer, pp. 366–381.

Taylor, Priest, et al., 2025. “Seasonal recurrence and modular assembly of an Arctic pelagic marine microbiome”. en. Nat. Commun. 16 (1), 1326. https://doi.org/10.1038/s41467-025-56203-3 issn: 2041-1723. https://www.nature.com/articles/s41467-025-56203-3 (visited on 02/25/2025).

The Human Microbiome Project Consortium, 2012. “Structure, function and diversity of the healthy human microbiome”. en. Nature 486 (7402), 207–214. https://doi.org/10.1038/nature11234 issn: 0028-0836, 1476-4687. https://www.nature.com/articles/nature11234 (visited on 02/12/2025).

Thomas, N Kipf, Welling, Max, 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Tian, Xi, 2015. The Use of high-throughput Amplicon Deep Sequencing to Explore Aquatic Virus Communities. University of British Columbia. PhD thesis.

Timmermans, Mary-Louise, Marshall, John, 2020. Understanding Arctic Ocean circulation: a review of ocean dynamics in a changing climate. J. Geophys. Res., Oceans 125 (4), e2018.

Tonner, Peter D., et al., 2017. Detecting differential growth of microbial populations with Gaussian process regression. Genome Res. 27 (2), 320–333.

Tsilimigras, Matthew C.B., Fodor, Anthony A., 2016. “Compositional data analy- sis of the microbiome: fundamentals, tools, and challenges”. en. Ann. Epidemiol. 26 (5), 330–335. https://doi.org/10.1016/j.annepidem.2016.03.002 issn: 10472797. https://linkinghub.elsevier.com/retrieve/pii/S1047279716300722 (visited on 02/24/2025).

Turnbaugh, Peter J., et al., 2007. “The Human Microbiome Project”. en. Nature 449 (7164), 804–810. https://doi.org/10.1038/nature06244 issn: 0028-0836, 1476-4687. https://www.nature.com/articles/nature06244 (visited on 02/12/2025).

Ushio, Masayuki, et al., 2017. Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. bioRxiv 113472.

Ushio, Masayuki, et al., 2018. Fluctuating interaction network and time-varying stability of a natural fish community. Nature 554 (7692), 360–363.

Valencia-Agami, Sonia S., et al., 2024. Structure and composition of microbial communities in the water column from Southern Gulf of Mexico and detection of putative hydrocarbon-degrading microorganisms. Environmental Microbiology Reports 16 (3), e13264.

Vincent, Warwick F., Laybourn-Parry, Johanna, 2008. Polar Lakes and Rivers: Limnol-Ogy of Arctic and Antarctic Aquatic Ecosystems. OUP Oxford.

Von Appen, Wilken-Jon, et al., 2021. Sea-ice derived meltwater stratification slows the biological carbon pump: results from continuous observations. Nat. Commun. 12.1, 7309.

Větrovský, Tomás, Baldrian, Petr, 2013. "The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses". en. PLoS One 8 (2). https://doi.org/10.1371/journal.pone.0057923. Ed. by Josh Neufeld, e57923. issn: 1932-6203. https://dx.plos.org/10.1371/journal.pone.0057923 (visited on 02/12/2025).

Wang, Jie, 2023. An intuitive tutorial to Gaussian process regression. Comput. Sci. Eng. 25 (4), 4–11.

Wang, Jiawei, 2024. Multivariate time series forecasting and classification via GNN and transformer models. Journal of Computer Technology and Software 3, 9.

Wang, Yongbin, et al., 2020. Secular seasonality and trend forecasting of tuberculosis incidence rate in China using the advanced error-trend-seasonal framework. In: Infection and Drug Resistance, pp. 733–747.

Weinreb, Caleb, et al., 2018. "Fundamental limits on dynamic inference from single-cell snapshots". en. Proc. Natl. Acad. Sci. 115.10, 1091–6490. https://doi.org/10.1073/pnas.1714723115 issn: 0027-8424. https://pnas.org/doi/full/10.1073/pnas.1714723115 (visited on 02/12/2025).

Weiss, Sophie, et al., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J. 10 (7), 1669–1681.

Weiss, Sophie, et al., 2017. "Normalization and microbial differential abundance strategies depend upon data characteristics". en. Microbiome 5.1, 27. https://doi.org/10.1186/s40168-017-0237-y issn: 2049-2618. https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-017-0237-y (visited on 02/23/2025).

Wessling, Erin G., et al., 2022. Evaluating the efficacy of a consumer-centric method for ecological sampling: using bonobo (Pan paniscus) feeding patterns as an instru- ment for tropical forest characterization. Ecol. Evol. 12 (12), e9606.

Widder, Stefanie, et al., 2016. Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J. 10 (11), 2557–2568.

Wright, Marvin N., Ziegler, Andreas, 2017. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Software 77, 1–17.

Wu, Zonghan, et al., 2020. A comprehensive survey on graph neural networks. IEEE Transact. Neural Networks Learn. Syst. 32 (1), 4–24.

Xue, Yuanyuan, et al., 2018. Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. ISME J. 12 (9), 2263–2277.

Ye, Hao, et al., 2015. Distinguishing time-delayed causal interactions using convergent cross mapping. Sci. Rep. 5 (1), 14750.

Zhang, Kaiyang, Nakaoka, Shinji, 2024. An energy landscape approach reveals the potential key bacteria contributing to the development of inflammatory bowel disease. PLoS One 19 (6), e0302151.

Zhang, Jinlun, et al., 2010. Modeling the impact of declining sea ice on the Arctic marine planktonic ecosystem. J. Geophys. Res., Oceans 115, C10.

Zhu, Qiang, et al., 2019a. Graph embedding deep learning guides microbial biomarkers' identification. Front. Genet. 10, 1182.

Zhu, Qiang, et al., 2019b. Graph embedding deep learning guides microbial biomarkers' identification. Front. Genet. 10, 1182.