

Promoters

Prof. Jana Asselman

Prof. Colin Janssen

Faculty of Bioscience Engineering, Ghent University

Dr. Maarten De Rijcke

Flanders Marine Institute (VLIZ)

Members of the examination committee

Prof. Katleen Raes (Chair)

Prof. Tim De Meyer

Dr. Maarten De Mol

Faculty of Bioscience Engineering, Ghent University

Prof. Jan Mees

Flanders Marine Institute (VLIZ)

Dr. Elisa Berdalet

Institute of Marine Sciences (ICM), Spanish National Research Council

Dean

Prof. Els Van Damme

Rector

Prof. Petra De Sutter

From Phytoplankton to Bioactivity: Sea Spray Aerosols as a Mechanistic Link between Ocean Ecology and Human Health

Zixia Liu

Thesis submitted in fulfilment of the requirements for the
degree of
Doctor of Bioscience Engineering

Submitted on 10th January 2026
Public Defence on 12th May 2026

Refer to this work as:

Liu, Z. (2026). *From phytoplankton to bioactivity: Sea spray aerosols as a mechanistic link between ocean ecology and human health* [Doctoral dissertation, Ghent University].

ISBN: 9789493542068

Copyright © 2026

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Acknowledgements of financing institutions

This work was supported by the China Scholarship Council (CSC) under Grant No. 201906100034. Additionally, the author acknowledges the financial support provided by the Special Research Fund (BOF) of Ghent University through the special COVID-19 extension funding.

Statement on the Use of Artificial Intelligence

During the preparation of this work, the author used Google Gemini and OpenAI ChatGPT in order to proofread the manuscript and improve sentence clarity. Semantic Scholar, Elicit AI, and Consensus AI were utilized to assist in the literature search and the identification of relevant references.

Additionally, Google NanoBanana was used to generate specific illustrative figures (Chapter 1, Figures 1–3); all other figures are original. The author has carefully reviewed and verified all AI-generated or AI-assisted content. The author assumes full responsibility for the final content of this thesis.

Preface

From its official start on October 1, 2019, to today, April 27, 2026, I have navigated this PhD journey for six and a half years. I never anticipated it would take this long to bring this thesis to a close.

Looking back on these six-plus years, the initial challenges came from nature. Between 2020 and 2022, I had to conduct field sampling once a month. To collect sea spray aerosol (SSA), I needed to wait by the sea for three hours facing strong winds, constantly adjusting the equipment according to the tides. The winter sea breeze was piercingly cold. Fortunately, my girlfriend Qianru would sometimes accompany me; we would huddle together in a tent for warmth and eat the lunch she had made herself. That has become my warmest memory.

Of course, the sea is not always gentle. It was extremely windy on May 11, 2020, and the beach was covered in thick foam produced by *Phaeocystis*. After setting up the equipment, I asked Yunmeng, who was also collecting samples, to keep an eye on it while I returned to MSO to filter seawater. Unexpectedly, just half an hour later, Yunmeng ran back to tell me the tide was rising exceptionally fast. When I rushed back to the beach, the equipment had already been swallowed by the waves. It took me several seconds to even locate it in the "sea of foam." By then, the water level had already risen more than 20 centimeters above the base, and the lower half of the air pump was completely submerged. I jumped into the waist-deep waves to drag the equipment back to the beach, but the air pump, worth 1,800 euros, was ruined. I contacted Jana with extreme trepidation. To my surprise, she didn't blame me at all and simply comforted me: "These things happen." I will never forget that level of understanding.

In the same year as that thrilling sampling experience, we faced another massive external challenge: COVID-19. The pandemic completely changed

the way I worked; wearing masks in public became mandatory, and working from home became the norm. Interestingly, though, during those "isolated" days with reduced social interaction, I was actually at my healthiest and rarely caught a fever.

Working from home gave me more time in front of the computer, but it also brought about the biggest "mental challenge" of my PhD: data analysis. During the data processing stages for Chapter 3 (RNA-Seq) and Chapter 4 (Metabarcoding), I found myself bogged down. Although I was very interested in upstream data quality control and statistical analysis, at the time, I could only piece together code line by line relying on Jana's example scripts, various literature, bioinformatics forums, and GitHub. The time spent purely on technical details amounted to nearly a year. It wasn't until the rise of AI technology in 2024 that my PhD research received a massive "boost." AI helped me clear the hurdles of tedious coding, making me profoundly realize: when technical issues are no longer a bottleneck, we can completely liberate our energy and focus on the "scientific questions" themselves. That feeling was truly mind-blowing.

As the data gradually became clearer, I also began to step onto a larger academic stage. In the fall of 2023, with generous funding from CWO and Jana, I attended the ACS conference and gave the first major international conference presentation of my life. During this time, I also had the honor of interacting with the renowned Professor Kim Prather and her team. If given the chance, I am still incredibly eager to visit her lab.

This conference and the entire PhD journey have greatly broadened my horizons. Although this thesis contains five independent studies that I have completed, in reality, I still have many unfinished projects on hand due to time constraints: such as the MART experiment conducted with Jing He; the weekly SSA samples collected on the MSO roof with Yunmeng Li in the summer of 2022; the SSA distribution map originally planned as Chapter 6;

and further optimization of the meta-analysis in the second part of Chapter 5... I once heard that a good PhD experience provides a scientist with research inspiration for years to come. From this perspective, I am immensely proud of my PhD career.

This six-plus-year journey would not have been possible without the companionship and support of many people. I would like to thank my supervisory team: Prof. Jana Asselman, Prof. Colin Janssen, and Dr. Maarten De Rijcke. Jana has always been my most solid foundation; whether in academics, finances, or daily life, she has given me unreserved support. I still remember when I was developing the seawater DNA extraction protocol for Chapter 4 and failed continuously for a whole week. It was she who patiently helped me identify the problem, ultimately spending an entire day with me in the lab to successfully complete the extraction. Colin's academic insight amazed me at every meeting, always getting straight to the heart of the matter. I still remember what he told me the first time we met: "What we want to do is not just publish papers, but conduct research of great significance to humanity; our goal is the Nobel Prize." Maarten is incredibly warm and friendly. In our previous monthly meetings with Yunmeng, he not only gave me many practical suggestions on phytoplankton research but also treated us to meals several times (I still remember once when he drove us to find the best fries in Ostend, but unfortunately, they were closed).

I also want to thank my colleagues, especially Emmanuel Van Acker, Ilias Semmouri, Eveline Diopere, and Yunmeng Li, who made important contributions to this thesis and received their well-deserved credit in their respective chapters. Thanks to Nancy De Saeyer for helping me coordinate sampling equipment and experimental arrangements; thanks to Emmy Pequeur for teaching me how to culture algae and perform chemical analyses; thanks to Jolien Depecker for helping extract DNA. Without your help, this thesis could not have been completed. Furthermore, I would like

to thank Miao Peng, Qiyun Zhang, João Barbosa, Zhiyue Niu, Silke Lambert, Wenxin Liu, Lotte Janssens, and many other colleagues; thank you for your encouragement and support in both work and daily life.

Finally, I want to save my deepest gratitude for my family. Over the past six-plus years, it is precisely because of the selfless psychological and financial support from my parents and my girlfriend that this thesis has finally come to fruition. Thank you all from the bottom of my heart!

Contents

Abbreviations 7

Chapter 1 Sea Spray Aerosols as a Link Between Ocean and Human Health.....10

 1. Oceans, aerosols, and human health: an emerging connection 10

 2. Formation and enrichment of SSAs 14

 2.1. Physical mechanisms of SSA production.....14

 2.2. Chemical enrichment and composition of SSAs 20

 3. From composition to function: bioactivity and exposure pathways.....22

 3.1. Chemical diversity and molecular carriers of bioactivity 26

 3.2. Epidemiological cases of harmful algal blooms and inhalation exposures.....30

 3.3. Emerging evidence for beneficial bioactivities 32

 4. Phytoplankton dynamics as a driver of SSA variability.....35

 4.1. Phytoplankton, carbon cycling, and surface organic reservoirs.....38

 4.2. Seasonal bloom regimes in the Southern Bight of the North Sea 39

 4.3. Environmental controls on bloom timing and magnitude 41

 5. Marine natural products as a reservoir of aerosolizable compounds 43

 6. Thesis rationale, objectives, and outline.....46

Chapter 2 Phospholipids in Sea Spray Aerosols Mediated by Phytoplankton and Environmental Drivers.....49

 Abstract.....49

 Keywords.....50

1. Introduction.....	50
2. Material and Methods.....	52
2.1. SSA Sampling	52
2.2. Chemical Analysis.....	55
2.3. Environmental Variables	57
2.4. Aerosol Source.....	58
2.5. DPPC Concentration Model	59
3. Results	61
3.1. Analytical Performance	61
3.2. Air Concentrations and SSA Composition	61
3.3. Aerosol Source.....	65
3.4. DPPC Concentration Model	69
4. Discussion.....	73
4.1. Source and Aerosolization of DPPC	73
4.2. DPPC Exposure and Human Health.....	77
Supporting Information.....	81
Chapter 3 Seasonal Variation in the Bioactivity of Sea Spray Aerosols Driven by Phytoplankton	97
Abstract.....	97
1. Introduction.....	98
2. Results	101
2.1. Differential gene expression in response to treatments.	102
2.2. Genes in mTOR pathway in response to SSA exposure.	110
2.3. Seasonal variations of SSA composition and exposure effects.....	115
3. Discussions	121
3.1. Bioactive chemicals in SSA may be released by	

phytoplankton.....	121
3.2. Implications of SSA exposure on key biological pathways.....	122
3.3. Limitations.....	123
4. Conclusions.....	126
5. Methods.....	127
5.1. Aerosol sampling and chemical extractions.....	127
5.2. Cell culture and exposure experiment.....	129
5.3. RNA extraction and sequencing.....	132
5.4. Laser-assisted ambient ionization mass spectrometry (LA-REIMS) analyses.....	132
5.5. Data analyses.....	133
Supplementary Information.....	135
Chapter 4 Seasonal Dynamics of Coastal Phytoplankton Communities Revealed by DNA Metabarcoding.....	161
Abstract.....	161
Keywords.....	162
1. Introduction.....	162
2. Materials and methods.....	166
2.1. Sample collection.....	166
2.2. DNA extraction, amplification, and sequencing.....	167
2.3. Bioinformatics.....	169
2.4. Data acquired from public databases.....	170
2.5. Statistics and visualization.....	172
3. Results.....	172
3.1. Sequencing results and phytoplankton community profile.....	172
3.2. Phytoplankton community responses to environmental factors.....	184

4. Discussion.....	196
4.1. Seasonal dynamics of environmental drivers in the Southern Bight of the North Sea.....	196
4.2. Earlier blooms observed in 2021	198
4.3. Limitations	201
Supplementary Information	204
Chapter 5 Marine Natural Products and Their Enrichment Potential in Sea Spray Aerosols.....	217
Part 1: MarineNP: A Comprehensive and Locally Deployable Database of Marine Natural Products	217
Graphical Abstract	217
Abstract.....	218
Keywords.....	218
1. Introduction.....	218
2. Materials and methods	222
2.1. Data collection.....	222
2.2. Data processing.....	222
3. Data description.....	223
3.1. Database content.....	223
3.2. Implementation	223
4. Discussion.....	224
4.1. User interface and functionality	224
4.2. Comparative Advantages Over Existing Resources.....	225
4.3. Conclusion	225
Part 2: A Quantitative Workflow for Assessing the Physicochemical Fate of Natural Compounds in Sea Spray Aerosol.....	227
Abstract.....	227
1. Introduction.....	228

2. Methods	233
2.1. Data sources and initial preprocessing	233
2.2. Descriptor definition and module-specific feature sets	234
2.3. Statistical feature cleaning and transformation	240
2.4. Assessment across mechanistic modules	241
2.5. Literature-based cluster annotation using reference compounds	241
2.6. Software and reproducibility	243
3. Results	243
3.1. Data integration and final descriptor spaces	243
3.2. Module results: enrichment, aerosolization, and atmospheric stability	247
3.3. Empirical anchoring of clusters using literature reference compounds	257
3.4. Prioritized candidate sets and overlaps	262
4. Discussion	266
4.1. Module spaces, clustering choices, and fate structure	266
4.2. Limits and uncertainty	269
4.3. Contribution, implications, and outlook	270
Supplementary Information	273
Chapter 6 Conclusions and Perspectives	279
1. Synthesis of the main findings	280
1.1. Q1: Phytoplankton derived surfactant lipids in natural SSAs	280
1.2 Q2: Seasonal variability in SSA composition and bioactivity	282
1.3 Q3: Changing phytoplankton phenology and implications for SSA mediated exposure	283
1.4 Q4: Marine natural products as candidate aerosolizable	

biogenics 285

1.5 Conceptual and methodological contributions..... 288

2. Limitations and perspectives 291

2.1 Limitations of the present work..... 291

2.2 Perspectives 294

3. Open science, data stewardship, and tool accessibility 297

References..... 301

Abbreviations

AMB	Ambient coastal aerosol
AMPK	AMP-activated protein kinase
ANOVA	Analysis of Variance
AOPWIN	Atmospheric Oxidation Program for Microsoft Windows
API	Application Programming Interface
BEAS-2B	Human bronchial epithelial cell line
BP	Boiling Point
	Collective benefit, Authority to control, Responsibility, and Ethics
CARE	
CAS	Chemical Abstracts Service
CCN	Cloud Condensation Nuclei
CMNPD	Comprehensive Marine Natural Products Database
CSV	Comma-Separated Values
DEG	Differentially Expressed Gene
DF	Dilution Factor
DHA	Docosahexaenoic acid
DMEM	Dulbecco's Modified Eagle Medium
DMPC	Dimyristoylglycerophosphocholine
DNA	Deoxyribonucleic acid
DPPC	Dipalmitoylphosphatidylcholine
EF	Enrichment Factor
EPA	Eicosapentaenoic acid
FAIR	Findability, Accessibility, Interoperability, and Reuse
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FWHM	Full Width at Half Maximum
GLM	Generalized Linear Model
HAB	Harmful Algal Bloom
HESI	Heated Electrospray Ionization
	Inductively Coupled Plasma Optical Emission
ICP-OES	Spectroscopy
InChI	International Chemical Identifier

IUPAC	International Union of Pure and Applied Chemistry
KEGG	Kyoto Encyclopedia of Genes and Genomes
LA-REIMS	Laser-Assisted Rapid Evaporative Ionization Mass Spectrometry
LC-MS	Liquid Chromatography–Mass Spectrometry
LOD	Limit of Detection
LOO	Leave-One-Out
LOQ	Limit of Quantification
MAD	Median Absolute Deviation
MNP	Marine Natural Product
MP	Melting Point
mTOR	Mammalian Target of Rapamycin
MW	Molecular Weight
NCE	Normalized Collision Energy
NCBI	National Center for Biotechnology Information
NEG	Negative control
NF- κ B	Nuclear Factor kappa-light-chain-enhancer of activated B cells
NZV	Near-Zero Variance
OBIS	Ocean Biodiversity Information System
PAR	Photosynthetically Active Radiation
PCA	Principal Component Analysis
PCSK9	Proprotein Convertase Subtilisin/Kexin Type 9
PhyC	Phytoplankton Carbon
PI3K	Phosphoinositide 3-kinase
PTFE	Polytetrafluoroethylene
QED	Quantitative Estimate of Drug-likeness
RDA	Redundancy Analysis
RDS	Respiratory Distress Syndrome
RIN	RNA Integrity Number
RNA	Ribonucleic acid
SMILES	Simplified Molecular Input Line Entry System
SPM	Suspended Particulate Matter
SRA	Sequence Read Archive

SRT	Surfactant Replacement Therapy
SSA	Sea Spray Aerosol
SSML	Sea Surface Microlayer
SWT	Seawater Temperature
TEP	Transparent Exopolymer Particles
TPSA	Topological Polar Surface Area
UHPLC	Ultra-High-Performance Liquid Chromatography
VIF	Variance Inflation Factor
VP	Vapor Pressure
WoRMS	World Register of Marine Species

Chapter 1 Sea Spray Aerosols as a Link Between Ocean and Human Health

1. Oceans, aerosols, and human health: an emerging connection

The world's oceans cover more than 70% of the Earth's surface and are a major source of atmospheric particles (Schulz et al., 2004). They act as both a sink and a source for gases and particulate matter that influence weather, climate, and ultimately life on land. Among the many materials exchanged across the sea–air boundary, sea spray aerosols (SSAs) stand out as one of the most abundant and chemically complex natural aerosol types, with estimated global emissions of up to 70 petagrams per year (Gantt and Meskhidze, 2013; Grythe et al., 2014). SSAs originate when breaking waves entrap air bubbles that burst at the ocean surface, ejecting microscopic droplets into the atmosphere. These droplets, although primarily saline, carry a diverse suite of organic and inorganic substances as well as micro-organisms such as bacteria, viruses, and their fragments derived from the uppermost ocean layer (Schiffer et al., 2018). Through their optical properties and role as cloud condensation and ice-nucleating particles, SSAs exert a major influence on the Earth's radiative balance and hydrological cycle (O'Dowd and De Leeuw, 2007; Quinn et al., 2015).

Beyond climate regulation, the ocean–atmosphere exchange is increasingly recognized as a vector of biological material and bioactive molecules to the atmosphere, and hence as a contributor to human exposure. The global population residing within 100 km of the coastline is expanding (Cosby et al., 2024), exposing hundreds of millions of people to marine air on a daily basis (Pendergraft et al., 2023). Populations in coastal areas encounter SSAs under highly variable conditions: wind speed, fetch, wave state, and biological productivity can all modulate near-shore concentrations by several orders of magnitude over hours to seasons (Lim et al., 2023a; Quinn et al., 2015; Van Acker et al., 2021b). Yet, despite decades of atmospheric research, the potential physiological implications of inhaling natural marine

aerosols remain only partially understood. Traditionally, this exposure has been viewed as benign or even therapeutic, yet the molecular complexity of SSAs suggests a subtler reality.

Early clues emerged from observations during harmful algal blooms (HABs). Episodes of respiratory irritation among beachgoers exposed to aerosolized marine toxins (such as brevetoxins produced by *Karenia brevis*) demonstrated that marine biogenic material can reach the human respiratory tract in biologically active form (Fleming et al., 2007, 2009; Lavery et al., 2023; Lim et al., 2023a). However, such events are not a universal outcome of algal blooms. They are confined in space and time, and typically involve extreme toxin concentrations. In contrast, chronic background exposure to non-toxic marine aerosols is continuous for coastal residents and recreational users.

Proximity to “blue spaces” correlates with improved overall health, reduced cardiovascular and respiratory mortality, and enhanced mental wellbeing (Gascon et al., 2017; Moore, 2015). Coastal-health epidemiology provides a population-level context for this mechanistic question. In England, Wheeler et al. (2012) analysed 2001 Census data for 48.2 million residents across 32,482 Lower-layer Super Output Areas and found that self-reported good health increased with residential coastal proximity; after adjustment for green space and socioeconomic deprivation, the prevalence of “good” health in urban areas within 1 km of the sea was 1.13% higher than in areas more than 50 km inland, with the association strongest in the most deprived communities. A subsequent longitudinal analysis of British Household Panel Survey data addressed part of the selection problem by comparing the same individuals across time and place: across 109,844 observations from 15,471 individuals for general health and 114,133 observations from 15,361 individuals for mental health, living within 5 km of the coast was associated with small but significant improvements in general and mental health, although not life satisfaction (White et al., 2013). These findings sit within a broader transition in environmental epidemiology

from focusing solely on harmful exposures toward research that also includes salutogenic influences of natural environments, including blue spaces, through pathways such as stress reduction, attention restoration, physical activity, social contact, immune regulation, and improved environmental quality (Silva et al., 2018). At the same time, coastal settings remain mixed exposure environments: Sandifer (2023) notes that eight of the world's ten most populous cities are coastal, that more than 200 million people in Europe live in coastal or river-front cities, and that coastal residents may encounter both health-supporting blue spaces and hazards such as HAB toxins, *Vibrio* bacteria, sewage-associated microbes, oil spills, storms, floods, and rip currents. Thus, coastal epidemiology does not identify SSAs as the causal factor behind coastal-health associations, but it provides the public-health rationale for examining whether SSA composition and bioactivity could form one mechanistic link between coastal ecosystem state and human physiology. These associations suggest that marine aerosol exposure may have context-dependent effects, varying with aerosol composition, dose, and environmental conditions.

In 2015, Moore proposed the “biogenics hypothesis”, suggesting that natural airborne exposure to certain marine compounds may downregulate key cellular signaling pathways, such as PI3K/Akt/mTOR, thereby promoting anti-inflammatory, anticancer, and longevity-associated effects (Moore, 2015). This hypothesis reframed the ocean not only as a climatic regulator but also as a potential source of bioactive molecules that may interact with human physiology. Subsequent *in vitro* studies supported this concept: human bronchial epithelial cells exposed to natural SSA extracts exhibited downregulation of mTOR-related and pro-inflammatory pathways and activation of energy-regulating AMPK signaling (Asselman et al., 2019; Van Acker et al., 2020). These responses resemble those elicited by pharmacological agents targeting chronic inflammation and cancer metabolism, raising the possibility that ambient SSA exposure may contribute, however subtly, to respiratory and systemic homeostasis.

In parallel with Moore's work, Rook advanced the "old friends" hypothesis (Rook, 2023), which proposes that reduced contact with evolutionarily familiar environmental microorganisms impairs the development of immunoregulatory networks and thereby increases susceptibility to chronic inflammatory and allergic diseases (Rook et al., 2003). In this view, regular low-level exposure to diverse environmental microbes and their molecular products helps train and maintain regulatory immune circuits. Recent work on natural SSAs indicates that coastal marine aerosols can carry microbial cells and immune active components that engage innate immune receptors in human cells, consistent with the idea that such exposures contribute to immune education rather than acting solely as stressors (Li et al., 2024). Nevertheless, whether marine aerosol microorganisms represent "old friends" in the strict co-evolutionary sense remains unresolved. Together, the biogenics hypothesis and the broader old friends hypothesis provide a complementary framework in which SSA-borne compounds and microorganisms are not only potential toxicants during extreme events, but also candidates for maintaining balanced immune and stress responses under typical coastal conditions.

The emerging picture is therefore two-sided. Under harmful algal bloom conditions, such as *Karenia brevis* red tides that release brevetoxins into the air (Fleming et al., 2005), or in the presence of surface pollution, such as oil slicks at the air-sea interface (Fontana, 1976), marine aerosols may deliver harmful toxins; under normal ecological conditions, they may convey beneficial biogenics. Determining where and when each outcome dominates requires understanding not only the physical production of SSAs but also their variable chemical composition and biological origin under changing seasonal and meteorological conditions. The ocean surface is a dynamic chemical reactor in which sunlight, microbial activity, and organic exudation, together with physical drivers such as wind speed, wave state, and mixing, continuously alter the pool of aerosolizable molecules (Lewis and Schwartz, 2004;

O'Dowd and De Leeuw, 2007; Quinn et al., 2015). Which of these compounds are preferentially transferred into the atmosphere, how their abundance varies with ocean ecology as well as with weather and season, and how they interact with human cells remain fundamental open questions.

Addressing these questions demands a cross-disciplinary approach linking marine biogeochemistry, aerosol physics, molecular biology, and environmental health. This thesis is built around that interface. The subsequent chapters explore how oceanic biological processes shape the composition and bioactivity of SSAs, how these aerosols interact with human lung cells at the molecular level, and how data-driven informatics can map the diversity of marine natural products underlying these effects. Together, these investigations aim to establish sea spray aerosols as a tangible mechanistic link between ocean ecology and human health.

2. Formation and enrichment of SSAs

2.1. Physical mechanisms of SSA production

The history of research on sea spray aerosol (SSA) production dates back to the late 1940s, with some of the earliest investigations carried out by Woodcock in 1948, who proposed that the water bubbles generated by whitecap waves could lead to the formation of SSA particles (Woodcock, 1948). This hypothesis was derived from a mechanism previously suggested by Stuhlman, who observed that water droplets could be produced through the bursting of air bubbles at the surface of agitated water (Stuhlman, 1932). In the years that followed, the whitecap hypothesis put forth by Woodcock gained further validation through the research conducted by Boyce and Kientzler (Boyce, 1951; Kientzler et al., 1954). Their experiments provided empirical evidence that a significant number of SSA particles were generated when the bubbles of whitecap waves burst, thus lending credence to Woodcock's initial proposal.

Throughout the 1960s and 1970s, several studies continued to explore the dynamics of whitecap formation and their role in SSA production (Blanchard, 1964; Cipriano et al., 1987). These investigations expanded upon the knowledge of the conditions under which SSA particles were generated, such as wind speed, atmospheric stability, and seawater temperature. It was during this period that researchers began to elucidate the role of surface-active organic matter in the formation of SSA particles and the enrichment process that takes place at the bubble-water interface (Thorpe and Humphries, 1980). Subsequent studies in the 1980s and 1990s built upon the earlier findings, refining the understanding of the processes involved in SSA production and the factors that influenced them (Lewis and Schwartz, 2004; O'Dowd et al., 1997). These studies also shed light on the composition and size distribution of SSA particles, as well as the impact of various environmental factors, such as the presence of organic matter and phytoplankton, on their formation (Facchini et al., 2008).

Subsequent decades of research refined this mechanism into a quantitative physical description involving two dominant droplet formation pathways:

- 1) Film drops, formed from the disintegration of the bubble cap film, and
- 2) Jet drops, generated by the upward recoil of the collapsing cavity after bubble bursting.

Film drops are typically submicron in size and constitute the majority of SSA particle number emissions (the total count of aerosol particles), whereas jet drops are coarser (1–10 μm) and dominate the mass flux (i.e. the total mass of aerosol emitted). The partitioning between these modes is controlled by wind speed, water temperature, salinity, and the presence of surface-active materials that modify film elasticity and drainage (Facchini et al., 2008; Lewis and Schwartz, 2004; O'Dowd and De Leeuw, 2007).

The full sequence of SSA formation can be summarized as follows:

- 1) Wind stress and wave breaking entrain air into the upper ocean, generating bubble plumes of varying sizes (Figure 1 Step 1: Breaking wave & plume formation).
- 2) As these bubbles rise, they scavenge surface-active organic molecules at the bubble–water interface, altering film viscosity and drainage dynamics (Figure 1 Step 2: Bubble rise & organic scavenging).
- 3) When bubbles reach the surface, their films merge with the sea surface microlayer, forming transient whitecaps (Figure 1 Step 3: Surface microlayer & foam/whitecap formation).
- 4) Upon bursting, bubbles eject fine film drops and coarser jet drops; film drops inherit the compositional fingerprint of the sea surface microlayer (SSML), whereas jet drops more closely reflect subsurface seawater (Figure 1 Step 4: Bubble bursting mechanism).
- 5) These droplets are lofted into the marine boundary layer, the lowest part of the atmosphere directly influenced by exchanges with the ocean surface, where evaporation, coagulation, and condensation further transform their size and chemical state (Figure 1 Step 5: Aerosol transport & cloud processing).

Laboratory and field measurements demonstrate that SSA particle number increases approximately with the cube of wind speed (U_{10}^3) up to moderate conditions of 12–15 m s⁻¹ (Lewis and Schwartz, 2004), but composition changes even more dramatically because organic enrichment is not linearly related to bubble generation (Facchini et al., 2008; Quinn et al., 2015). The resulting aerosol population spans over six orders of magnitude in particle diameter (10 nm–100 μm), with distinct chemical regimes across this spectrum. Submicron particles (< 1 μm) are typically organically enriched and more surface-active, while supermicron particles are dominated by sea salt and mineral phases (Aller et al., 2017). These physical principles underpin the transfer of biogenic matter from the ocean to the atmosphere, establishing the

mechanical foundation for the biogenics hypothesis discussed earlier.

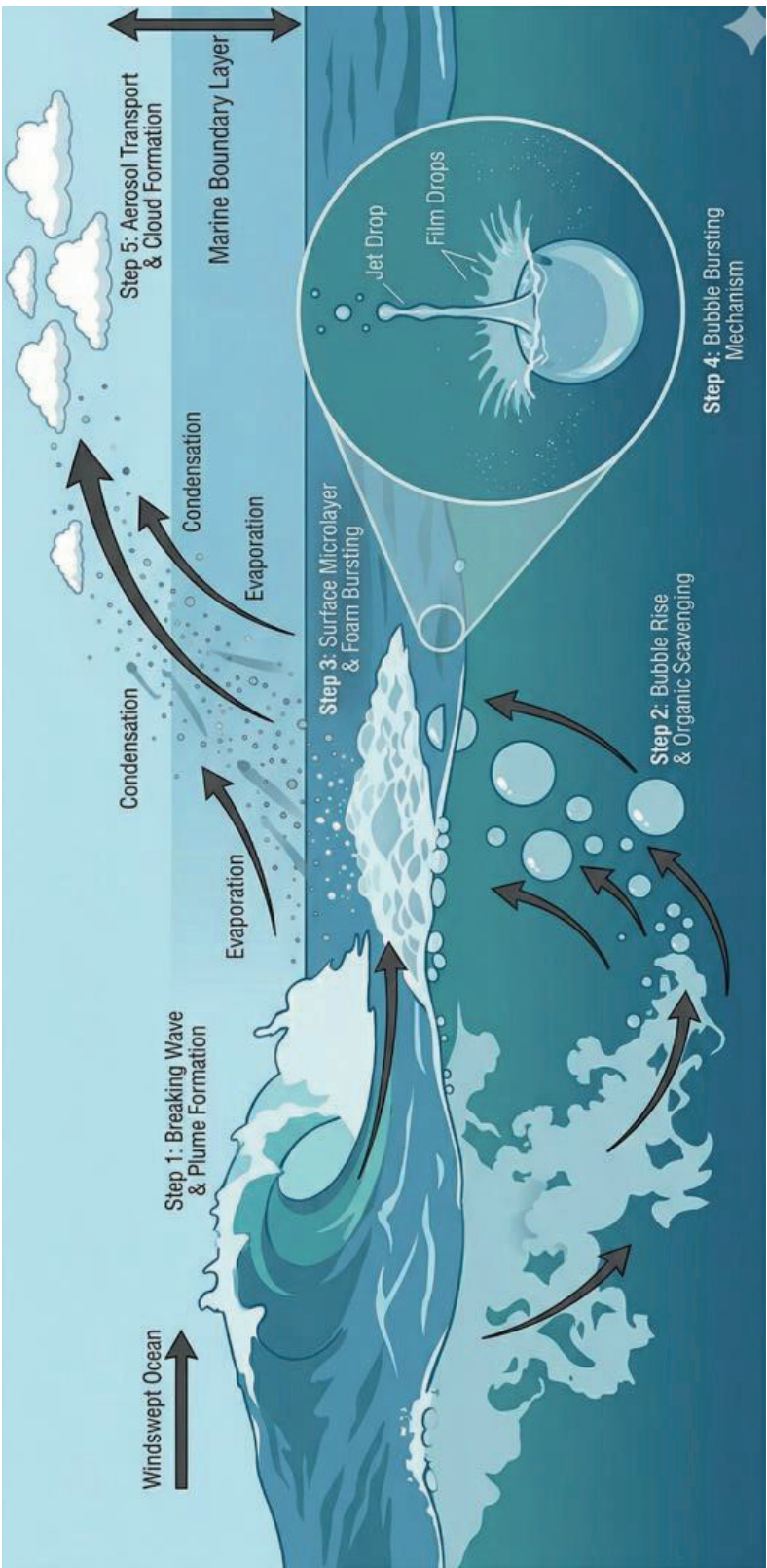


Figure 1. Schematic illustration of the Sea Spray Aerosol (SSA) generation and lifecycle. The process begins with wind-driven wave breaking that entrains air into the water column, creating turbulent bubble plumes (Step 1). As these submerged bubbles

rise against buoyancy, they scavenge surface-active organic matter from the surrounding bulk water, accumulating an organic coating on their interface (Step 2). At the surface, bubbles aggregate within whitecaps and interact with the Sea Surface Microlayer (SSML) (Step 3). The circular inset details the bubble bursting mechanism (Step 4), where the film cap shatters to produce fine film drops, followed by the collapse of the bubble cavity which ejects larger, vertically oriented jet drops. These particles are lofted into the Marine Boundary Layer (Step 5), where they undergo evaporation and condensation cycles, serving as cloud condensation nuclei (CCN) for marine cloud formation.

2.2. Chemical enrichment and composition of SSAs

Although SSAs originate from seawater, their chemical composition is far from a simple reflection of the ocean's bulk chemistry. Instead, selective enrichment processes occurring at the ocean–atmosphere boundary strongly modify the relative abundances of organic and inorganic constituents. This enrichment arises because surface-active and amphiphilic molecules preferentially accumulate at interfaces (both in the SSML and at bubble films), from which they are directly transferred to the atmosphere (Cunliffe et al., 2013; Engel et al., 2017). The enrichment of a compound is commonly quantified by an enrichment factor (EF):

$$EF_i = \frac{(C_i/C_{Na})_{SSA}}{(C_i/C_{Na})_{seawater}} \quad (1)$$

Where C_i is the concentration of compound i measured in SSA or in *seawater*, and C_{Na} is the concentration of sodium (Na^+) measured in the same matrix.

Sodium serves as a conservative tracer of sea because it is abundant and relatively stable during SSA generation, making it a useful normalization factor to compare enrichment relative to bulk sea-salt content rather than absolute concentration (Bertram et al., 2018). For many organic species, enrichment factors range from 10^2 to over 10^4 , indicating selective transfer mechanisms that depend on compound polarity, molecular weight, and particle size (Cravigan et al., 2020; Jayarathne et al., 2016).

Organic enrichment. Organic matter constitutes a significant and variable fraction of SSA mass, ranging from $< 5\%$ in open-ocean conditions to $> 90\%$ during intense biological activity (Quinn et al., 2015). Lipids, proteins, polysaccharides, and humic-like substances dominate this organic fraction. Lipids, especially fatty acids and phospholipids, are highly surface-active and partition strongly to the

bubble film, promoting their ejection as fine particles. Saturated phosphatidylcholines such as dipalmitoylphosphatidylcholine (DPPC), known from marine microalgae, are key examples of amphiphilic molecules abundant in the SSML (Cochran et al., 2016; Hayakawa et al., 1996; Suzumura, 2005). Proteins and amino acids exhibit enrichment factors of 10^3 – 10^6 in submicron aerosols (Triesch et al., 2021). They influence aerosol hygroscopicity and can act as nitrogen carriers to the atmosphere (Chan et al., 2005). Polysaccharides and transparent exopolymer particles (TEP), exuded by phytoplankton, contribute to the gel-like matrix of the SSML that entraps other molecules and microorganisms (Azetsu-Scott and Passow, 2004). These polysaccharidic gels accumulate in SSML and at bubble interfaces and can be transferred into both submicron and supermicron SSA fractions, thereby contributing to aerosol size distributions. When aerosolized, their high adhesivity and cation-bridged network structure may also promote the co-transfer and enrichment of other compounds by retaining them within interfacial organic matrices, while also influencing particle aggregation, mixing state, and potentially cloud activation (Aller et al., 2017).

Inorganic enrichment. Although sea salt is dominated by sodium and chloride, divalent cations such as calcium (Ca^{2+}) and magnesium (Mg^{2+}) display (SSA particle) size-dependent enrichment that reflects complexation with organic ligands and surface-film interactions (Salter et al., 2016; Van Acker et al., 2021b). Trace elements potassium, zinc, copper, iodine, and others also exhibit selective transfer under certain biogeochemical states (Barker and Zeitlin, 1972; Piotrowicz et al., 1972). These compositional shifts can influence aerosol pH, optical properties, and catalytic reactivity in heterogeneous atmospheric reactions (Schiffer et al., 2018).

Size dependence and phase partitioning. As discussed above, SSA enrichment is highly size-dependent: the smallest SSA particles ($< 0.3 \mu\text{m}$, dominated by film drops), are richest in organics, while larger

particles ($> 1 \mu\text{m}$, dominated by jet drops) reflect bulk seawater composition, as indicated by $\delta^{13}\text{C}$ measurements during a phytoplankton bloom mesocosm experiment (Crocker et al., 2022). The submicron fraction of SSA is therefore particularly important for both long-range atmospheric transport and human respiratory exposure, as these fine particles remain airborne over long distances and efficiently deposit in the lower respiratory tract, reaching the alveolar region where gas exchange occurs (Vu et al., 2015). The phase behavior of these particles, which can range from liquid to semi-solid or glassy states, depends on their organic load and the surrounding relative humidity, both of which influence their atmospheric lifetime and chemical reactivity (Koop et al., 2011).

In summary, the chemical enrichment and composition of SSAs arise from a finely tuned interaction between physical, chemical, and biological processes operating at the ocean surface. The SSML selectively concentrates amphiphilic and surface-active molecules, which are then transferred into the atmosphere through the bursting of bubbles and formation of fine droplets. Consequently, the emitted aerosol population spans a gradient from salt-rich coarse particles to highly organic submicron particles enriched in lipids, proteins, and polysaccharides. This selective enrichment governs the hygroscopicity, optical behavior, and atmospheric lifetime of SSAs, while reflecting the biological and ecological state of the source waters. Understanding these coupled enrichment processes is therefore crucial not only for characterizing marine aerosol chemistry but also for assessing how ocean-derived compounds influence the biological responses and potential health effects associated with human exposure to SSAs.

3. From composition to function: bioactivity and exposure pathways

The complex chemical composition of sea spray aerosols (SSAs) raises an essential question: *what biological effects do these airborne mixtures*

exert on humans who inhale them in natural environments? Because SSAs carry an evolving combination of inorganic salts, biogenic organics, and microbial materials, their potential influence on respiratory health spans a continuum from beneficial to harmful (Figure 2). Understanding this continuum requires integrating knowledge from toxicology, molecular biology, and epidemiology.

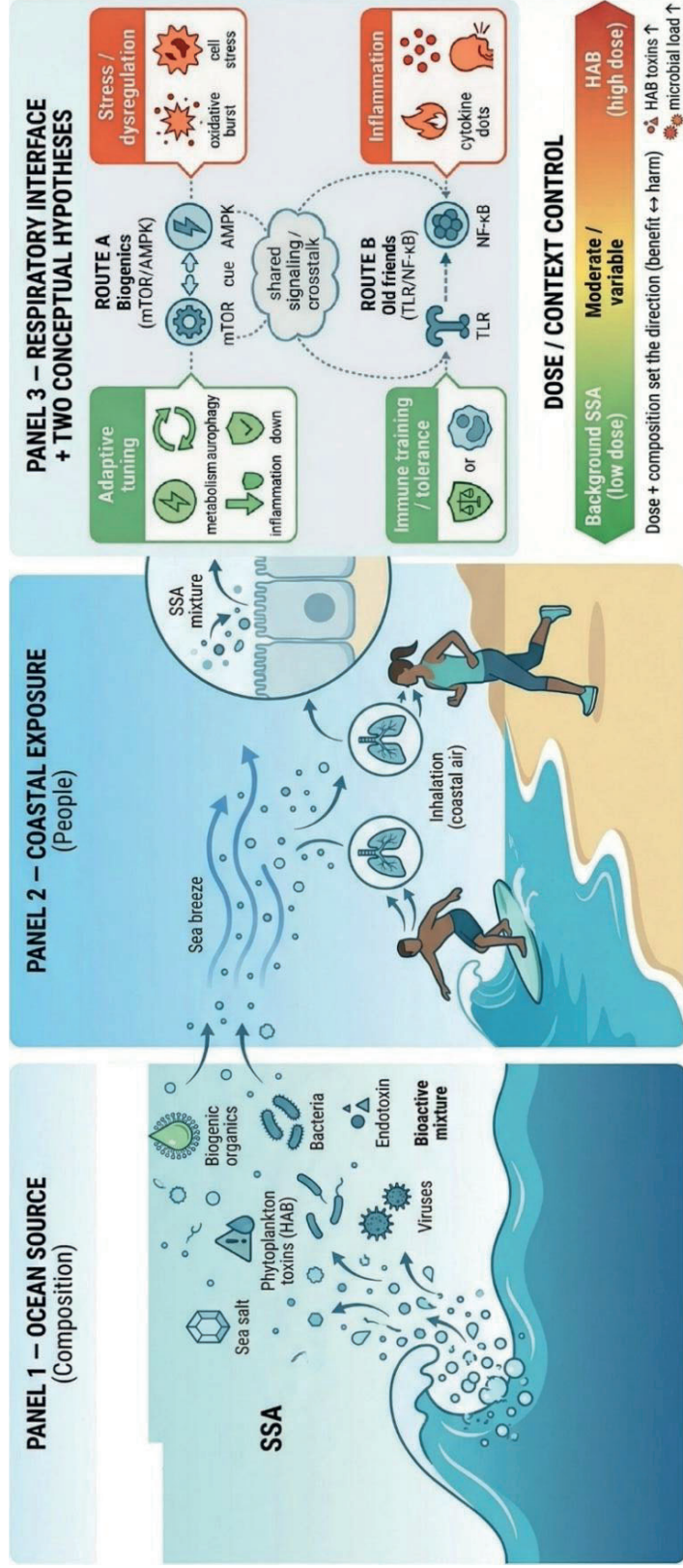


Figure 2. Conceptual model linking sea spray aerosol (SSA) composition, coastal exposure, and context-dependent bioactivity. Panel 1: SSAs are generated by bubble bursting at the sea surface microlayer and carry a mixture of sea salt, biogenic organics, microbial material, endotoxin, and phytoplankton toxins. Panel 2: Sea breezes transport SSA inland, where people are exposed

to aerosol particles in the coastal environment. Panel 3: SSA constituents may influence human health via two non-exclusive routes: Route A, Biogenics hypothesis (Moore, 2015) links mTOR/AMPK modulation to adaptive tuning ($\uparrow$ autophagy, improved metabolic balance, $\downarrow$ inflammatory tone) but can shift to stress/dysregulation at higher or more noxious exposures; Route B, “Old Friends” hypothesis (Rook, 2023) links TLR/NF- κ B sensing to immune training/tolerance at low-level, diverse exposures, but to inflammation when microbial or toxin loads are elevated. The dose/context bar emphasizes that dose and composition, especially during HAB high-dose exposure, shift effects from potentially beneficial to harmful.

3.1. Chemical diversity and molecular carriers of bioactivity

Although the physical and chemical features of sea spray aerosols (SSAs) determine what is transferred from the ocean to the atmosphere, the biological relevance of SSAs ultimately depends on whether the aerosolized molecules can interact with human cellular systems in a meaningful way (Ching and Kajino, 2018). The emerging evidence indicates that several classes of ocean-derived compounds possess intrinsic bioactivity, including the ability to modulate signaling pathways, cellular metabolism, immune responses, and oxidative balance. Importantly, these effects often occur at concentrations substantially lower than those typically associated with toxicity.

A key feature of SSA bioactivity is the ability of certain aerosolized molecules to engage with conserved cellular signaling pathways, particularly those governing cell growth, stress responses, and metabolism across diverse cell types. In vitro studies have shown that coastal SSA extracts can suppress PI3K, Akt, and mTOR signaling, a triad of pathways central to inflammation, autophagy, and tumorigenesis in human bronchial epithelial cells (Asselman et al., 2019; Van Acker et al., 2020). In parallel, SSA exposure can activate AMPK, a master regulator of cellular energy status (Asselman et al., 2019). These pathway shifts mirror the effects of known pharmacological modulators that promote metabolic health and reduce chronic inflammation (for example metformin and salicylates, respectively) (Schultze et al., 2012). This distinction may explain why background coastal exposure is associated with improved respiratory symptoms and general wellbeing, despite occasional harmful events during algal blooms.

The biological activity of natural SSAs is mediated through several interconnected mechanisms that enable aerosolized marine compounds to reach and influence human epithelial cells, most likely through a combination of membrane level effects, intracellular signaling modulation, immune interactions, and mixture synergies.

Membrane level effects and proximal signaling. Several classes of molecules that have been detected in SSAs, including saturated phosphatidylcholines and long-chain fatty acids (Bertram et al., 2018; Cochran et al., 2016), are well known to incorporate into or reorganize lipid monolayers and bilayers in other experimental systems, for example pulmonary surfactant and model membrane studies (Wauthoz and Amighi, 2014). Dipalmitoylphosphatidylcholine (DPPC), which has been identified in coastal SSA and will be discussed in Chapter 2, is the dominant phospholipid in human lung surfactant and critically controls film packing, surface tension, and spreading behavior at the air–liquid interface in the alveoli, thereby influencing the physical environment experienced by membrane proteins and receptors (Daniels and Orgeig, 2003). By analogy, deposition of trace amounts of such amphiphilic molecules from SSAs onto the respiratory surface could, in principle, alter local membrane structure or surfactant organization and thereby modulate the sensitivity of downstream signaling processes, although this has not yet been demonstrated directly for natural SSA mixtures.

Intracellular signaling, antioxidant and anti-inflammatory actions.

Transcriptomic analyses of human bronchial epithelial cells exposed to natural SSA extracts have revealed consistent changes in the expression of genes associated with the previously mentioned PI3K, Akt, mTOR, NF- κ B, and AMPK pathways, which are central regulators of inflammation, autophagy, cellular metabolism, and proliferation (Asselman et al., 2019; Moore, 2015; Van Acker et al., 2020). In these studies, environmentally scaled doses of SSA were associated with a general suppression of pro-growth and pro-inflammatory signaling and a concurrent activation of AMPK-linked stress responses, a pattern that resembles the action of several marine-derived and synthetic small molecules with chemopreventive or metabolic benefits (Dyshlovoy, 2021; Talero et al., 2015). A further, partly overlapping line of evidence comes from individual marine compounds with antioxidant and anti-inflammatory properties that are known to occur in bulk sea waters and,

for some classes, in SSAs themselves (Ghallab et al., 2024). Omega-3 polyunsaturated fatty acids such as EPA (Eicosapentaenoic acid) and DHA (Docosahexaenoic acid), derived from algal lipids, are well known to modulate eicosanoid production and to suppress pro-inflammatory transcription factors including NF- κ B, which in turn reduces the expression of cytokines and other inflammatory mediators (Calder, 2017; T et al., 2015). However, for many of these compound classes, their actual abundance, persistence, and biological relevance in natural SSAs remain insufficiently resolved. Carotenoids such as fucoxanthin from brown algae display strong free-radical scavenging activity and attenuate oxidative damage and inflammatory signaling in a range of *in vitro* and *in vivo* disease models (Ávila-Román et al., 2021; Peng et al., 2011), and certain marine sterols and related terpenoids have been investigated for their ability to suppress inflammatory responses and protect cells against oxidative injury (Hannan et al., 2020). Recent work has further shown that natural SSA samples can activate TLR4, TLR2/6, NF- κ B, and IRF signaling in human reporter cells, demonstrating that inhalable marine aerosols are capable of engaging redox- and inflammation-related pathways at low doses (Li et al., 2024). Taken together, these findings indicate that natural SSA mixtures can influence kinase networks and redox-sensitive transcriptional programs at low concentrations and support the view that at least part of SSA bioactivity operates through low-dose modulation of conserved signaling pathways rather than through overt cytotoxicity, although the specific compounds responsible and the exact causal chains remain to be fully resolved.

Immune-modulatory effects of microbial and macromolecular components. Beyond the intracellular inflammation-related signaling described above, natural SSA may also influence respiratory health through subtle, low-level immune stimulation by microbial and macromolecular components. The sea surface microlayer is enriched in bacteria, viruses, and extracellular polymeric substances, and fragments of these materials can be aerosolized and incorporated into SSAs (Aller et al., 2017; Cunliffe et al., 2013). Such components include

pathogen-associated molecular patterns that are capable of engaging pattern-recognition receptors on airway epithelial cells and resident immune cells, for example Toll-like receptors and C-type lectins (Moore, 2015). In other environmental settings, chronic exposure to complex bioaerosols has been associated with altered immune profiles and reduced risk of certain allergic diseases, suggesting that low-level stimulation of innate immunity can promote regulatory or tolerance-like responses (Rook et al., 2003). By analogy, repeated inhalation of marine bioaerosols could plausibly contribute to immunoregulatory effects at the respiratory surface. Until recently, direct mechanistic evidence for immune modulation by natural SSA was limited, but new work has shown that natural and artificial SSA with marine bacteria and endotoxins activate the NF- κ B pathway in human reporter cells in a dose-dependent manner, identifying bacterial components and endotoxins as key immunostimulatory ligands (Li et al., 2024). Even so, under chronic, low-level exposure, the relative contributions of different microbial taxa and macromolecular fractions, and cell-type-specific effects in primary airway and immune cells remain largely unresolved. At present, immune-related effects should therefore be regarded as a promising but insufficiently characterized component of SSA bioactivity that warrants targeted investigation.

Mixture and synergy effects. The mixture nature of SSAs provides an additional layer of complexity that is likely to influence their biological impact. Marine aerosols contain surfactants, lipids, polysaccharides, peptides, trace metals, and microbial fragments, which may interact in ways that modify cellular uptake and potency. Insights from pulmonary drug delivery indicate that phospholipids such as DPPC can enhance the deposition, spreading, and bioavailability of co-formulated compounds in the deep lung by altering surface tension and particle–surface interactions (Chono et al., 2009). Comparable processes may occur for natural SSAs, where marine surfactants could facilitate the transport of other bioactive molecules present in the same particles. These considerations support the view that SSA bioactivity is likely to

emerge from the combined action of multiple constituents rather than from single dominant compounds and underscore the need for mixture-oriented experimental designs and analytical approaches.

3.2. Epidemiological cases of harmful algal blooms and inhalation exposures

Although recent work has highlighted potential neutral or beneficial effects of non-bloom SSAs, the epidemiological literature is dominated by evidence of risk associated with aerosolized harmful algal bloom (HAB) toxins (Dahlin et al., 2024). Under bloom conditions, toxins produced by marine and freshwater microalgae can be transferred to the atmosphere through wave breaking and bubble bursting at the surface, where they become incorporated into SSAs and inhalable marine aerosols (Fleming et al., 2007, 2005). Field investigations, surveillance programs, and experimental studies have documented respiratory impacts related to HAB events, namely, to the *Karenia brevis* in Florida red tides (Fleming et al., 2005), *Ostreopsis* blooms in the Mediterranean (Berdalet et al., 2022), and cyanobacterial proliferations in freshwater (Labohá et al., 2023). These microalgal taxa produce different toxins, brevetoxins by *Karenia* (Fleming et al., 2005), palytoxin-like compounds by *Ostreopsis*, and microcystins by cyanobacteria, potentially involved in the respiratory problems, although not completely proven yet in the case of *Ostreopsis* proliferations (Berdalet et al., 2022).

The *K. brevis* red tide along the west coast of Florida is the best-studied case of inhalation exposure to marine algal toxins. *K. brevis* produces brevetoxins, polyether neurotoxins that bind to voltage-gated sodium channels, causing them to open at near-resting potentials and impairing normal inactivation (Abraham et al., 2005; Fleming et al., 2007). In experimental animal models of asthma, short-term exposures to aerosolized brevetoxins at environmentally relevant concentrations induce bronchoconstriction and changes in mucus transport, demonstrating that inhalation alone can perturb airway physiology

(Abraham et al., 2005; Dahlin et al., 2024). Human field studies during active blooms show consistent increases in eye, nose, and throat irritation, cough, and wheeze among beachgoers and coastal residents, with spirometry revealing modest but measurable decrements in lung function, particularly in people with pre-existing asthma (Abraham et al., 2005; Fleming et al., 2005; Lim et al., 2023a). Similar patterns of upper and lower airway symptoms have been reported during blooms of benthic *Ostreopsis* species in the Mediterranean, despite implicated compounds have not been indicated yet (Berdalet et al., 2022), and during freshwater cyanobacterial harmful algal blooms, where microcystins and other cyanotoxins are detected in aerosol above affected lakes (Labohá et al., 2023). In vitro exposure of primary human airway epithelium to aerosolized microcystin-LR confirms that such toxins can elicit inflammatory responses at the cellular level (Breidenbach et al., 2022).

At the population scale, HAB events are associated with quantifiable shifts in health care utilization. Retrospective emergency room analyses from Florida indicate that during a three month *K. brevis* bloom in 2001, residents living within 1.6 km of the coast experienced a 54 % increase in ER admissions for all respiratory diagnoses compared to inland residents (Kirkpatrick et al., 2006), along with an approximately 40 % increase in visits for gastrointestinal complaints during the same red tide period (Kirkpatrick et al., 2010). HAB illness surveillance in New York State similarly recorded an average of 31 hospital visits per year with reported bloom exposure between 2008 and 2014, of which about 5 % resulted in hospitalization (Figgatt et al., 2016). Occupational cohorts reinforce this picture: lifeguards and seaside workers report more frequent respiratory and eye symptoms during *K. brevis* events and take more sick and vacation days during bloom seasons, with associated productivity losses (Backer et al., 2005). Susceptible subgroups appear particularly affected; for instance, in one small survey of asthmatic beachgoers, 80 % of 15 participants reported asthma attacks following exposure to *K. brevis* aerosols (Fleming et al., 2005). Collectively, these

data demonstrate that HAB-related aerosols contribute not only to individual symptoms but also to measurable increases in health care demand and occupational impact at the community level.

3.3. Emerging evidence for beneficial bioactivities

Taken together with the HAB literature, the experimental and observational data introduced above suggest that inhalation of marine aerosols should not be viewed as uniformly harmful or uniformly beneficial, but rather as a context dependent exposure whose net effect depends on composition, dose, and timing. Against the backdrop of clearly documented risks during bloom events, there is a growing body of work indicating that under non bloom, background coastal conditions, SSAs can elicit cellular responses that resemble those induced by dietary bioactives or low dose pharmacological agents, rather than classical toxicants.

Several in vitro studies using natural coastal aerosol samples collected outside HAB episodes report transcriptional and signaling changes consistent with a mild, regulatory type of response at environmentally scaled doses (Asselman et al., 2019; Van Acker et al., 2020). Instead of overt cytotoxicity, these exposures lead to subtle shifts in pathways that are central to cell growth control, inflammatory tone, and metabolic regulation, in line with the mechanistic themes described in Section 3.1. The magnitude of these changes is generally modest, often involving tens to hundreds of differentially expressed genes or small fold changes in pathway activity, yet they occur at doses meant to approximate realistic alveolar deposition. This pattern aligns conceptually with the biogenics hypothesis, which proposes that repeated low level contact with aerosolized natural compounds may promote more favorable baseline set points in cellular stress and growth regulation (Moore, 2015). Importantly, the compounds implicated in this hypothesis are not necessarily the same toxins responsible for the acute respiratory effects reported during HABs. Nevertheless, biological responses may be

dose-dependent, with increasing exposure associated with progressively larger pathway-level transcriptomic effects across multiple processes (e.g., lysosome, spliceosome, glycogenesis, steroid biosynthesis), as shown in an exposure experiment for aerosolized homoyessotoxin and laboratory SSA extracts (Asselman et al., 2019).

A second emerging theme is the possibility of hormetic or non-monotonic dose–response relationships. Experimental work on individual marine compounds, as well as on other environmental bioactives, frequently shows that concentrations several orders of magnitude below those that cause toxicity can instead induce adaptive or protective responses, for example by triggering antioxidant defenses, enhancing autophagic clearance, or gently restraining mitogenic signaling (Ghallab et al., 2024; Zhang et al., 2025). Early SSA exposure studies hint at similar behavior, with low or intermediate concentrations producing pathway modulation without loss of viability, and higher concentrations or strongly bloom influenced mixtures tipping toward inflammatory or damaging effects (Asselman et al., 2019; Van Acker et al., 2020). Although the available data remain sparse and heterogeneous, they support the idea that the same broad class of aerosolized biogenics could contribute to health promoting or health impairing outcomes depending on where real-world exposures fall along the dose–response curve and on the presence of co-occurring stressors such as urban air pollution.

At the population level, evidence for beneficial bioactivity remains largely indirect, yet it is noteworthy that epidemiological studies consistently associate residential proximity to coasts with better self-reported health and reduced all cause and cardiovascular mortality, even after adjusting for socioeconomic status and green space access (Gascon et al., 2017). The dominant explanations for these patterns emphasize psychological restoration, increased physical activity, and cleaner air relative to urban cores. However, a small number of analyses have reported improved respiratory outcomes or reduced prevalence of

certain respiratory symptoms among coastal residents compared to inland populations when controlling for major confounders (Hooyberg et al., 2020; Mueller et al., 2022). These findings do not prove a causal contribution of SSA bioactivity, and they cannot disentangle the roles of exercise, humidity, temperature, and pollutant dilution. They do, however, demonstrate that coastal living is not associated with the systematic respiratory impairment that might be expected if chronic inhalation of marine aerosols were broadly harmful.

When viewed across scales, from molecules to human populations, the picture remains complex. Marine aerosols can clearly act as acute stressors during HAB events, yet under typical coastal conditions they appear capable of engaging cellular circuitry in ways that are more reminiscent of low dose nutritional or pharmacological cues. Existing evidence is strongest at the cellular level, where pathway-level responses can be measured directly, whereas links to clinically meaningful outcomes remain speculative. Key uncertainties include the identity and stability of the aerosolized compounds that drive beneficial responses, the extent to which they reach relevant tissue compartments *in vivo*, and how inter individual variability in age, genetics, co morbidities, and co exposures modulates any potential effects.

The work presented in this thesis is situated within this emerging landscape. By applying environmentally realistic exposure schemes, high resolution molecular readouts, and detailed characterization of aerosol composition, the subsequent chapters aim to refine our understanding of how natural SSAs interact with human bronchial epithelial cells, how these interactions vary with season and coastal ecosystem state, and to what extent they can be interpreted as potentially beneficial modulation rather than simple toxicity. In doing so, the thesis seeks to move the debate on ocean derived aerosols and health from a binary framing of “good” versus “bad” air toward a more nuanced, mechanistic perspective that integrates composition, dose, ecological context, and host biology.

4. Phytoplankton dynamics as a driver of SSA variability

The previous sections focused on how SSAs are formed and how their composition can carry bioactive signatures to the human respiratory system. A central question that now arises is what controls the temporal and spatial variability in that composition. In the surface ocean, phytoplankton and their associated microbiota and food webs are the main architects of dissolved and particulate organic matter and, therefore, of the aerosolizable pool present in the SSML (Figure 3). Understanding how phytoplankton communities grow, collapse, and shift over time is thus essential for interpreting when and where coastal SSAs are most enriched in bioactive compounds.

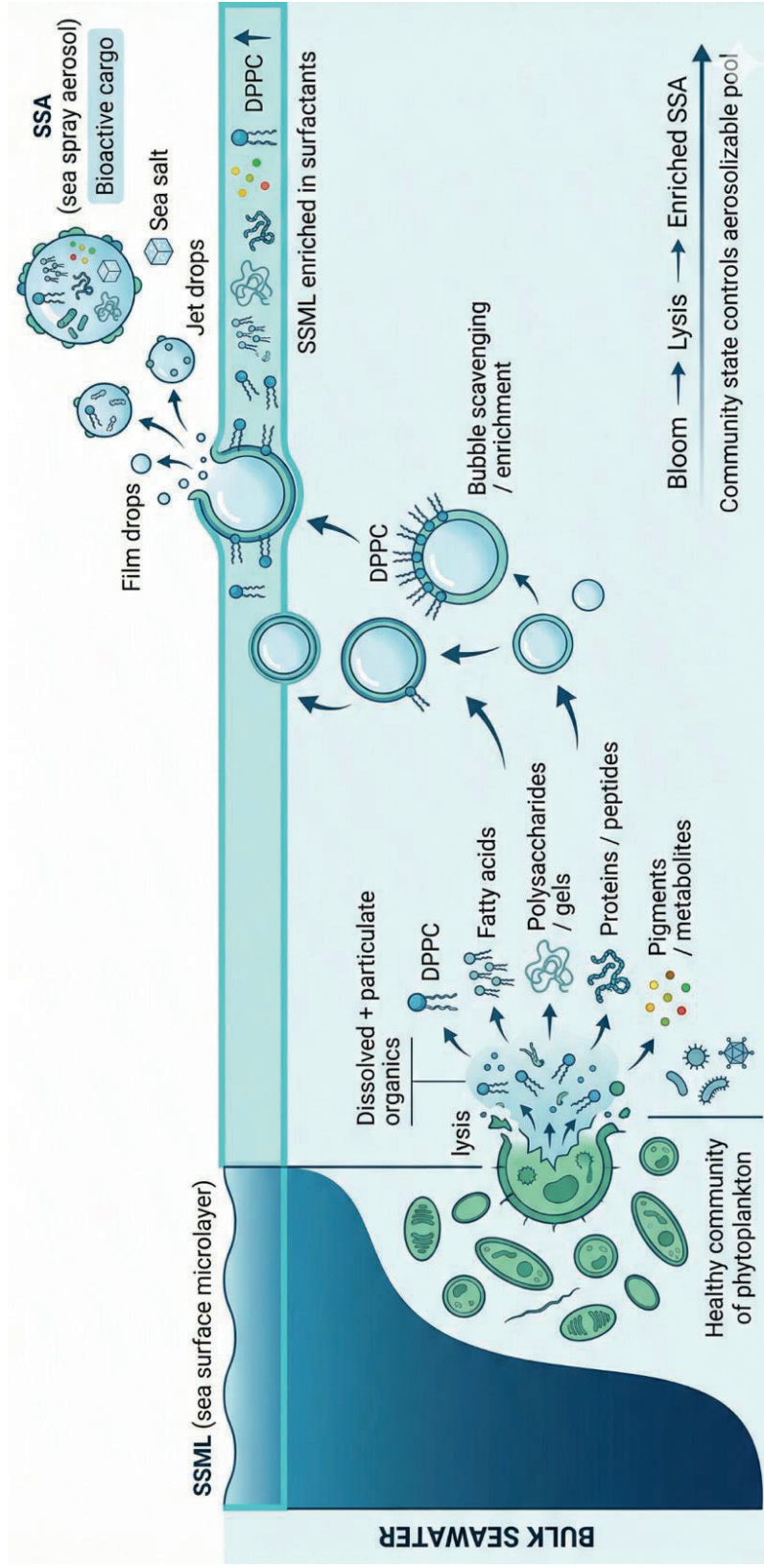


Figure 3. Phytoplankton-derived organic matter transfers bioactive compounds into sea spray aerosols. Phytoplankton-

produced compounds are released via metabolic exudation and/or cell breakdown. Dissolved and particulate organics (including DPPC) into bulk seawater, which associate with rising bubble surfaces and accumulate in the sea surface microlayer (SSML). Bubble bursting then carries this material into sea spray aerosols (SSA) as film and jet drops, shifting SSA bioactive cargo as blooms progress toward senescence and lysis.

4.1. Phytoplankton, carbon cycling, and surface organic reservoirs

Phytoplankton are responsible for roughly half of global net primary production, converting dissolved inorganic carbon into organic matter and oxygen through photosynthesis (Behrenfeld et al., 2006). A portion of this newly fixed carbon, as particulate organic matter, is exported to the deep ocean through the biological pump (as sinking cells and as part of zooplankton faecal pellets), while phytoplankton released dissolved organic carbon is recycled in the surface layer by heterotrophic bacteria, protists, and viral processes (Naselli-Flores and Padišák, 2023). These transformations determine the quantity and quality of organic matter that accumulates near the air–sea interface.

During active growth phases, phytoplankton release a variety of labile organic molecules into the surrounding water as part of normal metabolism, including low molecular weight acids and alcohols, lipids, and precursor compounds for exopolymeric substances. As blooms develop and transition across growth and decline phases, cells can aggregate with excreted polysaccharides to form gel-like particles and transparent exopolymer particles (TEP), which can rise and concentrate in the SSML (Azetsu-Scott and Passow, 2004; Jayarathne et al., 2016). In effect, phytoplankton convert inorganic nutrients and dissolved carbon into an amphiphile rich matrix at the surface that is particularly susceptible to bubble scavenging and aerosolization.

Microbial processing further reshapes this surface reservoir. Bacterioplankton degrade complex macromolecules into more surface-active fragments, for example fatty acids, alcohols, and phospholipid derivatives, which influence bubble film elasticity, drainage rates, and droplet size distributions (Cravigan et al., 2020; Triesch et al., 2021). Viral lysis can rapidly release intracellular components, including lipids and proteins, into the dissolved pool, which then feed back into SSML composition. In parallel, interactions between these biogenic surfactants and divalent cations such as magnesium and calcium alter

the relative partitioning of inorganic ions into fine SSA through complexation and co-enrichment (Salter et al., 2016; Van Acker et al., 2021b).

Taken together, these processes imply that phytoplankton are not only central to global carbon budgets, they also regulate the chemical identity of the ocean's "skin". When and where carbon is fixed, exuded, aggregated, and remineralized determines the concentration and character of the surface-active substrates that are most efficiently transferred into nascent SSAs.

4.2. Seasonal bloom regimes in the Southern Bight of the North Sea

The Southern Bight of the North Sea is a shallow, tidally energetic shelf system that exhibits pronounced seasonal and spatial variability in phytoplankton dynamics. Strong tidal mixing, riverine nutrient inputs from major estuaries, and gradients in salinity and turbidity create a mosaic of biogeochemical regimes, ranging from highly eutrophic nearshore waters to more oligotrophic offshore domains (Otero et al., 2022). Within this setting, recurring phytoplankton bloom sequences structure the timing and composition of organic matter at the surface, and thereby the pool of aerosolizable material available for SSA formation. Although phytoplankton provide the central seasonal framework considered here, this surface organic pool is also influenced by associated biological processing, including bacterial transformation of bloom-derived material and zooplankton-mediated release of organic matter.

A characteristic feature of the Southern Bight is the development of extensive spring blooms dominated by colony-forming taxa such as *Phaeocystis globosa* (Rousseau et al., 2013). These blooms typically arise in late winter to early spring, when increasing light levels and nutrient-replete conditions favor rapid growth. The colonies are

embedded in polysaccharide-rich mucilage, which enhances aggregation and promotes the formation of gel-like particles and TEP that accumulate in the upper water column and the SSML (Azetsu-Scott and Passow, 2004; Jayarathne et al., 2016). During and shortly after *Phaeocystis* peaks, the surface layer in coastal North Sea waters is therefore expected to be particularly enriched in carbohydrate-rich exopolymers, providing a substrate for strong organic enrichment of nascent SSAs and potentially altering their hygroscopic properties and cloud activation behavior. Some of this bloom-derived material may additionally be reworked by bacteria or released through grazing processes, but these broader food-web transformations are not treated here in detail because the emphasis of this thesis remains on phytoplankton-associated seasonal variability.

As the spring season progresses and nutrient ratios, grazing pressure, and light conditions evolve, diatom-dominated phases frequently follow. These diatom blooms contribute silica-associated biomass and a distinct suite of lipids and other amphiphilic compounds to the surface organic pool. Laboratory and field observations from North Sea and adjacent shelf regions indicate that long-chain fatty acids and alcohols produced during such diatom phases can be selectively transferred into submicron SSAs, where enrichment factors are highest and where particles play a disproportionate role in both atmospheric transport and respiratory deposition (Triesch et al., 2021). Toward summer and early autumn, smaller flagellates and mixed communities often prevail, sustaining lower-level but persistent inputs of dissolved organics and surfactants that maintain a background of aerosolizable biogenic material even outside major bloom events.

These seasonal regimes are superimposed on pronounced cross-shore gradients. Near the Belgian and Dutch coasts, riverine inputs and resuspension under strong tidal currents lead to turbid, nutrient-rich waters where *Phaeocystis* and other opportunistic taxa can attain high biomass, whereas more offshore stations experience clearer waters,

different nutrient limitation patterns, and distinct community compositions (Behrenfeld and Boss, 2014; González-Gil et al., 2022). As a result, the chemical signature of the SSML and SSAs are expected to vary not only over the annual cycle but also along short spatial transects between inshore and offshore locations. This spatial heterogeneity is particularly relevant for coastal sampling stations along the Southern Bight, where the balance between locally produced and advected organic matter, as well as the influence of estuaries and harbor systems, can modulate the composition of the aerosolizable pool on timescales of days to weeks (Liénart et al., 2018).

The Southern Bight of the North Sea provides a well-defined natural laboratory in which recurrent *Phaeocystis* and diatom blooms, shaped by regional hydrography and nutrient regimes, impose characteristic patterns on surface organic matter. These bloom sequences and spatial gradients are expected to imprint corresponding variability on SSA composition, including the relative abundance of polysaccharide-rich gels, lipids, and other surface-active compounds. This context is essential for interpreting seasonal changes in SSA chemistry and bioactivity at North Sea coastal sites, and it underpins the links between bloom phenology, aerosol composition, and cellular responses that are explored in Chapter 4 of this thesis.

4.3. Environmental controls on bloom timing and magnitude

The timing, intensity, and taxonomic composition of blooms in temperate coastal systems emerge from a balance between physical forcing and biogeochemical conditions (Lancelot, 1983). Light, nutrient availability, temperature, and mixed-layer dynamics all play key roles, and their interactions determine when surface phytoplankton can achieve sustained net growth (Alvarez-Fernandez et al., 2012). Variability in any of these drivers can shift bloom onset by weeks (Anderson et al., 2021; Hjerne et al., 2019), alter peak biomass, and change the dominant functional groups, with potential consequences for both SSML

composition and resulting SSA properties.

Nutrient regimes are a primary control. Winter and early spring water column vertical mixing typically replenish nitrate and phosphate in surface waters, setting the stage for rapid biomass accumulation as light levels increase (Desmit et al., 2020; González-Gil et al., 2018). However, local riverine inputs, salinity gradients, and anthropogenic nutrient loading can modify this baseline by changing the ratio and absolute concentrations of nitrogen, phosphorus, and silica, favouring some taxa over others (Groß et al., 2022; Lacroix et al., 2007). Reductions in salinity associated with freshwater plumes can also influence stratification and micronutrient delivery, thereby affecting community structure and bloom trajectories (Höfer et al., 2019). Light and mixing determine whether phytoplankton cells remain in the euphotic zone long enough to maintain positive net growth. Shoaling of the mixed layer under calm and sunny conditions can trigger bloom initiation, whereas storm-driven deep mixing can re-distribute cells below the critical depth, delaying bloom onset or causing partial collapse (González-Gil et al., 2022). These disturbance–recovery cycles, in which short-lived physical perturbations are followed by periods of stabilization, can shift phenology and succession patterns, sometimes advancing bloom phases earlier in the year or favoring fast-growing opportunistic species (Behrenfeld and Boss, 2014).

On shorter timescales, wind events and frontal passages can modulate the thickness and integrity of the SSML itself. Strong winds can disrupt and dilute surface films, temporarily reducing local surfactant concentrations and altering bubble lifetimes, whereas calmer periods favor re-accumulation and vertical convergence of surface active material (Quinn et al., 2015; Sellegri et al., 2006). These physical changes feed back on SSA production efficiency and enrichment, even if total phytoplankton biomass changes only modestly. In combination with the ecological processes described above, they create a highly dynamic environment in which the aerosolizable signal of phytoplankton

activity is intermittently amplified or suppressed.

In summary, phytoplankton dynamics provide the temporal and ecological backdrop against which variability in SSA composition and bioactivity should be interpreted. The ways in which blooms start, evolve, and terminate, and the environmental drivers that shift their timing or dominant taxa, all shape the pool of surface-active compounds available for transfer to the atmosphere. How this oceanic variability translates into concrete exposure scenarios is taken up in the subsequent chapters. The role of specific phytoplankton derived lipids in SSAs, and in particular the occurrence of dipalmitoylphosphatidylcholine (DPPC) in ambient coastal aerosols, will be discussed in **Chapter 2**. The extent to which seasonal changes in coastal ecosystem state and phytoplankton activity are mirrored in the chemical diversity and molecular effects of natural SSAs on human bronchial epithelial cells will be explored in **Chapter 3**. Finally, **Chapter 4** will examine long term changes in the timing of phytoplankton blooms in the Southern Bight of the North Sea, providing a regional phenological context for potential shifts in the magnitude and seasonality of SSA mediated exposure to marine biogenics.

5. Marine natural products as a reservoir of aerosolizable compounds

Marine ecosystems are prolific sources of structurally diverse natural products with documented bioactivities, ranging from antibiotics and anticancer agents to anti-inflammatory and neuroprotective compounds (Dyshlovoy, 2021; Talero et al., 2015; Zhang et al., 2025). These marine natural products (MNPs) are produced by a wide variety of organisms, including macroalgae, phytoplankton, bacteria, fungi, sponges, and other invertebrates. Many MNPs are secondary metabolites that mediate ecological interactions such as competition, grazing deterrence, and signaling (Petersen et al., 2020). At the same time, they often target highly conserved molecular pathways in eukaryotic cells, for example

kinases, transcription factors, ion channels, and metabolic enzymes (Song et al., 2018), which explains why they have attracted strong interest as leads for human therapeutics.

Quantitatively, curated resources such as the Comprehensive Marine Natural Products Database (CMNPD) currently list more than 31 000 distinct marine natural products with annotated structures, sources, and bioactivities (Lyu et al., 2021). Yet only a small fraction of these molecules has progressed from discovery to approved medicines. An assessment in 2022 indicate that approximately 20 drugs derived from, or inspired by, marine compounds are currently in clinical use, most of them antitumor agents, together with a smaller number used for chronic neuropathic pain, cardiovascular indications, and omega-3 supplementation (Haque et al., 2022). In addition, several dozen marine derived compounds are in Phase I–III clinical trials, reflecting an active but selective development pipeline in which only a tiny subset of known MNPs advances to late-stage evaluation. Taken together, these figures underscore both the richness of the marine metabolite pool and the substantial translational bottlenecks that still limit its exploitation in medicine (Haque et al., 2022).

Phytoplankton occupy a special place within this landscape. They synthesize a broad repertoire of lipids, carotenoids, sterols, polyphenols, and other specialized metabolites with demonstrated bioactivity to human health (Ávila-Román et al., 2021; Ghallab et al., 2024). Examples include omega-3 fatty acids such as EPA and DHA, sulfated polysaccharides, fucoxanthin and related carotenoids, and a variety of terpenoids and phlorotannins, many of which modulate inflammatory signaling, oxidative stress responses, and cell cycle regulation (Calder, 2017; Ghallab et al., 2024; Song et al., 2018). Some of these compounds have already been developed or proposed as nutraceuticals, cosmeceuticals, or drug candidates, for example fucoxanthin-rich formulations such as FucoVital® (Bermont et al., 2025), a microalgae-derived ingredient under clinical evaluation for metabolic

and liver health, and microalgal EPA/DHA supplements marketed as “Vegan Omega-3” (Bagchi et al., 2025) that are used as plant-based alternatives to fish oil in cardiovascular and cognitive support. From the perspective of this thesis, they are also relevant because phytoplankton are abundant in the surface ocean and directly influence the composition of the sea surface microlayer, where aerosolizable material accumulates.

Despite the extensive literature on MNPs in the context of drug discovery and chemical ecology, relatively little is known about which of these compounds actually reach the atmospheric compartment via sea spray aerosols. A few proof-of-concept studies have demonstrated aerosolization of specific biogenic molecules, for example certain phycotoxins during harmful algal blooms and selected lipids and surfactants under non bloom conditions (Labohá et al., 2023; Lim et al., 2023b), but for the vast majority of known marine metabolites their presence or absence in SSAs has never been systematically assessed. In principle, many MNPs share the physicochemical properties that favor transfer to the aerosol phase, including amphiphilicity, surface activity, and association with particulate or colloidal organic matter in the SSML. It is therefore plausible that a subset of pharmacologically interesting MNPs, especially those derived from phytoplankton and surface-dwelling microbes, are intermittently present in coastal aerosols, yet they remain largely invisible in most atmospheric measurements, which have traditionally focused on bulk organic fractions or targeted a small number of compounds.

Bridging this knowledge gap requires connecting marine natural product chemistry with aerosol science and informatics. Databases of MNPs provide curated information on molecular structures, producing taxa, geographic distribution, and reported bioactivities, but they have rarely been used to ask which molecules are candidates for aerosolization and which biological pathways they might affect if inhaled. This topic is explored further in **Chapter 5**.

6. Thesis rationale, objectives, and outline

The preceding sections outline how SSAs arise from ocean–atmosphere exchange, how they become enriched in biogenic material, and how inhalation of marine aerosols can, in principle, lead to both harmful and beneficial health outcomes. Despite this conceptual framework, the mechanistic chain that links ocean ecology to human respiratory physiology via SSAs remains incomplete. Several key uncertainties persist. First, the molecular identity of the aerosolized compounds that modulate cellular signaling is still poorly constrained, since most atmospheric measurements focus on bulk organic fractions or a small set of marker molecules. Second, even though harmful algal bloom toxins are known to be aerosolized, it remains insufficiently demonstrated that non-toxic compounds produced by phytoplankton during ambient, non HAB conditions, in particular surfactant lipids such as dipalmitoylphosphatidylcholine (DPPC), are efficiently transferred from the sea surface microlayer into natural SSAs and can be traced back to phytoplankton sources. Third, the way in which phytoplankton dynamics, microbial processing, and physical forcing jointly shape the pool of aerosolizable compounds at relevant temporal and spatial scales is not well resolved, nor is the extent to which satellite-based indicators of phytoplankton biomass and community structure can be linked to SSA composition. Fourth, there is limited information on how realistic mixtures of natural SSAs affect human lung cells at environmentally scaled doses across seasons and varying ecosystem states. Finally, current evidence for health effects is strongest at the cellular level, while links to long term respiratory or systemic outcomes remain largely speculative.

Addressing these gaps requires a cross-scale approach that connects processes occurring at the sea surface microlayer to compositional variability in SSAs, and from there to molecular responses in human cells and, ultimately, to hypotheses about population level health.

Methodologically, this entails combining in-situ aerosol sampling with high resolution chemical analysis, controlled in-vitro exposure experiments using environmentally realistic doses, and advanced molecular profiling such as transcriptomics. Conceptually, it also calls for integrating oceanography, aerosol physics, toxicology, remote sensing, and cheminformatics, so that specific marine natural products can be traced from their biological sources to their potential roles as airborne biogenics.

Against this backdrop, the overarching aim of this thesis is to investigate how natural SSAs can act as a mechanistic link between ocean ecology and human health, with a particular focus on the coastal atmosphere of the Southern Bight of the North Sea. In light of the gaps outlined above, the work is organized around the following central research questions:

Q1. To what extent can phytoplankton derived surfactant lipids, in particular DPPC, produced during ambient, non-HAB conditions be detected in natural coastal SSAs, and can their occurrence be quantitatively linked to ocean surface ecological state, for example through in situ indicators and satellite derived proxies of phytoplankton biomass and community composition?

Q2. How does the composition and bioactivity of natural SSAs vary over seasonal and environmental gradients in a temperate coastal system, and how do these changes translate into pathway level responses in human bronchial epithelial cells exposed to environmentally scaled doses of SSA derived material?

Q3. How do the timing, magnitude, and taxonomic composition of phytoplankton blooms in the Southern Bight of the North Sea vary on interannual timescales, and what are the implications of these shifts for the seasonality and intensity of SSA-mediated exposure to marine biogenics?

Q4. Which classes of marine natural products are both likely to be produced in surface ocean ecosystems and have physicochemical properties that favour aerosolization and inhalation, and what health relevant molecular pathways do these candidate aerosolizable compounds target according to current bioactivity knowledge and cheminformatics analyses?

These questions structure the thesis as follows. **Chapter 2** addresses Q1 by investigating the occurrence of DPPC and related surfactant lipids in ambient coastal SSAs, and by examining how their concentrations co vary with indicators of phytoplankton activity, including remotely sensed proxies of ocean color. **Chapter 3** addresses Q2 by characterizing the seasonal variability in SSA composition and by using RNA sequencing to quantify how key signaling pathways in human bronchial epithelial cells respond to natural aerosol samples collected under different environmental conditions. **Chapter 4** addresses Q3 by analyzing multi-year phytoplankton data from the Southern Bight of the North Sea in order to detect shifts in bloom phenology and community structure, thereby providing an ecological and climatic context for potential changes in SSA mediated exposure. **Chapter 5** addresses Q4 by introducing a comprehensive marine natural products database and using it to identify and prioritize classes of compounds that are both plausible constituents of SSAs and associated with bioactivities relevant to respiratory and systemic health.

Taken together, these chapters move from tracing a specific phytoplankton derived lipid into the atmospheric compartment, through characterizing the seasonal bioactivity of complex SSA mixtures, to placing these patterns within a longer term ecological framework and a broad chemical space of marine natural products. In doing so, the thesis aims to advance a more nuanced, mechanistic view of sea spray aerosols as carriers of ocean derived biogenics with the potential to influence human health.

Chapter 2 Phospholipids in Sea Spray Aerosols Mediated by Phytoplankton and Environmental Drivers

This chapter is adapted from a paper published in *Environmental Science & Technology* (2021, 55 (23), 15989–16000; <https://doi.org/10.1021/acs.est.1c04075>). As the paper did not include a CREDIT Authorship Statement, my contribution is specified here. It includes collecting and extracting phytoplankton abundance data from remote sensing products, constructing statistical models and performing all modelling analyses, and contributing substantially to the interpretation of the results and discussion.

Sea Spray Aerosols Contain the Major Component of Human Lung Surfactant

Abstract

Marine phytoplankton influence the composition of sea spray aerosols (SSAs) by releasing various compounds. The biogenic surfactant dipalmitoylphosphatidylcholine (DPPC) is known to accumulate in the sea surface microlayer, but its aerosolization has never been confirmed. We conducted a 1-year SSA sampling campaign at the Belgian coast and analyzed the SSA composition. We quantified DPPC at a median and maximum air concentration of 7.1 and 33 $\mu\text{g m}^{-3}$, respectively. This discovery may be of great importance for the field linking ocean processes to human health as DPPC is the major component of human lung surfactant and is used as excipient in medical aerosol therapy. The natural airborne exposure to DPPC seems too low to induce direct human health effects but may facilitate the effects of other marine bioactive compounds. By analyzing various environmental variables in relation to the DPPC air concentration, using a generalized linear model, we established that wave height is a key environmental predictor and that it has an inverse relationship. We also demonstrated that DPPC content in SSAs is positively correlated with enriched aerosolization of

Mg²⁺ and Ca²⁺. In conclusion, our findings are not only important from a human health perspective but they also advance our understanding of the production and composition of SSAs.

Keywords

sea spray aerosols, sea surface microlayer, human health, airborne exposure, UHPLC-HRMS, dipalmitoylphosphatidylcholine (DPPC),

1. Introduction

Sea spray aerosols (SSAs) are best known for their crucial role in diverse meteorological and climatological processes. They act as cloud condensation nuclei (Quinn et al., 2015) and ice nucleating particles (Quinn et al., 2015; Wilson et al., 2015) and increase the albedo of the atmosphere (Finlayson-Pitts and Pitts, 2000; Li et al., 2011). SSAs are primarily produced by breaking waves that entrap air and create underwater bubbles. When these air bubbles reach the seawater surface and burst, they eject aerosol drops into the atmosphere by means of two consecutive mechanisms. First, up to several hundreds of fine SSAs (i.e., film drops) are produced by each bursting bubble film (Lewis and Schwartz, 2004; Saltzman, 2009). Second, a few coarser SSAs (i.e., jet drops) are formed out of the overshooting water jet, which fills the bubble cavity (Lewis and Schwartz, 2004; Saltzman, 2009). SSAs are more than merely aerosolized seawater. Both the inorganic (i.e., Ca²⁺, K⁺, Mg²⁺, and trace metals) (Jayarathne et al., 2016; Schwier et al., 2017; Witt et al., 2010) and organic (Jayarathne et al., 2016; Quinn et al., 2014) composition of SSAs can be enriched relative to the subsurface seawater. Such enrichment processes are usually determined relative to sodium (Na⁺), which is the most common proxy to quantify SSA densities and is considered not to be enriched during SSA formation (Lewis and Schwartz, 2004).

The abundance, diversity, and life phase of marine phytoplankton (and

their associated food webs) can affect both the organic and inorganic composition of SSAs (Jayarathne et al., 2016; Quinn et al., 2014; Schwier et al., 2017). The literature suggests that phytoplankton dynamics (i.e., increase and decrease in phytoplankton biomass and changes in specific taxa) are more important in determining the SSA composition than merely the abundance or magnitude of phytoplankton biomass. Phytoplankton photosynthesis (i.e., productivity) was shown to induce increased excretion of marine surfactants (Long et al., 2014), while higher organic contents have been observed in SSAs upon the demise of phytoplankton communities or blooms (Cravigan et al., 2020). The latter may result from the lysis of phytoplankton cells, bacterial degradation, and the release of intracellular biogenic compounds (O'Dowd et al., 2015; Quinn et al., 2014; Rinaldi et al., 2013; Schwier et al., 2017; Wang et al., 2015). Biogenic or natural surfactants accumulate on the sea surface and influence the SSA production processes. As such, a chemically distinct film (20–400 μm) is formed at the air–water interface, known as the sea surface microlayer (SSML) (Burrows et al., 2014; Hardy, 1982). Surfactants decrease the surface tension at this interphase and increase the bubble surface lifetime, that is, the time an air bubble resides at the sea surface before bursting. With increasing surface lifetime, more seawater drains from the bubble film, making it thinner and eventually producing smaller film drops with a higher organic enrichment (Modini et al., 2013; Sellegri et al., 2006). The phospholipid dipalmitoylphosphatidylcholine (DPPC) is one such surfactant released upon phytoplankton cell lysis and is known to be abundant in the SSML (Cochran et al., 2016; Hayakawa et al., 1996). Although it has been suggested that DPPC is likely to be found in SSAs (Adams et al., 2016), its presence has never been quantified nor detected.

Despite the fact that the marine biosphere contains a large number of natural compounds with known or suspected pharmacological effects (Arrieta et al., 2010; Hamed et al., 2015), little research has been performed on the potential human health effects resulting from airborne

exposure to SSAs. In this context, the potential DPPC content of SSAs can be very important. DPPC is the major component ($\pm 40\%$) of the pulmonary surfactant in human alveoli and is essential for lung functioning (Hidalgo et al., 2015; Orgeig and Daniels, 2001). Conversely, lung surfactant abnormalities are associated with obstructive lung diseases (e.g., bronchiolitis and asthma) (Griese, 1999). As discussed in detail below, phospholipids like DPPC can also enhance the delivery and activity in the alveoli of aerosolized bioactive compounds (Hidalgo et al., 2015; Wauthoz and Amighi, 2014). Drug formulations containing DPPC as an excipient are therefore used and widely investigated in medical aerosol therapies (Ghadiri et al., 2019; Iasella et al., 2018; Velino et al., 2019).

The current study aimed to investigate the natural sea spray aerosolization of DPPC. To this end, a sampling campaign was conducted at the Belgian coast. The SSA samples were analyzed for their organic and inorganic composition. This allowed the potential quantification of DPPC concentrations in coastal air ($[DPPC]_{\text{coastal air}}$) and the characterization of the SSA composition. By analyzing our results relative to key environmental variables, we aimed to reveal the meteorological and/or ecological processes determining the $[DPPC]_{\text{coastal air}}$.

2. Material and Methods

2.1. SSA Sampling

A 1-year SSA sampling campaign was conducted from March 19 2018 until March 19 2019. A SSA collection system was installed, ± 300 m from the high-water mark on the beach, at the laboratory facilities of the Flanders Marine Institute (VLIZ) in Ostend, Belgium (Figure 1). Every week, a new pretreated quartz fiber membrane filter (QM-A, $\varnothing$ 47 mm, Whatman) was inserted in a precleaned stainless-steel in-line filter holder (Pall Corporation) and connected to an AirCube HE high-volume sampler (AMS Analytica). This pump was set at a constant airflow of 25

L min⁻¹ and registered a sampled volume of $\pm 250 \text{ m}^3 \text{ week}^{-1}$. The quartz filters have a pore size of $2.2 \text{ }\mu\text{m}$, but more importantly, their air retention (according to the product certificate) is $>99.95\%$ for $0.3 \text{ }\mu\text{m}$ particles. They thus collect total suspended particles. The sampling inlet was oriented in the north direction (i.e., toward sea) and fixed in place. This sampling setup is shown in Figure S1 (Supporting Information). To confirm absence of contamination and account for any filter background, 10 field blank samples were taken over the course of the campaign. These field blank filters underwent the same manipulations as the real samples but were not connected to the sampling pump. In short, they were also inserted in a precleaned filter holder, carried to the sampling pump, and stored (after 30 min instead of 1 week) in a sealed Petri dish at $-20 \text{ }^\circ\text{C}$.



Figure 1. SSA sampling station (x) situated ± 300 m from the high-water mark in Ostend, Belgium ($51^{\circ}14'09.7''\text{N}$, $2^{\circ}55'42.4''\text{E}$). The inland marine water bodies (i.e., the harbor and the sluice dock “Spuikom”; both manmade saltwater bodies) are indicated with a red dashed line. The dashed black lines at the edge of the dark blue semicircle indicate the wind directions ($^{\circ}\text{N}$), which was considered to be the boundary between sea and land wind. The meteorological station measuring the wind and air temperature (yellow ∇) and the buoy measuring the waves and water temperature (red Δ) are also indicated on the map, together with a 7×7 km square (centered at $51^{\circ}16'2''\text{N}$, $2^{\circ}53'19''\text{E}$), wherein the phytoplankton biomass ([PhyC]) data were selected.

The DPPC extraction method used was based on the dedicated extraction protocols of Ullah et al. (2015). (DPPC from human aerosols on polymer filters) and Van Acker et al (Van Acker et al., 2021b). (phycotoxins from SSAs on quartz fiber filters). In short, filters were inserted in glass tubes and spiked with 2.4 ng of dimyristoylglycerophosphocholine (DMPC; Avanti Polar Lipids) as an internal standard (Colantonio et al., 2011; Panda et al., 2004). The filters were subsequently eluted three times using 4 mL of analytical grade methanol and a 5 min sonication and 20 s vortex step (per elution). These sonication and vortex steps were included to improve the extractability of the collected DPPC, by providing better contact of the eluent with the filter membrane. Extracts were combined (± 12 mL), centrifuged (10 min, 2250 g), and the supernatant was transferred to a new glass tube. Extracts were then evaporated to incipient dryness, under a gentle N₂ stream at 40 °C, reconstituted in 120 μ L of 90% LC–MS grade methanol in water containing 3.5 mM ammonium acetate, and vortexed for 10 s. After centrifugation (10 min, 2500 g), the supernatant was transferred to a glass insert in a LC–MS vial prior to analysis. To avoid contamination, the following precautionary measures were taken: (1) the extraction was performed using new glass tubes that were rinsed with chloroform and methanol, (2) the other accessories (tips, beakers, and Pasteur pipets) were also made of glass or metal and were cleaned extensively, and (3) a face mask was worn by the analyst during the extraction procedure.

2.2. Chemical Analysis

2.2.1. DPPC

Analysis of DPPC in the final filter extracts was performed in a targeted manner using high-performance liquid chromatography coupled to high-resolution Q-Orbitrap hybrid mass spectrometry (UHPLC-HR-Q-Orbitrap-MS/MS). The DPPC reference standard was purchased from Avanti Polar Lipids. Chromatographic separation was achieved on a

Dionex Ultimate 3000 XRS UHPLC system (Thermo Fisher Scientific) equipped with an Acquity ethylene-bridged hybrid phenyl column (1.7 μm and 150×2.1 mm; Waters). A constant flow rate of $300 \mu\text{L} \cdot \text{min}^{-1}$ and a column oven temperature of 40°C were applied. A C18 DGM Symmetry HPLC column (5 μm and 150×2.1 mm; Waters) was used as a trapping column and installed between the LC pump and injector valve of the injector. The applied analytical method was adapted from the UHPLC-HR-Q-Orbitrap-MS lipidomics method of Van Meulebroek et al. (2017). Their UHPLC gradient elution program using a binary solvent system consisting of ultrapure water (solvent A) and methanol (solvent B), both acidified with 3.5 mM ammonium acetate, was shortened from 20 to 15 min. The following proportions (v/v) of solvent B were now used: 0–1 min at 75%, 1–2 min from 75 to 90%, 2–5 min from 90 to 98%, 5–10 min from 98 to 100%, and 10–12 min at 100%, followed by 3 min equilibration at 75% (i.e., initial conditions). Detection was carried out using a benchtop HR-Q-Orbitrap-mass spectrometer (Thermo Fisher Scientific) preceded by heated electrospray ionization (HESI-II). The full-scan method of Van Meulebroek et al (Van Meulebroek et al., 2017). was modified by adding a parallel reaction monitoring approach that operated in the positive-ion mode and selected the precursor ion of DPPC (i.e., $[\text{M} + \text{H}]^+$; m/z of 734.56943) in an isolation window of 1 m/z for fragmentation. Fragmentation was performed using high-energy collision dissociation, with an optimized (Supporting Information Section 1; Figure S2) normalized collision energy of 52 eV. Furthermore, the targeted HRMS/MS analysis used an automatic gain control of 2×10^5 ions, a maximum injection time of 100 ms, and the mass resolution was set at 17,500 full width at half maximum (FWHM).

Data processing, including identification and quantification, was performed using XCalibur 3.0 software (Thermo Fisher). Compound identification was realized based on the dedicated method of Ullah et al. (Ullah et al., 2015), using (1) the relative chromatographic retention time ($\Delta \leq 2.5\%$), that is, the ratio of the retention time of DPPC to that of the internal standard, (2) the accurate masses (i.e., within an extraction

window of 3 ppm) of the three main DPPC fragments, and (3) the correct product ion ratios (i.e., relative peak intensities) of these fragments ($\Delta \leq 10\%$). Quantification was performed using the area ratio (i.e., the ratio of the analyte and internal standard peak area) of the most abundant fragment and a calibration curve prepared with the extracts of seven spiked filters (i.e., 0, 0.3, 0.6, 0.9, 1.2, 1.8, and 2.4 ng of DPPC per 1/2 filter). The limits of detection (LODs) and quantification (LOQs) were determined using the standard deviation of the response, that is, the error of estimate ($S_{y/x}$), and the slope of the calibration curve (ICH, 2005; Krueve et al., 2015).

2.2.2. Inorganics

For every sample, half of the filter was analyzed for its inorganic content. This was performed using inductively coupled plasma optical emission spectroscopy (ICP-OES; Thermo Scientific iCAP 7000 series). Instrument RF power was set at 1.2 kW; coolant, auxiliary, and nebulizer flows were 12, 0.5, and 0.5 L Ar min⁻¹, respectively, and replicate ($n = 3$) read time was 7 s. Cranberry 05 and TMDA 25.5, that is, certified reference materials purchased from Environment Canada, were used to verify the calibration curves (0.1 to 100 μg per mL extract) of the major (Na^+ , Mg^{2+} , Ca^{2+} , and K^+) and minor inorganic constituents (Cu^{2+} , Zn^{2+} , Pb^{2+} , and Cd^{2+}), respectively. The majors were analyzed in the radial-view modus and the minors in the axial-view modus.

2.3. Environmental Variables

2.3.1. Meteorological Data

To assess the quantified [DPPC]_{coastal air}, SSA densities (i.e., $[\text{Na}^+]_{\text{coastal air}}$), and SSA composition relative to environmental variables that can influence the SSA production and enrichment of DPPC in the SSML and SSAs, we collected meteorological data. We used the air temperature (T_{air}) and wind data measured at 10 m height at a meteorological station situated 330 m northwest from our sampling station and the water temperature (T_{water}) and wave data measured with a monitoring buoy

situated 1 km north from our sampling station (Figure 1). The measurements of the meteorological station and buoy have respective time resolutions of 10 and 30 min and originate from the database “Meetnet Vlaamse Banken” of the Flemish Government (Flemish Government, 2021). For every sampling week, we calculated (1) the time during which the wind actually blew from the direction of the sea (% time_{seawind}; $\leq 58^\circ\text{N}$ and $\geq 235^\circ\text{N}$; Figure 1), that is, the wind directions that were assumed to contain the highest SSA densities and (2) the mean value during periods with sea wind for the wave height (WH_{seawind}), wave period (WP_{seawind}), and wind speed (WS_{seawind}).

2.3.2. Phytoplankton Data

To investigate the influence of phytoplankton on the [DPPC]_{coastal air}, we used data of the daily phytoplankton biomass in the surface seawater (depth ≤ 1 m) expressed as concentrations of carbon ([PhyC], $\mu\text{mol C L}^{-1}$ or $\mu\text{M C}$). These [PhyC] data were taken from the “ocean biogeochemistry reanalysis for the North-West European Shelf” data set (Ciavatta et al., 2018) of the Copernicus program, which were generated using the European regional seas ecosystem model (ERSEM 15.06). This model uses both satellite measurements and in-situ observations and produces data with a horizontal spatial resolution of 7×7 km. Given this spatial resolution, the ERSEM-derived [PhyC] values were interpreted as a regional proxy for phytoplankton biomass around the sampling area rather than as an exact local measurement at the aerosol sampling site. The [PhyC] data of the geographical “point” closest to our sampling station (Figure 1) were selected to perform further analysis via modeling.

2.4. Aerosol Source

As is common practice in SSA studies (Lewis and Schwartz, 2004), we used Na^+ as a proxy to quantify SSA densities and potential enrichment processes. To confirm that the Na^+ in our air samples indeed originated from seawater and thus from SSAs, we analyzed the correlation

between the $[Na^+]_{\text{coastal air}}$ and the $\% \text{ time}_{\text{seawind}}$ using a linear regression. We repeated this exercise for the other analyzed inorganic ions (Mg^{2+} , Ca^{2+} , K^+ , and trace metals). Next, by analyzing the Mg^{2+}/Na^+ mass ratio in our air samples, relative to this ratio in seawater (11.9%) (Dickson and Goyet, 1994; Pilson, 2012), we assessed the seawater and thus SSA origin (May et al., 2018) of Na^+ and Mg^{2+} . While characterizing the inorganic SSA composition, we also determined the SSA enrichment factor (EF_{SSA}) for the major inorganic ions (X) relative to the subsurface water (SSW). This was performed using their theoretical seawater ratios (Pilson, 2012) (i.e., per Na^+) and the following equation

$$EF_{SSA} = \frac{[X]_{SSA}/[Na^+]_{SSA}}{[X]_{SSW}/[Na^+]_{SSW}} \quad (1)$$

To examine if the DPPC in our air samples originates from SSAs and not from a terrestrial aerosol source, we performed a similar linear regression analysis between $[DPPC]_{\text{coastal air}}$ and $\% \text{ time}_{\text{seawind}}$. A higher $\% \text{ time}_{\text{seawind}}$ should in this case, under comparable environmental conditions, lead to a higher collection of DPPC on the filter membranes. Our sampling station is, however, surrounded with “inland” marine water bodies (see [Figure 1](#)), which may—to an extent—also produce SSAs containing DPPC. As such, the potential relation between $[DPPC]_{\text{coastal air}}$ and $\% \text{ time}_{\text{seawind}}$ can be partially masked.

2.5. DPPC Concentration Model

By exploring the relation between the $[DPPC]_{\text{coastal air}}$ and other environmental variables, we aimed to get a better understanding on the meteorological and/or ecological processes determining the $[DPPC]_{\text{coastal air}}$. As phytoplankton cell lysis and/or active metabolite excretion are potentially involved in the release of DPPC, we expected a time lag between the manifestation of phytoplankton biomass (i.e., $[PhyC]$) and their highest DPPC excretion/release, after which DPPC becomes available for aerosolization. During the current modeling exercise, we included a time lag correction on the $[PhyC]$ data. The optimal correction was found by testing different time lags (from 0 to 14

days), as further explained below. Using these time lag-corrected [PhyC] data, we calculated the total weekly decrease ([PhyC]_{decrease}) and increase in phytoplankton biomass ([PhyC]_{increase}). The [PhyC]_{decrease} was calculated as the summation of all daily decreases in [PhyC] during each sampling week. The [PhyC]_{increase} was calculated alike with all daily increases in [PhyC]. In addition, weekly mean values were also calculated ([PhyC]_{mean}) using the time lag-corrected [PhyC] data.

A generalized linear model (GLM) was made to reveal the relations between the [DPPC]_{coastal air} and the environmental predictors. We transformed the [DPPC]_{coastal air} using a natural logarithm and then performed z-transformations on all variables to obtain homoscedasticity. These z-transformed data were used to construct the GLM. This was performed by first using a full model containing all individual environmental predictors or independent variables (see Table S1) and their interactions. The latter was included to test the potential interaction effects between each combination of two predictors (Finney, 1948). Then, we conducted stepwise regressions to end up with the best regression formula, containing only the significant predictors and interactions. Backward selection was performed based on the Akaike information criterion and adjusted R^2 of the regressions. $WP_{seawind}$ and $WS_{seawind}$ showed high covariances (>0.75) with $WH_{seawind}$. To avoid multicollinearity, we only used $WH_{seawind}$ in the GLM as it showed the highest covariance with [DPPC]_{coastal air} (Table S1). The final GLM was accepted after the following confirmatory analyses: (1) the linear model assumptions were verified using the “gvlma” R package (Peña and Slate, 2006), (2) the absence of multicollinearity was confirmed by calculating the variance inflation factor (VIF, <5 threshold) (Fox and Monette, 1992) of each predictor, and (3) component residual plots (Fox and Weisberg, 2018) were used to verify that no significant trends remained in the residual data. Finally, to analyze the relative importance of each environmental predictor to [DPPC]_{coastal air}, linear regressions were performed for each environmental predictor individually.

3. Results

3.1. Analytical Performance

3.1.1. DPPC

The product-ion ratios (i.e., relative peak intensities of the DPPC fragments; Figure S2C) of the sampled DPPC deviated on average less than 2.3% from the DPPC reference standard and showed in every individual case a deviation below the tolerance level of 10%. As such, DPPC was successfully identified in the air samples when present at detectable concentrations. The major fragment (m/z 86.09681, Figure S2C) was used for quantification. The calibration curve showed excellent linearity ($R^2 = 0.998$) and no lack-of-fit (F -test, $p > 0.05$; 95% confidence level). The derived LOD and LOQ were 138 and 418 pg per 1/2 filter, respectively. In a 250 m³ air sample, these values correspond to air concentrations of 1.1 and 3.3 pg m⁻³. The field blank samples never contained detectable DPPC concentrations ($\geq$ LOD).

3.1.2. Inorganics

The LOD and LOQ values for the targeted inorganic constituents were determined using the standard deviation of spiked samples with a concentration near the LOD (Lipps et al., 2023). For the minor and major constituents, the LOQs (per filter) ranged from 30 to 240 ng and from 1.5 to 3 μ g, respectively. All LOD and LOQ values for the analyzed inorganic ions are available in Table S2, together with their respective background levels in the field blank samples. When the concentrations of the analyzed inorganic constituents in the field blank samples reached values above the LOD (see Table S2), these concentrations were subtracted from the concentrations of the actual field samples.

3.2. Air Concentrations and SSA Composition

Six of the 52 SSA samples of the 1-year campaign (i.e., 52 weeks) were impaired due to various interferences (e.g., electricity interruptions and

clogging of the sampling inlet) and could not be analyzed. DPPC was found in detectable concentrations ($>LOD$) in all but one of the 46 remaining samples and 38 of these contained quantifiable ($>LOQ$) concentrations. Figure 2A shows the DPPC and Na^+ concentrations of the sampled air ($[DPPC]_{\text{coastal air}}$ and $[Na^+]_{\text{coastal air}}$) as a function of time. In Figure 2B, the $[DPPC]_{\text{coastal air}}$ is plotted together with the DPPC/ Na^+ ratios. As $[Na^+]_{\text{coastal air}}$ were used to quantify SSA densities, these DPPC/ Na^+ ratios represent the DPPC content of the SSAs. Conversely, the SSA density (i.e., $[Na^+]_{\text{coastal air}}$) had an inverse correlation with the $[DPPC]_{\text{coastal air}}$ ($R^2 = 0.24$ and $p < 0.001$; Figure S3B) and with the DPPC content of SSAs ($R^2 = 0.77$ and $p < 0.001$; Figure S3C). These inverse relations with the $[Na^+]_{\text{coastal air}}$ seemed to result from the inverse relationship between the WH_{seawind} and the DPPC content of SSAs (Figure S3D). As discussed below (in section 4.1), these data do not only represent the first measurements of DPPC in SSAs but they also contribute to our understanding of the factors controlling the composition of SSAs.

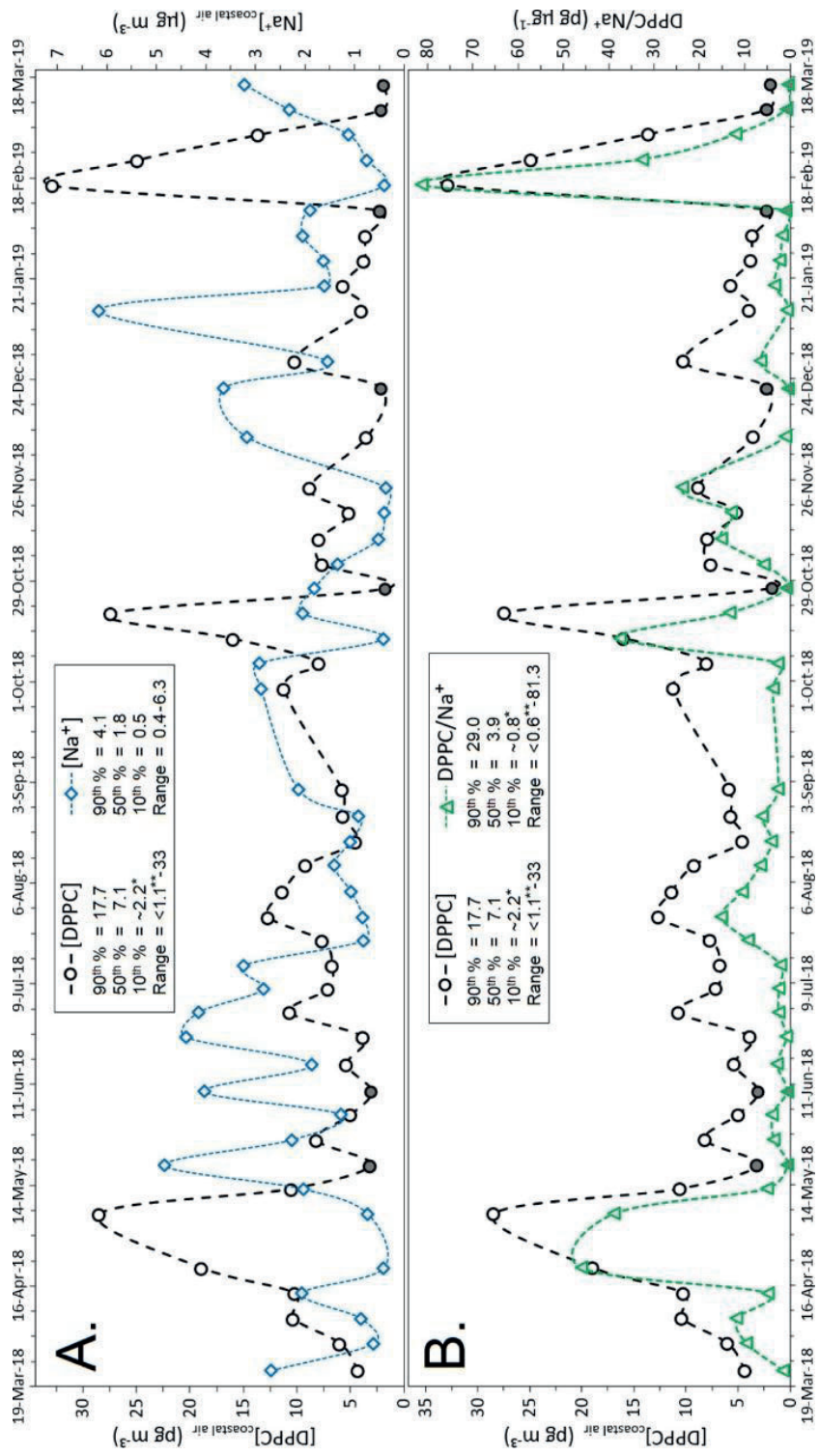


Figure 2. Results of the SSA sampling campaign. (A) DPPC and Na^+ concentrations ($[DPPC]$ and $[Na^+]$) of the sampled coastal

air are plotted. (B) $[DPPC]_{\text{coastal air}}$ is plotted next to the corresponding $DPPC/Na^+$ ratios, which illustrate the DPPC content of the sampled SSAs. The filled dots (gray $\circ$, green Δ) indicate $[DPPC] < LOQ$. The ranges and the 10th, 50th, and 90th percentiles of these data are also given. The (*) and (**) indicate the values based on concentrations below the LOQ and LOD, respectively.

3.3. Aerosol Source

Next to $[\text{Na}^+]_{\text{coastal air}}$, we analyzed the coastal air concentrations of three other major inorganic seawater constituents and four trace metals. By analyzing the correlation with the % time_{seawind} (Figure S4), we identified their main aerosol source. All trace metals (i.e., Cu^{2+} , Zn^{2+} , Pb^{2+} , and Cd^{2+}) predominantly originated from a terrestrial aerosol source or from the anthropogenic activities in the harbor ($p < 0.001$). Ca^{2+} had a weak but significant ($p < 0.05$; 95% confidence level) positive correlation with the % time_{seawind}. This implies that it cannot be completely excluded that other (terrestrial) aerosol sources influenced the $[\text{Ca}^{2+}]_{\text{coastal air}}$ to a certain extent. The $[\text{K}^+]_{\text{coastal air}}$ did not significant correlate with the % time_{seawind}, implying that neither the SSA nor inland aerosol sources were dominant. Finally, Na^+ and Mg^{2+} clearly originated from the sea wind ($p < 0.001$). The $\text{Mg}^{2+}/\text{Na}^+$ ratios of our air samples, additionally confirmed the SSA origin of the collected Na^+ and Mg^{2+} . Indeed, Figure S5 shows that at low to average DPPC contents ($\text{DPPC}/\text{Na}^+ < 7 \text{ pg } \mu\text{g}^{-1}$), the sampled $\text{Mg}^{2+}/\text{Na}^+$ mass ratios ($12.4 \pm 1.4\%$) were consistent with those of seawater (11.9%) (Dickson and Goyet, 1994; Pilson, 2012). At higher DPPC contents ($\text{DPPC}/\text{Na}^+ \geq 7 \text{ pg } \mu\text{g}^{-1}$), however, an increasing $\text{Mg}^{2+}/\text{Na}^+$ ratio could be observed. This does not mean that in these cases, the collected Na^+ and Mg^{2+} have an alternative origin. In fact, Figure 3 shows that the SSA enrichment (EF_{SSA}) of Mg^{2+} and Ca^{2+} is strongly related to the enriched content of DPPC (i.e., DPPC/Na^+). For comparison, the sampled coastal air concentrations of the major inorganic constituents are shown as a function of time in Figure S6.

Additional support for the SSA origin of the sampled DPPC is provided by the analysis of the effect of the wind direction on the sampled $[\text{DPPC}]_{\text{coastal air}}$ (alike this analysis for inorganics). As discussed above, the $[\text{DPPC}]_{\text{coastal air}}$ is largely dependent on the SSA content (i.e., DPPC/Na^+). The relation between $[\text{DPPC}]_{\text{coastal air}}$ and % time_{seawind} was therefore examined in four different DPPC/Na^+ ranges (Figure S7). In all but one range, the % time_{seawind} had a significant positive correlation with

the $[\text{DPPC}]_{\text{coastal air}}$ ($p < 0.05$). As discussed below, the SSA origin of DPPC was further supported by our final GLM and by the strong relationship of the enriched DPPC content in SSAs with the EF_{SSA} for Mg^{2+} and Ca^{2+} .

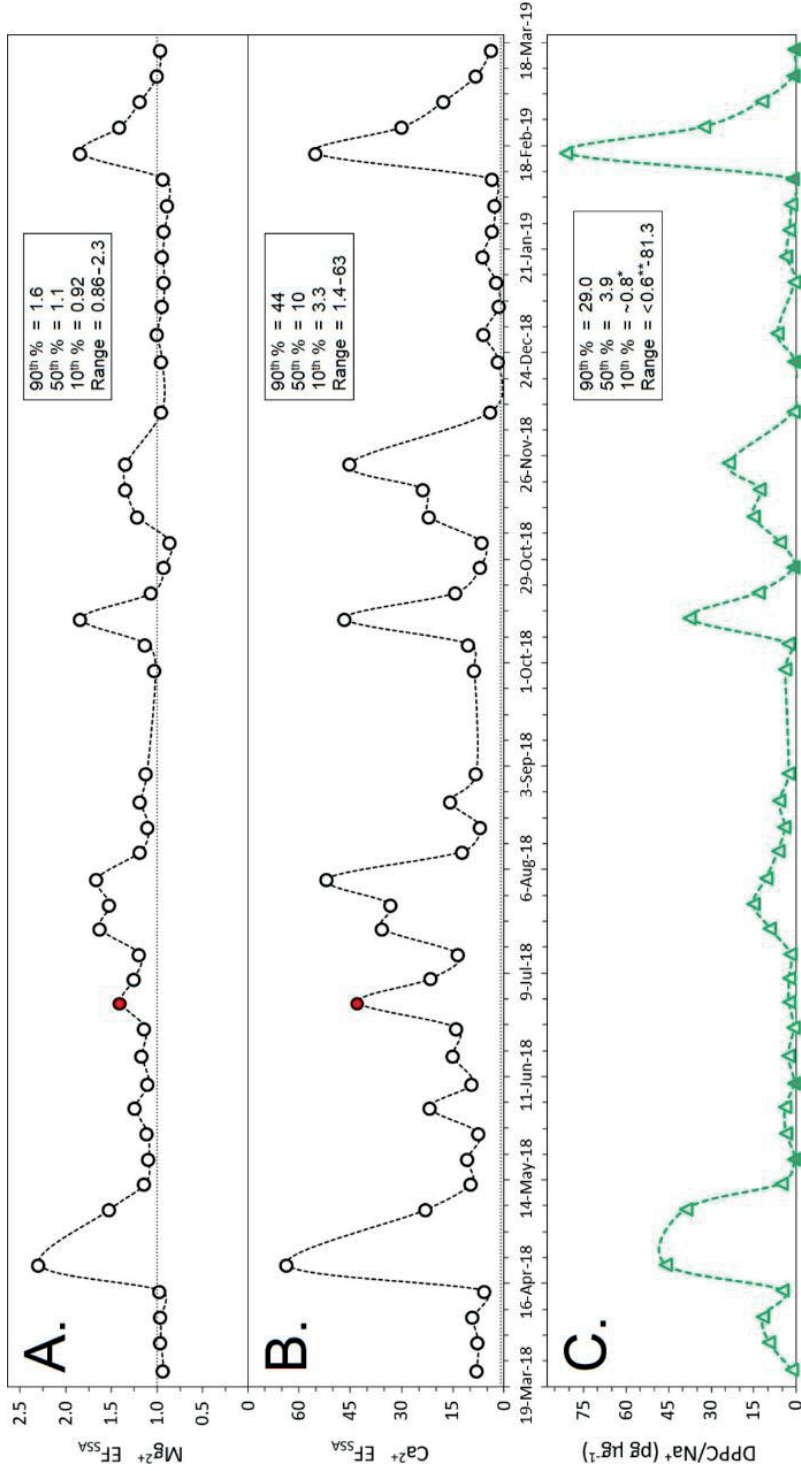


Figure 3. SSA enrichment factors (EF_{SSA}) for (A) Mg^{2+} and (B) Ca^{2+} , calculated using eq 1 and the theoretical seawater ratios

for Mg^{2+}/Na^+ (11.9%) and Ca^{2+}/Na^+ (3.82%) (Pilson, 2012). (C) DPPC/ Na^+ ratios are plotted to illustrate the relationship of the DPPC content of the sampled SSAs and their respective EF_{SSA} for Mg^{2+} and Ca^{2+} . The red $\circ$ and green $\triangle$ indicate a sampling week during which nearby construction activity was observationally recorded and may have contributed non-marine dust input. and $[DPPC] < LOQ$, respectively. The dotted gray lines indicate the non-enrichment levels ($EF_{SSA} = 1$). The ranges and 10th, 50th, and 90th percentiles of these data are also given. (*) and (**) indicate the values based on concentrations below the LOQ and LOD, respectively.

3.4. DPPC Concentration Model

The model fitness was examined for different time lag corrections on the [PhyC] variable (from 0 to 14 days). The best time lag correction, based on the adjusted R^2 , was found to be 6 days (Figure S3). While a 6-day lag was selected based on the best adjusted R^2 among the tested lag times (0–14 days), a more detailed sensitivity analysis of model robustness across neighboring lag values was beyond the scope of the present study. The environmental predictors that were retained in the final GLM are plotted in [Figure 4](#). The variables that were not retained in the final GLM (i.e., T_{air} , T_{water} , $WP_{seawind}$, and $WS_{seawind}$) and the % $time_{seawind}$ are plotted in [Figure S8](#). This GLM achieved an adjusted R^2 of 0.529 (F -test; $p < 0.001$) and showed a good accuracy (see [Figure S9](#)). Four environmental predictors ([Table 1](#)) had significant influence on the $[DPPC]_{coastal\ air}$, but no significant interaction between the predictors was found. The final model equation with standardized (z-transformed) coefficients a , b , c , and d ([Table 1](#)) is as follows

$$\ln[DPPC]_{coastal\ air} = a \times [WH]_{seawind} + b \times [PhyC]_{mean} + c \times [PhyC]_{decrease} + d \times [PhyC]_{increase} \quad (2)$$

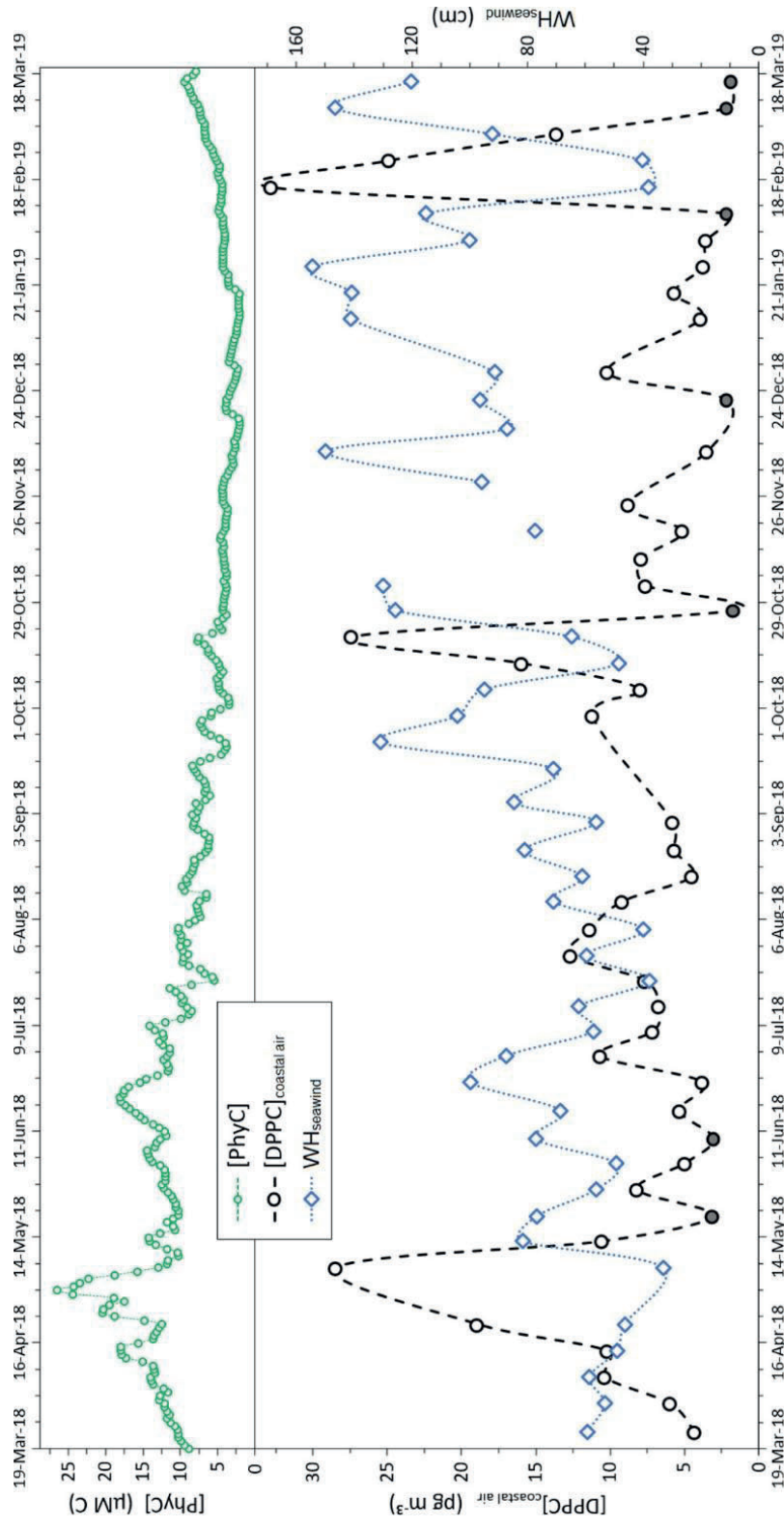


Figure 4. Top: The daily phytoplankton biomass expressed as concentrations of carbon ([PhyC], $\mu\text{M C}$). The [PhyC] here

presented are the (6 day) time lag-corrected data. Bottom: The DPPC concentrations of the (weekly) coastal air samples ($[DPPC]_{\text{coastal air}}$) and the weekly mean wave height during periods with sea wind (WH_{seawind}). The two lacking data points for the WH_{seawind} indicate the two weeks when the wind was only coming from across land ($\% \text{ time}_{\text{seawind}} = 0$). The filled dots (green○) indicate $[DPPC] < \text{LOQ}$.

Table 1. Overview of the Independent Variables in the GLM for [DPPC]_{coastal air}^a

Independent variable	VIF	Standardized coefficient	Standard error	P-value	Adjusted R ² of regression using only this single predictor	linear the
(intercept)		<0.001	0.105	1 not significant		
WH _{seawind}	1.506	-0.747	0.113	<0.001	0.382 (<i>p</i> < 0.001)	
[PhyC] _{mean}	2.054	-0.576	0.122	<0.001	not significant	
[PhyC] _{decrease}	1.395	0.316	0.055	<0.05	0.049 (<i>p</i> < 0.10)	
[PhyC] _{increase}	1.509	0.319	0.067	<0.05	0.061 (<i>p</i> < 0.10)	

^aThe description and details of these variables are given in Table S1. The adjusted R² results for the (simple) linear regressions performed for each independent variable individually are also given and indicate their relative importance to [DPPC]_{coastal air}.

The WH_{seawind} had major negative influence (Table 1) on the $[DPPC]_{\text{coastal air}}$, indicating that physical disturbances of the SSML strongly reduced the $[DPPC]_{\text{coastal air}}$. Among the different phytoplankton predictors, $[PhyC]_{\text{mean}}$ had a negative influence on the $[DPPC]_{\text{coastal air}}$, while the $[PhyC]_{\text{decrease}}$ and $[PhyC]_{\text{increase}}$ had positive influences. The linear regressions using each environmental predictor individually showed similar results (Table 1). The adjusted R^2 for WH_{seawind} was 0.382, which is much higher than that observed for the other predictors. Compared to the adjusted R^2 of our final GLM, the WH_{seawind} accounted for 72.3% of the total explained variance. Note that this model was developed to analyze the environmental predictors determining the observed $[DPPC]_{\text{coastal air}}$ and not to predict actual $[DPPC]_{\text{coastal air}}$ as the model was not validated using an additional subset of $[DPPC]_{\text{coastal air}}$ data.

4. Discussion

4.1. Source and Aerosolization of DPPC

The $[Na^+]_{\text{coastal air}}$ we quantified (i.e., from 0.4 to 6.3 $\mu\text{g m}^{-3}$; see Figure 2) are realistic. This can be concluded based on the meta-analysis of Lewis and Schwartz (Lewis and Schwartz, 2004), which shows (Figure S10) that the 10th and 90th percentile of WS_{seawind} in our study would generate $[Na^+]_{\text{coastal air}}$ on an open sea of ± 4 to $\pm 9 \mu\text{g m}^{-3}$. In addition, a previous field study (Van Acker et al., 2021b) where we sampled SSAs at a smaller distance (50 m) from the high-water mark also showed slightly higher $[Na^+]_{\text{coastal air}}$ from 0.23 to 12 $\mu\text{g m}^{-3}$.

Our results combined with the available environmental data confirm that DPPC originated from the sampled SSAs. This is concluded based on: (1) the positive relation of $[DPPC]_{\text{coastal air}}$ with % $\text{time}_{\text{seawind}}$ (Table 1 and Figure S7), (2) the strong relationship of the DPPC aerosol content ($DPPC/Na^+$) with the enriched aerosol contents (Figure 3) of two SSA-based inorganic constituents Mg^{2+} and Ca^{2+} (Figure S4 B/C), and (3) the significant inverse impact of WH_{seawind} and (changes in)

phytoplankton biomass on the $[DPPC]_{\text{coastal air}}$, as determined by the GLM (eq 2; Table 1). Although the first argument is straightforward, the second and third are further discussed in the next two paragraphs. ERSEM-derived [PhyC] is a regional proxy, as its 7×7 km resolution may miss fine-scale coastal variability. Higher-resolution FlowCam observations, local station data, and future hyperspectral or coastal ocean-colour products would help resolve site-level variability and test the [PhyC]–DPPC relationship more directly. We do recognize, however, that our study did not investigate possible degradation of DPPC after its collection on the filters. An additional study would allow to exclude that the $[DPPC]_{\text{coastal air}}$ measured here may in reality be even higher. To investigate this matter, one could, for example, analyze parallel series of short and long sampling periods (e.g., 15×1 day vs 3×5 days).

It was long assumed that the inorganic composition of SSAs is similar to that of seawater (Cravigan et al., 2020; Lewis and Schwartz, 2004). Since 2004, however, a number of studies monitoring ambient air (Leck and Svensson, 2015; Salter et al., 2016; Sievering et al., 2004) or using SSA generating systems (Ault et al., 2013; Jayarathne et al., 2016; Keene et al., 2007; Prather et al., 2013, 2013; Salter et al., 2016; Schwier et al., 2017) have reported enriched SSA concentrations for K^+ , Mg^{2+} , and Ca^{2+} . Alike the SSA enrichment of organic compounds (Facchini et al., 2008; Jayarathne et al., 2016; Quinn et al., 2014; Rastelli et al., 2017), inorganic ions seem to be increasingly enriched with decreasing SSA sizes (Leck and Svensson, 2015; Salter et al., 2016; Sievering et al., 2004). In this study, SSAs were collected at ± 300 m from the high-water mark (i.e., shoreline at high tide). This distance favors smaller-sized SSAs having a higher (in)organic enrichment as compared to nascent SSAs. Jayarathne et al. (2016) performed experiments using natural seawater and analyzed the EF_{SSA} for both fine ($< 2.5 \mu\text{m}$) and coarser SSAs ($2.5\text{--}10 \mu\text{m}$). Their results (EF_{SSA}) for the fine SSAs are comparable (1.1–2.1) to the ambient EF_{SSA} range for Mg^{2+} we observed (Figure 3) but are considerably lower (1.0–3.4) for Ca^{2+} . Ca^{2+} is generally found to be the most enriched inorganic

constituent in SSAs (Keene et al., 2007; Salter et al., 2016; Schwier et al., 2017). The highest EF_{SSA} reported for Ca^{2+} , in very fine (0.029–0.060 μm) nascent SSAs, reaches values higher than 100 (Salter et al., 2016). The origin of the enrichment processes for Ca^{2+} and Mg^{2+} is thought to result from the complexes they form with surface active organic compounds in the SSML (Leck and Svensson, 2015; Nielsen and Bilde, 2020). This cation binding effect has also been reported for DPPC in air–water interphases (Adams et al., 2016; Casper et al., 2016). As biogenic surfactants (e.g., DPPC) are enriched in the SSML (Casper et al., 2016; Gašparović et al., 2008; Marty et al., 1979), they can enhance the enrichment of Ca^{2+} and Mg^{2+} concurrently with their own enrichment in SSAs (Cochran et al., 2016). Figure 3 demonstrates such strong relation between the DPPC SSA content and the EF_{SSA} of Mg^{2+} and Ca^{2+} . Our study therefore indicates that biogenic surfactants (e.g., DPPC), produced and released by marine phytoplankton (without discarding possible influences of their associated food webs), not only control the enriched organic (Hoffman and Duce, 1976; Quinn et al., 2015) but also the inorganic composition of SSAs.

Interestingly the $[DPPC]_{\text{coastal air}}$ we sampled was largely dependent on the DPPC content of SSAs (i.e., $DPPC/Na^+$) and much less on SSA density (i.e., $[Na^+]_{\text{coastal air}}$; Figure S3A/B), although the former relationship is not interpreted mechanistically because $DPPC/Na^+$ is a derived ratio. In fact, significantly lower $[DPPC]_{\text{coastal air}}$ and $DPPC/Na^+$ were found at higher SSA densities (Figure S3B/C). To explain these findings, the SSA production processes and the involvement of surfactant effects need to be considered. It is clear that a lower SSA density results from a lower SSA production and thus from a lower wave action (WH_{seawind} ; see Figure S3D). However, in accordance with the concept of forced bubble bursting (Sellegrì et al., 2006), not only the SSA density but also the SSA size and content change as a function of WH_{seawind} . The lifetime of air bubbles at the sea surface increases with a decreasing wind speed and wave height (Sellegrì et al., 2006). The bubble film then has lower tendency to burst

and has more time to drain out water and become thinner (Blanchard, 1963). As such, at lower wave heights (and wind speeds), bursting bubbles produce smaller and more enriched (e.g., higher DPPC/Na⁺) film drop SSAs (Sellegrì et al., 2006). In addition, surfactants like DPPC stabilize the bubble film, prolong the bubble surface lifetime (Modini et al., 2013; Sellegrì et al., 2006), and increase the draining rate of bubble films (Modini et al., 2013; Sellegrì et al., 2006). When high DPPC concentrations are present in the seawater and SSML, the produced SSAs will thus be smaller and more enriched. Surfactants can therefore stimulate their own SSA enrichment. A stronger wave action, however, increasingly disturbs the SSML as breaking waves force surfactants back down in the SSW (Cunliffe et al., 2011; Hardy, 1982). As such, a strong wave action can influence the SSA size and SSA content via various mechanisms. As a consequence of the production of smaller film drops, SSAs will have a longer atmospheric residence time and will be transported further (Saltzman, 2009).

The developed GLM indicated that $WH_{seawind}$ is much more important in determining $[DPPC]_{coastal\ air}$, as compared to the phytoplankton biomass ($[PhyC]$) and its changes. The significant positive influences of the $[PhyC]_{decrease}$ and $[PhyC]_{increase}$ on the $[DPPC]_{coastal\ air}$ were expected, as the literature indicates that phytoplankton productivity stimulates excretion of marine surfactants (Long et al., 2014) and the demise of phytoplankton biomass increases the organic SSA content (Cravigan et al., 2020). The latter can be explained by release of DPPC upon cell lysis of phytoplankton cells (Adams et al., 2016; Cochran et al., 2016; Hayakawa et al., 1996), but may also reflect grazing-induced cell disruption and subsequent microbial degradation of bloom-derived organic matter, which could further increase the pool of surface-active compounds available for aerosolization. The negative influence of the $[PhyC]_{mean}$, on the other hand, seems contradictory at first sight. A possible explanation for this can be found in nutrient-phytoplankton-zooplankton-detritus models of the North Sea (Daewel et al., 2019; Daewel and Schrum, 2013). During winter and early spring, the

phytoplankton biomass is not under strong grazing control by zooplankton, which may allow the phytoplankton to release relatively more DPPC into the seawater instead of entering the food chain. This potential ecological explanation of the indirect inverse effect of the $[\text{PhyC}]_{\text{mean}}$ on the $[\text{DPPC}]_{\text{coastal air}}$ can currently not be proven or disproven and is only introduced as a hypothesis at this time. We also examined potential temperature effects, as especially the water temperature is known to influence air entrainment, bubble plume formation, and eventually SSA production upon the breaking of waves (Callaghan et al., 2014; Christiansen et al., 2019; Nielsen and Bilde, 2020). Significant temperature effects on the $[\text{DPPC}]_{\text{coastal air}}$ were, however, not observed. As such, these predictors were not retained in our final model.

This study also contributes to the knowledge on the production and behavior of SSAs in a meteorological context. Our results agree with the literature (Modini et al., 2013; Sellegri et al., 2006), suggesting that increasing surfactant concentrations and decreasing forced bubble bursting processes induce a higher surfactant SSA content and smaller SSAs. Smaller (film drop) aerosols have a larger atmospheric retention time and an increased meteorological impact. Additionally, a higher surfactant content in SSAs enhances the hygroscopic cloud droplet growth processes (Frossard et al., 2018). The concurrent enrichment of cations (Figure 3) with marine surfactants (DPPC) can also introduce meteorological effects. Indeed, the presence of charges on the surfaces of SSAs influences their activity as cloud condensation nuclei and play a role in lightning and thunderstorms (Blanchard, 1963; Casper et al., 2016; Jungwirth et al., 2005).

4.2. DPPC Exposure and Human Health

Although DPPC had never been detected in coastal air before, it has been suggested that DPPC is likely to be present in SSAs (Adams et al., 2016). Here, we quantified the range of $[\text{DPPC}]_{\text{coastal air}}$ ($<1.1\text{--}33\text{ pg m}^{-3}$

³; Figure 2) to which coastal populations are exposed. As such, the maximum airborne exposure via coastal air, assuming an extremely high inhalation flow rate (100 L min^{-1}) (Clarenbach et al., 2005), equals $4.8 \text{ ng of DPPC day}^{-1}$. DPPC plays a crucial role in the human respiratory system as it is the major component of the lung surfactant present in alveoli (Hidalgo et al., 2015; Orgeig and Daniels, 2001), and as lung surfactant abnormalities are associated with obstructive lung diseases (e.g., bronchiolitis and asthma) (Griese, 1999).

We suggest, however, that the quantified $[\text{DPPC}]_{\text{coastal air}}$ is too low to induce human health effects through the additional supply of pulmonary surfactant. Medical therapies for pulmonary surfactant-related diseases use much larger doses. Respiratory distress syndrome (RDS) in preterm newborns, for example, is treated with (modified) animal-derived lung surfactant using doses ranging from 50 to 200 mg of DPPC per kg bodyweight (Hentschel et al., 2020). As lung surfactant contains 40% DPPC (Hidalgo et al., 2015; Orgeig and Daniels, 2001) and the preterm birth weight can range from 0.75 to 2.5 kg, the deployed DPPC doses of a treatment range from 15 to 200 mg. Such surfactant replacement therapies (SRT) have also been widely examined to treat acute RDS (Meng et al., 2019), wherein even higher surfactant doses were used. Intubation methods for SRT are still the predominant treatment strategy but aerosol delivery systems are increasingly examined and introduced (Heching et al., 2021; Hentschel et al., 2020; Samsudin, 2013; Syedain et al., 2015). Another argument indicating that the determined $[\text{DPPC}]_{\text{coastal air}}$ is too low to induce direct health effect is given by Ullah et al. (Ullah et al., 2015) They determined that humans can exhale 20 to 700 ng DPPC m^{-3} via aerosols upon (strong) breathing. This exhaled concentration range is up to three orders of magnitude higher than that found in coastal air. Although it cannot be completely excluded upon chronic exposure, it is unlikely that $[\text{DPPC}]_{\text{coastal air}}$ have direct health effects.

There is, however, another fascinating mechanism by which DPPC can

induce indirect health effects through airborne exposure via SSAs. Surfactants like fatty acids and phospholipids coat the surface of SSA droplets (Schiffer et al., 2018). This SSA coating can play a facilitating role in the delivery of other marine bioactive chemicals. Indeed, DPPC is one of the excipients, which is currently authorized and used for medical pulmonary drug delivery via aerosol therapy (Velino et al., 2019; Wauthoz and Amighi, 2014). These pulmonary drug delivery methodologies are rapidly evolving and have already been used for a range of drugs (e.g., antibiotics, vaccines, peptides, and chemotherapeutics) (Muralidharan et al., 2015; Velino et al., 2019). Depending on the type of aerosol formulation (e.g., micelles, liposomes, and solid particles) (Wauthoz and Amighi, 2014), phospholipids like DPPC may (1) improve the transport of poorly water-soluble drugs (Hidalgo et al., 2015; Wauthoz and Amighi, 2014), (2) reduce the recognition of exogenous aerosols by alveolar macrophages (Hadinoto et al., 2007; Poelma et al., 2004; Wauthoz and Amighi, 2014), (3) serve as a post-deposition spreading agent (Iasella et al., 2018), and (4) stimulate pharmacokinetics via enhancement of the drug permeability through the lung epithelium (Chono et al., 2009; Wauthoz and Amighi, 2014). As such, the DPPC content of SSAs draws an interesting parallel with medical aerosol therapy. It could facilitate the effects of other bioactive compounds present in SSAs. We know that the marine biosphere contains a whole spectrum of natural compounds with known or suspected pharmacological effects (Arrieta et al., 2010; Hamed et al., 2015). The medical/pharmaceutical research investigating the facilitating effects of DPPC and other phospholipids may thus enhance our understanding of the mechanisms occurring upon inhalation of SSAs.

It should be noted that this facilitating effect of marine surfactants like DPPC could also hold true for toxic chemicals. It has been shown, for example, that some marine phycotoxins can be found in coastal air (Ciminiello et al., 2014; Pierce et al., 2005; Van Acker et al., 2021b) and induce respiratory syndromes during harmful algal blooms (HABs)

(Fleming et al., 2011; Poli et al., 2018; Vila et al., 2016). Such harmful exposures to aerosolized phycotoxins are rather rare events, requiring a combination of favorable weather conditions for SSA production and a severe toxin-producing HAB (Ciminiello et al., 2014). Although this potential negative effect cannot be ignored, we do not expect it to overrule the potential positive effect. If the adverse effects of DPPC would outweigh its positive effects, one would expect the health status of coastal populations to be lower than that of inland populations, and this is not the case. On the contrary, a positive coastal proximity health effect is increasingly observed in epidemiological research (Brereton et al., 2008; Hooyberg et al., 2020; Wheeler et al., 2012; White et al., 2013). In this context, Moore (Moore, 2015) published his biogenics hypothesis, suggesting that regular airborne exposure to aerosolized natural compounds may be health-promoting via the downregulation of specific cell signaling pathways. Although our previous work using in vitro studies on lung cells (Asselman et al., 2019; Van Acker et al., 2020) supported the principles of this hypothesis, our new findings on DPPC presented here further highlight the potential health-promoting effect of airborne exposure to SSAs in coastal air.

Supporting Information

Fragmentation collision energy

The normalized collision energy (NCE) used for the fragmentation of DPPC was optimized to receive three main fragments for identification of which one could be used for quantification. 10 different collision energies between 20 eV and 80 eV were tested. At the lowest collision energy (i.e., 20 eV) the precursor ion of DPPC ($[M+H]^+$; m/z of 734.56943) was still detectable, but there was only one main fragment (Figure S2A). At a collision energy of 33 eV, an intermediate scenario was found where two smaller fragments started to appear (Figure S2B). Fragmentation using a collision energy of 52 eV was found to give the best balance between identification and quantification of DPPC. Three main fragments could be used for identification and one fragment was dominant and could be used for quantification purposes (Figure S2C). The fragments with a m/z of 86 and 184 were also found and used by Ullah et al.¹ to identify and quantify DPPC. According Ullah et al. (Ullah et al., 2015) the formation of these ions originate from the polar headgroup of choline and the phosphocholine headgroup, respectively.

Supportive figures

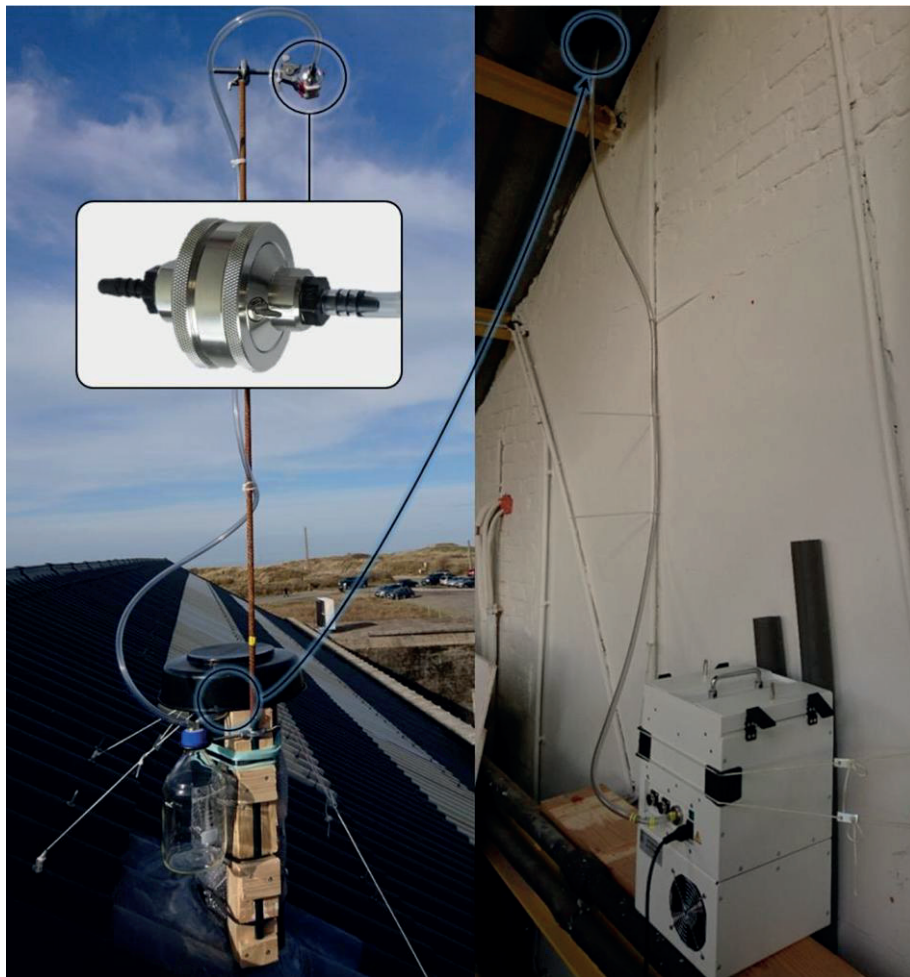


Figure S1: SSA sampling set-up on the roof (10m above ground level) of the lab facilities of the Flanders Marine institute (VLIZ) at 300 m from the high water mark. The filter holder containing the quartz fiber membrane filter is connected via tubing to the sampling pump, which was placed inside the building (close to the roof). Halfway the tubing connection we inserted a glass bottle to protect the pump against any liquid (i.e., condensation or potential rain intake).

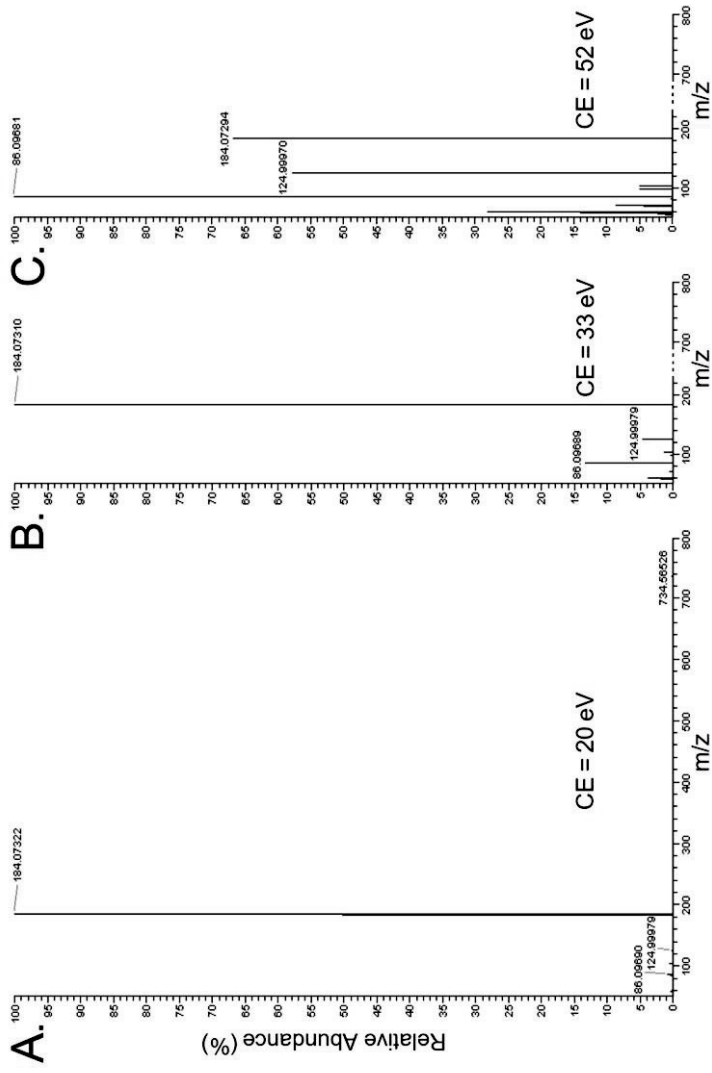


Figure S2: HR-Q-Orbitrap™-MS/MS fragmentation spectra showing the product ion ratios (i.e., relative peak intensities)

of dipalmitoylphosphatidylcholine (DPPC) at different collision energies (CE). At the lowest CE the parent ion (m/z 734.56526) is still detectable

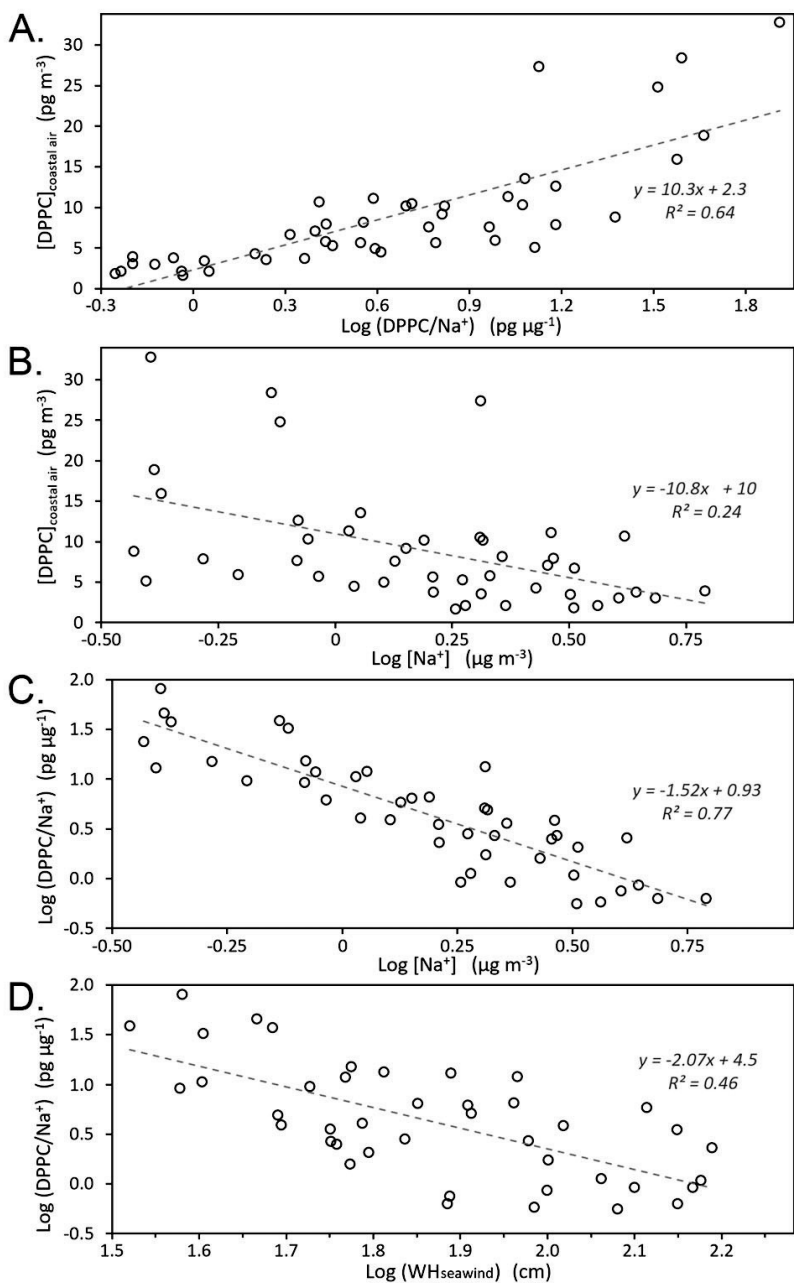


Figure S3: Significant correlations ($p < 0.001$; 99.9% confidence level) between (A) the DPPC air concentrations ($[DPPC]_{\text{coastal air}}$) and ($\log_{10}$ -

transformed) DPPC/ Na^+ ratios, which represent the DPPC content of the SSAs, in the sampled coastal air, (B) the $[\text{DPPC}]_{\text{coastal air}}$ and ($\log_{10}$ -transformed) Na^+ concentrations ($[\text{Na}^+]$), (C) the ($\log_{10}$ -transformed) DPPC/ Na^+ ratio and ($\log_{10}$ transformed) $[\text{Na}^+]_{\text{coastal air}}$, and (D) the ($\log_{10}$ -transformed) DPPC/ Na^+ ratio and ($\log_{10}$ -transformed) mean wave height of the sampling week during seawind periods wave height ($\text{WH}_{\text{seawind}}$). The gray dashed trendlines illustrate the linear correlation between the two variables, of which the equations and coefficients of determination (R^2) are given in the graphs.

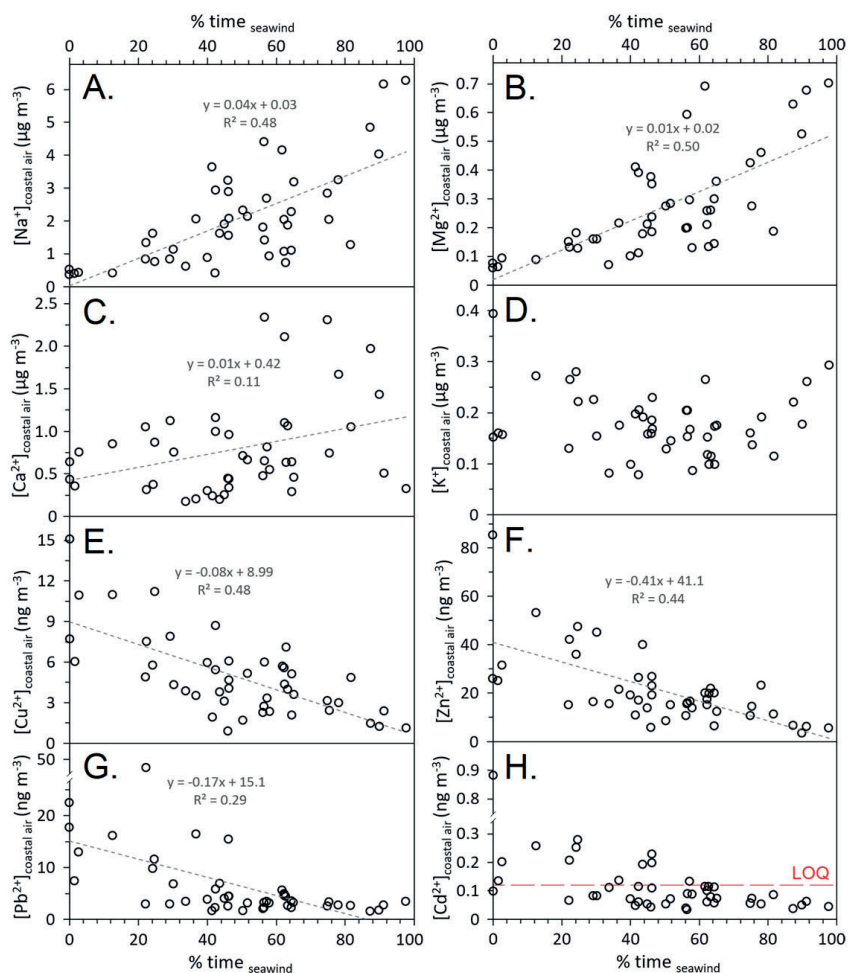


Figure S4: Sampled air concentrations of the diverse analyzed inorganic ions as a function of the % time with seawind. The gray (linear) trendlines illustrate the significant correlations between the two variables ($p < 0.05$ for (C) and $p < 0.001$ for (A), (B), (E), (F), (G)). The equations and coefficients of determination (R^2) for these linear models are also given in the graphs. The Cd^{2+} concentrations, shown in (H), were in almost 70% of the samples below the LOQ (i.e., 0.12 ng m^{-3}) which is shown here as a red dashed line. All other analyzed inorganic ions were always above their respective LOQ values (see Table S2).

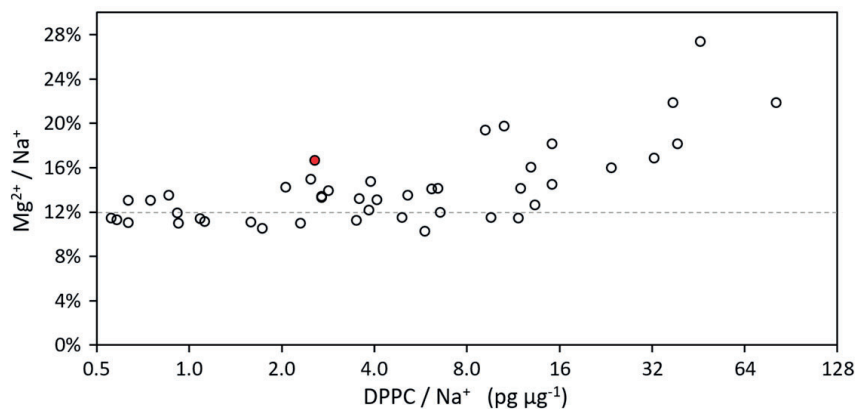


Figure S5: $\text{Mg}^{2+}/\text{Na}^+$ ratio ($\mu\text{g } \mu\text{g}^{-1}$) as a function of the DPPC/Na^+ ratio (DPPC SSA content) of the sampled SSAs. The red filled dot indicates nearby construction works (i.e., dust production), and the dashed gray lines indicates the theoretical $\text{Mg}^{2+}/\text{Na}^+$ mass ratio in seawater (11.9%).

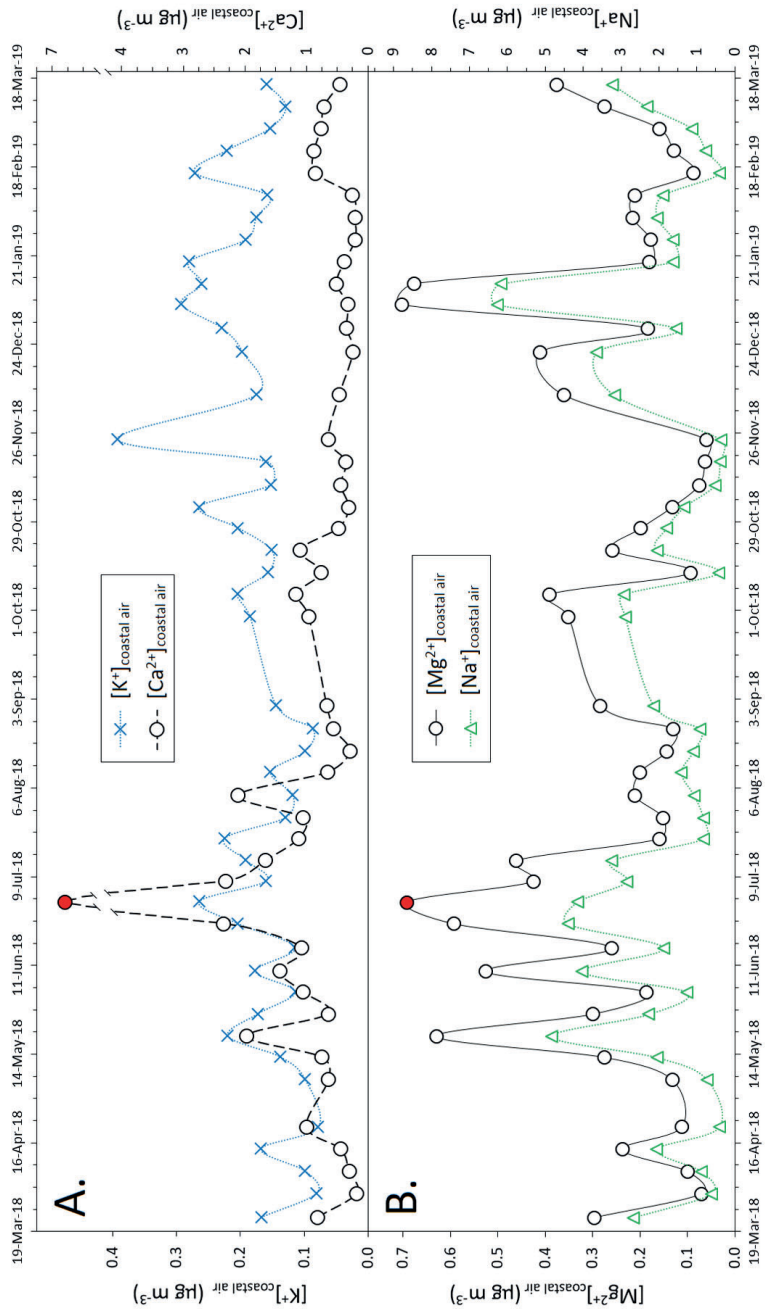


Figure S6: Sampled coastal air concentrations of the major inorganic constituents as a function of the time. In (A) the results for K^+ and Ca^{2+} are shown and in (B) those for Mg^{2+} and Na^+ . The red filled dots indicate a sampling week with nearby construction works (i.e., dust production).

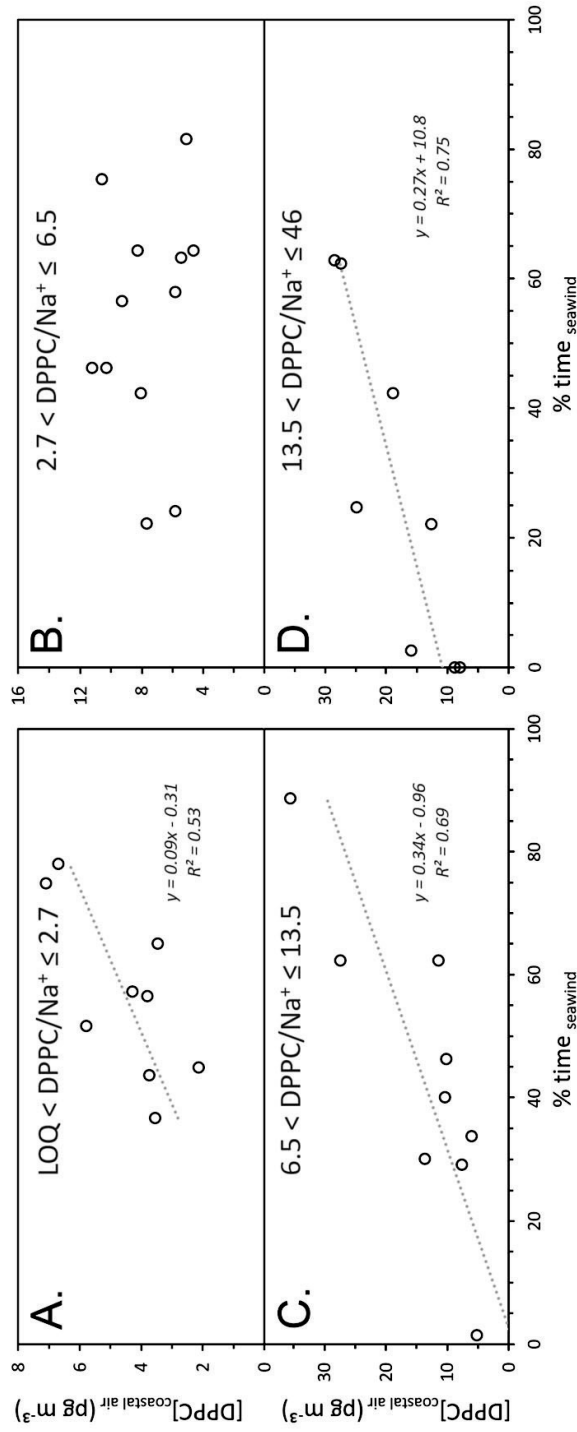


Figure S7: DPPC concentration of the sampled coastal air ($[DPPC]_{\text{coastal air}}$) as a function of the % time with seawind (Figure 1). Because the $[DPPC]_{\text{coastal air}}$ is largely dependent on the aerosol content ($DPPC/Na^+$; see Figure S3 and Figure 2), we examined this relation in different classes of $DPPC/Na^+$, ranging over (A) ≤ 2.7, (B) 2.7 – 6.5, (C) 6.5 – 13.5, and (D) 13.5 – 46. The gray (linear) trendlines illustrate the significant positive correlations between the two variables ($p < 0.05$; 95% confidence level). The equations and coefficients of determination (R^2) for the linear models are also given in the graphs.

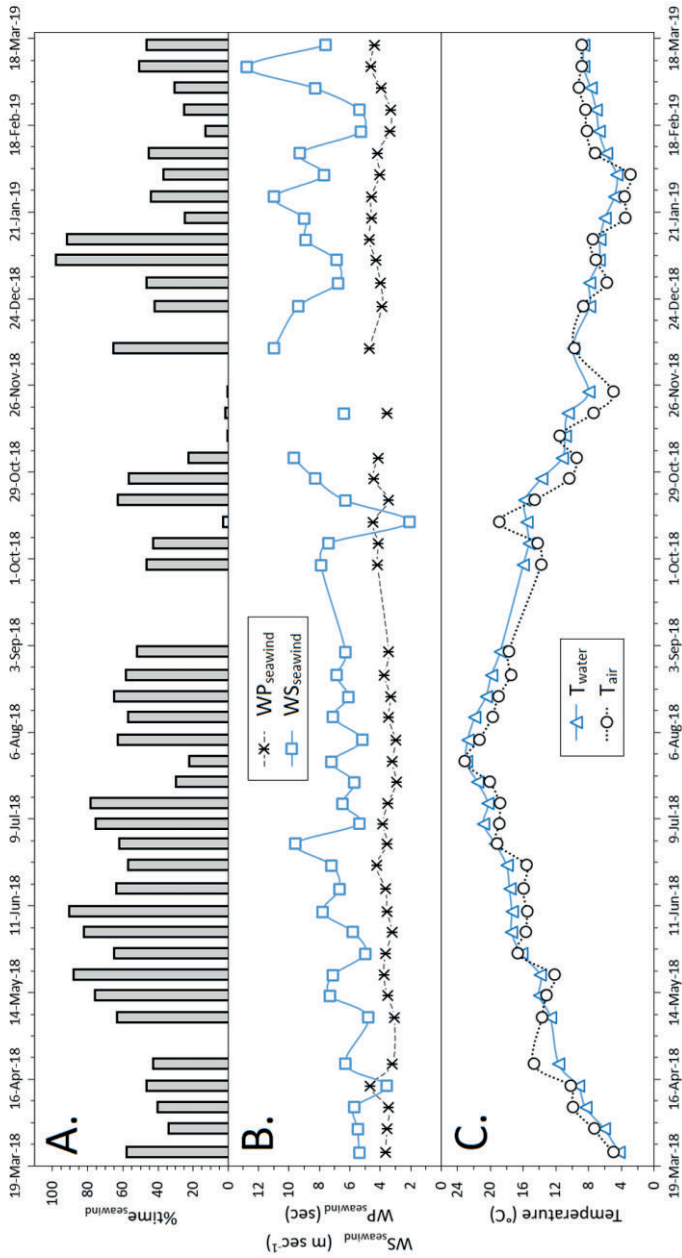


Figure S8: Graphs showing for each (successful) sampling week the (A) %time_{seawind}, and (B,C) the variables that were not used in the final GLM. In (B) the mean wave period (WP_{seawind}) and wind speed (WS_{seawind}) during seawind periods are depicted, and in (C) the mean air (T_{air}) and water temperature (T_{water}). As discussed in the 'Environmental variables' section under materials and methods, the %time_{seawind} represents the fraction of time during which the wind actually blew from across the sea ($\leq 58^\circ\text{N}$ and $\geq 235^\circ\text{N}$; see Figure 1), i.e., the wind directions that were assumed to primarily contain SSAs.

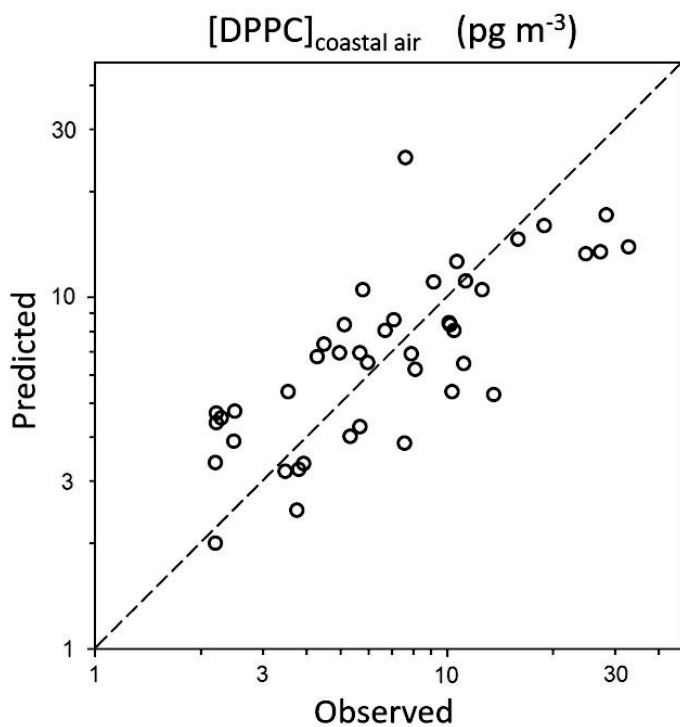


Figure S9: Scatter plot of the observed vs the predicted (modeled) values for the DPPC concentration of the sampled coastal air ($[DPPC]_{\text{coastal air}}$). The dashed line indicates the 1:1 line.

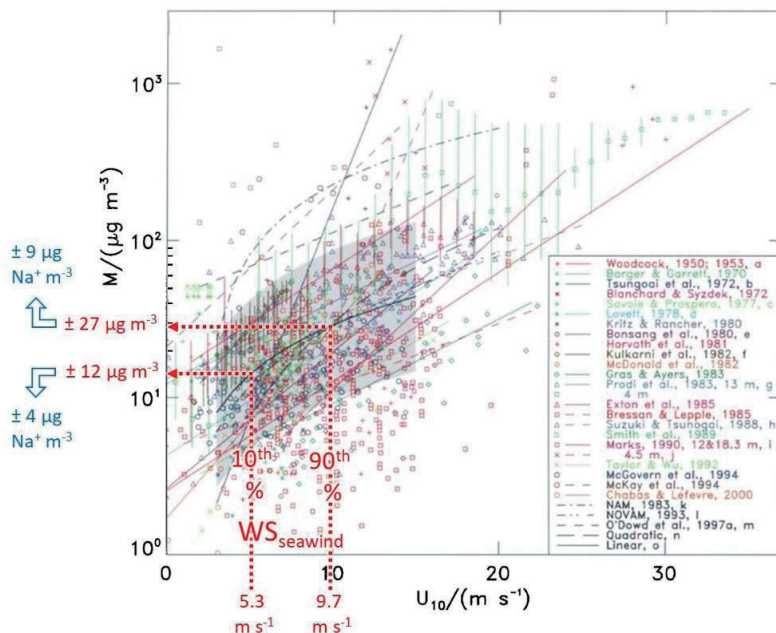


Figure 17. Measurements of total SSA dry mass concentration M , mostly in the lower marine boundary layer, as a function of wind speed at 10 m above sea surface, U_{10} , averaged over the sample period. Individual data sets are described in Table 11. Symbols denote measurements; symbols with vertical range bars denote means and standard deviations. Symbols with upward pointing arrows denote values above the maximum on the figure on which they were presented [Fig. 7 of *Savoie and Prospero*, 1977]. Data of *Exton et al.* [1985] and *Smith et al.* [1989] are offset slightly from each other for greater clarity. Lines denote fits by investigators to their data except where noted in Table 11. Black lines and curves denote results of parameterizations of size distributions of SSA concentration discussed in §4.1.4.4: NAM, the original Navy Aerosol Model [Gathman, 1983a,b], NOVAM, Navy Ocean Vertical Aerosol Model [Gathman and Davidson, 1993], O'Dowd et al. [1997a], and quadratic (based on canonical SSA size distribution of concentration: §2.1.5; §4.1.4.4). Linear relation denotes an empirical linear dependence on wind speed; shaded gray area is a factor of ≈ 3 about this relation.

Notes to Fig. 17:

^a Fit to data of Woodcock [1953] is that presented by Tsunogai et al. [1972].

^b Data shown for Tsunogai et al. [1972] consist of 55 values from Fig. 5 of that reference; line is fit to these data by those investigators.

^c Data shown for Savoie and Prospero [1977] are those from Barbados and Sal Island (Cape Verde Islands); data from Miami, not presented in that reference, were stated to be lower by roughly a factor of three at any given wind speed.

^d Data shown for Lovett [1978] consist of 65 values from Fig. 2 of that reference representing a single cruise; fit is that given for all data (Table 7 of that reference).

Figure S10: meta-analysis of Lewis and Schwartz (Lewis and Schwartz, 2004) showing the total SSA dry mass concentration on open sea (M) as a function of the wind speed at 10 m above the sea surface. In red, we have indicated on their figure the 10th and 90th percentile of the mean wind speed of the sampling week during seawind periods ($WS_{seawind}$) and the corresponding SSA dry mass concentrations (M). In blue, we showed the calculated maximum $[Na^+]_{coastal\ air}$ you would expect, as Na^+ represents (at maximum) about one third of the dissolved matter in seawater.

Supportive tables

Table S1: Overview of the independent variables whose influence on DPPC coastal air concentrations ($[DPPC]_{\text{coastal air}}$) was examined by constructing a generalized linear model (GLM). The phytoplankton biomass data discussed here represent the (6 day) time-lag corrected $[PhyC]$ data. The variance before transformation (i.e., square of the standard deviation) has the same units as the independent variable, but to the power 2. Note that only four of these variables were retained in the final GLM.

Independent variable (units)	Description	Variance before transformation	Covariance with $[DPPC]_{\text{coastal air}}$ after transformation
WH_{seawind} (cm)	Mean wave height of the sampling week during seawind periods.	1403	-0.63
WS_{seawind} (m sec ⁻¹)	Mean wind speed of the sampling week during seawind periods.	4.07	-0.52
WP_{seawind} (sec)	Mean wave period of the sampling week during seawind periods.	0.26	-0.47
T_{air} (°C)	Mean air temperature of the sampling week.	26.5	0.22
T_{water} (°C)	Mean water temperature of the sampling week.	33.1	0.14
$[PhyC]_{\text{mean}}$ (μM C)	Mean value of $[PhyC]$ of the sampling week.	17.9	0.13
$[PhyC]_{\text{decrease}}$ (μM C)	Accumulated daily decreases of $[PhyC]$ of the sampling week.	5.02	0.27
$[PhyC]_{\text{increase}}$ (μM C)	Accumulated daily increases of $[PhyC]$ of the sampling week.	2.07	0.29

Table S2: The limit of detection (LOD), limit of quantification (LOQ), and the background levels (in field blank samples; n=10) for the various analyzed major and minor inorganic ions. The filter content of the field blanks is reported as the mean content and standard deviation (n=10).

	Na ⁺	Mg ²⁺	Ca ²⁺	K ⁺	Cu ²⁺	Zn ²⁺	Pb ²⁺	Cd ²⁺
LOD per filter (µg)	0.9	0.45	0.45	0.9	0.060	0.015	0.075	0.009
per 250 m ³ (ng/m ³)	3.6	1.8	1.8	3.6	0.24	0.06	0.3	0.036
LOQ per filter (µg)	3.0	1.5	1.5	3.0	0.15	0.060	0.24	0.030
per 250 m ³ (ng/m ³)	12	6.0	6.0	12	0.60	0.24	0.96	0.12
Mean content field blank filters (µg)	22 ± 8	0.48 ± 0.30	6.1 ± 3.6	2.0 ± 0.66	0.083 ± 0.012	0.38 ± 0.15	< LOD	< LOD

Table S3: Model performance metrics across lag times from 0 to 14 days. The Gaussian generalized linear model (GLM) was fitted using log-transformed DPPC concentration as the response variable and wave height, variable decrease, variable increase, and variable mean as predictors. Bold values indicate the best-performing lag for each metric. Abbreviations: DPPC, dipalmitoylphosphatidylcholine; GLM, generalized linear model; RMSE, root mean square error; MAE, mean absolute error; AIC, Akaike information criterion; BIC, Bayesian information criterion; R^2 , coefficient of determination.

Lag Time (days)	Adjusted R^2	RMSE	MAE	AIC	BIC	Residual Deviance	Dispersion
0	0.464	0.688	0.610	101.914	112.481	20.376	0.536
1	0.464	0.688	0.601	101.919	112.486	20.378	0.536
2	0.488	0.673	0.590	99.936	110.504	19.460	0.512
3	0.497	0.667	0.593	99.171	109.738	19.116	0.503
4	0.479	0.678	0.602	100.647	111.214	19.784	0.521
5	0.483	0.676	0.583	100.339	110.906	19.643	0.517
6	0.529	0.645	0.537	96.334	106.901	17.896	0.471
7	0.467	0.686	0.584	101.658	112.225	20.255	0.533
8	0.438	0.705	0.607	103.939	114.506	21.358	0.562
9	0.435	0.707	0.606	104.199	114.766	21.488	0.565
10	0.431	0.709	0.603	104.468	115.036	21.623	0.569
11	0.444	0.701	0.603	103.446	114.013	21.115	0.556
12	0.437	0.706	0.605	104.046	114.613	21.411	0.563
13	0.520	0.652	0.562	97.193	107.760	18.257	0.480
14	0.524	0.649	0.553	96.809	107.377	18.095	0.476

Chapter 3 Seasonal Variation in the Bioactivity of Sea Spray Aerosols Driven by Phytoplankton

This chapter is adapted from a paper published in *Environment International* (2025, 195, 109255; <https://doi.org/10.1016/j.envint.2025.109255>).

CRedit authorship contribution statement

Zixia Liu: Methodology, Data curation, Software, Writing- Original draft preparation. Emmanuel Van Acker: Conceptualization, Methodology, Data curation. Maarten De Rijcke: Conceptualization, Methodology. Filip Van Nieuwerburgh: Methodology. Colin Janssen: Conceptualization, Supervision. Jana Asselman: Conceptualization, Methodology, Supervision. All: Writing- Reviewing and Editing.

Exploring Seasonal Dynamics of Sea Spray Aerosol Bioactivity: Insights into Molecular Effects on Human Bronchial Epithelial Cells

Abstract

Sea spray aerosol (SSA) is a complex mixture of natural substances that can be inhaled by coastal residents. Previous studies have suggested that SSA may have positive effects on human health, but the molecular mechanisms and the factors influencing these effects are poorly understood. In this study, we exposed human bronchial epithelial cells (BEAS-2B) to natural SSA samples, collected monthly using quartz microfiber filters mounted on tripods within 15 meters of the waterline, with air drawn through pumps, throughout a one-year period at the Ostend coast, Belgium, and measured cellular gene expression changes using RNA sequencing. To simulate environmentally relevant exposure conditions, SSA extracts were applied at scaled doses equivalent to human alveolar exposure levels (multiplicative factors $M = 10, 20, 40, \text{ and } 80$). We found that SSA exposure influenced the

expression of genes involved in critical signaling pathways: mTOR, PI3K, Akt, and NF- κ B were downregulated, while AMPK was upregulated. Downregulation of mTOR, PI3K, Akt, and NF- κ B potentially indicates a protective response against tumor-promoting and inflammatory signals, whereas upregulation of AMPK may confer a beneficial effect on metabolic regulation. The number and direction of differentially expressed genes (DEGs) varied depending on the SSA sampling time and correlated with the phytoplankton density and chemical diversity of the SSA samples. Our results suggest that SSA contains bioactive compounds that may originate from marine algae and modulate cellular processes related to human health. We provide novel insights into the molecular effects of SSA exposure and highlight its potential as a source of natural therapeutics. To our knowledge, this is the first study to expose human lung cells to natural SSA at environmentally relevant levels, presenting a pioneering exploration of seasonal variations in exposure effects.

Keywords: sea spray aerosol (SSA); bronchial epithelial cells; RNA sequencing; biogenics hypothesis; mTOR pathway; seasonal variation; oceans and human health

1. Introduction

The coastal environment, an important element of "blue spaces," is increasingly recognized not only for its psychological benefits but also for its potential physiological impacts on human health (Bell et al., 2015; Gascon et al., 2017; Wheeler et al., 2012). Beyond the well-documented restorative effects on mental well-being (Gascon et al., 2017), recent research suggests that aerosolized biogenic compounds present in sea spray aerosols (SSAs) might play a significant role in influencing human health (Moore, 2015). These compounds, which include a variety of bioactive substances, are hypothesized to interact with cellular signaling pathways such as mTOR, Akt, or PI3K (Asselman et al., 2019; Moore, 2015; Van Acker et al., 2020). This interaction could have both positive

and negative health implications for individuals residing in coastal areas. Such a perspective expands the traditional understanding of the health benefits associated with coastal environments, moving beyond psychological effects to encompass a potential biochemical impact.

The major transport route of marine biogenics is via SSA (Lang-Yona et al., 2014), primarily generated by the bursting of bubbles taking place in the sea surface microlayer (SSML) (Keene et al., 2007). The SSML, a unique interface between the air and ocean, is rich in various hydrophobic and amphiphilic substances, including biogenic compounds (Aller et al., 2005). These compounds are often found in much higher concentrations in SSAs compared to bulk seawater (Van Acker et al., 2021b), highlighting the significance of SSAs as a carrier for these substances.

Among the diverse sources of marine biogenics, phytoplankton stand out due to their substantial biomass (Bar-On and Milo, 2019) and the wide array of bioactive compounds they release into the ocean. These biogenic compounds, encompassing polysaccharides, lipids, amino acids, and enzymes (Menaar et al., 2021), not only play a crucial role in the marine ecosystem but also have the potential to impact human health when aerosolized, either in positive or negative ways. Several studies have reported the negative health effects of aerosolized marine biogenic compounds during harmful algal blooms (HABs) (Backer et al., 2005; Fleming et al., 2011). The phycotoxins produced during HABs were confirmed to be aerosolized and exist in SSAs of coastal regions (Fleming et al., 2005), and the relationship between respiratory symptoms of people living in coastal regions after HAB events has been quantified by an epidemiological study (Hoagland et al., 2009). Notably, due to their hydrophobic nature, some phycotoxins were reported to be enriched up to thousands of times in SSAs (Van Acker et al., 2021b), intensifying their potential impact. Beyond HAB events, a recent study has shown that a biogenic compound produced by marine algae can be detected in coastal air, even at ambient algal concentrations (Van Acker

et al., 2021a). This indicates the coastal residents may be continuously exposed to marine biogenic compounds throughout the year, suggesting a broader potential health impact of SSAs on coastal populations.

Furthermore, certain marine algal compounds have demonstrated anti-cancer or anti-inflammatory properties, which are studied for their pharmaceutical and cosmeceutical applications (Dong et al., 2020; Pereira, 2018). While characterization of SSA composition has mainly focused on inorganic fractions (Bertram et al., 2018), some biogenic compounds have already been identified (Ault et al., 2013; Bertram et al., 2018; Cochran et al., 2017; DeMott et al., 2016; Miyazaki et al., 2020). The biogenic compound dipalmitoylphosphatidylcholine (DPPC), likely produced by marine algae, has been detected in SSA samples (Van Acker et al., 2021a). Given DPPC is also the main constituent of pulmonary surfactants and has been used as an inhalable surfactant and inhalation enhancer (Hidalgo et al., 2015). As such, DPPC in SSA may have implications for respiratory health. At the molecular level, artificial SSA produced from media containing the phycotoxin homoyessotoxin, as well as natural SSA samples, have been shown to interact with the mTOR pathway at both gene and pathway level (Asselman et al., 2019).

The studies mentioned above lead us to this expanded biogenics hypothesis: marine algae can release bioactive compounds that get enriched in the SSA; to which coastal residents are continuously exposed. From the exposure side, the detection of marine toxins and other biogenic compounds in SSA (Cheng et al., 2010; Fleming et al., 2009; Van Acker et al., 2021a, 2021b) provides direct evidence that these substances can become airborne and potentially inhaled by people living near the coast. Daily respiratory exposure to SSA may inhibit signaling pathways such as mTOR, Akt or PI3K and thereby positively influence human health. Studies that exposed human cell lines to coastal air (Asselman et al., 2019) or aerosolized marine toxins

(Van Acker et al., 2020) have supported this hypothesis from the effect side. Each piece of the knowledge puzzle is supported by one or more studies, yet no single study has verified the entire story at an environmentally relevant dose, particularly with regards to the seasonal variation in the production and composition of the SSAs, given the well-recognized variations of phytoplankton composition across seasons.

In this study, we exposed human bronchial epithelial cells (BEAS-2B) to nine natural SSA samples collected at the Belgian coast throughout a one-year campaign. The SSA samples were collected during conditions with strong winds capable of generating white-cap waves, with the wind blowing from the sea to the coast, ensuring the aerosols were primarily composed of SSA. Additionally, we included ambient coastal aerosol samples (AMB, collected during conditions of little wind, ensuring no locally generated SSA was present in the sample), city aerosol samples (CTY), positive controls (POS, mTOR inhibitor PP242), and negative controls (NEG) in the exposure experiments, with each serving as a treatment group. The RNA expressions of each BEAS-2B cell line sample were quantified, compared across treatments, and fitted against environmental factors and the chemical diversity found within the SSA to support the biogenics hypothesis at the molecular level using environmentally realistic conditions.

2. Results

Across all treatment groups, we quantified the expression of 58,233 genes and 199,234 transcripts. Hierarchical clustering (**Figure 1.a**) displayed no obvious outliers, suggesting consistent sample quality. A distinct clustering pattern emerged when we applied a cut-off at a height of 40,000 on the dendrogram. This division led to three major clusters (as indicated in SI.11): Cluster 1 comprising samples exposed to SSA during temperate conditions (air temperature 8.39°C - 16.12°C), Cluster 2 with control samples and those exposed to SSA in cold weather, and Cluster 3 containing samples from hot weather SSA exposures.

Intriguingly, SSA collected in cold weather exhibited significantly lower Na⁺ content compared to other conditions. A correlation with phytoplankton density (see **Discussion** for details), was noted.

2.1. Differential gene expression in response to treatments.

Out of all the genes and transcripts analyzed, 19,768 protein-coding genes and 79,757 transcripts were further examined. Gene-level DEGs reflect changes at the gene level, while transcript-level DEGs capture specific transcript variants. Analyzing both provides a comprehensive view of cellular responses. Comparing the expression levels of these genes between the NEG group and all the other treatment groups, the SSA group has (one of) the highest number of DEGs across all treatment groups: 91 gene-level DEGs and 425 transcript-level DEGs at a False Discovery Rate (FDR) ≤ 0.05 , and 32 gene-level DEGs and 327 transcript-level DEGs at FDR ≤ 0.01 , with a >2 -fold change. Notably, most DEGs were downregulated, except for the positive control group, which showed a higher count of upregulated DEGs. Although the numbers of transcript-level DEGs are generally higher than gene-level DEGs, the patterns of DEG numbers for each treatment group are comparable. To find as many DEGs as possible that respond to SSA exposure, an FDR ≤ 0.05 threshold was adopted. **Figure 1 b & Figure 1 c** illustrate the DEG counts and distributions (volcano plots) for each group. Disregarding the gene type of genome or transcriptome, the AMB, CTY, and SSA treatment groups showed similar patterns in volcano plots, but the dots in the similar position do not necessarily mean the same gene. There are 44 gene-level DEGs and 106 transcript-level DEGs in the SSA exposure group which are unique; 8 gene-level DEGs and 61 transcript-level DEGs are shared by the SSA group and POS group. To be noted, there are 27 gene-level DEGs and 206 transcript-level DEGs shared by the SSA group and the AMB group, which count for 29.7% and 61.2% of the SSA group's DEG number. This indicates the ambient coast air may also contain some bioactive compounds as in SSA. There are also some DEGs shared by the SSA group and CTY

group, which may be the effects of exposure to city air and were addressed in the discussion. The numbers of shared and unique DEGs across all treatments were visualized in **Figure 1 d**.

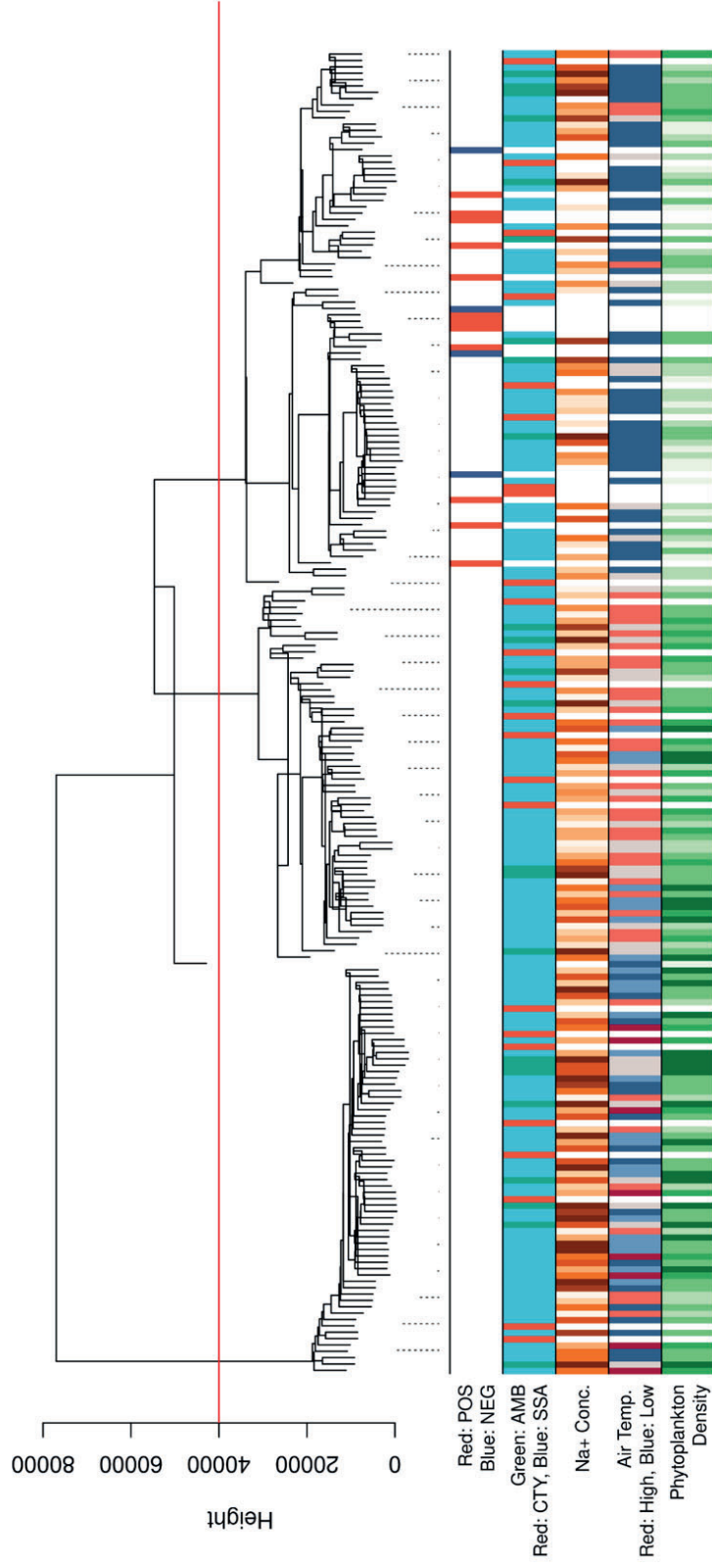


Figure 1 a Differential gene expression in response to treatments. Hierarchical clustering of gene expression in samples exposed to SSA under varying environmental conditions. The clustering of gene expression was performed using genome-

wide expression profiles and employed the centroid linkage method. Clusters are demarcated by a 40,000 height threshold (red line), with sample leaves colored by treatment: positive or negative control (POS/NEG), ambient air (AMB), city air (CTY), and SSA. Below the dendrogram, metadata corresponding to SSA collection parameters (Na^+ concentration, air temperature, and phytoplankton density) are represented by colors. These colors were assigned based on k-means clustering of the respective metadata values: Na^+ concentration (9 clusters, orange gradient), air temperature (5 clusters, red-blue diverging palette), and phytoplankton density (5 clusters, green gradient). White represents missing values for metadata.

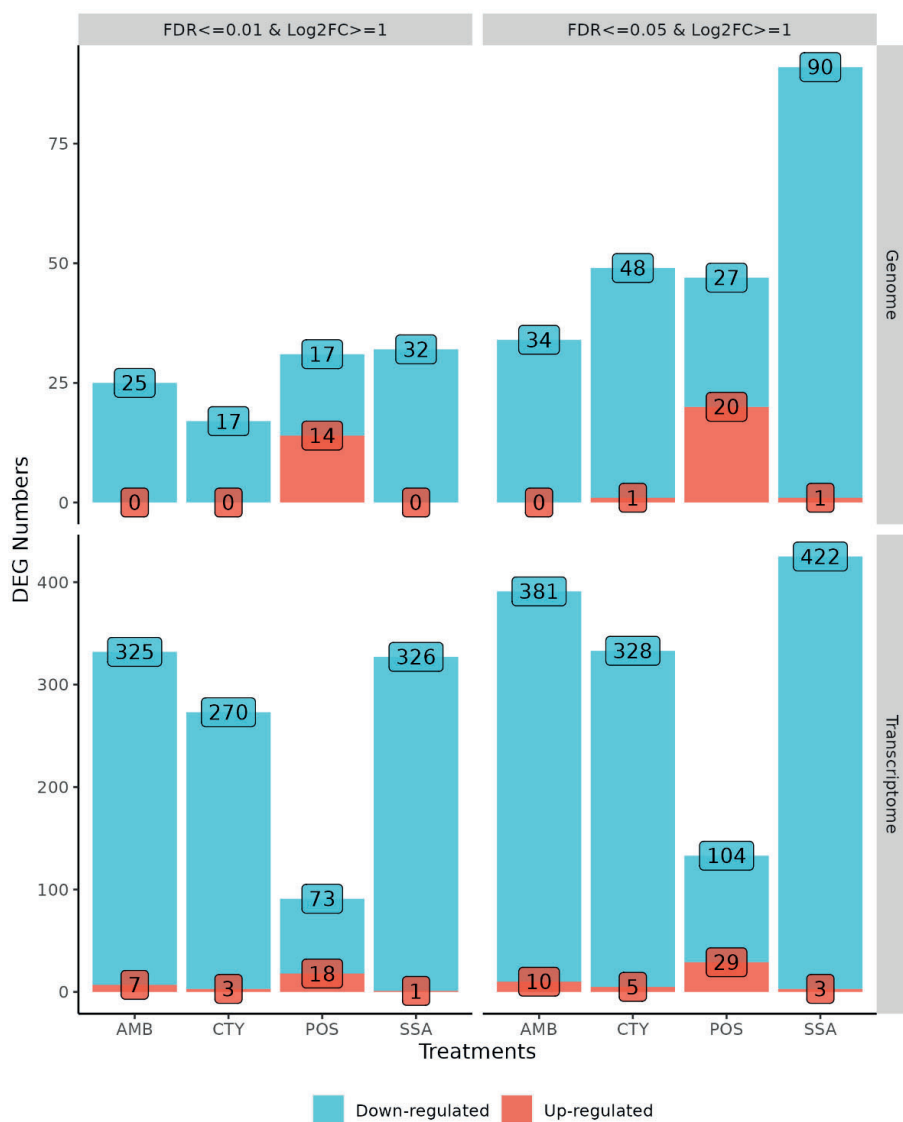


Figure 1 b Differential gene expression in response to treatments. Counts of genome and transcript-level DEGs categorized by downregulation (blue) and upregulation (red) across treatments. DEGs are identified using thresholds of $FDR \leq 0.05$ and $|\log_2FC| \geq 1$ (right panels) or stricter thresholds of $FDR \leq 0.01$ and $|\log_2FC| \geq 1$ (left panels).

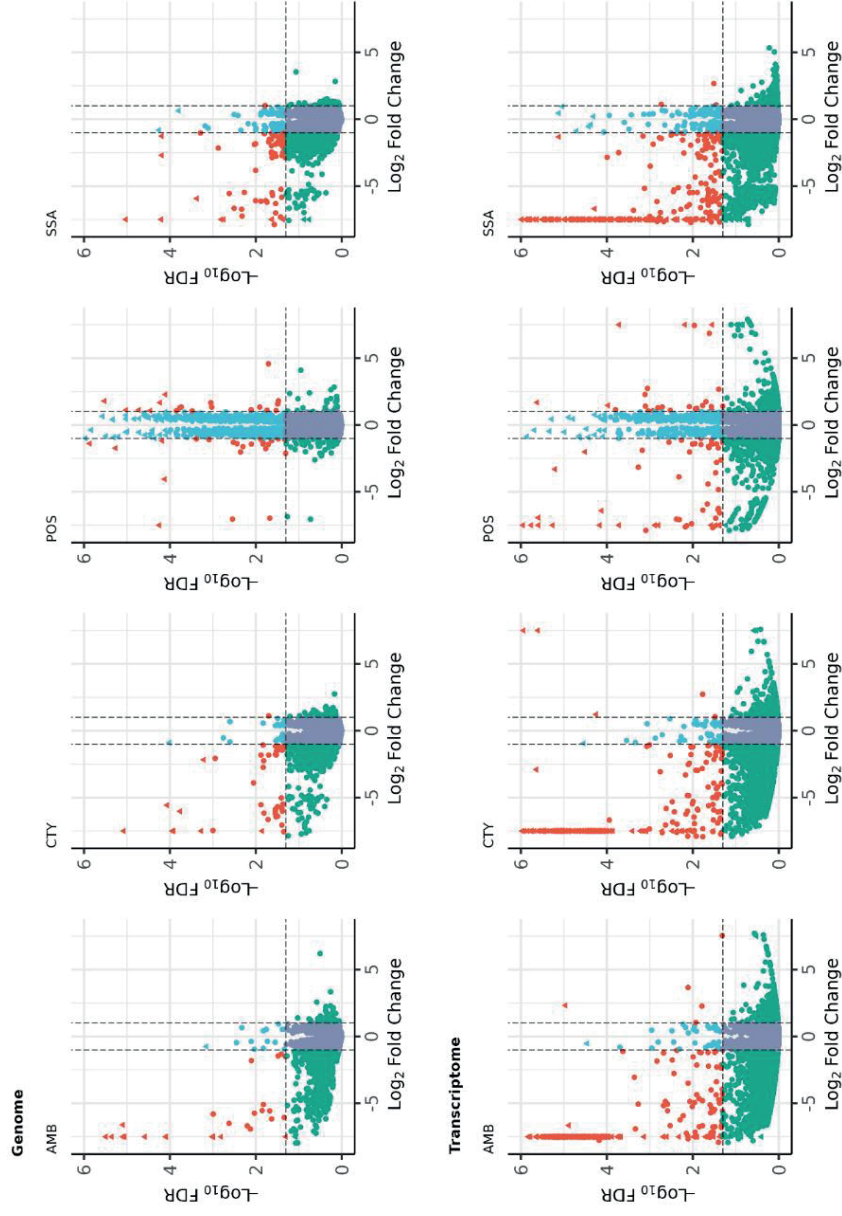


Figure 1 c Differential gene expression in response to treatments. Volcano plots illustrating the $\log_2$ fold change against the

negative $\log_{10}$ FDR for DEGs in genome (up) and transcriptome (down) analyses. Genes meeting both thresholds of $FDR \leq 0.05$ and $|\log_2 FC| \geq 1$ are highlighted in red, genes meeting only the FDR threshold are shown in blue, and genes meeting only the $\log_2 FC$ threshold are shown in green. Genes not meeting either threshold are depicted in gray. Dashed lines indicate the significance thresholds.

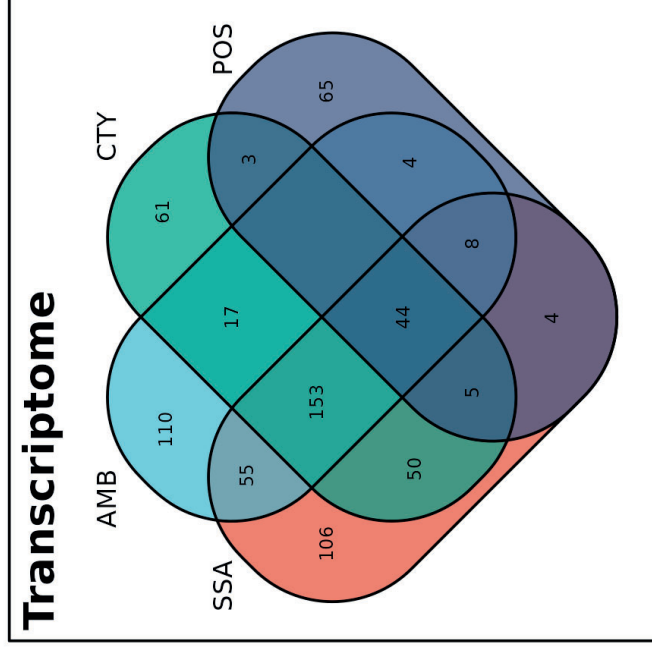
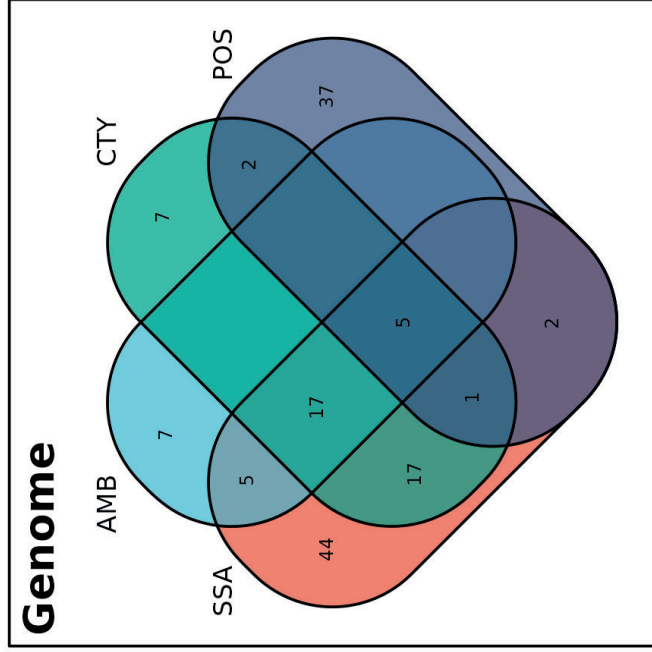


Figure 1 d Differential gene expression in response to treatments. Venn diagrams representing the overlap and uniqueness of differentially expressed genes (left) and transcripts (right) across four treatment groups. Each section is quantified to illustrate the number of shared and distinct DEGs.

2.2. Genes in mTOR pathway in response to SSA exposure.

The mTOR pathway has been reported to be regulated by SSA exposure. To verify this, we compared the expression of these genes across all samples. In the heatmap of the mTOR signal pathway (**Figure 2 a**), there are two clusters of sample that showed a clear difference in gene expression, which were either in the "temperate" or "hot" weather group. This cluster is composed of 64 samples, among which 48 were in the SSA treatment group, 8 were in the AMB group, and another 8 were in the CTY group. This cluster's phytoplankton densities were significantly higher than the other cluster (ANOVA, $p = 0.0289$). Samples exposed to different concentrations of the same SSA/coast air extract were in one cluster (SSA091, SSA110, SSA134, AMB119), indicating a lack of dose-dependent variation in terms of DEG number was observed. To further elucidate the impact of SSA on mTOR pathway gene expression, samples treated with abovementioned extracts were compared with NEG group. The resultant upregulation and downregulation of mTOR genes are depicted in a KEGG pathway plot (**Figure 2 b**), based on log2 fold changes. Our findings reveal that both complexes within the mTOR pathway are modulated following SSA exposure.

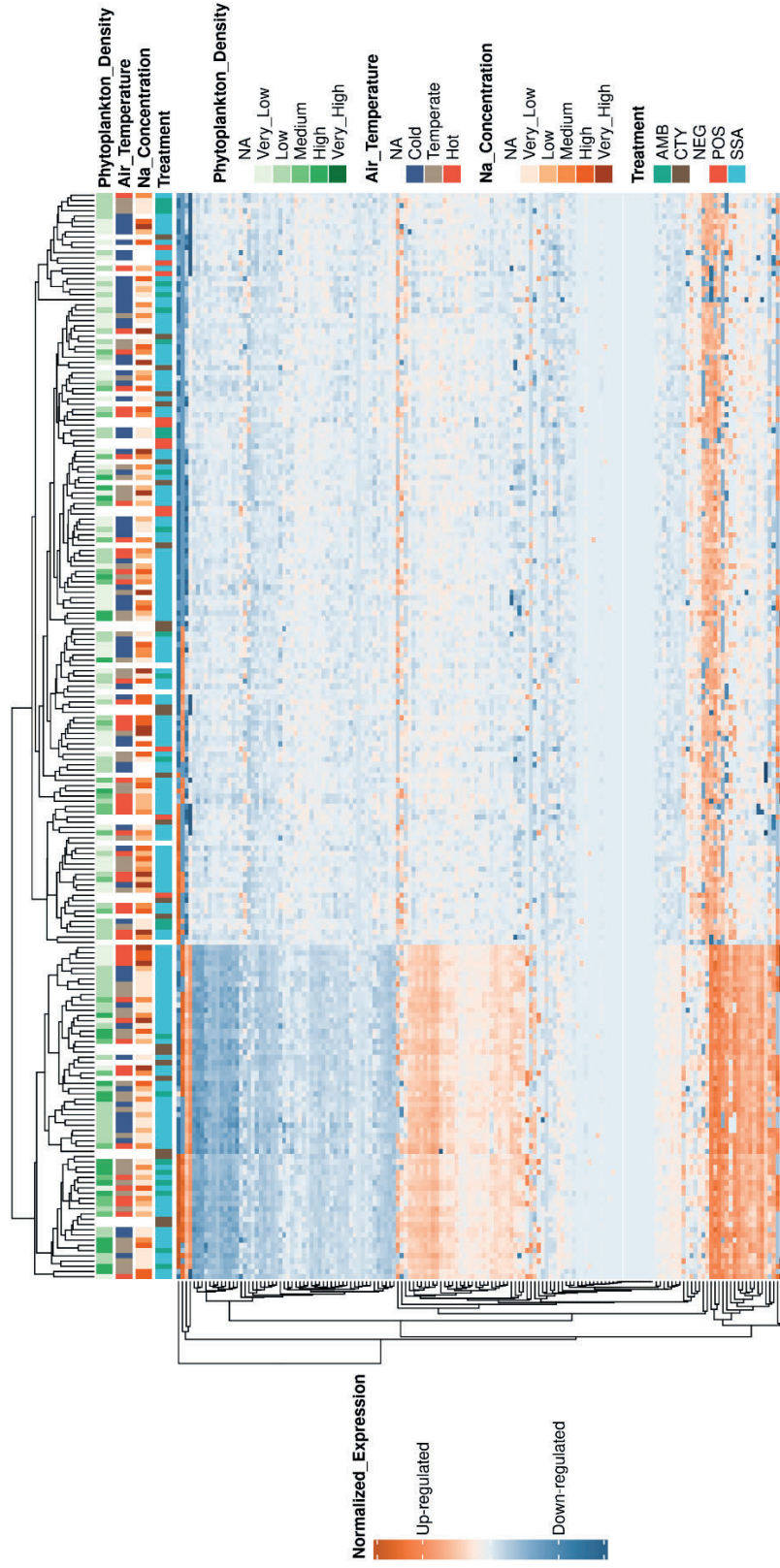


Figure 2 a *mTOR* pathway response to SSA exposure and environmental conditions. Heatmap illustrating gene expression

levels across samples, categorized by phytoplankton density, air temperature, sodium concentration, and treatment type. The clustering of gene expression was performed using mTOR-related gene expression profiles and employed the complete linkage method. Metadata annotations below the heatmap were grouped using k-means clustering for binning of sodium concentration (5 clusters), air temperature (3 clusters), and phytoplankton density (4 clusters). Color gradients represent expression levels (blue: downregulated, red: upregulated).

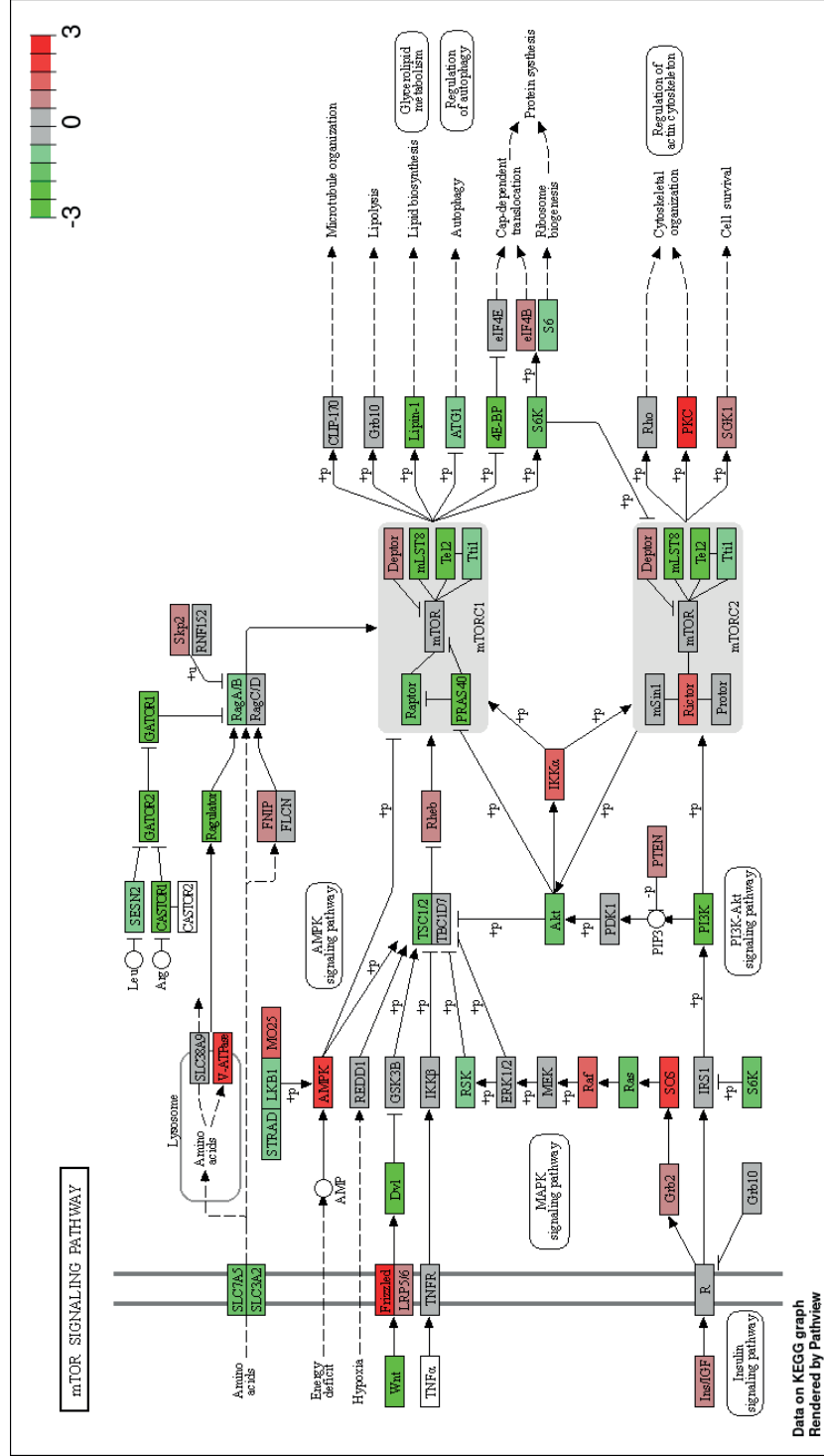


Figure 2 b mTOR pathway response to SSA exposure and environmental conditions. KEGG pathway plot indicating specific

mTOR pathway genes modulated by SSA exposure, showing both complexes affected, with changes annotated by log2 fold expression differences from controls.

2.3. Seasonal variations of SSA composition and exposure effects.

As shown in **Figure 1 a & SI.11**, there are distinct compositional differences between samples collected in colder (Cluster 3 in SI.11) and warmer months (Cluster 2 in SI.11), as indicated by blue and red color bands representing air temperatures within the orange rectangle of SI.11. This variation correlates with the seasonal trends in phytoplankton abundance (see **SI.4**) and air temperature. We can make a reasonable deduction, higher phytoplankton biomass during warmer periods led to increased concentrations of bioactive chemicals in SSA, and therefore had different expressions compared to the SSA samples collected in colder periods. To further elucidate the impact of seasonal variations on SSA exposure effects, DEGs were calculated again grouping by the SSA sampling date. The number of DEGs varied significantly, with differences up to four magnitudes between samples taken on different dates (**Figure 3 a**). Notably, there is a strong correlation between the number of DEGs and the phytoplankton density at the time of SSA collection (**Figure 3 b**). This relationship is particularly pronounced in transcriptome genes, where phytoplankton density accounts for approximately 69% of the variance in DEG numbers across all samples ($p < 0.01$). Nevertheless, this warrants further investigation to determine effective causality between phytoplankton biogenics in SSA and potential human health impact.

The chemical diversity within aerosol samples was quantified and analyzed to understand the variation in SSA exposure effects. Using ANOVA and post-hoc comparisons, we found that the SSA group samples exhibited significantly higher feature counts, indicating greater chemical diversity, compared to control groups (**Fig 3 c**; ANOVA, $p < 0.01$; Tukey HSD results in **SI.6**). Further, linear regression analysis supports this observation, with a t-value of 2.571 and a p-value of 0.0213, highlighting the distinctive chemical composition of SSA group.

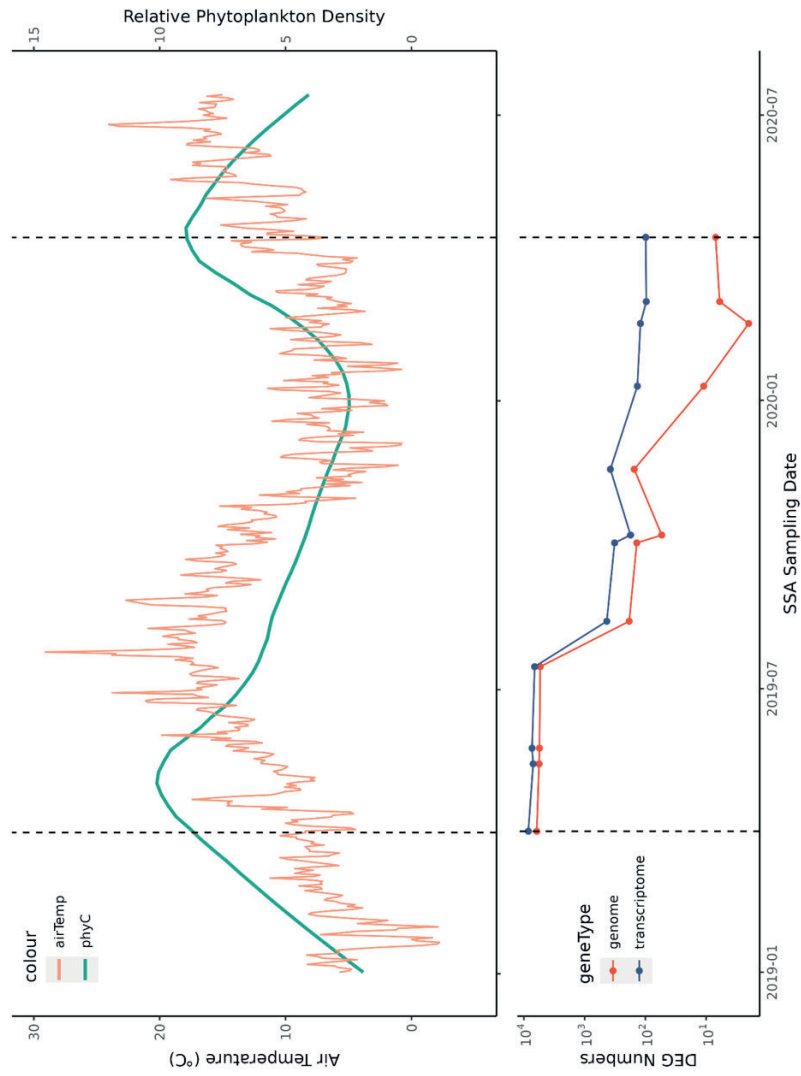


Figure 3 a Seasonal variation of SSA composition and exposure effects. Temporal changes in differential gene

expression (DEG) numbers over SSA sampling dates, showing distinct shifts along with relative phytoplankton density (PhyC) and air temperature (airTemp) between April 2019 and April 2020.

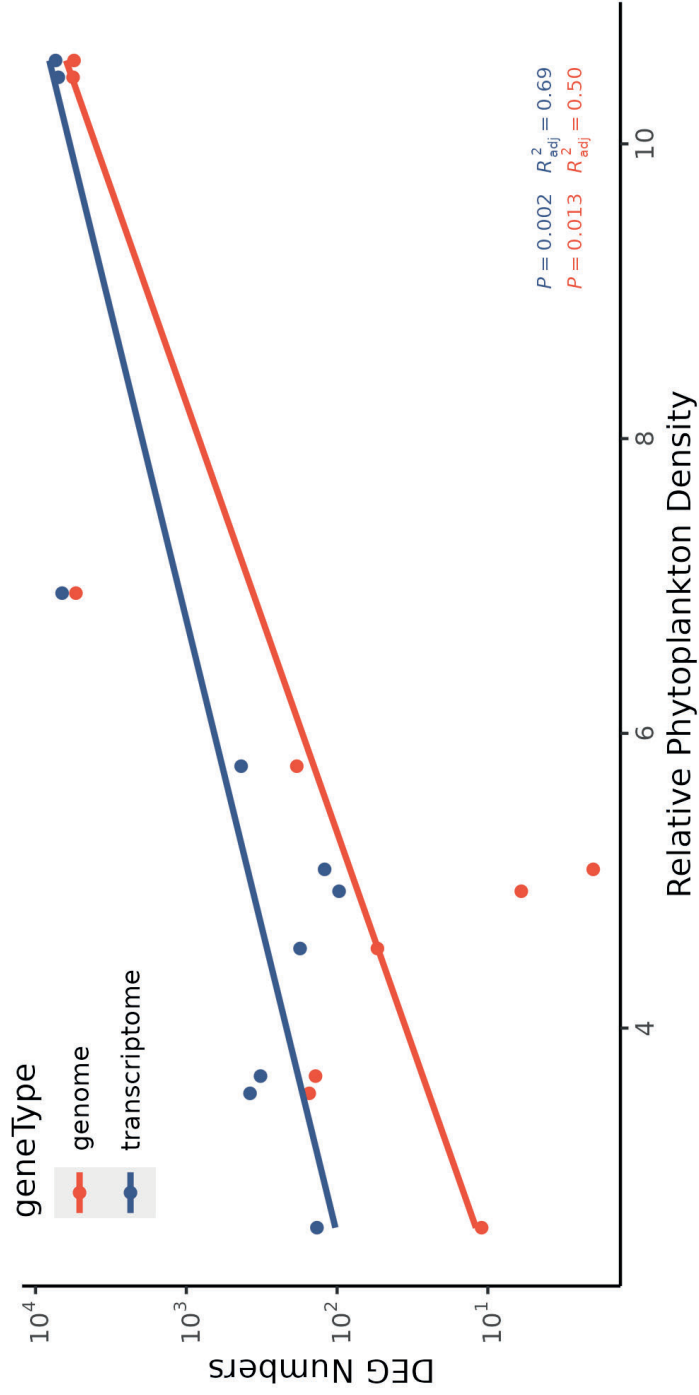


Figure 3 b Seasonal variations of SSA composition and exposure effects. Relationship between DEG numbers and relative phytoplankton density, indicating a strong correlation with both genomic and transcriptomic responses.

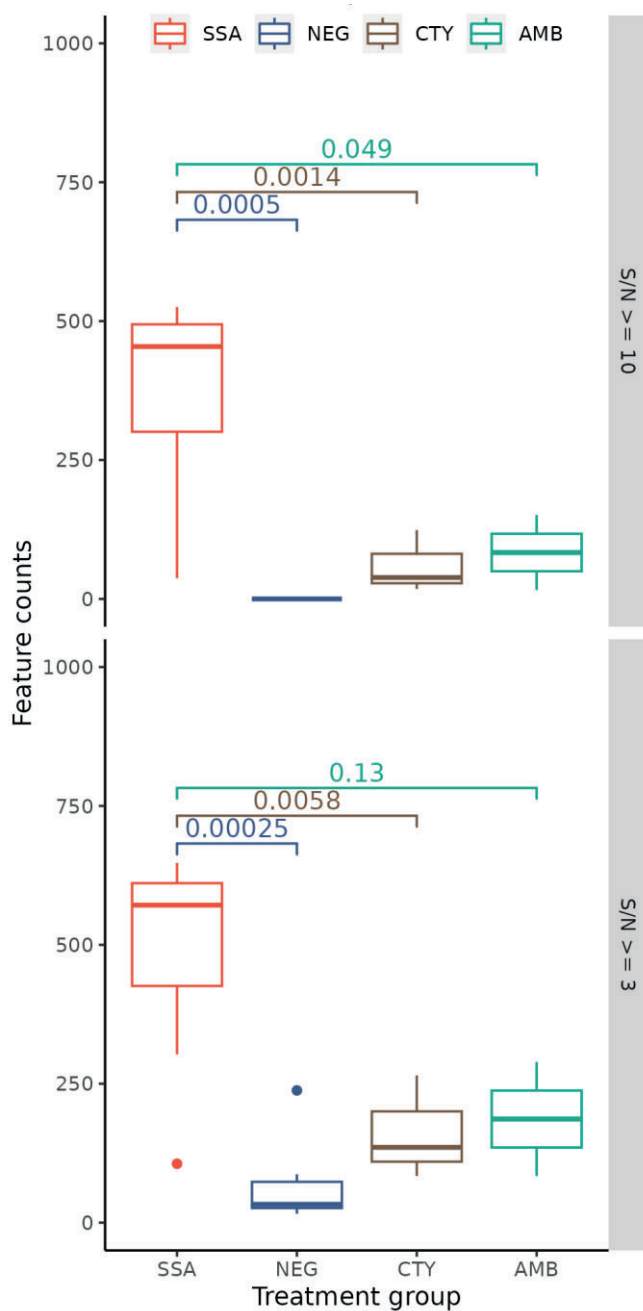


Figure 3 c Seasonal variations of SSA composition and exposure

effects. Feature counts (the number of distinct signals detected by the mass spectrometer that exceed specified signal-to-noise (S/N) thresholds) across treatment groups, with SSA showing significantly higher diversity in chemical composition than negative control (NEG), ambient air (AMB), and city air (CTY) groups.

3. Discussions

3.1. Bioactive chemicals in SSA may be released by phytoplankton.

Strong correlations between DEG numbers and the phytoplankton densities of the sampling area were observed in our results (**Figure 3 b**), suggesting the bioactive components in SSA which caused gene differential expressions might originate from the phytoplankton. It is noteworthy that the phytoplankton densities in April of 2019 and 2020 showed differences in terms of trend and values, likely due to annual environmental variations in the sampling area that affect biological diversity, as demonstrated in **Figure 3 a**. The DEG number trends seem to follow the phytoplankton density trends but with a lag phase around 4 days (**SI.7 & SI.8**, lag = 4 days, decrease of phytoplankton aligned to DEG number change). A similar pattern is reported in a one-year study (Van Acker et al., 2021a) on SSA chemical composition, it observed a 6-day lag phase between phospholipid (DPPC) concentration in SSA and the decreasing of phytoplankton abundance. This may be explained by the phytoplankton cell lysis a few days after the death, and it takes another few days to be enriched and transported to the coast air. If this holds true, the concentrations of bioactive chemicals in SSA may not be directly linked to the abundance of phytoplankton, but rather to the quantity of deceased phytoplankton from a few days prior. However, since the sampling dates of SSA are approximately one month apart rather than one week, we are not able to use modeling to confirm this hypothesis.

According to several systematic reviews on algal-derived anti-inflammatory or anti-cancer compounds, the bioactive chemical (s) responsible for the exposure effects may belong to the groups of phytosterols, polysaccharides, and polyphenols (Alves et al., 2018; Esposito et al., 2022; Tan et al., 2021). These groups provide us with directions for future exploration of beneficial bioactive substances in SSA, potentially leading to the discovery of substances or drugs that contribute to the improved health of individuals living in the “blue

spaces". The chemical diversity of SSA samples significantly decreased in May and June 2019 (see **SI.9**), coinciding with the spring blooming of 2019. We hypothesize that the decreased species evenness during the bloom led to a reduction in chemical diversity of the released bioactive chemicals, despite the (delayed) overall rise in the concentration of bioactive chemical (s) in both the water column and SSAs due to the increased biomass.

Although existing studies (May et al., 2018; Pierce et al., 2003; Zhang and Du, 2017) have demonstrated that higher concentrations of toxins are present in SSA when HABs occur, and DEG numbers after SSA exposure are correlated with the abundances of phytoplankton in seawater in this study, it is currently insufficiently demonstrated that the bioactive substances responsible for DEG number variations originate from phytoplankton. Studies (Nielsen and Bilde, 2020; Santander et al., 2022) have indicated that air temperature and sea surface temperature are important factors influencing the enrichment factors of hydrophobic components in SSML or SSA. Therefore, the observed interannual variations of DE may also be attributed to the temperature-induced changes in the enrichment factor of bioactive substances in SSA (**Figure 3 a & Figure 3 b**). Moreover, other marine organisms like bacteria and zooplankton, exhibiting similar temperature-correlated trends, cannot be discounted as alternative bioactive substance sources. Because they all track the same interannual variation trend, it is not possible to verify through the collection of natural samples and exposure experiments conducted in this study. Consequently, controlled Marine Aerosol Reference Tank (MART) experiments, varying marine organisms, are necessary to pinpoint the specific sources of these bioactive substances.

3.2. Implications of SSA exposure on key biological pathways.

Our findings from the spring 2019 SSA exposure (April to July) revealed distinct molecular responses compared to control groups (**Figure 2 a &**

Figure 3 a). Notably, there was a significant inhibition of the mTOR pathway (as shown in Figure 3.2.2), PI3K, Akt, and NF- κ B (detailed in the **SI.5**), alongside a marked activation of the AMPK signaling pathway (**SI.5**). These molecular changes mirror aspects of bioactive substances produced by algae, which have been reported to either inhibit PI3k/Akt/mTOR/NF- κ B (Kim et al., 2015; Walter et al., 2022; Zhao et al., 2020) or activation of AMPK (Eo et al., 2015). These five signaling pathways are often associated with cancer, diabetes, and obesity (Moore, 2015), and there are currently many inhibitors targeting these specific signaling pathways available as drugs or still in clinical trials (Peng et al., 2022). The mTOR, PI3K, and Akt pathways are heavily implicated in cancer development and progression, and their inhibition is often associated with reduced cell proliferation and increased apoptosis (Esposito et al., 2022; Peng et al., 2022). NF- κ B is a key regulator of inflammation, and its downregulation suggests that SSA exposure may have anti-inflammatory properties (Kim et al., 2015). AMPK is a central regulator of energy homeostasis that modulates numerous metabolic pathways. Activation of AMPK can enhance insulin sensitivity and is a target for diabetes and obesity treatments (Crunkhorn, 2021; Eo et al., 2015; Esposito et al., 2022; Liu et al., 2021; Peng et al., 2022). In summary, the gene-level responses to the SSA exposure demonstrated potential positive effects on cancer prevention, anti-inflammatory, and metabolic benefits.

3.3. Limitations.

We observed effects between AMB group and NEG control, which share the same trend as SSA group, i.e. the presence of wind or not during sampling period does not affect the effects of exposure, this result is not what we expected but can be explained. The smaller SSA particles (majorly developed from film drops) contain a much higher proportion of organic carbon, which contains the bioactive parts of SSA, while the bigger ones mainly contain sea salts (Bertram et al., 2018). Smaller SSA particles also have lower settling velocities, making them more prone to

suspension in the air. Literature reports indicate that they sampled SSA particles generated more than 72 hours earlier at a distance of 700 kilometers from the nearest seawater source (May et al., 2018). Analysis of backward trajectories (see **SI.10**) shows that, except for one SSA sampling date (2019-04-01), all sampling dates involved air masses passing over seawater regions. These trajectory results support the hypothesis that AMB samples may also contain the bioactive components of SSA. This suggests that even in windless conditions, we might collect film drops produced a day or even earlier before sampling. Therefore, the consistent results between the AMB and SSA group in our study are plausible. In future research, it is necessary to incorporate trajectory tracking of the sampling period and location into the model for comprehensive consideration, rather than relying solely on the in-situ wind speed or the content of Na^+ to determine the aerosol sampling concentration.

While our study found a significant correlation between the number of differentially expressed genes (DEGs) and phytoplankton densities in the sampling area, we acknowledge that this correlation alone does not provide strong enough evidence to conclusively support the impact of bioactive chemicals in sea spray aerosols (SSA) released by phytoplankton. The chemical composition of SSA is complex and not fully characterized in our study, making it a "black box" that limits our ability to pinpoint specific bioactive compounds responsible for the observed biological effects. Additionally, other environmental factors not accounted for could contribute to DEG variability. Without direct chemical analyses linking specific phytoplankton-derived compounds in SSA to gene expression changes, the evidence remains correlational rather than causal. Future studies should focus on isolating and identifying the bioactive components within SSA and assessing their direct effects on gene expression to strengthen the understanding of how phytoplankton activity may influence human health through aerosol exposure.

We are aware that methanol can influence cellular metabolism, and solvent extraction may impact results by enhancing or diminishing the bioactivity of chemical substances (Erukainure et al., 2016). Although our blank control group underwent the same extraction method on unsampled QMA filters, comparing the RNA expression levels between the treatment groups and the blank control group allows us to remove the effects caused by methanol itself. However, the blank control group cannot eliminate the interactive effects between methanol and other bioactive components, as well as the extraction bias caused by the polarity of the solvent. Since we don't know which bioactive chemical (s) substance is responsible for the observed effects, further research is required for clarification. In our research, we focused only on methanol extracted compounds because they are most likely to be enriched during the formation of SSA (Cravigan et al., 2020; Van Acker et al., 2021b) and are expected to be present in the fine, respirable fraction of SSA that more likely to be captured by human respiratory system (Van Acker et al., 2021a).

While our results provide insights into the molecular mechanisms underlying the exposure effects of SSA, it is also important to note that the effects on gene level do not necessarily translate to effects on cell level or above. The transition from in-vitro to in-vivo is a complex process that requires additional studies to determine whether the observed gene-level responses translate into meaningful physiological effects.

Cell viability assays were not conducted at the end of the treatment because the primary focus was the transcriptional profiling of SSA rather than assessing cytotoxicity or cell health per se. While cell viability can be a valuable parameter in certain experimental contexts, RNA integrity was prioritized here as a more direct and relevant measure of experimental validity. All RNA samples exhibited high-quality RIN scores with no degradation, indicating that the cells were in a suitable state for transcriptomic analysis. The experimental design incorporated appropriate controls and biological replicates to address potential

variability. Additionally, previous studies on individual compounds present in SSA included comprehensive analyses of cell viability for each compound. Previous study (Van Acker et al., 2020) reported effect concentrations (EC_{10}) ranging from approximately 1 $\mu\text{g/L}$ up to 100 $\mu\text{g/L}$ or even more than 1000 $\mu\text{g/L}$, depending on the compound. Furthermore, prior exposure data for some of these compounds showed their concentrations in the range of picograms per cubic meter of air (Van Acker et al., 2021b), which are much lower than any reported effect data. Taken together, these findings suggest that limited effects on cell viability were expected. RNA integrity was used as a measure to ensure high-quality extractions with minimal impact on cell viability. Nevertheless, potential mixture effects affecting cell viability cannot be entirely excluded. Therefore, no conclusions regarding the potential impact on cell viability are made, as this was not tested.

Another limit of our study is the time interval for our SSA sample collection, which is approximately one month in between each sampling date. This is insufficient to track the rapid changes in planktonic communities, especially during spring and summer algal blooming. If future studies can collect samples on a weekly basis during the spring algal bloom period and a few (e.g. four) days after, they may draw more compelling conclusions.

4. Conclusions

The research presented in this article is, to the best of our knowledge, the first study to expose human lung epithelial (BEAS-2B) cells to natural SSAs at environmentally relevant concentrations and investigate their molecular-level health effects. In this research we have the following findings: 1. Under normal environmental conditions, exposure to natural SSA can influence the RNA expression of BEAS-2B cells, including the inhibition of the PI3k/Akt/mTOR/NF- κ B signaling pathway and the activation of the AMPK signaling pathway. 2. The bioactive components of SSA responsible for the effects may be

influenced by algal activity, with the most pronounced effects observed during spring algal blooms. 3. Air samples collected during no wind conditions may also contain these bioactive components, potentially originating from previously generated SSA or from distant sources.

5. Methods

5.1. Aerosol sampling and chemical extractions.

The sea spray aerosol sampling campaign was conducted from 1st April 2019 until 14th April 2020 on the beach of Ostend, Belgium (51.239°N, 2.931°E). The aerosol sampling method followed Van Acker et al. (Van Acker et al., 2021a) except the collected air volume of each aerosol sample was fixed to approximately 2000 L instead of collecting for a fixed period. Briefly, for each aerosol sample, an acid-washed (1% HNO₃) QM-A quartz microfiber filter (Whatman, Ø 47mm, air retention 99.95% for 0.3 µm particle) was placed in a cleaned stainless-steel in-line filter holder (Pall Corporation) and connected to a constant airflow sample pump (Leland Legacy, SKC Inc.). During the sampling, the filter holders were placed on a tripod within 15m of the waterline and the tripod was moved frequently to maintain the same distance to waterline. (following the tide). The height of filter holders was around 1.6 m and the front openings of the filter holders were facing the direction from which the wind was blowing. The pumps were set at an airflow rate of 10~12 L min⁻¹ until 2000 L of air containing SSA was sampled on the filters. In-situ average and gust wind speeds were measured using a wind meter before and after the sampling. The sampling days were chosen based on the weather conditions. We only took SSA samples when there was no precipitation and the wind was blowing from the sea. According to literature, SSA production occurs at speeds above 4 m/s (Gantt and Meskhidze, 2013; Lewis and Schwartz, 2004; O'Dowd and De Leeuw, 2007). In this campaign, 12 SSA samples were collected, three of them were collected at wind speeds lower than 3 m/s when the wind was blowing from inland. These samples were considered as ambient coastal aerosol samples (AMB). For each sampling day, an

extra filter was taken through the exact same procedure as the aerosol sample filters except it was not connected to the pump and was used as a blank control. In addition to the SSA and AMB aerosol samples, three city aerosol samples (CTY) were taken with the same sampling system in Ghent University Campus Coupure, Ghent Belgium (51.053°N, 3.707°E) on the roof of a 12 m high building. This city's location is around 40 km away from the coastline. The detailed time and environmental information are listed in the **SI.1**.

The filter samples were first cut into two pieces with scissors rinsed with 1% HNO₃: 1/4 and 3/4 separately. (I) The 1/4 parts were used for Inductively Coupled Plasma - Optical Emission Spectroscopy (ICP-OES, Thermo Scientific iCAP 7000 series) analysis to measure the concentrations of Na⁺ and Mg⁺ in the SSA samples and Na⁺ was used as a proxy quantifier of the SSA content in a sample (Lewis and Schwartz, 2004) and Na⁺/Mg⁺ ratio was used to confirm the source of SSA. The filter parts were placed in falcon tubes with 5 mL 1% HNO₃, vortexed, and sonicated to elute Na⁺ and Mg⁺ from the filter. The elution step was repeated three times, resulting in a 15 mL extract solution for each sample. Extracts were then filtered over 0.45 µm (Supor filter, Cytiva) to remove filter debris generated during elution and stored at 4°C until analysis. (II) The 3/4 parts were extracted for the exposure experiment: The extraction method used was based on the dedicated extraction protocols of Ullah (Ullah et al., 2015) and optimized by Van Acker (Van Acker et al., 2021b). Methanol was used as eluent, and the elution step was repeated twice (10 mL and 3 mL), resulting in a 13 mL extract for each sample. Afterward, extracts were filtered by 0.2 µm polytetrafluoroethylene (PTFE) syringe filters (Phenex) and dried under a gentle N₂ stream at 40°C until around 1 mL extracts were obtained and reconstituted in 1.628 mL methanol. The filter extracts were subsequently stored at -20°C until they were used as treatments in the *vitro* experiment.

5.2. Cell culture and exposure experiment.

In the in vitro experiment, a human bronchial epithelial cell line (BEAS-2B) obtained from the Laboratory of Experimental Cancer Research at Ghent University was utilized. The use of this cell line was conducted in compliance with ethical standards, having received approval from the ethical commission under the reference number EC/2019/1856. Additionally, the cell line is registered with the Federal Agency for Medicines and Health Products of Belgium (FAGG) biobank with the identification number BB200014. Cells were cultured in Dulbecco's modified eagle medium (DMEM), supplemented with 10% fetal bovine serum (FBS), 100 units mL⁻¹ penicillin, and 100 µg mL⁻¹ streptomycin. Cell cultures were grown in a humidified incubator at 37°C, 5% CO₂, and >95% relative humidity. Confluent cultures were sub-cultured, using 0.5% Trypsin–EDTA, and split two times per week in a 1:6 ratio. Passage numbers of the cell culture were limited to less than 50. Confluent cell cultures were trypsinized and seeded in six-well plates at a density of 320,000 cells well⁻¹ with 3 mL fresh DMEM, followed by 10 hours incubation at the same conditions as mentioned above.

The cells were exposed to one of the treatments listed in **Table 1**. and then incubated for another 48 hours prior to RNA extraction, the time length of exposure was the maximum duration without splitting the cells. To simulate environmentally relevant exposure conditions in our in vitro experiments, we calculated the doses of aerosol treatments by proportionally scaling the concentration of aerosol extracts based on the ratio of the total human alveolar surface area to the alveolar surface area exposed in our cell culture model under different environmental exposure conditions including inhalation volume and increased aerosolization. This method was also used by Asselman (Asselman et al., 2019). The doses of exposure treatments were calculated based on the human Alveolar surface as **Formula (1)** using the principle in toxicology of geometric progression with a constant factor of 2.

$$M = \frac{(T_s \times DF) / A_e}{T_e / A_h} \quad (1)$$

Calculation of dilution factors used for aerosol extracts.

M: the multiplicative factor by which the environmental condition is multiplied to achieve the desired exposure dose, in this research we have $M = 10, 20, 40, \text{ or } 80$. **T_e**: the total duration for which the cells were exposed to the aerosol extracts. **T_s**: the duration for which the aerosol sample was collected. **A_e**: the calculated alveolar surface area exposed, in this research $A_e = 9.6 \text{ cm}^2$. **A_h**: the total human alveolar surface area which is $4 \times 10^5 \text{ cm}^2$. **DF**: the dilution factor, in this research we used 289, 145, 72, or 36 for different M .

A known mTOR kinase inhibitor Torkinib or PP242 (MedChemExpress) was used as the positive control (POS) at concentration of $0.1 \mu\text{M}$. The negative control (NEG) also contained 2% methanol to exclude a solvent effect as all other treatments were extracted, diluted, or dissolved in methanol.

In the in vitro experiment, we exposed BEAS-2B cells to (1) SSA samples (SSA) (2) ambient coast aerosol samples (AMB, where there was no wind during the sampling), (3) city aerosol samples (CTY), (4) positive controls (POS), and (5) negative controls (NEG). For each aerosol/control sample, several different concentrations were used for exposure, resulting in 50 treatments (aerosol sample - concentration combination) with 17 different aerosol samples or chemicals. Every treatment consists of four replicas (with two exceptions noted in the remarks). The number of cell line samples in this research is 208. A summary of treatment groups and sample information is listed in **Table 1**. A detailed cell line sample list is attached in the **SI.2**.

Table 1 Summary of treatment groups and sample information

Exposure type	Concentration levels	Num. of aerosol samples/chemicals	Num. of treatments	celline samples	Num. of samples
Sea spray aerosol (SSA)	x10 x20 x40 x80	9	36		144
Ambient coast aerosol (AMB)	x10 x80	3	6		24
City aerosol (CTY)	x10 x80	3	6		24
Negative control (NEG)	x1	1	1		4
Positive control (POS)	x1	1	1		12
Total		17	50		208

Note: Concentration levels refer to the calculated exposure doses, expressed as multiplicative factors ($M = 10, 20, 40, \text{ or } 80$), which simulate environmentally relevant exposure conditions *in vitro*. These factors were derived by proportionally scaling aerosol extract concentrations based on the ratio of total human alveolar surface area to the alveolar surface area exposed in the cell culture model. The dose calculation is detailed in the methods section under "Formula (1)". For the positive control (POS), the concentration represents the use of $0.1 \mu\text{M}$ mTOR kinase inhibitor PP242. All treatments were performed in four replicates unless specified otherwise.

5.3. RNA extraction and sequencing.

RNA was extracted using the Qiagen RNEasy kit following the manufacturer's instructions including DNase digestion. Quality and concentration of the input total RNA was checked with 'Quant-it ribogreen RNA assay' (Life Technologies, Grand Island, NY, USA) and the RNA 6000 nano chip (Agilent Technologies, Santa Clara, CA, USA), respectively. The RIN values were 9.01 ± 0.85 (average $\pm$ standard deviation). Subsequently, 48 ng of RNA was used to perform an Illumina sequencing library preparation using the QuantSeq 3' mRNA-Seq Library Prep FWD Kit (Lexogen, Vienna, Austria) using sample-specific dual indices for library multiplexing according to the manufacturer's protocol using 14 enrichment PCR cycles. Briefly, this protocol generates 3' end-focused mRNA libraries using oligo(dT) priming of naturally polyadenylated transcripts (no separate rRNA depletion step was applied). Libraries were quantified by qPCR, according to Illumina's protocol 'Sequencing Library qPCR Quantification protocol guide'. A High sensitivity DNA chip (Agilent Technologies, Santa Clara, CA, US) was used to control the library's size distribution and quality. Based on Illumina sequencing guidelines, the libraries are first normalized to the same concentration, then equal volumes are pooled to yield a pool of the same concentration of each of the original normalized libraries. Libraries were then sequenced on an Illumina HiSeq sequencer together with 2% PhiX Sequencing Control library, generating 50 bp single-end reads.

5.4. Laser-assisted ambient ionization mass spectrometry (LA-REIMS) analyses.

LA-REIMS was used for the chemical analysis of our aerosol samples. This untargeted approach was chosen given the undefined nature of potential compounds of interest, providing a comprehensive fingerprinting of the samples, which may be useful in future studies. Initial analyses were performed by ProDigest BV (Belgium) using established methodology (Van Meulebroek et al., 2020). The LA-REIMS

platform utilized a MID infrared laser system (OpoletteTM HE2940, OPOTEK, LLC, USA) that consisted of a Q-switched Nd:YAG laser pumping an Optical Parametric Oscillator (OPO). The transmission of the laser energy into the sample was achieved through free space optics, which included a series of metallic-coated mirrors and a Plano-convex lens. The process of laser ablation, performed for 3 seconds per sample, is initiated based on the laser-emitted infrared wavelength regime that excites the most intense vibrational band of the water molecules contained by the sample. This leads to matrix-assisted desorption and ionization of intact biomolecules. The produced aerosol is then transferred to the REIMS platform. Mass analysis was conducted using a Xevo G2-XS Quadrupole Time-of-Flight (QToF) mass spectrometer, operated in negative ionization mode with an m/z scan range of 50–1200 Da, obtaining relative intensities for each sample. An external standard (leucine-enkephalin, 0.25 ng/ μ L) was used for lock mass correction and signal intensity drift correction. For the purpose of distinguishing true signals from noise, a noise value was established using the mean relative intensities of blank samples augmented by 0.001. Features were then distinguished based on signal-to-noise (S/N) ratios, employing thresholds of both 3 and 10, thereby representing the chemical diversity of each sample.

5.5. Data analyses.

Poly-A and Illumina Truseq RNA adapters are trimmed according to the manufacturer (Lexogen QuantSeq 3'mRNA-Seq Data Analysis) using bbdutck.sh in BBMap package version 35.85 (Bushnell et al., 2017). Poly(A) trimming is performed because oligo(dT)-primed 3' mRNA library construction can yield reads containing poly(A) stretches originating from polyadenylated transcripts. We used STAR version 2.7.9a (Dobin et al., 2013) to align the sequences, using the human Genome Reference Consortium build 38 (GRCh38) (Cunningham et al., 2015) as alignment reference genome. We used the same parameters as the manufacturer suggested, except we did not specify the SAM

attributes (outSAMattributes), so the output BAM files have the default attributes (NH HI AS nM). We used StringTie version 2.1.6 (Pertea et al., 2015) for the assembly of the aligned read sequences. In this process, we supplied only the reference genome, specifically the gene annotation file (Homo_sapiens.GRCh38.89.gtf) obtained from Ensembl (Martin et al., 2023), along with the designated file paths for input and output. All other parameters were left at their default settings.

We use R version 4.1.2 (R Core Team, 2014) as the general software platform for statistical analyses and visualizations. The ggplot2 version 3.3.5 (Wickham, 2016) was used as the main tool for generating figures. Differential gene expression analysis between treatment groups (SSA, AMB, CTY, POS) and negative control (NEG) was performed using the DESeq function in the package DESeq2 version 1.32.0 (Love et al., 2014) with built-in batch effect calibration and default parameters. Count matrices were used as input for DESeq2, and statistical testing was therefore based on DESeq2's default internal normalization and model fitting. For unsupervised hierarchical clustering, count data were TMM-normalized, and genes were ranked by median absolute deviation (MAD) to prioritize the most variable features. The False Discovery Rate (FDR) and Log2 Fold Change (Log2FC) were used to screen for DEGs, thresholds were determined according to the applications and were mentioned in related figures. Filters for protein-coding genes were generated based on the same reference genome GRCh38 as for alignments. The Euclidean distances between each sample's gene expression were calculated and subsequently clustered using the centroid algorithm for unsupervised hierarchical clustering. We used the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata et al., 1999) database for annotating the gene functions and signaling pathways. Enrichment analyses were performed by clusterProfiler version 4.0.5 (Yu et al., 2012) with P-value cutoff at 0.01. The data for plotting the dendrogram was computed with hierarchical cluster analysis using R's built-in *hclust* function with centroid algorithm. Analysis of variance (ANOVA) was conducted on phytoplankton densities and

chemical diversity with the *aov* function. Homogeneity of variance was checked with the *vif* function from the *car* package (version 3.1-2) and normality was assessed with Shapiro-Wilk tests using *shapiro.test* function. For chemical diversity, post-hoc comparisons between different treatments were made using the *TukeyHSD* function. A generalized linear model was fitted using *glm* function to determine the sources of variation in the number of DEGs. Aerosol backward trajectories were calculated using the SplitR package in R with the HYSPLIT split model to identify potential sources of aerosol particles impacting the sampling site.

Supplementary Information

SI.1. SSA sampling details and environmental information

#	Aerosol ID	type	Sampling Date	naCon (ug/filter)	airTemp (°C)	phyC
1	0	NEG	NA	NA	NA	NA
2	1	POS	NA	NA	NA	NA
3	91	SSA	2019-04-01	67.04	8.39	5.34
4	110	SSA	2019-05-14	3.38	10.57	10.45
5	119	AMB	2019-05-24	4.26	13.36	10.57
6	125	CTY	2019-06-17	0.00	16.12	7.26
7	134	SSA	2019-07-15	36.62	15.36	6.95
8	143	SSA	2019-08-13	40.81	14.74	5.78
9	149	CTY	2019-09-30	0.00	15.02	3.21
10	155	SSA	2019-10-02	79.65	12.40	3.67
11	162	AMB	2019-10-07	1.71	12.04	4.54
12	191	SSA	2019-11-18	96.23	6.20	3.56
13	206	SSA	2020-01-10	59.66	6.80	2.65
14	216	SSA	2020-02-19	44.59	6.54	5.08
15	223	AMB	2020-03-04	2.54	5.76	4.93
16	230	CTY	2020-03-10	0.00	10.65	5.71
17	235	SSA	2020-04-14	NA	6.78	10.34

SI.2. List of cell line samples

#	Aerosol Sample ID	Cellline Sample ID	Treatment Group	Concentration Level
1	119	BEAS_AMB_119_10_1	AMB	x10
2	119	BEAS_AMB_119_10_2	AMB	x10
3	119	BEAS_AMB_119_10_3	AMB	x10
4	119	BEAS_AMB_119_10_4	AMB	x10
5	119	BEAS_AMB_119_80_1	AMB	x80
6	119	BEAS_AMB_119_80_2	AMB	x80
7	119	BEAS_AMB_119_80_3	AMB	x80
8	119	BEAS_AMB_119_80_4	AMB	x80
9	162	BEAS_AMB_162_10_1	AMB	x10
10	162	BEAS_AMB_162_10_2	AMB	x10
11	162	BEAS_AMB_162_10_3	AMB	x10
12	162	BEAS_AMB_162_10_4	AMB	x10
13	162	BEAS_AMB_162_80_1	AMB	x80
14	162	BEAS_AMB_162_80_2	AMB	x80
15	162	BEAS_AMB_162_80_3	AMB	x80
16	162	BEAS_AMB_162_80_4	AMB	x80
17	223	BEAS_AMB_223_10_1	AMB	x10
18	223	BEAS_AMB_223_10_2	AMB	x10
19	223	BEAS_AMB_223_10_3	AMB	x10
20	223	BEAS_AMB_223_10_4	AMB	x10
21	223	BEAS_AMB_223_80_1	AMB	x80
22	223	BEAS_AMB_223_80_2	AMB	x80
23	223	BEAS_AMB_223_80_3	AMB	x80
24	223	BEAS_AMB_223_80_4	AMB	x80
25	125	BEAS_CTY_125_10_1	CTY	x10
26	125	BEAS_CTY_125_10_2	CTY	x10
27	125	BEAS_CTY_125_10_3	CTY	x10
28	125	BEAS_CTY_125_10_4	CTY	x10
29	125	BEAS_CTY_125_80_1	CTY	x80
30	125	BEAS_CTY_125_80_2	CTY	x80

#	Aerosol Sample ID	Cellline Sample ID	Treatment Group	Concentration Level
31	125	BEAS_CTY_125_80_3	CTY	x80
32	125	BEAS_CTY_125_80_4	CTY	x80
33	149	BEAS_CTY_149_10_1	CTY	x10
34	149	BEAS_CTY_149_10_2	CTY	x10
35	149	BEAS_CTY_149_10_3	CTY	x10
36	149	BEAS_CTY_149_10_4	CTY	x10
37	149	BEAS_CTY_149_80_1	CTY	x80
38	149	BEAS_CTY_149_80_2	CTY	x80
39	149	BEAS_CTY_149_80_3	CTY	x80
40	149	BEAS_CTY_149_80_4	CTY	x80
41	230	BEAS_CTY_230_10_1	CTY	x10
42	230	BEAS_CTY_230_10_2	CTY	x10
43	230	BEAS_CTY_230_10_3	CTY	x10
44	230	BEAS_CTY_230_10_4	CTY	x10
45	230	BEAS_CTY_230_80_1	CTY	x80
46	230	BEAS_CTY_230_80_2	CTY	x80
47	230	BEAS_CTY_230_80_3	CTY	x80
48	230	BEAS_CTY_230_80_4	CTY	x80
49	0	BEAS_NEG_000_00_1	NEG	x0
50	0	BEAS_NEG_000_00_2	NEG	x0
51	0	BEAS_NEG_000_00_3	NEG	x0
52	0	BEAS_NEG_000_00_4	NEG	x0
53	1	BEAS_POS_001_01_1	POS	x1
54	1	BEAS_POS_001_01_2	POS	x1
55	1	BEAS_POS_001_01_3	POS	x1
56	1	BEAS_POS_001_01_4	POS	x1
57	1	BEAS_POS_001_03_1	POS	x1
58	1	BEAS_POS_001_03_2	POS	x1
59	1	BEAS_POS_001_03_3	POS	x1
60	1	BEAS_POS_001_03_4	POS	x1
61	1	BEAS_POS_001_03_5	POS	x1

#	Aerosol Sample ID	Cellline Sample ID	Treatment Group	Concentration Level
62	1	BEAS_POS_001_03_6	POS	x1
63	1	BEAS_POS_001_03_7	POS	x1
64	1	BEAS_POS_001_03_8	POS	x1
65	91	BEAS_SSA_091_10_1	SSA	x10
66	91	BEAS_SSA_091_10_2	SSA	x10
67	91	BEAS_SSA_091_10_3	SSA	x10
68	91	BEAS_SSA_091_10_4	SSA	x10
69	91	BEAS_SSA_091_20_1	SSA	x20
70	91	BEAS_SSA_091_20_2	SSA	x20
71	91	BEAS_SSA_091_20_3	SSA	x20
72	91	BEAS_SSA_091_20_4	SSA	x20
73	91	BEAS_SSA_091_40_1	SSA	x40
74	91	BEAS_SSA_091_40_2	SSA	x40
75	91	BEAS_SSA_091_40_3	SSA	x40
76	91	BEAS_SSA_091_40_4	SSA	x40
77	91	BEAS_SSA_091_80_1	SSA	x80
78	91	BEAS_SSA_091_80_2	SSA	x80
79	91	BEAS_SSA_091_80_3	SSA	x80
80	91	BEAS_SSA_091_80_4	SSA	x80
81	110	BEAS_SSA_110_10_1	SSA	x10
82	110	BEAS_SSA_110_10_2	SSA	x10
83	110	BEAS_SSA_110_10_3	SSA	x10
84	110	BEAS_SSA_110_10_4	SSA	x10
85	110	BEAS_SSA_110_20_1	SSA	x20
86	110	BEAS_SSA_110_20_2	SSA	x20
87	110	BEAS_SSA_110_20_3	SSA	x20
88	110	BEAS_SSA_110_20_4	SSA	x20
89	110	BEAS_SSA_110_40_1	SSA	x40
90	110	BEAS_SSA_110_40_2	SSA	x40
91	110	BEAS_SSA_110_40_3	SSA	x40
92	110	BEAS_SSA_110_40_4	SSA	x40

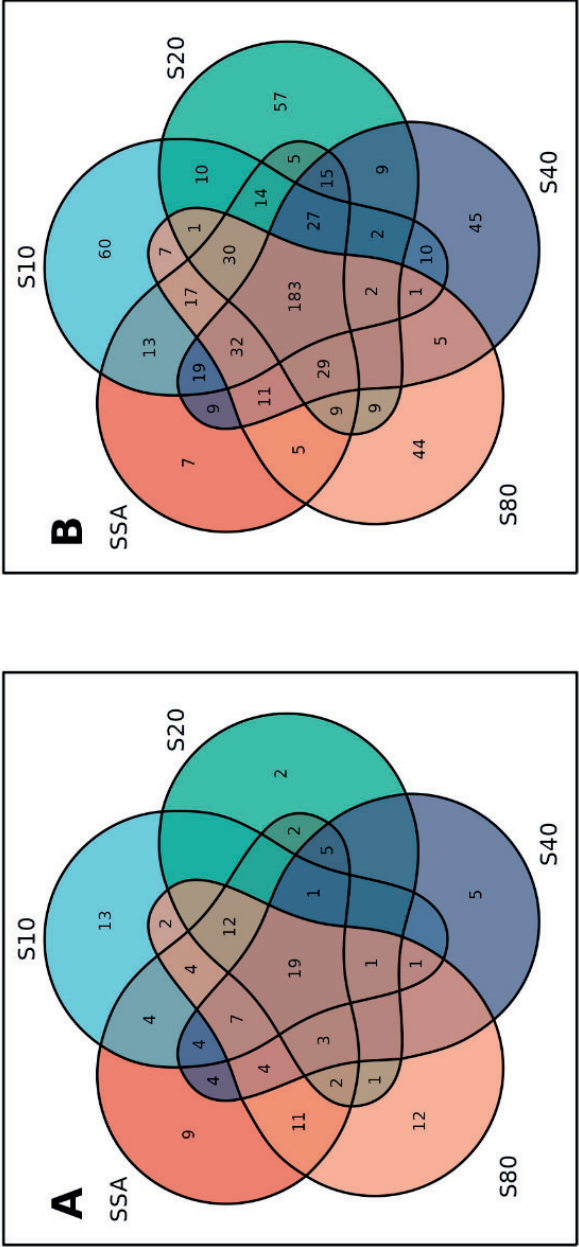
#	Aerosol Sample ID	Cellline Sample ID	Treatment Group	Concentration Level
93	110	BEAS_SSA_110_80_1	SSA	x80
94	110	BEAS_SSA_110_80_2	SSA	x80
95	110	BEAS_SSA_110_80_3	SSA	x80
96	110	BEAS_SSA_110_80_4	SSA	x80
97	134	BEAS_SSA_134_10_1	SSA	x10
98	134	BEAS_SSA_134_10_2	SSA	x10
99	134	BEAS_SSA_134_10_3	SSA	x10
100	134	BEAS_SSA_134_10_4	SSA	x10
101	134	BEAS_SSA_134_20_1	SSA	x20
102	134	BEAS_SSA_134_20_2	SSA	x20
103	134	BEAS_SSA_134_20_3	SSA	x20
104	134	BEAS_SSA_134_20_4	SSA	x20
105	134	BEAS_SSA_134_40_1	SSA	x40
106	134	BEAS_SSA_134_40_2	SSA	x40
107	134	BEAS_SSA_134_40_3	SSA	x40
108	134	BEAS_SSA_134_40_4	SSA	x40
109	134	BEAS_SSA_134_80_1	SSA	x80
110	134	BEAS_SSA_134_80_2	SSA	x80
111	134	BEAS_SSA_134_80_3	SSA	x80
112	134	BEAS_SSA_134_80_4	SSA	x80
113	143	BEAS_SSA_143_10_1	SSA	x10
114	143	BEAS_SSA_143_10_2	SSA	x10
115	143	BEAS_SSA_143_10_3	SSA	x10
116	143	BEAS_SSA_143_10_4	SSA	x10
117	143	BEAS_SSA_143_20_1	SSA	x20
118	143	BEAS_SSA_143_20_2	SSA	x20
119	143	BEAS_SSA_143_20_3	SSA	x20
120	143	BEAS_SSA_143_20_4	SSA	x20
121	143	BEAS_SSA_143_40_1	SSA	x40
122	143	BEAS_SSA_143_40_2	SSA	x40
123	143	BEAS_SSA_143_40_3	SSA	x40

#	Aerosol Sample ID	Cellline Sample ID	Treatment Group	Concentration Level
124	143	BEAS_SSA_143_40_4	SSA	x40
125	143	BEAS_SSA_143_80_1	SSA	x80
126	143	BEAS_SSA_143_80_2	SSA	x80
127	143	BEAS_SSA_143_80_3	SSA	x80
128	143	BEAS_SSA_143_80_4	SSA	x80
129	155	BEAS_SSA_155_10_1	SSA	x10
130	155	BEAS_SSA_155_10_2	SSA	x10
131	155	BEAS_SSA_155_10_3	SSA	x10
132	155	BEAS_SSA_155_10_4	SSA	x10
133	155	BEAS_SSA_155_20_1	SSA	x20
134	155	BEAS_SSA_155_20_2	SSA	x20
135	155	BEAS_SSA_155_20_3	SSA	x20
136	155	BEAS_SSA_155_20_4	SSA	x20
137	155	BEAS_SSA_155_40_1	SSA	x40
138	155	BEAS_SSA_155_40_2	SSA	x40
139	155	BEAS_SSA_155_40_3	SSA	x40
140	155	BEAS_SSA_155_40_4	SSA	x40
141	155	BEAS_SSA_155_80_1	SSA	x80
142	155	BEAS_SSA_155_80_2	SSA	x80
143	155	BEAS_SSA_155_80_3	SSA	x80
144	155	BEAS_SSA_155_80_4	SSA	x80
145	191	BEAS_SSA_191_10_1	SSA	x10
146	191	BEAS_SSA_191_10_2	SSA	x10
147	191	BEAS_SSA_191_10_3	SSA	x10
148	191	BEAS_SSA_191_10_4	SSA	x10
149	191	BEAS_SSA_191_20_1	SSA	x20
150	191	BEAS_SSA_191_20_2	SSA	x20
151	191	BEAS_SSA_191_20_3	SSA	x20
152	191	BEAS_SSA_191_20_4	SSA	x20
153	191	BEAS_SSA_191_40_1	SSA	x40
154	191	BEAS_SSA_191_40_2	SSA	x40

#	Aerosol Sample ID	Cellline Sample ID	Treatment Group	Concentration Level
155	191	BEAS_SSA_191_40_3	SSA	x40
156	191	BEAS_SSA_191_40_4	SSA	x40
157	191	BEAS_SSA_191_80_1	SSA	x80
158	191	BEAS_SSA_191_80_2	SSA	x80
159	191	BEAS_SSA_191_80_3	SSA	x80
160	191	BEAS_SSA_191_80_4	SSA	x80
161	206	BEAS_SSA_206_10_1	SSA	x10
162	206	BEAS_SSA_206_10_2	SSA	x10
163	206	BEAS_SSA_206_10_3	SSA	x10
164	206	BEAS_SSA_206_10_4	SSA	x10
165	206	BEAS_SSA_206_20_1	SSA	x20
166	206	BEAS_SSA_206_20_2	SSA	x20
167	206	BEAS_SSA_206_20_3	SSA	x20
168	206	BEAS_SSA_206_20_4	SSA	x20
169	206	BEAS_SSA_206_40_1	SSA	x40
170	206	BEAS_SSA_206_40_2	SSA	x40
171	206	BEAS_SSA_206_40_3	SSA	x40
172	206	BEAS_SSA_206_40_4	SSA	x40
173	206	BEAS_SSA_206_80_1	SSA	x80
174	206	BEAS_SSA_206_80_2	SSA	x80
175	206	BEAS_SSA_206_80_3	SSA	x80
176	206	BEAS_SSA_206_80_4	SSA	x80
177	216	BEAS_SSA_216_10_1	SSA	x10
178	216	BEAS_SSA_216_10_2	SSA	x10
179	216	BEAS_SSA_216_10_3	SSA	x10
180	216	BEAS_SSA_216_10_4	SSA	x10
181	216	BEAS_SSA_216_20_1	SSA	x20
182	216	BEAS_SSA_216_20_2	SSA	x20
183	216	BEAS_SSA_216_20_3	SSA	x20
184	216	BEAS_SSA_216_20_4	SSA	x20
185	216	BEAS_SSA_216_40_1	SSA	x40

#	Aerosol Sample ID	Cellline Sample ID	Treatment Group	Concentration Level
186	216	BEAS_SSA_216_40_2	SSA	x40
187	216	BEAS_SSA_216_40_3	SSA	x40
188	216	BEAS_SSA_216_40_4	SSA	x40
189	216	BEAS_SSA_216_80_1	SSA	x80
190	216	BEAS_SSA_216_80_2	SSA	x80
191	216	BEAS_SSA_216_80_3	SSA	x80
192	216	BEAS_SSA_216_80_4	SSA	x80
193	235	BEAS_SSA_235_10_1	SSA	x10
194	235	BEAS_SSA_235_10_2	SSA	x10
195	235	BEAS_SSA_235_10_3	SSA	x10
196	235	BEAS_SSA_235_10_4	SSA	x10
197	235	BEAS_SSA_235_20_1	SSA	x20
198	235	BEAS_SSA_235_20_2	SSA	x20
199	235	BEAS_SSA_235_20_3	SSA	x20
200	235	BEAS_SSA_235_20_4	SSA	x20
201	235	BEAS_SSA_235_40_1	SSA	x40
202	235	BEAS_SSA_235_40_2	SSA	x40
203	235	BEAS_SSA_235_40_3	SSA	x40
204	235	BEAS_SSA_235_40_4	SSA	x40
205	235	BEAS_SSA_235_80_1	SSA	x80
206	235	BEAS_SSA_235_80_2	SSA	x80
207	235	BEAS_SSA_235_80_3	SSA	x80
208	235	BEAS_SSA_235_80_4	SSA	x80

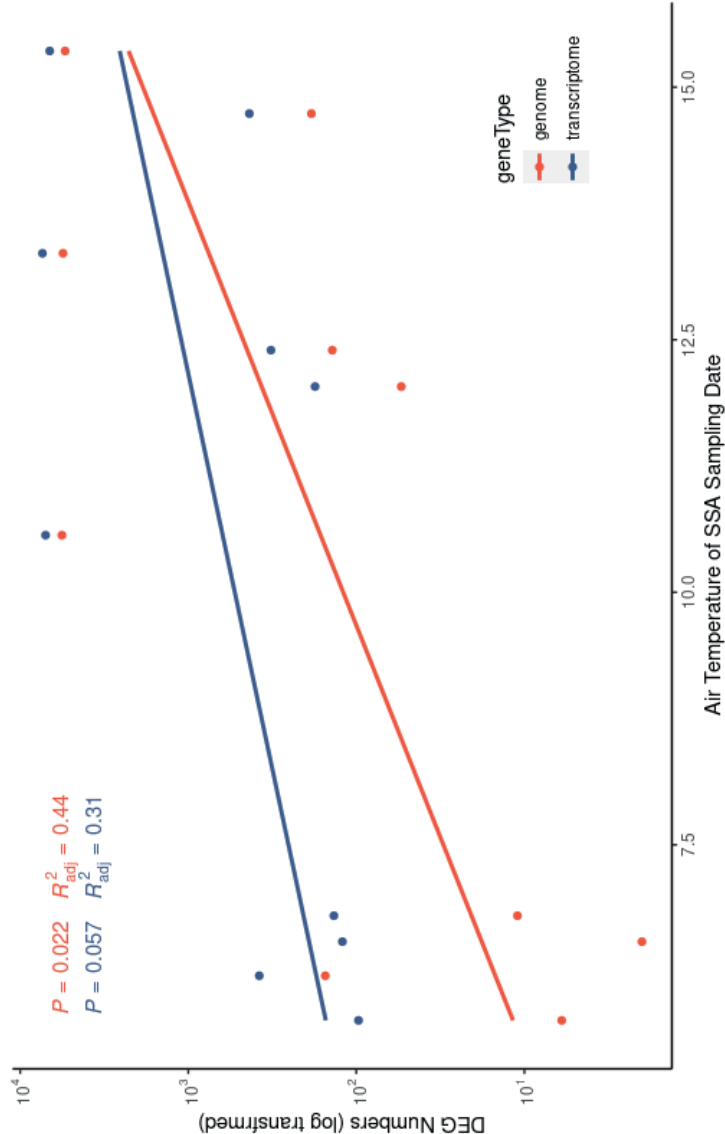
Sl.3. Differential expressed gene (DEG) numbers across SSA treatment groups showing no significant dose effect



This Venn diagram illustrates the unique and shared DEGs among different SSA treatment groups. The SSA group represents the entire set of samples exposed to sea spray aerosol. Within this, subgroups S10, S20, S40, and S80 correspond to samples exposed to SSA at concentrations of 10, 20, 40, and 80 units, respectively. The numbers in each section of the diagram indicate

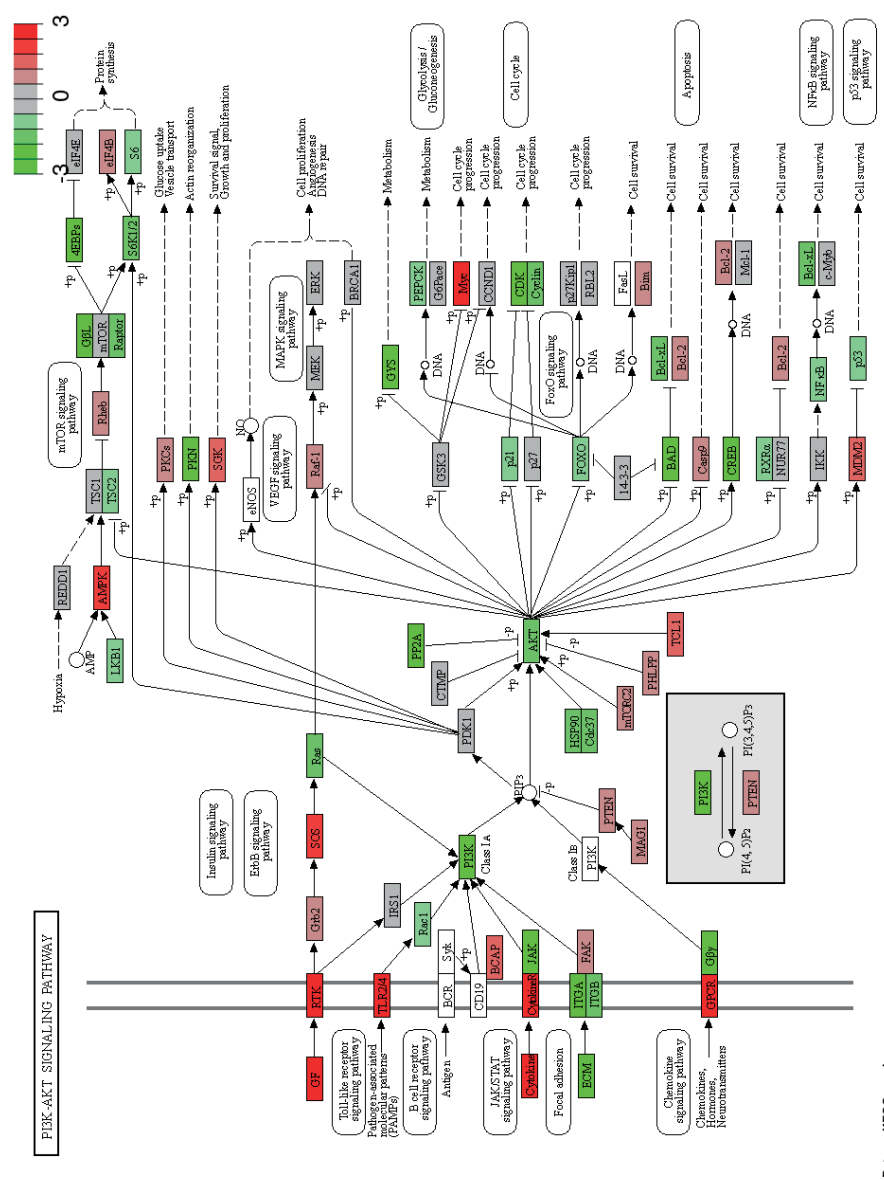
the count of DEGs unique to or shared between these groups, providing insight into the dose-dependent genetic response to SSA exposure. A. Genome B. Transcriptome.

Sl.4. Correlation between differential expressed gene (DEG) numbers and air temperature on SSA sampling days

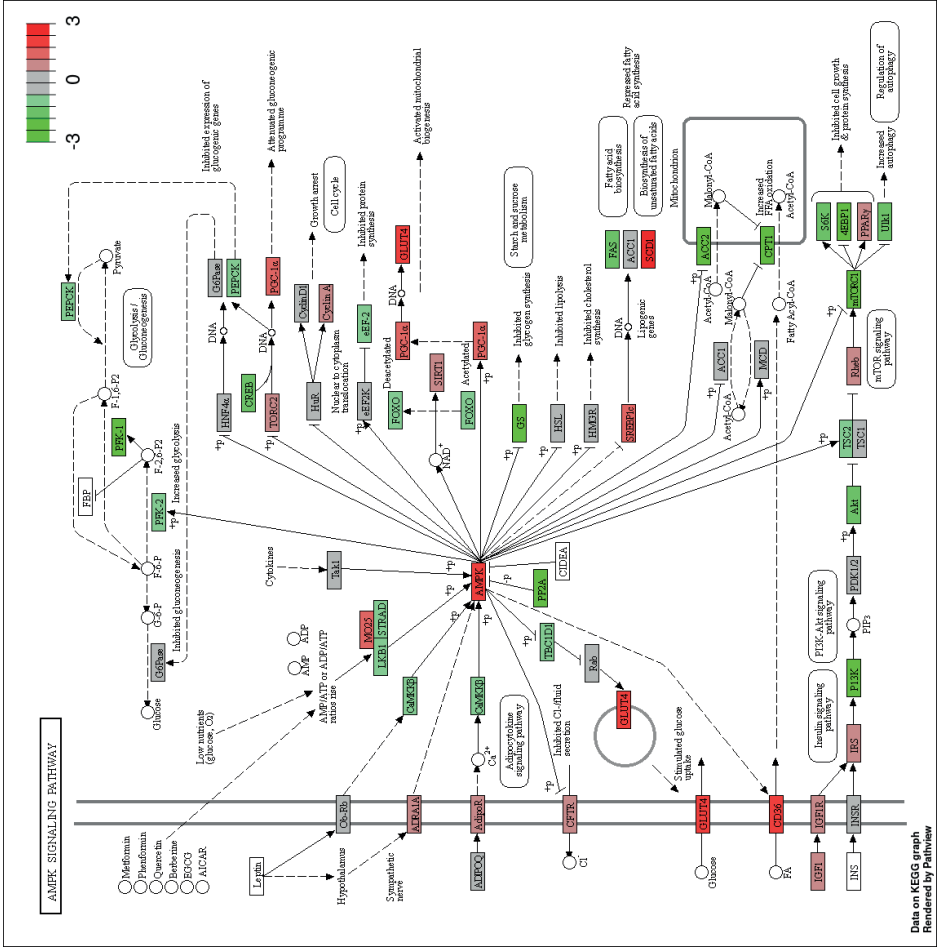


This scatter plot with fitted regression lines shows the relationship between the log-transformed number of DEGs and the air temperature at the time of SSA sampling. The red line represents DEGs identified in the genome, while the blue line

corresponds to DEGs in the transcriptome. The P-values and adjusted R-Squared values are indicated for each regression, suggesting a significant (for the genome) and a marginally non-significant (for the transcriptome) correlation with temperature. Points represent individual SSA samples.



Data on KEGG graph
Rendered by Pathview



Data on KEGG graph
Rendered by Pathview

SI.6. Tukey HSD test results for comparing the feature counts of aerosol samples between treatment groups

Group	diff	lwr	upr	p adj
NEG-AMB	-114.7	-462.2	232.8	0.778
CTY-AMB	-25.0	-413.5	363.5	0.998
SSA-AMB	300.2	-36.3	636.7	0.089
CTY-NEG	89.7	-211.3	390.6	0.826
SSA-NEG	414.9	185.0	644.7	0.001
SSA-CTY	325.2	37.0	613.3	0.025

Tukey Honest Significant Difference (HSD) Test results for feature counts across treatment groups: negative control (NEG), ambient air (AMB), city air (CTY), and SSA. The table details pairwise comparisons of feature counts at a 95% family-wise confidence level. 'diff' is the mean difference in feature counts between groups, 'lwr' and 'upr' are the confidence interval bounds, and 'p adj' shows the adjusted p-values for multiple testing. Notably, significant differences are found when comparing SSA with NEG and CTY, demonstrating the unique feature counts associated with SSA exposure.

Sl.7. The regression model was tested using 1 to 10 days as “Lag Phase” in order to find a suitable time span

Lag phase (day)	Gene		Transcript		
	P-value	Adj R-sqr	P-value	Adj R-sqr	
1	0.6	-0.08	0.51	-0.06	p ≤ 0.1
2	0.09	0.23	0.07	0.28	p ≤ 0.05
3	0.08	0.25	0.05	0.33	
4	0.09	0.23	0.04	0.35	
5	0.12	0.18	0.05	0.32	
6	0.31	0.02	0.15	0.14	
7	0.98	-0.12	0.69	-0.1	
8	0.46	-0.05	0.68	-0.1	
9	0.89	-0.12	0.71	-0.1	
10	0.58	-0.08	0.31	0.02	

The upper limit of 10 days was determined because the water residence time in this area is within 10 days¹.

Evaluation of Regression Model Lag Phases for Gene and Transcript-Level Data Over a 10-Day Period. This table displays the outcomes of testing the regression model across lag phases ranging from 1 to 10 days, intended to identify the optimal time span for assessing the environmental impact on gene and transcript levels. The 'Lag phase (day)' column specifies the day of the lag phase, while the 'Gene' and 'Transcript' columns present the p-values and adjusted R-squared values obtained for each respective day. Highlighted cells indicate statistical significance with green shading for p-values less than 0.05 and lighter green for those less than 0.1. The 10-day upper limit was

1

https://dredging.org/documents/ceda/html_page/msfd_art_8_initialassessment_be.pdf

chosen based on the local water residence time. Note that the model shows increasing adjusted R-squared values with longer lag phases for transcripts, suggesting a potential delay in transcriptional response to environmental changes.

SI.8. Regression analysis showing the lag phase of 4 days between the DEG numbers and the phytoplankton density change

Call:

```
lm(formula = log(num) ~ deltaPhyC, data = countDEG[which(countDEG$geneType ==  
"genome"), ])
```

Residuals:

Min	1Q	Median	3Q	Max
-4.7637	-0.7395	-0.1221	1.0168	4.4026

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.8499	0.9985	3.856	0.00484 **
deltaPhyC	-2.2143	1.1467	-1.931	0.08959 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.545 on 8 degrees of freedom

Multiple R-squared: 0.3179, Adjusted R-squared: 0.2326

F-statistic: 3.729 on 1 and 8 DF, p-value: 0.08959

Call:

```
lm(formula = log(num) ~ deltaPhyC, data = countDEG[which(countDEG$geneType ==  
"transcriptome"), ])
```

Residuals:

Min	1Q	Median	3Q	Max
-1.91432	-0.80662	-0.08678	0.19880	2.99057

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.5729	0.5612	9.931	8.94e-06 ***
deltaPhyC	-1.5670	0.6445	-2.431	0.0411 *

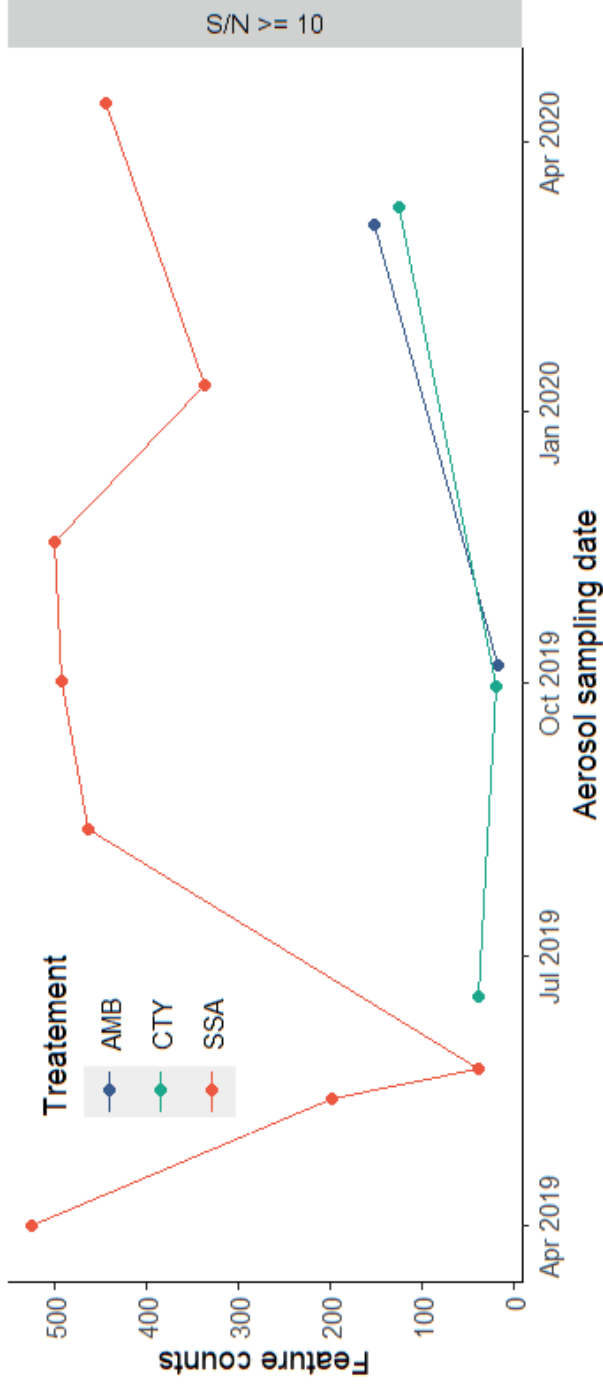
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.43 on 8 degrees of freedom

Multiple R-squared: 0.425, Adjusted R-squared: 0.3531

F-statistic: 5.912 on 1 and 8 DF, p-value: 0.04111

SI.9. Temporal variation of feature counts across different treatments.

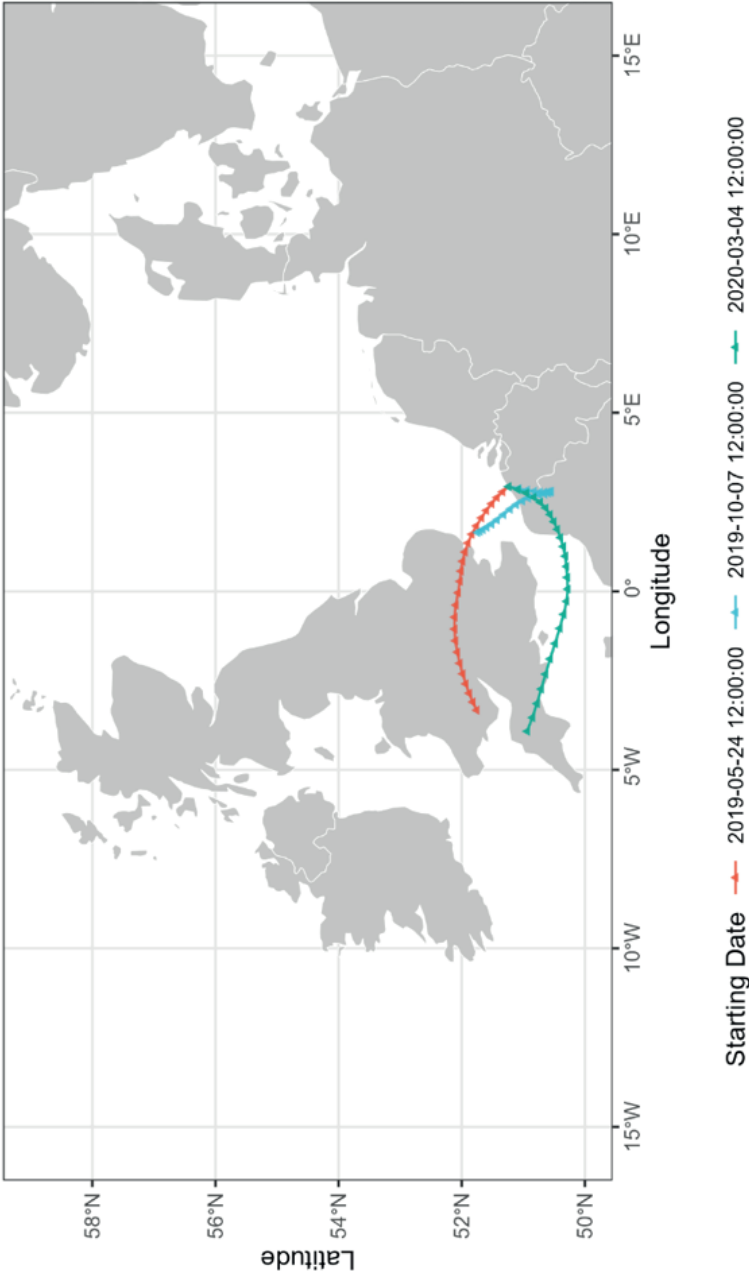


Temporal Variation in Feature Counts by Sample Type During the Sampling Period. The line graph illustrates the number of features counted in aerosol samples collected from April 2019 to April 2020, segmented by sample types: ambient air (AMB), city air (CTY), and sea spray aerosol (SSA). A notable decrease in feature counts for SSA is observed in May and June 2019, aligning with the period of spring bloom as detailed in the

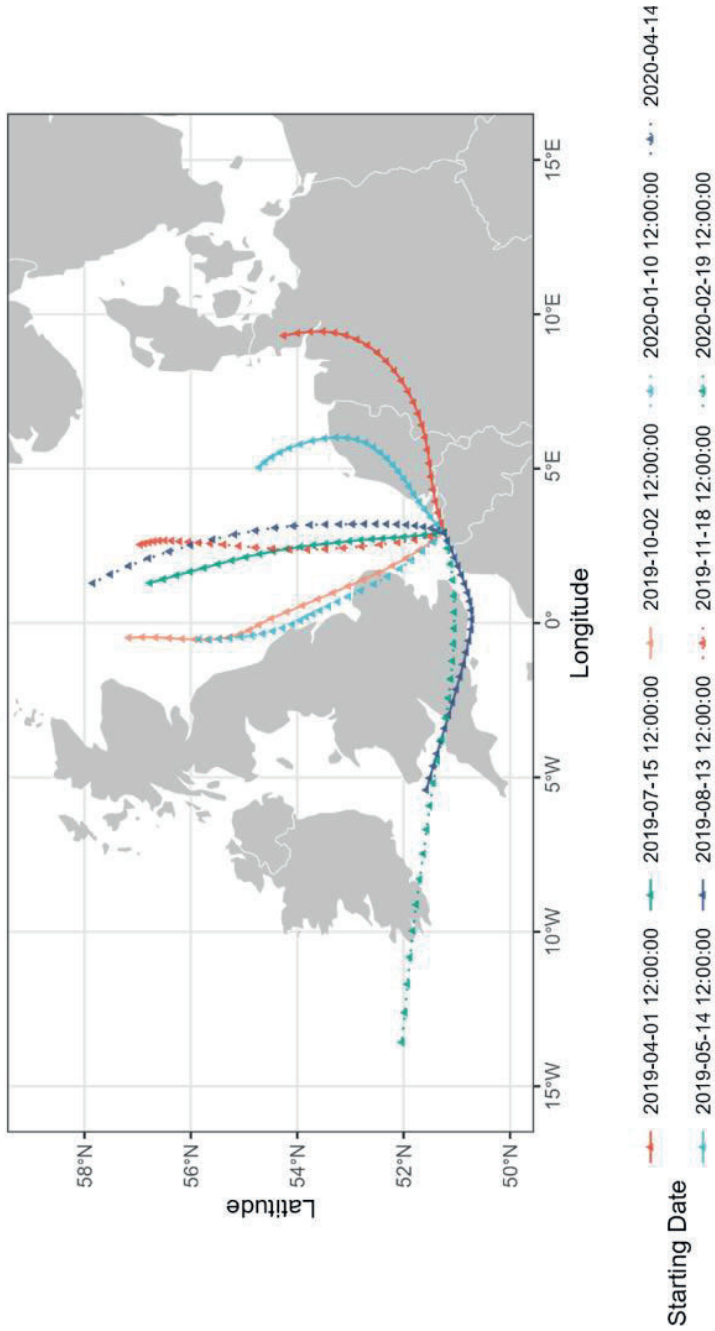
Results section. This temporal pattern suggests a significant influence of phytoplankton event on the chemical diversity of SSA samples.

SI.10. Backward trajectory of AMB and SSA samples

a 24 Hours Backward Trajectory: AMB



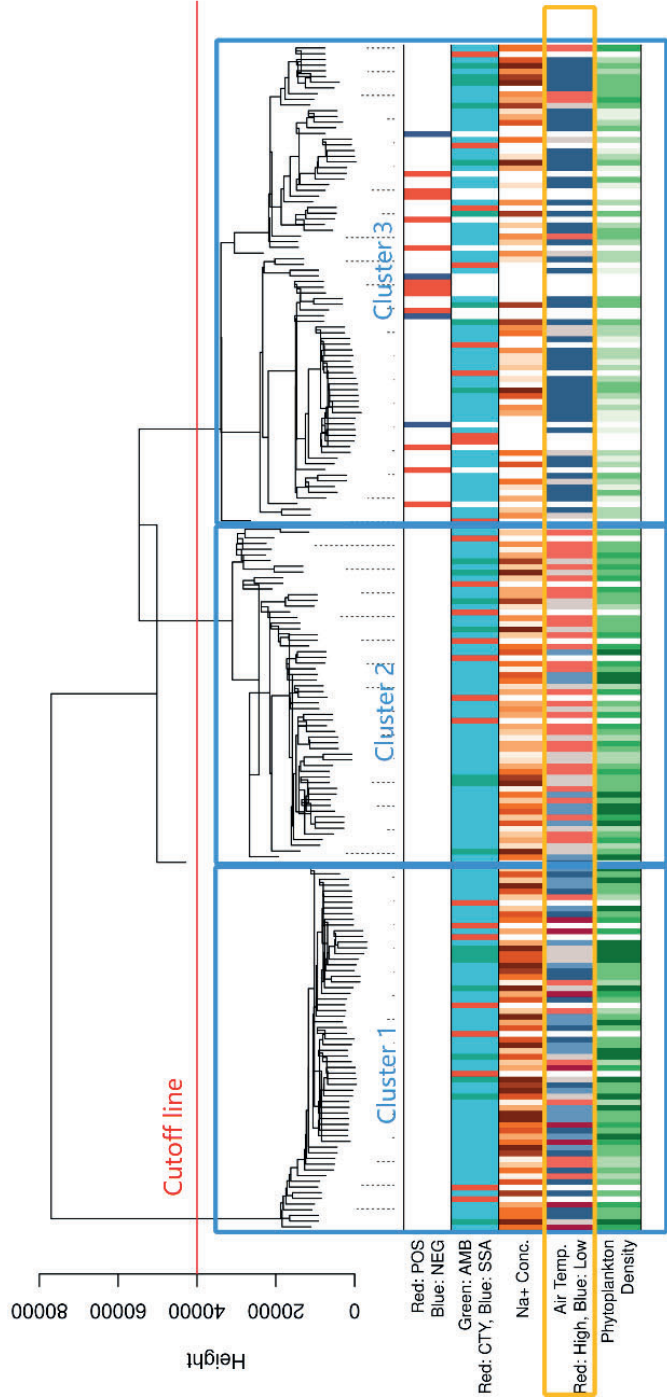
b 24 Hours Backward Trajectory: SSA



(a) 24-hour backward trajectories for AMB samples, and (b) 24-hour backward trajectories for SSA samples

calculated at 50 m AGL (Above Ground Level). Each triangle marker along the trajectories represents a 1-hour interval. The trajectories indicate that both ambient air (AMB) and sea spray aerosol (SSA) samples passed over seawater areas before reaching the sampling location. This suggests that the air masses influencing both sample types were affected by marine environments, highlighting the potential impact of seawater on the sample composition.

SL.11 Hierarchical clustering of gene expression with highlighted clusters and metadata (annotated version of Figure 1 a)



This figure annotates Figure 1.a by adding blue rectangles under the cutoff line (red line at height 40,000) to

indicate three main clusters, labeled as Cluster 1, Cluster 2, and Cluster 3 from left to right. An orange rectangle highlights the metadata color band corresponding to air temperature below the dendrogram.

Chapter 4 Seasonal Dynamics of Coastal Phytoplankton Communities Revealed by DNA Metabarcoding

This chapter is adapted from a paper published in Marine Environmental Research (2025, 212, 107570; <http://doi.org/10.1016/j.marenvres.2025.107570>).

CRedit authorship contribution statement

Zixia Liu: Conceptualization, Methodology, Data curation, Software, Writing- Original draft preparation. Ilias Semmouri: Conceptualization, Methodology. Yunmeng Li: Conceptualization, Methodology. Maarten De Rijcke: Conceptualization, Methodology. Emmanuel Van Acker: Data curation, Methodology. Colin Janssen: Conceptualization, Supervision. Jana Asselman: Conceptualization, Methodology, Supervision. All: Writing- Reviewing and Editing.

Earlier onset of phytoplankton bloom in the Southern Bight of the North Sea in response to climate variability

Abstract

Phytoplankton are essential to marine ecosystems, driving biogeochemical cycles and supporting diverse marine species and systems. Shifts in bloom timing of these species can disrupt food webs and carbon cycling. This study investigates changes in the timing and composition of phytoplankton blooms in the Belgian Part of the North Sea (BPNS) and examines the environmental drivers behind these shifts. We analyzed data from 2018, 2019, and 2021, combining remote sensing, *in situ* measurements, flow imaging microscopy, and DNA metabarcoding to assess phytoplankton community structure and bloom phenology. Typically, two major annual blooms occur in this region: a mid-spring bloom dominated by *Phaeocystis globosa* and an early-summer diatom bloom primarily consisting

of *Thalassiosira* species. In 2021, both blooms occurred approximately one month earlier than in 2018 and 2019. Contrary to expectations, this advance in blooming was not associated with higher sea temperatures or increased light availability (quantified by photosynthetically active radiation). Instead, we observed increases in phosphate concentrations and reductions in salinity preceding the blooms. These changes suggest that alterations in nutrient dynamics and salinity acted as bottom-up controls, triggering the earlier onset of phytoplankton blooms and supporting the Disturbance Recovery Hypothesis.

Keywords

Phytoplankton bloom, Phenology shifts, 18S rDNA metabarcoding, Nanopore sequencing, Disturbance recovery hypothesis, *Phaeocystis globosa*

1. Introduction

Phytoplankton, as primary producers within marine ecosystems, perform nearly half of the global photosynthetic activity, contributing approximately 49 % of the global net primary production ($\approx 108 \text{ Pg C year}^{-1}$) (Naselli-Flores and Padisák, 2023). They play a crucial role in sequestering carbon dioxide, releasing oxygen, and shaping both marine and atmospheric chemistry (Behrenfeld et al., 2006). By forming the base of aquatic food webs, they support a diverse array of marine life (Karlusich et al., 2020). Phytoplankton communities possess the ability to rapidly adapt to environmental changes, thereby driving seasonal cycles and ecological processes within marine ecosystems (Hutchins and Fu, 2017). Their bloom dynamics - characterized by timing, magnitude, and species composition - are intricately controlled by environmental variables such as nutrient availability, light conditions, and temperature (Behrenfeld and Boss, 2014; Bopp et al., 2005; Henson et al., 2018; Mortelmans et al., 2024). In temperate regions like the southern North Sea, seasonal phytoplankton blooms are particularly

significant due to strong seasonal light and nutrient fluctuations (Behrenfeld, 2010; Goffart et al., 2015), potentially influencing biogeochemical cycles and contributing to the overall health of marine environments (Wiltshire et al., 2008). However, recently observed shifts in the phenology of these blooms, likely driven by changes in environmental conditions, have prompted concerns regarding their potential impacts on marine ecosystems and broader climatic systems (Henson et al., 2018).

In recent decades, significant shifts in the phenology of phytoplankton blooms have been observed, with many studies attributing these changes to global climate change and the resulting changes in bottom-up drivers such as temperature, light availability, and nutrient dynamics (Anderson et al., 2021; Winder and Sommer, 2012), as well as top-down controls including grazing pressure (Kagami et al., 2002; Zheng et al., 2022) and overwintering mortality of zooplankton (Sommer and Lewandowska, 2011). In particular, the earlier onset of spring blooms has been increasingly reported across various marine ecosystems, raising concerns about potential mismatches between primary production and the life cycles of other marine organisms. For instance, Haëck (Haëck et al., 2023) reported earlier phytoplankton blooms by 1–2 weeks over a period of 20 years along fronts in the Gulf Stream, potentially disrupting the timing of zooplankton and other marine organisms that rely on these blooms for food. Similarly, in the Baltic Sea, experimental research shows that warming temperatures can alter the timing and magnitude of phytoplankton blooms, leading to possible mismatches with the reproductive cycles of zooplankton (Sommer et al., 2012). As for top-down controls, mesocosm experiments demonstrated that warming advanced the spring phytoplankton peak, while overwintering zooplankton reduced bloom magnitude, highlighting the combined role of temperature and grazing in bloom timing (Sommer and Lewandowska, 2011). Likewise, modeling work demonstrated that selective grazing by zooplankton can stabilize phytoplankton communities under moderate conditions but promote harmful algal

blooms when grazing selectivity is high (Zheng et al., 2022). Despite these findings, the mechanisms underlying these phenological shifts remain inadequately understood, particularly in temperate coastal regions like the Southern North Sea, where complex interactions among multiple environmental factors are involved (Desmit et al., 2020; Friedland et al., 2016).

Within the Southern North Sea, a number of studies have examined phytoplankton responses to changing environmental drivers, yet most have focused either on total phytoplankton biomass (Lavigne et al., 2022; Lefebvre and Dezécache, 2020; Nohe et al., 2020) or on a limited range of taxa (Peperzak et al., 1998; Rousseau et al., 2013). Studies focusing on phytoplankton aggregate biomass measures, such as chlorophyll concentrations, have documented shifts in bloom timing and chlorophyll *a* levels concurrent with de-eutrophication and warming trends (Desmit et al., 2020). However, relying solely on these bulk indicators provides only a partial perspective as total biomass may appear stable even as underlying species compositions and competitive interactions undergo profound reorganization (Sommer, 1989). This highlights the limitations of aggregate biomass metrics for capturing the nuances of phytoplankton community adaptation to environmental changes. Similarly, targeted investigations into specific phytoplankton groups, such as diatoms or the haptophyte *Phaeocystis*, have yielded valuable insights into how particular taxa respond to hydroclimatic forcing and altered nutrient regimes (Breton et al., 2006; Gypens et al., 2007). A study has extended this perspective by examining phytoplankton communities at finer taxonomic resolution in relation to riverine nutrient inputs, showing that freshwater discharge can trigger early blooms and alter species composition (Groß et al., 2022). Yet, it is restricted to short-term incubations and thus does not capture the interannual variability in community-wide responses. Yet, these approaches overlook the broader spectrum of phytoplankton diversity that plays a critical role in shaping ecosystem structure and functioning. As a result, both aggregate and taxon-specific studies fall short of

representing the shifts in bloom phenology and community composition. Addressing this gap requires integrative approaches, for example by combining high-resolution taxonomic data with overall phytoplankton biomass observations.

The Belgian Part of the North Sea (BPNS) serves as an ideal case study for investigating these shifts, as this region has been subject to significant and well-documented changes in plankton dynamics over recent decades (Desmit et al., 2020; Nohe et al., 2020; Semmouri et al., 2023). Moreover, the North Sea has experienced significant warming trends over recent decades, demonstrating its status as a hotspot for climate change (Holt et al., 2012). Past research has highlighted several key trends, such as the long-term impacts of rising temperatures on plankton communities (Beaugrand et al., 2003), significant declines in copepod populations (Semmouri et al., 2023), and the emergence of new phytoplankton blooms, such as the *Bellerochea* sp. (Heterokontophyta) summer blooms that have been observed since 2018 (Mortelmans et al., 2024). Shifts in plankton dynamics in this area have had cascading effects on the marine food web, from the survival of larval fish species to the overall structure of the ecosystem (Lynam et al., 2017). These documented changes make the BPNS a valuable region for studying how changing environmental factors, especially temperature and nutrient dynamics, influence coastal marine ecosystems and primary producers such as phytoplankton.

This study thus seeks to disentangle a selection of key environmental drivers that shape phytoplankton bloom phenology in the Belgian Part of the North Sea (BPNS). We employed a combination of publicly available datasets, including those derived from remote sensing (e.g., phytoplankton concentrations), *in situ* measurements (e.g., sea surface temperatures), and modeling outputs (e.g., photosynthetically active radiation), along with DNA metabarcoding techniques to profile the phytoplankton community structure over multiple years. By integrating these approaches, we provide a comprehensive analysis that captures

both these environmental factors and the responses of the phytoplankton community at a taxon-specific level. In particular, this research addresses how phytoplankton species shift in regions like the BPNS can inform global climate models and future predictions of primary production, by linking observed bloom phenology and community composition changes to multiple environmental drivers including nutrient dynamics, salinity, temperature, and light availability. Therefore, this study enhances our understanding of the complex interactions between such environmental factors and phytoplankton dynamics, offering critical insights into the potential implications of climate-driven changes on marine ecosystems. The findings have important implications for predicting future shifts in primary production, with broader impacts on marine food webs and biogeochemical cycles in temperate coastal regions.

2. Materials and methods

2.1. Sample collection

Seawater samples were collected periodically between March 2018 and October 2021 from the knee-deep surf zone (≤ 1 m depth) of a recreational sandy beach in Ostend, Belgium ($51^{\circ}14'28''\text{N}$, $2^{\circ}55'56''\text{E}$). A total of 24 water samples were collected in the years 2018 (nine samples), 2019 (nine samples), and 2021 (six samples) for further analysis, mostly focusing on the spring-summer period which is characterised by increased phytoplankton densities (Wiltshire et al., 2008). Due to the COVID-19 pandemic in 2020, no samples could be collected due to restrictions on travel and lab work. The specific collection dates are detailed in **Supplementary Table S1**. Each seawater sample was obtained using a 1000 ml DURAN glass bottle by immersing the bottle inlet beneath the seawater surface until completely filled. Immediately after sampling, the seawater was transported to the laboratory within 30 min for filtration. Initially, each sample was passed through a sterilized 60 μm nylon mesh to eliminate larger debris and subsequently through a 10 μm PTFE filter ($\varnothing$ 47 mm, Omnipore) to

capture the phytoplankton. After filtration, the PTFE filters containing phytoplankton were promptly sealed in plastic petri dishes and stored at -80°C until DNA extraction. All materials that came into contact were either single-use or sterilized with 70 % ethanol and thoroughly rinsed with deionized water before reuse.

2.2. DNA extraction, amplification, and sequencing

Each PTFE filter was quartered using a sterile lancet and allocated to four individual 2 ml Eppendorf tubes. A bead-beating step was included to facilitate cell detachment from the filter (Djurhuus et al., 2017). This involved adding 0.5 g of 0.1 mm glass beads (BioSpec Products), along with two portions of 165 μl ATL Buffer (Qiagen) and 20 μl of Proteinase K to each tube. The tubes were then stirred and incubated in a VWR Incubating Mini Shaker at 56°C for 2 h. Following incubation, each tube was centrifuged for 1 min at 4000 g at room temperature, and the resulting supernatant was collected in new 2 ml tubes and centrifuged again for 1 min at 13000 g at room temperature. DNA extraction was conducted using the Qiagen DNeasy Blood & Tissue Kit, following the manufacturer's protocol with some modifications: 320 μl AL Buffer and 320 μl ethanol (96 %–100 %, molecular biology grade, Sigma-Aldrich) were added, and final elution was carried out with 20 μl AE Buffer per sample. The DNA extracts from the four subsamples, each derived from a quarter of the original filter, were pooled into a single DNeasy Mini spin column. A Nanodrop 2000 Spectrophotometer (Thermo Scientific) was used for quality control and rough estimation of obtained DNA concentrations.

Prior to amplification, primers were evaluated and selected *in silico* by assessing local species coverage using the LifeWatch database (Martínez et al., 2020) and comparing against the SILVA 138 database (Quast et al., 2013) via BLAST version 2.12.0. The primer set F-566/R-1200 (f-CAGCAGCCGCGGTAATTCC, r-CCCGTGTGAGTCAAATTA AGC) was chosen for 18S rRNA gene amplification (Hadziavdic et al.,

2014). PCR amplification mixtures were prepared in 50 μ l volumes according to the manufacturer's protocol, comprising 25 μ l ThermoFisher Scientific Platinum™ Hot Start PCR Master Mix (2X), 1 μ l each of 20 μ M forward and reverse primer stock solutions, DNA template depending on concentration (refer to **Table S1**), and DNase-free water. Amplifications were conducted using a Bio-Rad C1000 Thermal Cycler, with an initial denaturation step at 95 °C for 4 min to activate the DNA polymerase. The PCR reaction consisted of 35 cycles of 95 °C for 30 s (denaturation), 44 °C for 30 s (annealing), and 72 °C for 1 min (extension), followed by a final 10 min extension at 72 °C. PCR products (2 μ l per sample) were verified via electrophoresis on a 1.5 % agarose gel stained with ThermoFisher Scientific DNA gel loading dye. Amplicon size was evaluated by comparison with 6 μ l of the 100 bp DNA Ladder (ThermoFisher Scientific GeneRuler). Successful amplification was confirmed by observing distinct bands around 500–700 base pair amplicons on the agarose gel. DNA quantity was determined subsequently using the Qubit double-stranded DNA Broad Range (BR) Assay Kit (Life Technologies) and read on a Qubit 2.0 Fluorometer (Life Technologies). Samples were stored at –80 °C until library preparation.

Library preparation was conducted using the SQK-LSK109 kit (Oxford Nanopore Technologies) in combination with the PCR barcoding expansion EXP-PBC096 kit (Oxford Nanopore Technologies) to enable the pooling of multiple libraries. A total of 52 sequencing libraries were included in this batch, with 24 being the samples for this study. All barcoded libraries were equimolarly pooled in a 1.5 mL DNA LoBind Eppendorf tube. The manufacturer's protocol was followed with slight modifications (Semmour et al., 2023). The prepared libraries were carefully loaded onto a FLOWMIN106 flow cell (R9.4.1) (Oxford Nanopore Technologies), following the recommended loading procedure to ensure optimal sample distribution. DNA sequencing was performed in a darkened room using a MinION device for approximately 25 h 52 min under standard MinION settings. The raw DNA reads are deposited in the NCBI Sequence Read Archive (SRA) database under

accession number PRJNA1167470.

2.3. Bioinformatics

The basecalling of raw nanopore signal data (.fast5 files) was performed using the Guppy (v6.4.6) basecaller (Oxford Nanopore Technologies) with the dna_r9.4.1_450bps_hac.cfg model, converting the electrical signals into nucleotide sequences. The basecalling produced FASTQ files containing the raw sequence reads. Porechop (v0.2.4) was then employed for demultiplexing and trimming, using barcodes to separate pooled samples and remove low-quality sequences, retaining only those suitable for downstream analysis. Quality control of the sequences was conducted with PycoQC (v2.5.2) (Leger and Leonardi, 2019), providing comprehensive reports on read quality, length distribution, and sequencing performance. Reads with a length between 400 and 650 bp were retained for further analysis. The curated sequences were processed through the wf-metagenomics pipeline (v2.9.4) on the EPI2ME platform (v5.1.10). Taxonomic classification was performed with the integrated Kraken2 classifier (v2.1.3) (Wood et al., 2019) against the SILVA 138 database (Quast et al., 2013), using default configuration settings. Taxonomic assignments were resolved up to the genus level. The resulting taxonomic assignment table and abundance table were exported for subsequent statistical analysis in R.

To focus on the phytoplankton counts relevant to our study, we applied additional filtering steps to the obtained abundance count table. The procedure involved the following steps: (1) We utilized the WoRMS taxon match service (WoRMS Editorial Board, 2025) from genus to phylum level. For each run, matched records were retained, and unmatched records were processed at the next higher taxonomic level. This iterative approach resulted in 75.27 % of records being successfully matched. (2) Approximately 25 % of the remaining records did not correspond to the correct taxonomic level. To expedite the process, these were grouped by their recorded class and manually

searched in WORMS. A small subset (2.24 %, $n = 29$ out of 1297 records) remained unmatched, primarily comprising taxa labeled as unknown, incertae sedis, uncultured. (3) With the taxonomic classifications confirmed, we aggregated the count data at the phylum level. Phytoplankton phyla were determined through search on AlgaeBase (Guiry et al., 2014), and any phylum containing records in AlgaeBase was flagged as a phytoplankton phylum. Notably, the phylum Tracheophyta, while present in AlgaeBase, was manually excluded as it comprises seagrasses and terrestrial plants.

2.4. Data acquired from public databases

To better understand the environmental drivers affecting phytoplankton dynamics, we incorporated a wide range of available environmental factors: ammonium (NH_4^+), nitrite (NO_2^-), nitrate (NO_3^-), phosphate (PO_4^{3-}), silicate (SiO_4^{4-}), suspended particulate matter (SPM), salinity, seawater temperature (SWT), and photosynthetically active radiation (PAR). Summary statistics for these variables are provided in **Table 1**. These factors were further selected in the data analysis and will be addressed in detail in the next paragraph. Daily Photosynthetically Active Radiation (PAR) data were obtained from NASA Prediction of Worldwide Energy Resources using the nasapower (v4.2.1) package (Sparks et al., 2025), with the ALLSKY_SFC_PAR_TOT parameter employed to measure total surface PAR. Additionally, SWT were sourced from the “Meetnet Vlaamse Banken” database (<https://meetnetvlaamsebanken.be/>), collected by the ‘Ostend eastern palisade – Buoy’ situated about 600 m from our sampling site ($51^\circ 14' 48'' \text{N}$, $2^\circ 55' 39'' \text{E}$). Nutrient concentration data (i.e. NO_3^- , NO_2^- , PO_4^{3-} , and SiO_4^{4-} concentrations) were sourced from the LifeWatch local observation dataset (Martínez et al., 2020). Records of Station 130 ($51^\circ 16' 14'' \text{N}$, $2^\circ 54' 15'' \text{E}$) were used, which is located 3.8 km from our sampling site. This station was the closest one to our sampling site and provided the most frequent records available within a 5 km radius of the sampling location.

Table 1. Summary statistics of environmental factors.

Variable	Description	Unit	Min	Max	Avg	SD
NH ₄ ⁺		μmol L ⁻¹	0.59	10.88	3.80	2.91
NO ₂ ⁻		μmol L ⁻¹	0.04	1.03	0.40	0.27
NO ₃ ⁻		μmol L ⁻¹	0.14	43.97	12.72	13.86
PAR	Photosynthetically active radiation	MJ m ⁻² day ⁻¹	8.16	142.73	74.07	37.78
PO ₄ ³⁻		μmol L ⁻¹	0.01	0.92	0.37	0.25
Salinity		PSU	30.70	34.78	32.43	1.09
SiO ₄ ⁴⁻		μmol L ⁻¹	0.44	17.05	7.62	4.68
SPM	Suspended particulate matter	mg L ⁻¹	3.20	151.00	49.11	38.38
SWT	Seawater temperature	°C	5.24	21.38	13.53	4.97

Phytoplankton biomass in the surface seawater (depth ≤1 m), expressed as carbon concentrations ([PhyC], μM C L⁻¹), was obtained from the “Ocean Biogeochemistry Reanalysis for the North-West European Shelf” dataset (Ciavatta et al., 2018) under the Copernicus program, utilizing the European Regional Seas Ecosystem Model (ERSEM 15.06) (Butenschön et al., 2016). To match our *in situ* sampling site at Ostend (51°14'28"N, 2°55'56"E), we extracted values from a 0 m depth layer within a 0.4° × 0.2° box (2.72–3.12°E, 51.14–51.34°N) surrounding the site and calculated daily mean values across this area. This ensured that the reanalysis data spatially represented the same coastal location as our field sampling.

Phytoplankton counts based on FlowCam analysis at the level of species, genus, or higher, were retrieved from the LifeWatch database

(Flanders Marine Institute (VLIZ), 2018). We used records from Station 130, similar to the nutrient concentration data. The “taxon” column was manually matched to WoRMS identifiers (AphiaID) at the phylum level, and phytoplankton phyla were then selected following the procedure described in the Bioinformatics section.

2.5. Statistics and visualization

Statistical analyses and visualizations were performed using the R programming language (version 4.3.3) (R Core Team, 2014). Specifically, the dplyr (v1.1.4) (Wickham et al., 2023) and tidyverse (v2.0.0) (Wickham and RStudio, 2023) packages were utilized for data manipulation and formatting. The biodiversity of the different samples was estimated by calculating the Berger-Parker index, the effective number of species, Fisher's alpha, inverse Simpson's index, Pielou's evenness, richness, and Shannon diversity index. These indices were calculated using the same wf-metagenomics pipeline (v2.9.4) on the EPI2ME platform (v5.1.10) as mentioned in the bioinformatics section above. Redundancy analysis (RDA), and principal coordinates analysis (PCoA) were conducted using the vegan package (v2.6-6) (Oksanen et al., 2025). Correlation analyses were carried out using the corrplot package (v0.92) (Wei et al., 2024). The primary visualization tool was ggplot2 (v3.5.1) (Wickham, 2016), with heatmaps generated using the pheatmap package (v1.0.12) (Kolde, 2025).

3. Results

3.1. Sequencing results and phytoplankton community profile

The sequencing yielded 3.82 million high-quality reads, totaling 2.51 gigabases, with a median read length of 652 and a median PHRED quality score of 9.9. After demultiplexing, we obtained a total of 763,711 high-quality reads with a minimum PHRED quality score of 10. This cut-off is commonly used for Nanopore sequencing data (Doorenspleet et al., 2025), which show lower per-base quality scores than short-read

technologies such as Illumina. The number of reads per sample averaged $31,821 \pm 1978$ (mean $\pm$ SE). From these, 412,893 reads (54.06 %) were assigned and annotated to the SILVA 138 database, with an average of $17,204 \pm 6600$ counts per sample (mean $\pm$ SE). In total, 65 phyla, 151 classes, 293 orders, 387 families, and 1297 genera were matched to SILVA's taxonomic assignments. After the cross-check step described in the Method section, 23 phyla were phytoplankton. The overall rarefaction curve almost been reached saturation (**Figure 1a** & **Figure S1**), indicating that the sequencing data is sufficient for subsequent analysis

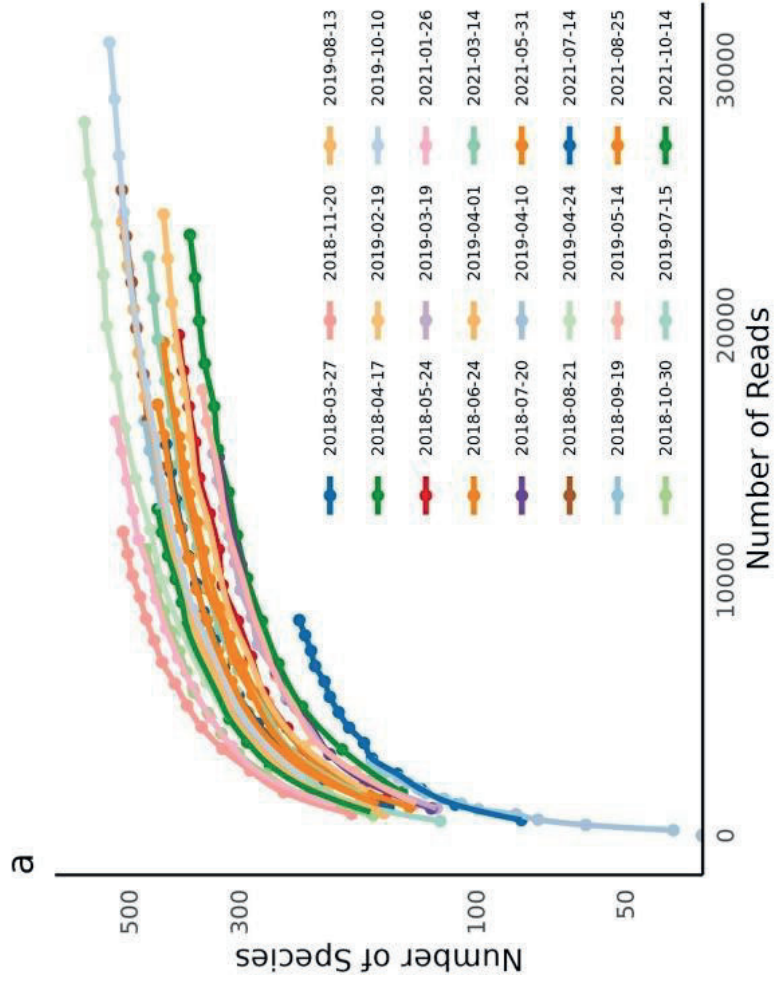


Figure 1. (a) Sample-based rarefaction curves to display observed taxa richness. Sample size shows the

number of reads sampled from the total number of reads analyzed. The Y-axis indicates the number of unique taxa at the last analyzed taxonomic rank in those subsampled reads.

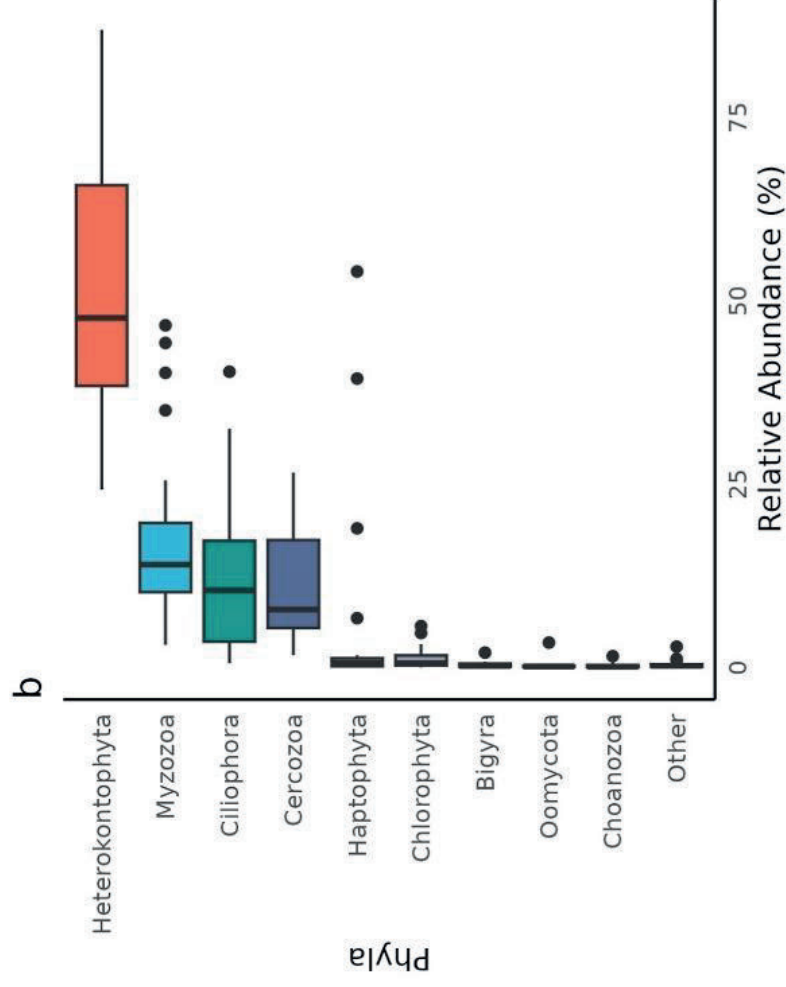


Figure 1. (b) Boxplot summarizing the relative abundance (%) of the nine most abundant phytoplankton phyla

across all samples. Each box represents the distribution of relative abundances for one phylum, with outliers shown as points. The category 'Other' groups all remaining, less abundant taxa to simplify visualization.

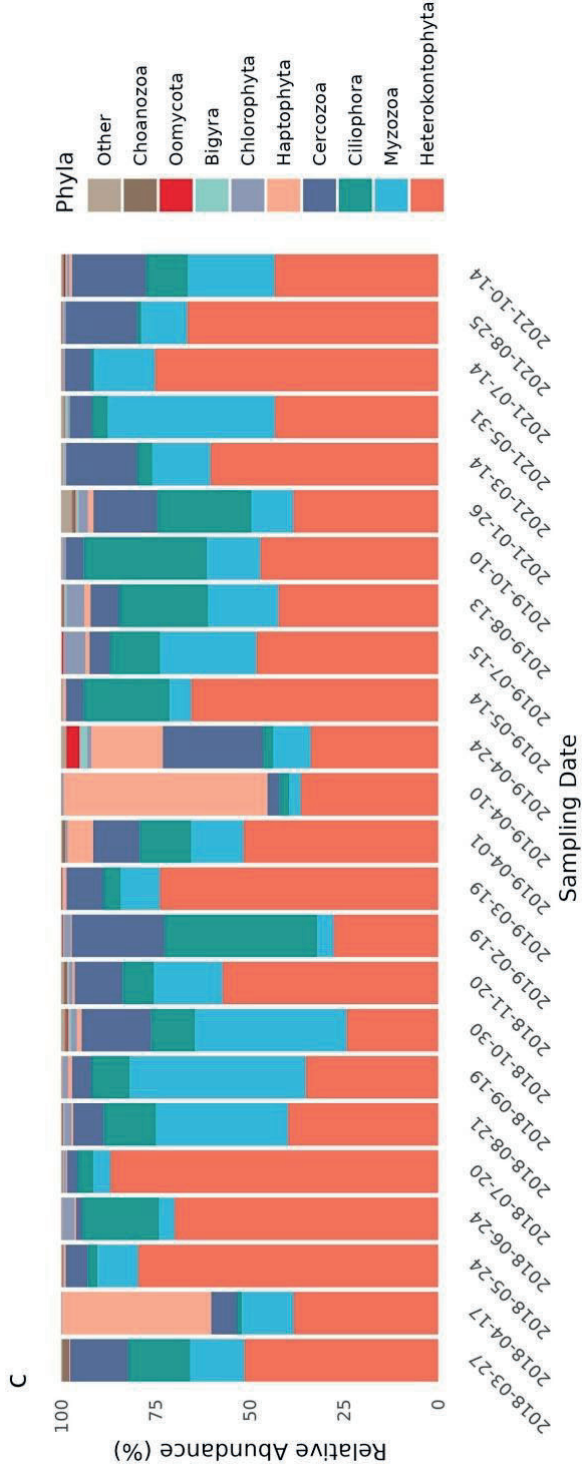


Figure 1. (c) Stacked barplot depicting the temporal dynamics of phytoplankton community composition across sampling dates. Each bar represents one sampling date and is segmented by phylum, showing the relative contribution of each of the nine dominant phytoplankton phyla, with less abundant taxa grouped under 'Other.'

Heterokontophyta was consistently the most dominant phylum throughout all years studied (**Figure 1b & c**), with its relative abundance ranging from 24.3 % to 86.9 % and averaging 51.5 %. It was mainly composed of the diatom genus *Thalassiosira* (**Figure S4a**). Myzozoa (comprising dinoflagellates among other groups) was also prevalent in samples of most sampling dates, with its maximum relative abundance at 46.7 % and averaging 17.3 %, except during periods of Haptophyta outbreaks. This likely corresponds to the spring blooms of *Phaeocystis globosa* each April (2018-04-17, 2019-04-10, and 2019-04-24), see **Figure S4b** for detailed community structure during the blooming dates. Additionally, Ciliophora and Cercozoa were also major phyla, maintaining relatively consistent abundances throughout the years, with average relative abundance of 12.1 % and 11.0 %, respectively.

In addition to the relative taxon abundance derived from metabarcoding, the absolute abundance of phytoplankton expressed as carbon concentrations (PhyC), was plotted in a time series in **Figure 2a**. It reveals clear seasonal variations in phytoplankton abundance, with peaks consistently occurring in the spring and early summer months. Despite consistent trends across the years 2018, 2019, and 2021, there were interannual variations in both the magnitude and the timing of these peaks. Specifically, the peak and average values of PhyC in 2018 were the highest, whereas those in 2019 were the lowest. The 2021 peaks, in addition to lower magnitudes, also appeared to be earlier in the season (28 March, 3 May) compared to 2018 (26 April, 13 June) and 2019 (24 April, 02 June) blooms. This observed variability in PhyC concentrations, combined with the relative abundance derived from metabarcoding, enables further analysis of environmental factors influencing phytoplankton communities. The *in situ* cell density of phytoplankton measured with the FlowCam, is presented in a time series in **Figure 2b**. The FlowCam results present the similar trend as those obtained from metabarcoding, with Heterokontophyta consistently emerging as the most dominant phylum across all sampling years.

Myzozoa, Ciliophora, and Cercozoa were also prevalent on most sampling dates. All these dominant phyla exhibited pronounced seasonal patterns, characterized by higher cell densities during the summer months -indicative of seasonal blooms -and substantially lower densities in the winter. To better integrate and compare the metabarcoding and FlowCam results, we extracted the density data of the most dominant phyla (Heterokontophyta and Myzozoa) from both datasets and then normalized them to account for differences in measurement scales and units. The normalized data were then plotted together in **Figure 2c**, providing a unified visual representation and highlighting the concordance between the two independent methodologies.

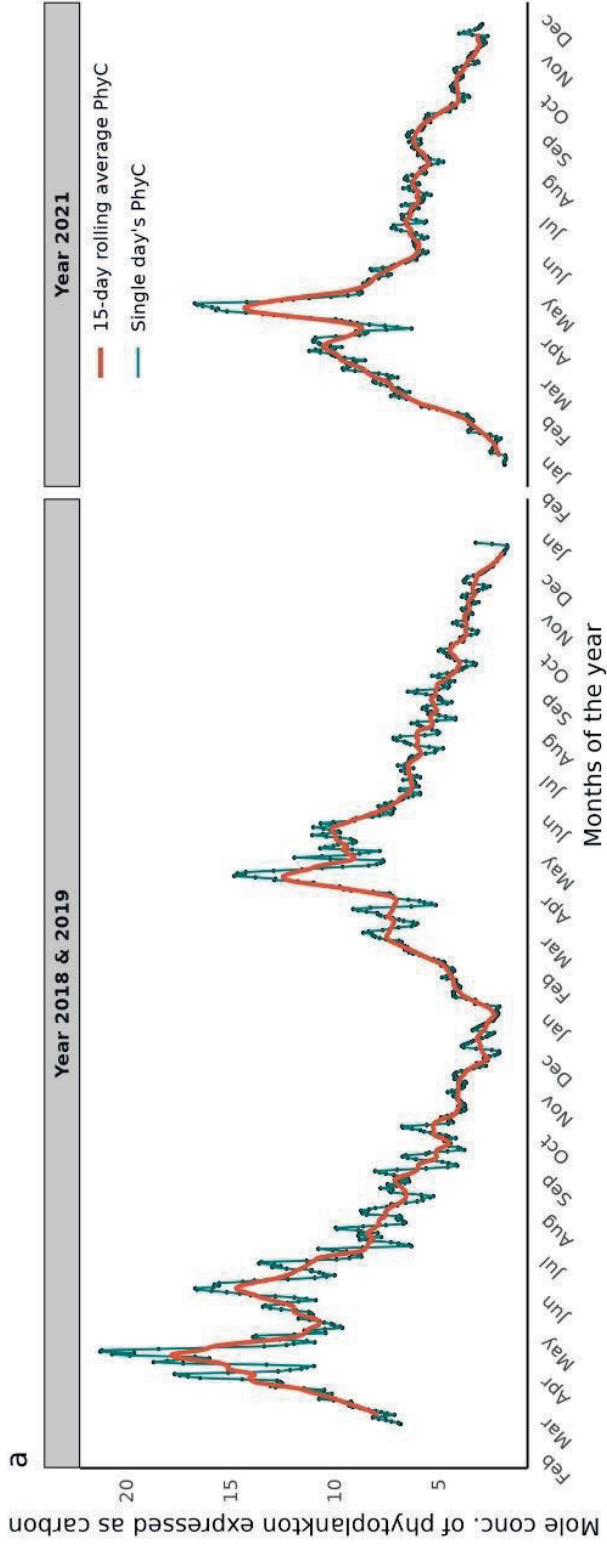


Figure 2. (a) Time series of phytoplankton absolute abundance. The X-axis spans years 2018, 2019, and 2021. The Y-axes represent phytoplankton molar concentration as carbon units ($\mu\text{M C L}^{-1}$).

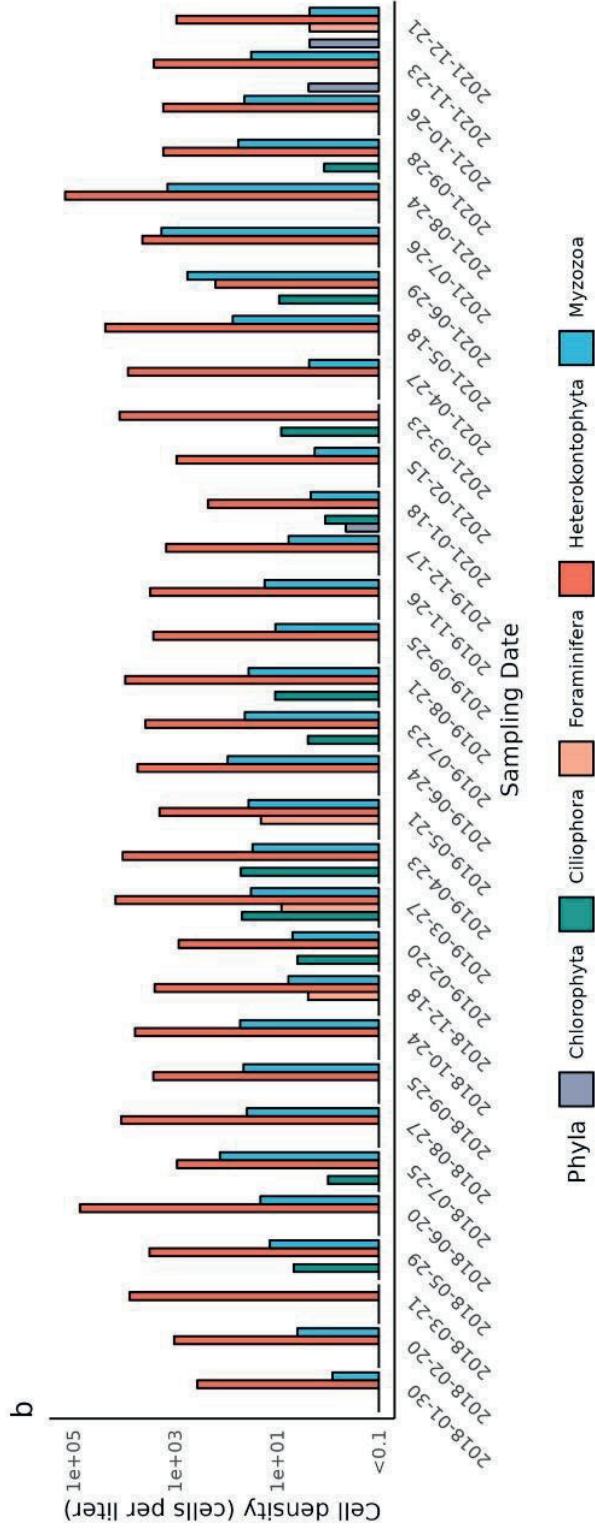


Figure 2. (b) Barplot of phytoplankton cell density ($\log_{10}$ -transformed), based on *in-situ* FlowCam measurements. The Y-axes represent cell density (cells L^{-1}).

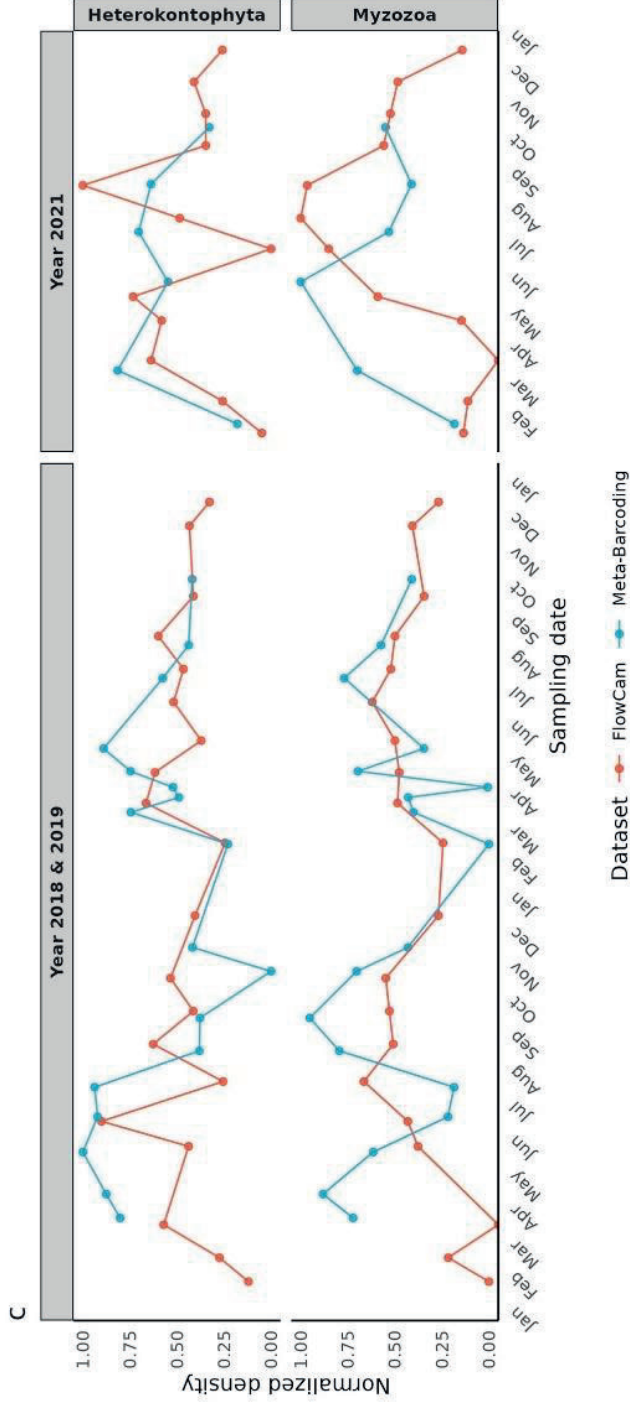


Figure 2. (c) Time series of normalized densities of dominant phyla Heterokontophyta and Myzozoa, comparing meta-barcoding and FlowCam results. The X-axis spans years 2018, 2019, and 2021. The Y-axes represent normalized density.

3.2. Phytoplankton community responses to environmental factors

Table 1 presents the summary statistics of various environmental factors used in this study. The correlations between different environmental factors are shown in **Figure 3a**. Nutrient levels in the sea surface are strongly influenced by SWT, which shows strong negative correlations with NO_3^- ($r = -0.775$, $p < 0.001$) and SiO_4^{4-} ($r = -0.557$, $p < 0.01$). Nutrient levels also relate to PAR, which has a strong negative correlation with PO_4^{3-} ($r = -0.855$, $p < 0.001$). Within nutrient levels, there are interdependencies: NH_4 is positively correlated with NO_2^- ($r = 0.384$, $p = 0.064$) and PO_4^{3-} ($r = 0.392$, $p = 0.058$), while NO_3^- shows strong positive correlations with SiO_4^{4-} ($r = 0.784$, $p < 0.001$) and PO_4^{3-} ($r = 0.551$, $p < 0.01$).

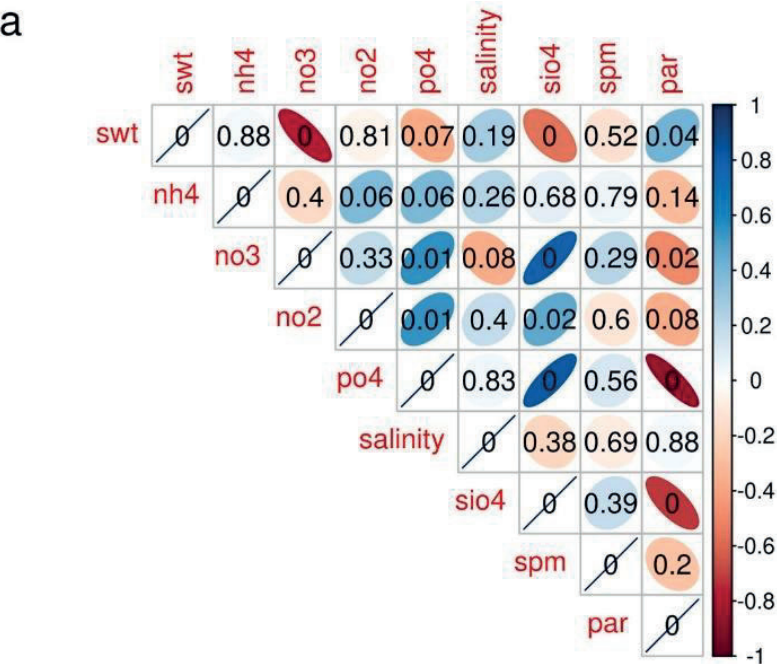


Figure 3. (a) Correlations between different environmental variables.

The color scale of the ellipses indicates correlation strength and direction, ranging from deep blue (strong positive correlation, $r \approx +1$) to deep red (strong negative correlation, $r \approx -1$). The shape of each ellipse further conveys the correlation: narrow, diagonally oriented ellipses represent stronger correlations, while more circular shapes indicate weaker or no correlation ($r \approx 0$). The numbers inside each box indicate the p -values from the correlation tests.

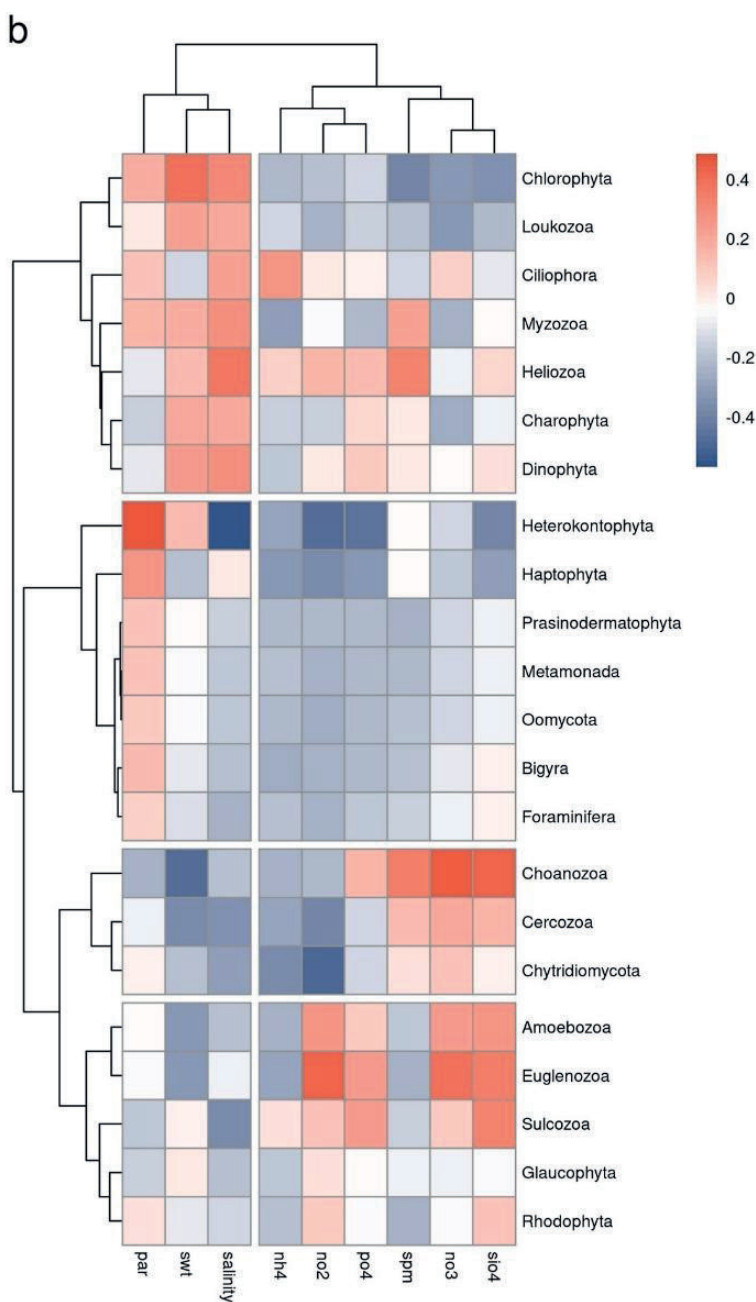


Figure 3. (b) Heatmap representing the correlation between the

absolute abundance of different phyla and various environmental factors. The dendrogram on the top and left sides of the heatmap shows the hierarchical clustering of the environmental factors and taxa based on their Euclidean distance. The rows correspond to different phyla, while the columns represent environmental factors. The two red lines indicate the cuts that define the hierarchical clusters.

C

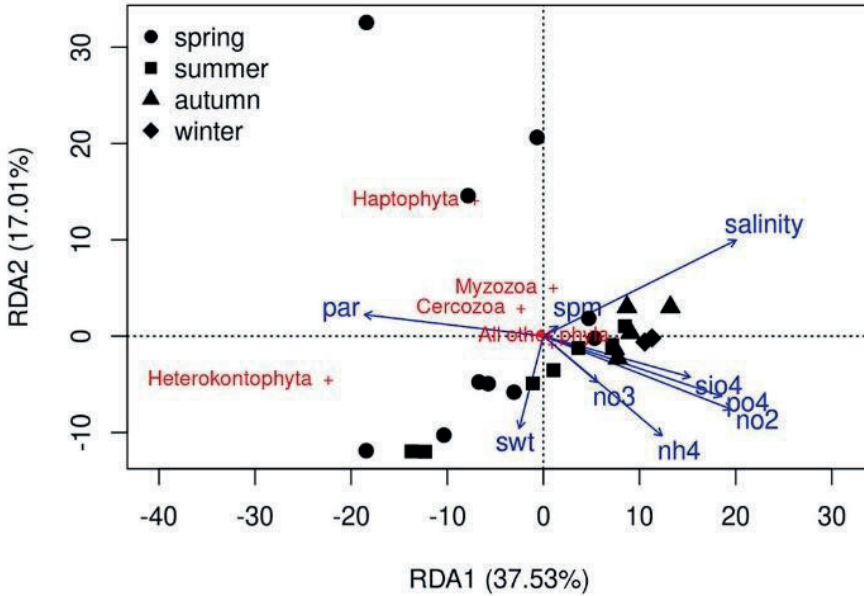


Figure 3. (c) Redundancy Analysis (RDA) plot visualizing the relationship between environmental factors and phytoplankton composition. Black points with various shapes represent samples collected in different seasons in the dataset. Blue arrows indicate environment (predictor) scores, showing the influence of environmental factors on the phytoplankton community. Red crosses indicate phylum (biotic) scores, showing the phylum's response to the environmental factors; overlapping phyla have been grouped as “All other phyla” for readability. Abbreviations used: par: photosynthetically active radiation, spm: suspended particulate matter, swt: seawater temperature, nh4: NH_4^+ , no2: NO_2^- , no3: NO_3^- , po4: PO_4^{3-} , sio4: SiO_4^{4-} .

The relationship between environmental factors and the absolute abundance of the different phyla is illustrated through a heatmap (**Figure 3b**). Hierarchical clustering analysis of these data reveals distinct patterns of association. The environmental factors, represented as columns in the heatmap, are grouped into two primary clusters: the cluster of nutrient concentrations, including NH_4^+ , NO_2^- , NO_3^- , PO_4^{3-} , SiO_4^{4-} ; and the cluster of other environmental factors, which include PAR, SWT, and Salinity. The phytoplankton phyla, shown as rows in the heatmap, are similarly categorized into four distinct clusters based on their correlations with these environmental conditions: Cluster 1: Dominated by the phyla Chlorophyta and Myxozoa, this cluster demonstrates positive correlations with the key physical parameters (PAR, SWT, and Salinity) alongside mixed correlations with nutrient concentrations. Cluster 2: Represented by Heterokontophyta, Haptophyta, Oomycota, and Bigyra, this cluster is characterized by negative correlations with nutrient concentrations. Its relationship with the physical parameters (PAR, SWT, and Salinity) is mixed, specifically showing a positive correlation with PAR and a negative correlation with Salinity. Cluster 3: With Choanozoa and Cercozoa as the dominant phyla, this cluster exhibits negative correlations with the key physical parameters (PAR, SWT, and Salinity) and mixed correlations with nutrient concentrations. Cluster 4: This cluster shows mixed relationships with both nutrient concentrations and the key physical parameters (PAR, SWT, and Salinity). According to metabarcoding results, no single phylum is dominant within this group.

The Redundancy Analysis (RDA) plot in **Figure 3c** visualizes the relationship between environmental factors and phytoplankton composition. RDA1 and RDA2 explain 37.53 % and 17.01 % of the variance, respectively. Nutrient factors (NH_4^+ , NO_2^- , NO_3^- , PO_4^{3-} , SiO_4^{4-}) generally point to similar directions, suggesting a correlated influence of these nutrients on the phytoplankton composition. The arrows for PAR, salinity, and SWT point to three different quadrants, indicating their distinct influences on the phytoplankton composition.

The biodiversity within the different samples was quantified through the calculation of several indices: the Berger-Parker index, the effective number of species, Fisher's alpha, inverse Simpson's index, Pielou's evenness, richness, and the Shannon diversity index. These biodiversity indices are detailed in **Table S2**. **Figure 4a & b** illustrate the temporal variations in Fisher's alpha, richness, and Shannon index across two time scales. In **Figure 4a**, the indices are plotted against dates spanning from 2018 to 2021 (excluding 2020). Shannon index and richness generally share the similar trends: both indices show declines during late spring (after April) in 2018 and 2019, but not in 2021, due to our sampling not covering the bloom period. Notably, both the Shannon index and richness experienced significant declines in July 2021. However, the declines in July 2018 are less pronounced, and in July 2019, a decline is only observed in richness. As for Fisher's alpha, a peak was observed in April of 2019, coinciding with declines in Shannon index and richness. In **Figure 4b**, the data are consolidated by month to reveal general seasonal patterns. Each diversity index exhibits clear seasonal fluctuations throughout the year: Fisher's alpha peaks at mid-spring and late-summer while the Shannon index and richness dip at that period. This difference likely reflects the fact that these indices capture different aspects of community diversity. While richness measures the number of detected taxa and the Shannon index is strongly influenced by both richness and evenness, Fisher's alpha is derived from the abundance distribution and may be more sensitive to the contribution of low-abundance taxa. Therefore, these indices do not necessarily vary in parallel, and opposite trends may emerge during periods of strong community dominance.

The relationship between environmental factors and biodiversity indices is illustrated through a heatmap (**Figure 4c**). The hierarchical clustering reveals distinct patterns of association between biodiversity indices and environmental conditions. The resulting two (row) clusters of biodiversity indices are (1) richness, effective number of species, Shannon index, Pielou's evenness, and Simpson's index, which show positive

correlations with nutrient availability factors (PO_4^{3-} , NH_4^+ , NO_2^- , SiO_4^{4-}), salinity and seawater temperature, and negative correlations with other nutrient availability factors (PAR, NO_3^- , SPM); (2) total counts, Fisher's alpha, Berger-Parker index, inverse Simpson's index which showed opposite correlation patterns with the environmental factors compared with the previous cluster.

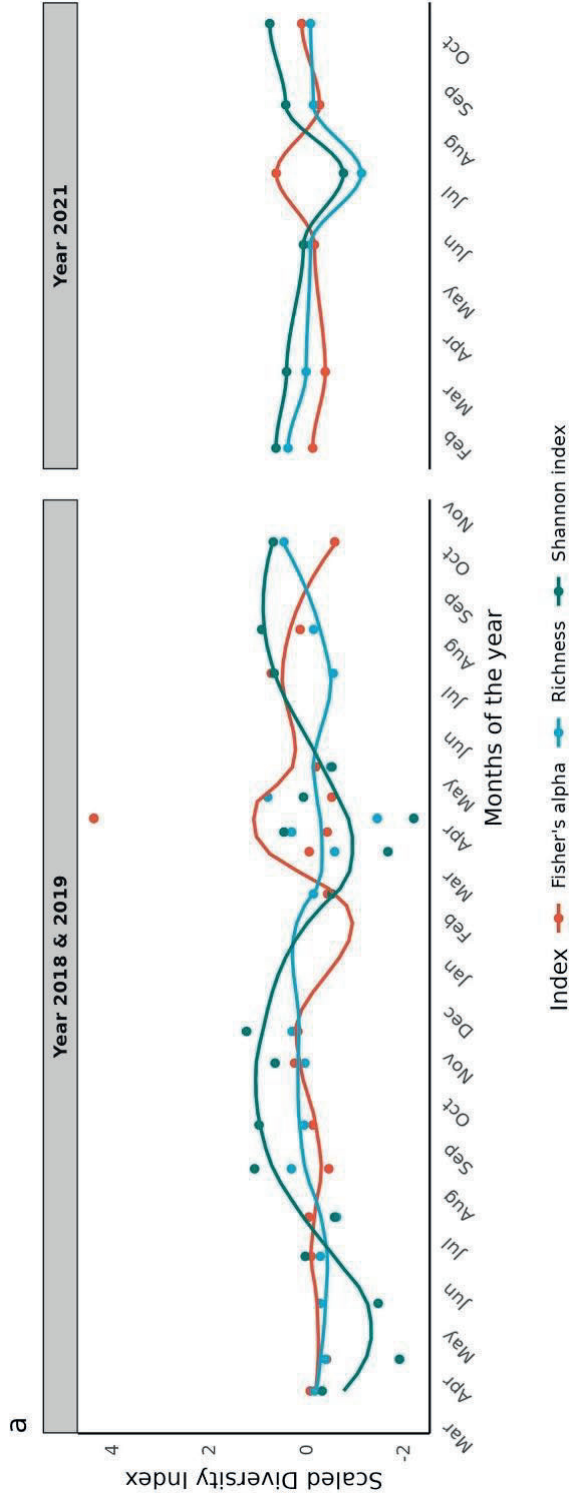


Figure 4. (a) Temporal trends of biodiversity indices from 2018 to 2021, excluding 2020 (vertical dashed red line). Data points represent individual observations, while smoothed lines show trends.

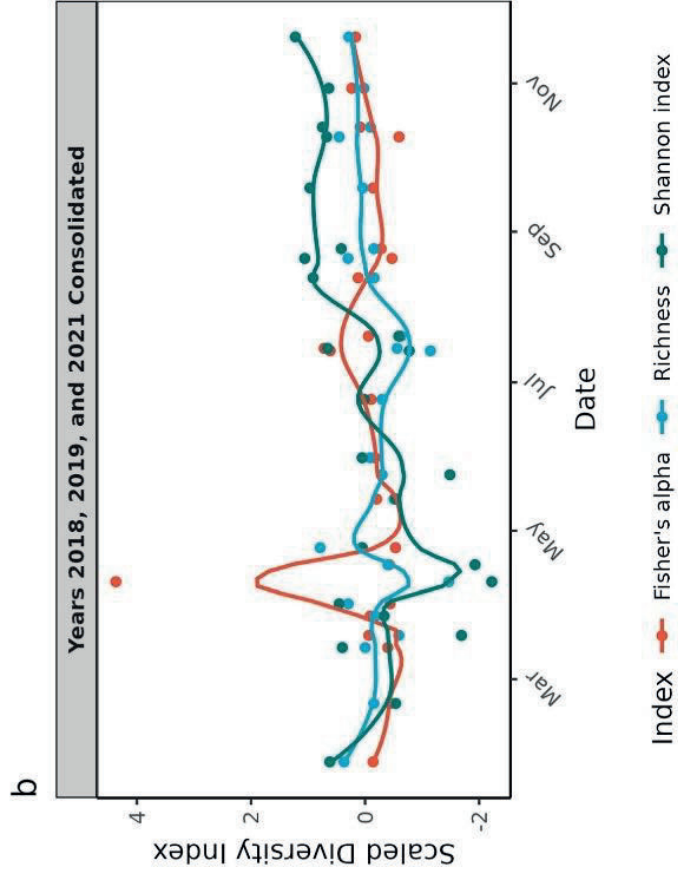


Figure 4. (b) Consolidated monthly data highlighting seasonal patterns of the three biodiversity indices.

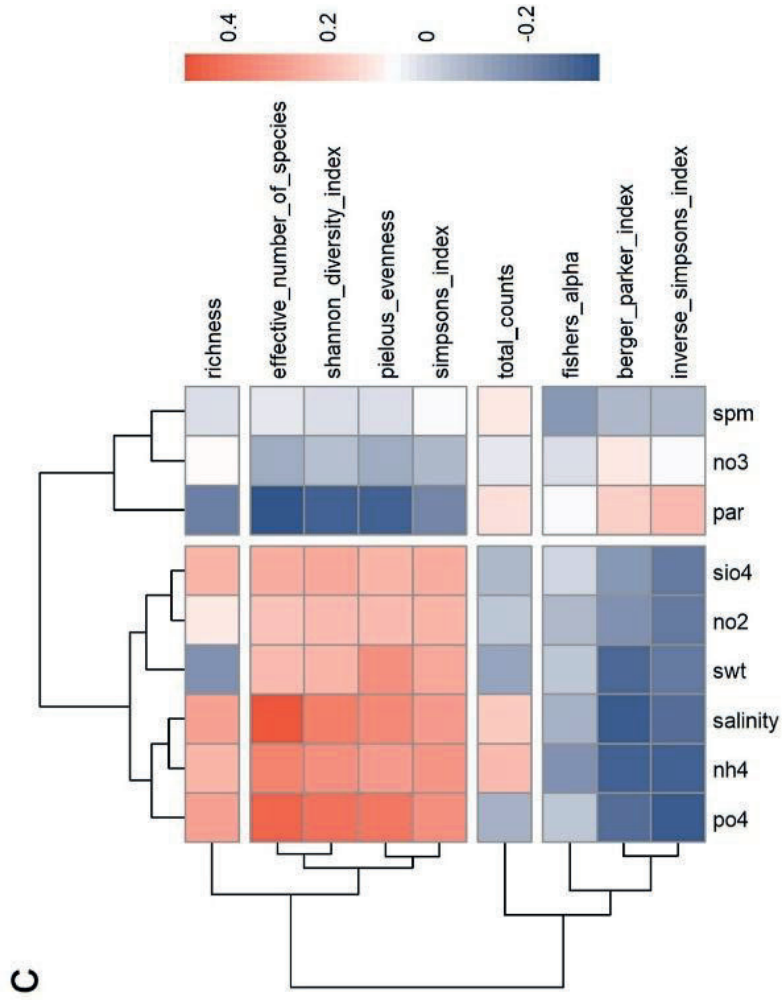


Figure 4. (c) Heatmap representing the correlation between various biodiversity indices across different

environmental factors. The dendrograms on the top and left sides of the heatmap show the hierarchical clustering of the biodiversity indices and environmental factors based on their Euclidean distance. The rows correspond to different biodiversity indices, while the columns represent the environmental factors. Color intensity indicates the strength of correlation, with red representing positive correlations and blue representing negative correlations. Abbreviations used: par: photosynthetically active radiation, spm: suspended particulate matter, swt: seawater temperature.

4. Discussion

4.1. Seasonal dynamics of environmental drivers in the Southern Bight of the North Sea

Seasonal variation in environmental drivers is a defining feature of temperate marine ecosystems, including the study area of the Southern Bight of the North Sea where a periodicity of solar irradiance and temperature govern the recurrence of phytoplankton growth and bloom timing (Scharfe and Wiltshire, 2019). We observed several interannual variations that may influence phytoplankton dynamics in the southern North Sea. While seawater temperature and PAR exhibited consistent seasonal cycles with minimal differences across the years studied (**Figure S2 b & c**), notable deviations were observed in nutrient concentrations and salinity levels (**Figure S3**). Specifically, elevated phosphate (PO_4^{3-}) concentrations were recorded from January to April in 2021, accompanied by lower seawater salinity, compared to the same periods in 2018 and 2019. These may be due to the increased freshwater influx, possibly from both precipitation (January 2021 **Figure S3c**) and riverine discharge from Scheldt/Rhine/Seine rivers (Lacroix et al., 2007), reducing the seawater salinity and introducing higher nutrient loads into the coastal system during that time (Groß et al., 2022; Lacroix et al., 2004). Besides precipitation and riverine input, wind direction and vertical mixing may also explain the observed shifts in phosphate and salinity in early 2021. By altering stratification and nutrient redistribution in shallow coastal waters (Cyr and Coman, 2012), wind-driven processes can create favorable conditions for earlier bloom initiation (Desmit et al., 2020). Although wind data were not analyzed in this study, their role as a physical driver of nutrient dynamics should be considered in future work.

Additionally, reduced phytoplankton absolute abundance during early 2021 (January and February) could have also contributed to the higher nutrient levels due to the reduced nutrient consumption. Moreover, our analysis (**Figure 3a**) revealed significant correlations among

environmental factors, indicating intricate interdependencies that further influence phytoplankton dynamics. The observed negative correlations between temperature and nutrients such as nitrate (NO_3^-) and silicate (SiO_4^{4-}) align with the premise that higher temperatures enhance phytoplankton metabolic activity, leading to increased nutrient uptake and subsequent depletion of these nutrients in the water column (Groß et al., 2022; Mortelmans et al., 2019). Similarly, higher light availability, as suggested by PAR levels, can stimulate primary production, accelerating nutrient assimilation by phytoplankton.

Figure 3b provides further insight into the diverse ecological strategies of the phytoplankton community, as distinct clustering patterns reveal that different phyla respond differently to the environmental factors. The environmental variables form two main groups: nutrient concentrations and physical factors (PAR, temperature, and salinity). The phytoplankton phyla also group into four clusters based on their correlations with these environmental drivers. The two major bloom-forming phyla, Haptophyta (*Phaeocystis*) and Heterokontophyta (diatoms), are both negatively correlated with nutrients, reflecting nutrient drawdown during their respective blooms. However, their distinct correlations with physical factors align with their seasonal succession. The mid-spring *Phaeocystis* bloom appears insensitive to salinity changes, whereas the subsequent early-summer diatom bloom shows a strong negative correlation with salinity, suggesting it may be more directly influenced by nutrient-rich freshwater discharge from rivers (Lacroix et al., 2007). In contrast, the triggers for the Haptophyta (*Phaeocystis*) bloom appear more nuanced; the data show its onset is positively correlated with light availability (PAR) but not with salinity changes, and it is negatively associated with the seawater temperature, suggesting its timing depends on a specific window where light is sufficient but temperatures have not yet peaked. As waters warm into early summer, diatom biomass typically increases while *Phaeocystis* declines, and this seasonal handover helps explain the observed negative correlation between *Phaeocystis* and seawater

temperature (Peperzak et al., 1998; Smith and Trimborn, 2024; Weisse et al., 1986). Furthermore, the positive correlation of Myzozoa (dinoflagellates) with sea temperature is consistent with this group becoming more prominent in the community during the warmest part of the summer, following the main spring and early-summer blooms.

These patterns highlight the complexity of environmental drivers in the Southern North Sea, characterized by both regular seasonal cycles and interannual variability. While our study addresses the key environmental factors (listed in **Table 1**) that are relevant to phytoplankton dynamics (Falkowski et al., 1998; Henson et al., 2018; Winder and Sommer, 2012), due to data availability and the focus of our research scope, we acknowledge that numerous other environmental variables also contribute to the phytoplankton dynamics.

4.2. Earlier blooms observed in 2021

In the study area of the Southern North Sea, two prominent phytoplankton blooms are observed annually: a mid-spring bloom typically occurring around April and an early-summer bloom around June (**Figure 1c**). According to our metabarcoding results (**Figure S4**), the mid-spring blooms are predominantly driven by Haptophyta (*Phaeocystis globosa*), whereas the early-summer blooms are mainly attributed to Heterokontophyta (*Thalassiosira* sp.). These seasonal bloom patterns align with findings from previous studies in the region, which have documented similar temporal dynamics in phytoplankton community composition. This alignment is further supported by recent work demonstrating the timing of *Phaeocystis globosa* blooms in late April to early May 2020 and diatom blooms in late June 2020 sampled at 500m away from our sampling site (Nohe et al., 2020). *Phaeocystis globosa* is known for forming dense colonial blooms in spring during April and May (Breton et al., 2006; Rousseau et al., 2013); while *Thalassiosira* sp. and other diatoms, exhibit a more persistent presence throughout the year with relative modest blooms in summer

(Meyer et al., 2000). However, a notable deviation in the phenology of these blooms was detected in 2021, where both the *Phaeocystis globosa* and *Thalassiosira* sp. blooms occurred approximately one month earlier compared to the years 2018 and 2019 (Nohe et al., 2020). Specifically, *Phaeocystis globosa* blooms were recorded on April 26, 2018, April 24, 2019, and March 28, 2021, while *Thalassiosira* blooms occurred on June 13, 2018, June 2, 2019, and May 3, 2021 (**Figure S2a**). The earlier bloom timing observed in 2021 aligns with long-term trends reported in the BPNS (Nohe et al., 2020): the analysis of phytoplankton dynamics in the BPNS between the 1970s and 2000s revealed significant changes. A shift in diatom bloom patterns from a bimodal cycle (spring and summer-autumn) to a single, extended season starting three weeks earlier in February and ending in mid-summer; dinoflagellate blooms also transitioned to a single, extended growing period from May to September. This shift towards an earlier bloom timing in 2021 suggests an alteration in the environmental conditions that regulate bloom dynamics. In addition to bottom-up factors such as nutrient supply and salinity, top-down controls can also play a role. In particular, a reduction in grazing pressure by overwintering zooplankton populations has been shown to facilitate earlier spring bloom onset (Sommer and Lewandowska, 2011).

Initially, we hypothesized that earlier warming of seawater or increased PAR, might be driving these earlier blooms, as widely reported (Breton et al., 2006; Mortelmans et al., 2019). However, our data did not support this hypothesis: neither seawater temperatures were warmer nor PARs were higher during the 2021 blooms compared to previous years (**Figure S2**). In fact, the seawater temperature and PAR were lower in 2021 during the blooms of both *Phaeocystis* (3.8 °C and 21.8 Mmol·m⁻²·s⁻¹ lower respectively) and *Thalassiosira* (6.7 °C and 6.8 Mmol·m⁻²·s⁻¹ lower respectively). These findings indicate that other factors, rather than temperature or light conditions, might be driving the observed phenological shifts.

To explore alternative explanations, we examined nutrient availability as a potential bottom-up control factor, particularly before the bloom periods, as nutrient levels typically peak prior to a bloom and decline as phytoplankton uptake increases (Huppert et al., 2002). **Figure S3a** reveals a significant increase in phosphate (PO_4^{3-}) concentrations in early 2021 compared to 2018 and 2019, which may account for the earlier bloom of *Phaeocystis*. However, due to limited data it was not possible to determine if a similar increase in PO_4^{3-} levels occurred before the *Thalassiosira* bloom in 2021. Previous research in the Southern Bight of the North Sea by Lefebvre and Dezécache (Lefebvre and Dezécache, 2020) found that changes in nutrient availability, especially dissolved inorganic nitrogen (DIN), affected the relative abundance and timing of *Phaeocystis* and diatom blooms. Lancelot et al. (Lancelot et al., 2014) reported similar findings, additionally they found the competitive advantage of diatoms over *Phaeocystis* has shifted in the past decades, with DIN concentrations in recent years becoming relatively more favorable to diatoms and less favorable to *Phaeocystis*. This appears to contradict our explanation for the earlier bloom of *Phaeocystis* in March 2021 due to increased nutrient concentrations. However, this discrepancy can be explained by phytoplankton succession where a robust early bloom of diatoms depletes silicate, subsequently paving the way for a stronger *Phaeocystis* bloom (Peperzak et al., 1998; Smith and Trimborn, 2024). Because we also observed a strong correlation between seawater salinity and the absolute abundance of *Thalassiosira* (**Figure 3b & c**), we compared the salinity levels prior to these blooms (**Figure S3b**). This revealed lower salinities in 2021 compared to 2018 and 2019 as the red rectangle of **Figure S3b** indicates (February–May 2021). Additionally, there were significant salinity reductions noted just before each *Thalassiosira* bloom event (**Figure S3b**). This pattern suggests that reduced salinity may act as a trigger or facilitator for the *Thalassiosira* blooms, possibly by affecting species competition.

There is another possible explanation for the earlier blooms, known as

the Disturbance Recovery Hypothesis, which suggests that ecosystems experiencing disturbances, such as changes in salinity or nutrient levels, may exhibit rapid biological responses and allow opportunistic phytoplankton to exploit favorable conditions (Behrenfeld and Boss, 2014; González-Gil et al., 2022). On one hand, the increases in PO_4^{3-} concentrations preceding the *Thalassiosira* blooms discussed above can possibly act as disturbances that reduced competition in May 2021; on the other hand, the nutrient level as represented by PO_4^{3-} was higher in 2021 (May) when *Thalassiosira* bloomed compared to 2018 (June) and 2019 (June). *Thalassiosira*, as a dominant phytoplankton genus in our study area, therefore bloomed almost one month earlier. This hypothesis aligns with the observed strong correlation between nutrient availability and phytoplankton abundance (**Figure 3a and b, & S3a**), as well as the significant reduction in salinity noted before each *Thalassiosira* bloom event (**Figure S3b**). This suggests that the phenological shifts observed in 2021 are likely driven by a complex interaction of environmental disturbances, rather than simple temperature or light availability changes, highlighting the intricate dynamics of phytoplankton community responses to changing environmental conditions. If the hypothesis is correct, then the delayed peak observed in 2021 may have been driven by a complex interaction of environmental disturbances, rather than simple changes in temperature or light availability. This possibility requires further investigation into the specific mechanisms by which nutrient levels and salinity fluctuations influence phytoplankton community dynamics and competitive interactions in this region.

4.3. Limitations

One of the primary limitations of this study is the temporal resolution of the sampling, which will have missed finer-scale bloom dynamics and rapid environmental fluctuations that occur between the monthly sampling intervals. Higher-frequency sampling could provide more detailed insights into the short-term drivers of phytoplankton dynamics,

especially during critical transition periods in bloom development. Moreover, while remote sensing and modeled products were instrumental in filling gaps where *in situ* data were unavailable, the spatial and temporal resolution of these products may not fully capture local variability, particularly in coastal regions like the BPNS where small-scale processes can play a significant role. Enhancing the availability and frequency of local *in situ* measurements, including high-resolution monitoring of salinity, nutrient concentrations, and phytoplankton abundance, for example stations with hyperspectral radiometer system (Lavigne et al., 2022), would improve our ability to discern fine-scale drivers of phytoplankton community changes.

Furthermore, our study did not investigate top-down controls such as zooplankton grazing and overwintering mortality. These processes are known to strongly influence phytoplankton bloom timing and magnitude (Kagami et al., 2002; Sommer and Lewandowska, 2011; Zheng et al., 2022), and their absence represents an important limitation of our work. In the Southern Bight, herbivorous zooplankton have been estimated to consume 35–50 % of daily primary production in summer (Baars and Fransz, 1984). Moreover, copepods can exert particularly strong grazing pressure at bloom onset, and warmer years may delay spring blooms because copepod populations remain active longer into autumn and winter (Deschamps et al., 2024), thereby increasing grazing pressure during the biomass-building phase (Friedland et al., 2016; Sommer and Lewandowska, 2011). While we focused primarily on bottom-up drivers, future studies integrating zooplankton abundance, grazing pressure, and overwintering survival with phytoplankton monitoring will be essential for disentangling the combined roles of top-down and bottom-up mechanisms in shaping bloom phenology.

Another limitation relates to the sampling strategy, in addition to the accuracy and taxonomic resolution of our sequencing data. By pre-filtering seawater through a 60 µm nylon mesh and subsequently retaining material on a 10 µm PTFE filter, our metabarcoding dataset

primarily represented the 10–60 μm size fraction rather than the full phytoplankton community. This likely underrepresented smaller taxa as well as some larger or long-chain taxa outside this window. Therefore, comparisons with literature based on unfractionated seawater samples or broader microscopy surveys should be made cautiously, especially for taxonomic composition. This limitation is less critical for the broader phenological patterns, because bloom timing was interpreted alongside FlowCam observations and phytoplankton biomass data, but it does constrain how comprehensively the metabarcoding results represent the full community. In addition, although Oxford Nanopore Technologies offers a cost-effective and rapid sequencing solution, its lower alignment rate compared to other next-generation sequencing platforms led to reduced taxonomic resolution in this study. Together with the PCR-related bias and the reference database used for taxonomic assignment, these factors likely impacted our ability to fully characterize the phytoplankton community, particularly for rare or less abundant taxa. However, since our study primarily focused on dominant species and supplemented sequencing data with morphological observations from FlowCam analysis, this issue was mitigated to some extent. To enhance taxonomic resolution and accuracy in future research, greater sequencing depth and the use of more comprehensive or specialized reference databases may help reduce the proportion of unknown taxa.

Supplementary Information

Table S1. Summary of Seawater Samples and DNA Processing Details

#	Collection Date	Library Barcode	DNA Concentration (ng/μl)		DNA Template Volume Used in PCR (μl)	
			First Round	Barcoding	First Round	Barcoding
1	2018-03-27	barcode03	204.3	49.9	1.0	1.0
2	2018-04-17	barcode04	604.4	59.4	0.5	0.8
3	2018-05-24	barcode05	194.6	76.1	1.0	0.7
4	2018-06-24	barcode06	340.3	49.2	1.0	1.0
5	2018-07-20	barcode07	331.2	44.9	1.0	1.1
6	2018-08-21	barcode08	320.4	42.1	1.0	1.2
7	2018-09-19	barcode09	121.9	50.7	2.0	1.0
8	2018-10-30	barcode22	140.0	38.3	1.0	1.3
9	2018-11-20	barcode10	208.8	93.1	1.0	0.5
10	2019-02-19	barcode11	135.0	64.3	1.0	0.8
11	2019-03-19	barcode12	68.4	43.6	3.0	1.1
12	2019-04-01	barcode24	108.5	84.3	2.0	0.6
13	2019-04-10	barcode61	347.0	49.4	2.0	1.0
14	2019-04-24	barcode13	83.0	43.0	2.0	1.2
15	2019-05-14	barcode62	254.9	67.2	2.0	0.7
16	2019-07-15	barcode14	320.9	30.6	2.0	1.6
17	2019-08-13	barcode15	190.1	50.7	1.0	1.0
18	2019-10-10	barcode23	73.0	97.9	1.0	0.5
19	2021-01-26	barcode16	121.4	48.5	1.0	1.0
20	2021-03-14	barcode17	421.9	60.0	1.0	0.8
21	2021-05-31	barcode18	155.8	46.1	1.0	1.1
22	2021-07-14	barcode19	537.7	39.6	0.5	1.3
23	2021-08-25	barcode20	240.5	75.7	1.0	0.7
24	2021-10-14	barcode21	145.5	37.6	1.0	1.3

Table S2. Biodiversity Indices Calculated from abundance table

#	Sampling Date	Berger Parker Index	Effective Number of Species	Fishers Alpha	Inverse Simpsons Index	Pielous Evenness	Richness	Shannon Diversity Index	Simpsons Index
1	2018-03-27	0.28	33.7	338.12	1.11	0.58	421	3.52	0.9
2	2018-04-17	0.38	14.13	295.42	1.27	0.45	377	2.65	0.79
3	2018-05-24	0.52	17.93	312.05	1.38	0.48	397	2.89	0.72
4	2018-06-24	0.15	41.03	336.01	1.06	0.62	397	3.71	0.94
5	2018-07-20	0.2	29.39	342.1	1.1	0.58	330	3.38	0.91
6	2018-08-21	0.12	72.17	289.39	1.04	0.68	517	4.28	0.97
7	2018-09-19	0.1	69	331.5	1.03	0.69	465	4.23	0.97
8	2018-10-30	0.22	57.32	379.28	1.07	0.66	461	4.05	0.94
9	2018-11-20	0.14	78.89	370.78	1.04	0.7	514	4.37	0.96
10	2019-02-19	0.19	30.28	292.55	1.09	0.56	426	3.41	0.91
11	2019-03-19	0.5	16.09	340.97	1.35	0.48	338	2.78	0.74
12	2019-04-01	0.23	51.96	293.58	1.07	0.63	516	3.95	0.93
13	2019-04-10	0.5	12.11	907.37	1.37	0.49	162	2.49	0.73
14	2019-04-24	0.17	41.63	281.43	1.08	0.58	615	3.73	0.93
15	2019-05-14	0.25	30.36	323.74	1.11	0.58	355	3.41	0.9
16	2019-07-15	0.13	58.19	441.1	1.05	0.7	344	4.06	0.96
17	2019-08-13	0.11	66.38	364.89	1.04	0.69	425	4.2	0.96
18	2019-10-10	0.12	58.4	273.55	1.04	0.64	548	4.07	0.96
19	2021-01-26	0.15	56.85	331.7	1.06	0.64	531	4.04	0.94
20	2021-03-14	0.1	50.32	298.67	1.05	0.64	456	3.92	0.96
21	2021-05-31	0.2	41.65	327.36	1.07	0.61	438	3.73	0.93
22	2021-07-14	0.24	26.69	427.85	1.1	0.61	227	3.28	0.91
23	2021-08-25	0.09	50.87	313.53	1.04	0.65	426	3.93	0.96
24	2021-10-14	0.15	61.02	360.67	1.05	0.68	438	4.11	0.95
Total	NA	0.18	78.78	165.85	1.05	0.61	1297	4.37	0.96

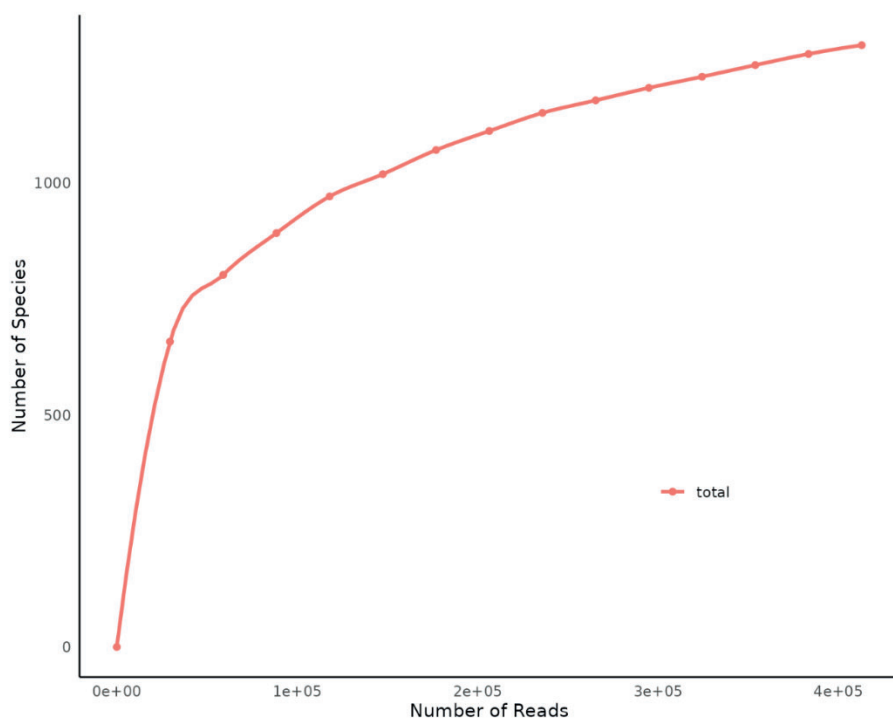
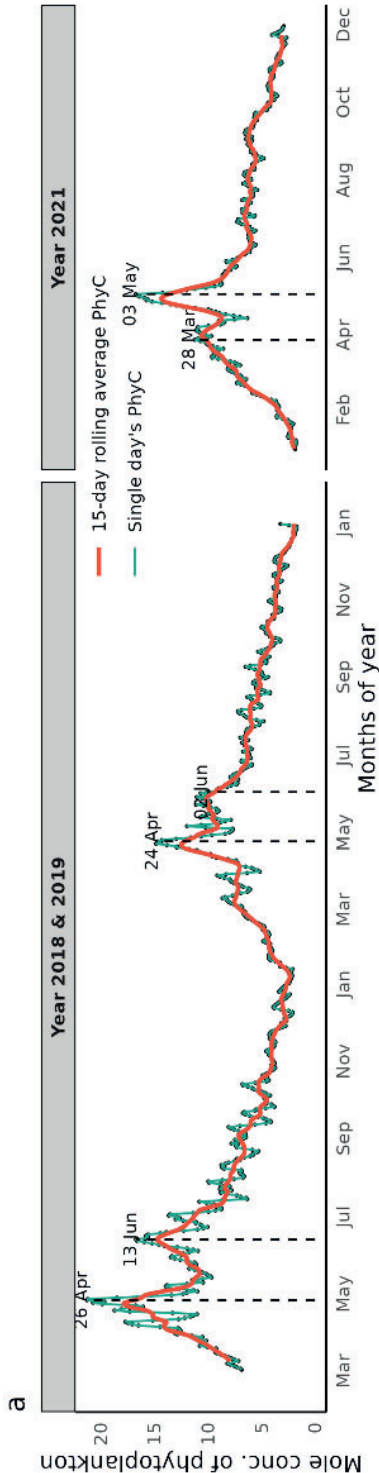
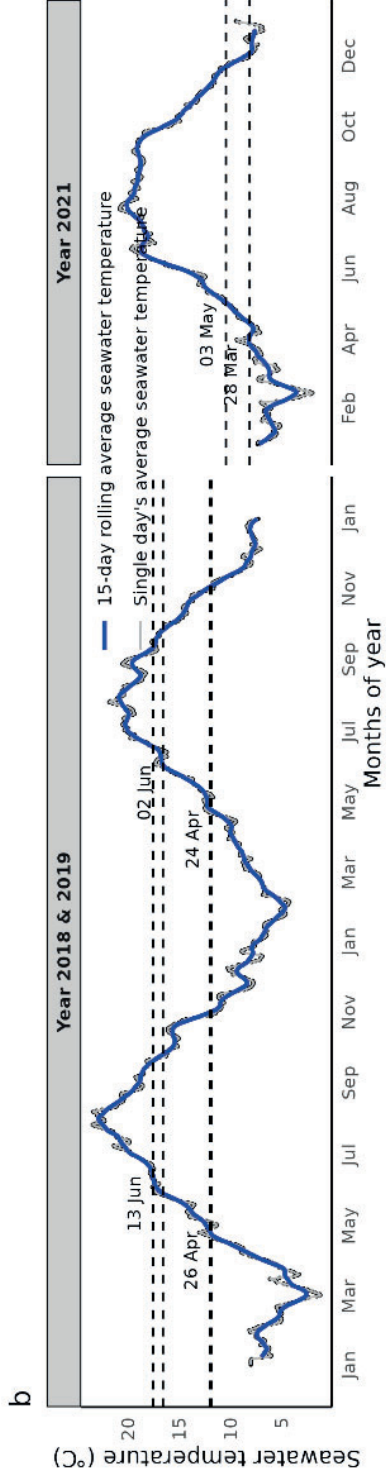


Figure S1: Rarefaction curve showing species richness as a function of sampling depth. The curve reaches saturation, indicating that the observed species richness stabilizes with increasing sampling efforts. This suggests that the majority of species within the community have been captured, with minimal new species being observed beyond the current sampling depth.

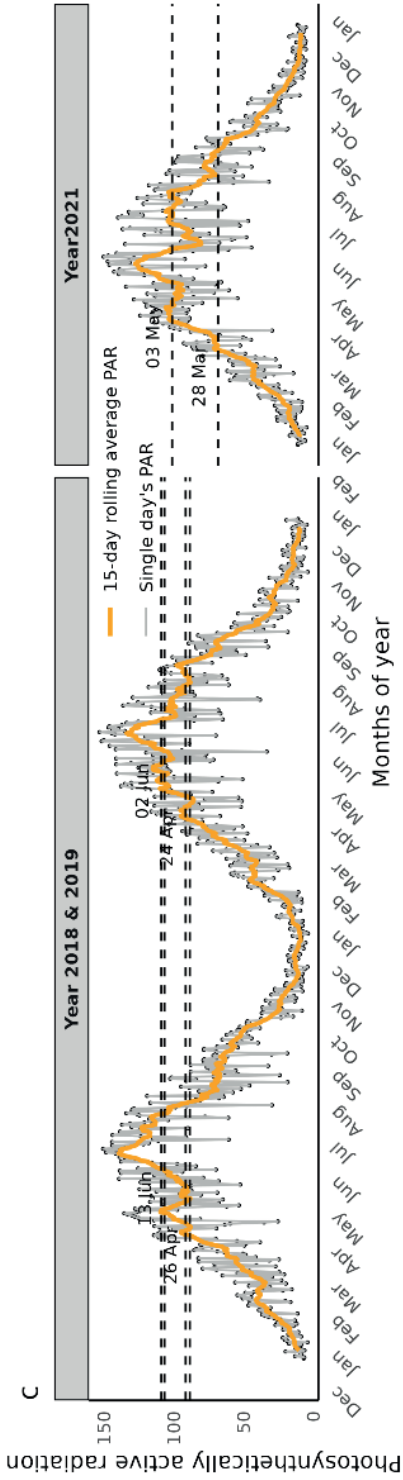
Figure S2: Time series of environmental variables and phytoplankton abundance over the years 2018, 2019, and 2021, with 2020 omitted (indicated by a vertical black dashed line).



(a) Phytoplankton absolute abundance (Phyc) shown as a 15-day rolling average (red line) alongside daily values (green line). Dashed vertical lines indicate the bloom dates for each year, capturing two primary bloom events per year.

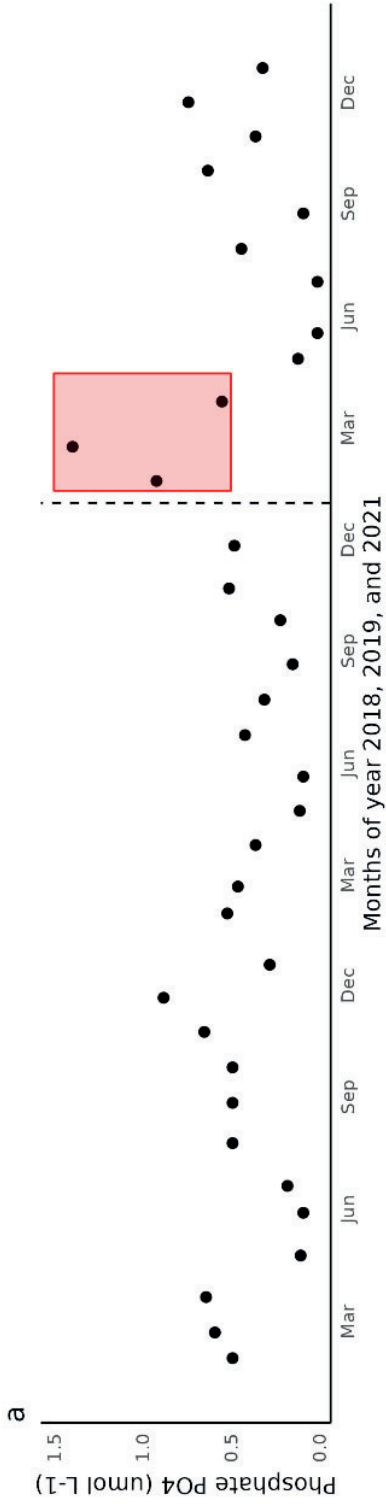


(b) Seawater temperature depicted with a 15-day rolling average (blue line) and daily averages (grey points). Horizontal dashed lines align with the bloom dates from panel (a), highlighting the seawater temperature corresponding to each bloom event.

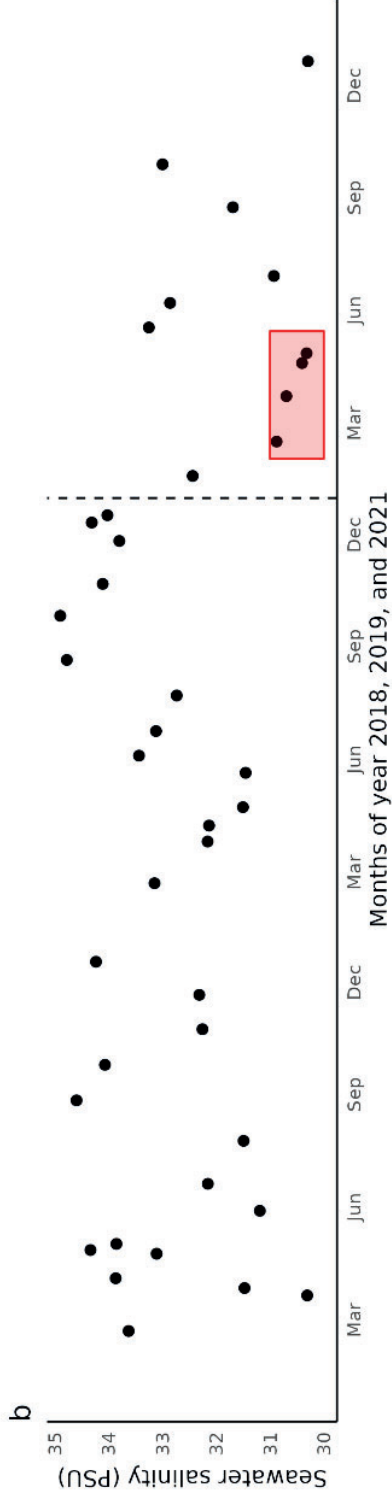


(c) Photosynthetically Active Radiation (PAR) represented by a 15-day rolling average (orange line) and daily values (grey points). Horizontal dashed lines align with the bloom dates from panel (a), showing the PAR levels associated with each bloom period.

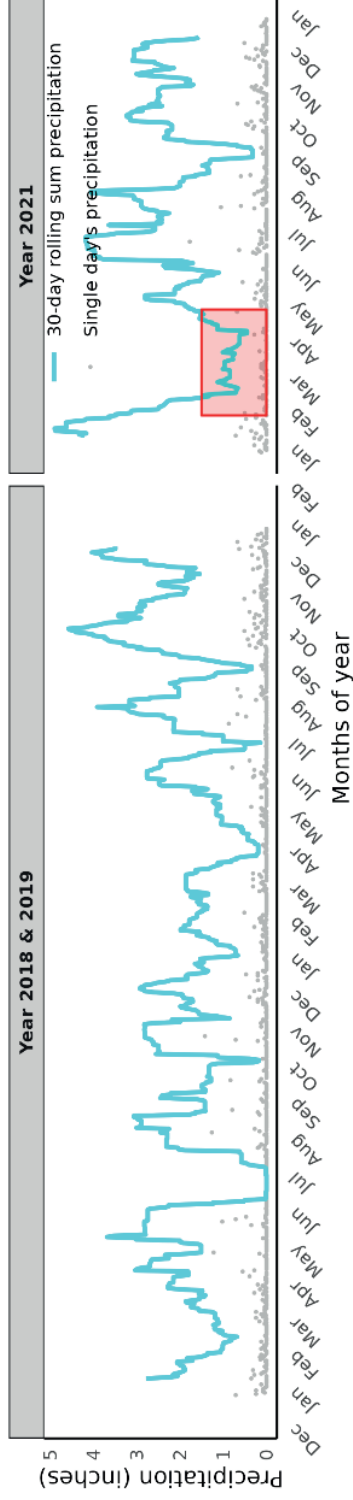
Figure S3: Time series of environmental variables in the research area for the years 2018, 2019, and 2021, with 2020 omitted (indicated by a vertical black dashed line).



(a) Seawater phosphate (PO_4) concentration ($\mu\text{mol L}^{-1}$) measured periodically throughout the study period. Data were obtained from the National Oceanic and Atmospheric Administration's Integrated Surface Database (NOAA-ISD) of the USAF-064070 station, which is located approximately 800 meters from the sampling site ($51^\circ 14' 25''\text{N}$, $2^\circ 55' 14''\text{E}$). Highlighted in red are regions with notable deviations in values when compared to previous years, indicating potential shifts in environmental conditions, further discussed in the manuscript's discussion section.



(b) Seawater salinity (PSU) across the same timeframe, showing variability over the seasons. Data were obtained from the National Oceanic and Atmospheric Administration's Integrated Surface Database (NOAA-ISD) of the USAF-064070 station, which is located approximately 800 meters from the sampling site ($51^{\circ}14'25''\text{N}$, $2^{\circ}55'14''\text{E}$). Highlighted in red are regions with notable deviations in values when compared to previous years, indicating potential shifts in environmental conditions, further discussed in the manuscript's discussion section.



(c) Precipitation levels recorded as daily values (grey points) and a 30-day rolling sum (cyan line), capturing seasonal precipitation patterns. Data were obtained from the National Oceanic and Atmospheric Administration's Integrated Surface Database (NOAA-ISD) of the USAF-064070 station, which is located approximately 800 meters from the sampling site ($51^{\circ}14'25''\text{N}$, $2^{\circ}55'14''\text{E}$). Highlighted in red are regions with notable deviations in values when compared to previous years, indicating potential shifts in environmental conditions, further discussed in the manuscript's discussion section.

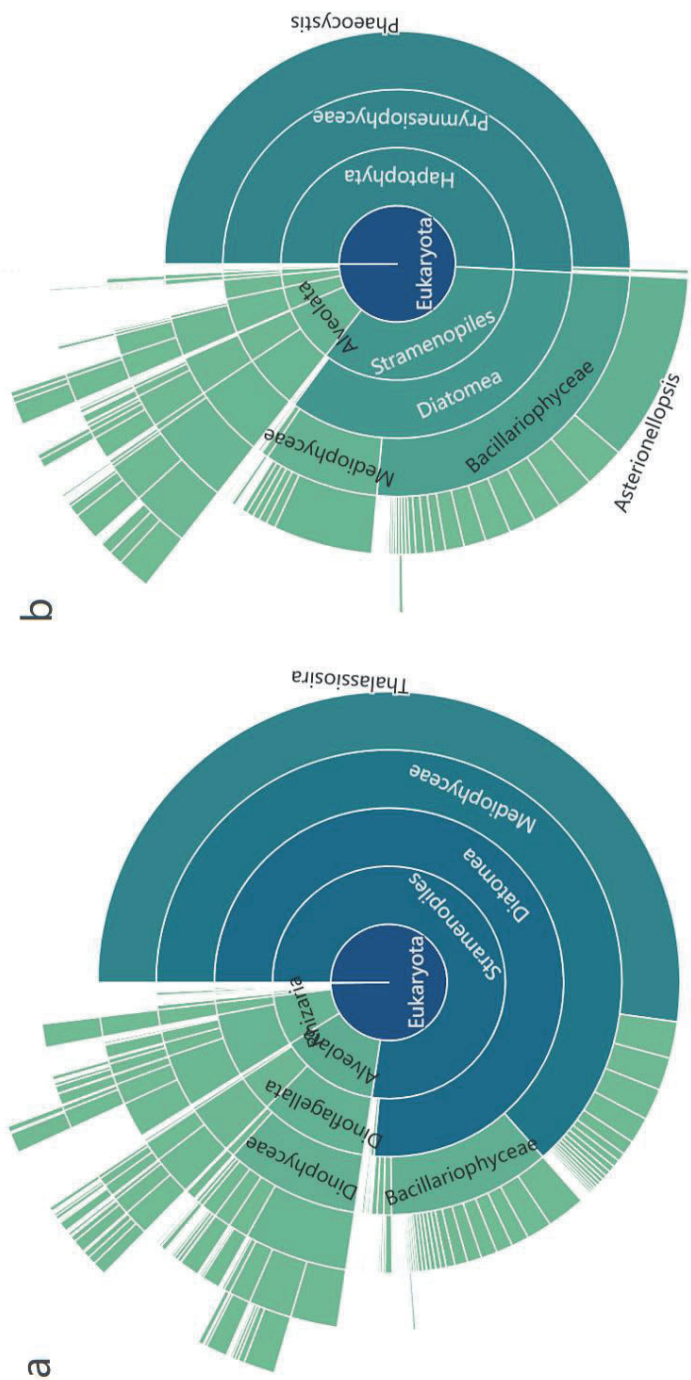


Figure S4. Sunburst plots showing the community composition of plankton samples collected on (a) 2018-05-24 and (b) 2019-04-10. The hierarchical structure of the plots represents taxonomic classifications, with

Eukaryota at the core expanding outward to specific groups. Key blooming species, Phaeocystis and Thalassiosira, are highlighted in each sample, illustrating their dominant presence during the bloom events on the respective dates. These plots provide a visual confirmation of the species responsible for significant bloom events in each year.

QC: Reads per barcode

Dashed line = median; highlighted = lowest barcode

Group ● Lowest-depth ● Others

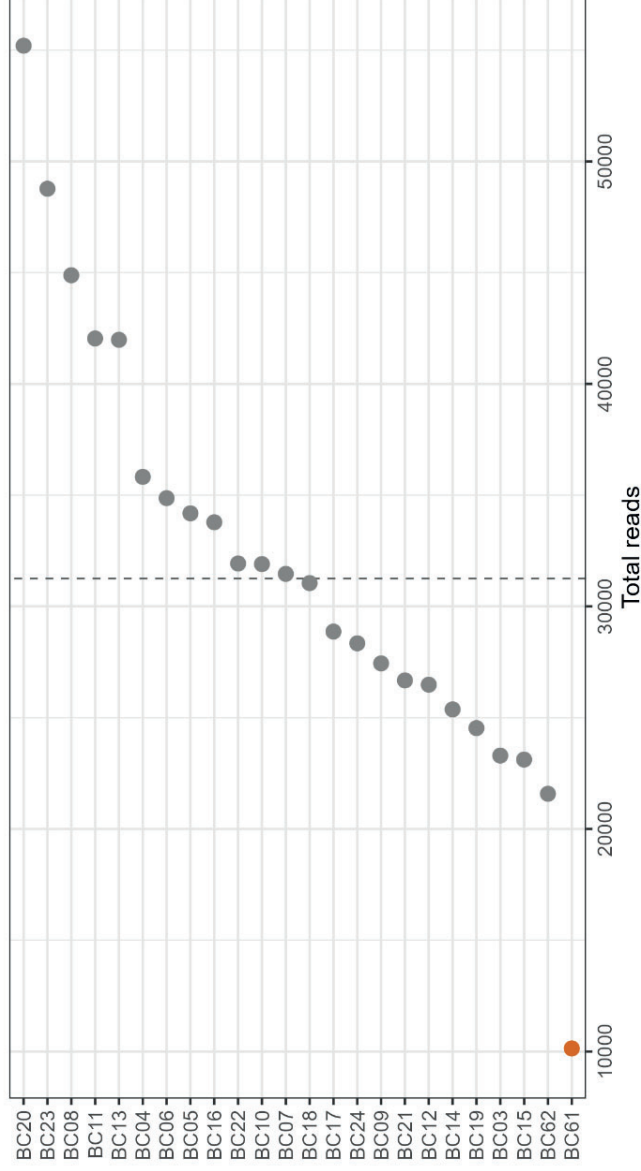


Figure S5. Sequencing depth per barcoded sample. Dot plot showing the total number of demultiplexed

Nanopore reads obtained for each sample. The dashed vertical line indicates the median read count across all barcodes. The lowest-depth barcode is highlighted to facilitate identification of potentially under-sequenced sample. While sequencing depth varies among samples, the lowest-depth barcode (BC61) remains within the same order of magnitude as the other barcodes (approximately three-fold lower than the median), indicating no extreme under-sequencing.

Chapter 5 Marine Natural Products and Their Enrichment Potential in Sea Spray Aerosols

This chapter is adapted from two papers under review.

Part 1: MarineNP: A Comprehensive and Locally Deployable Database of Marine Natural Products

Zixia Liu^{a,*}, Eveline Diopere^a, Colin R. Janssen^a, Jana Asselman^a

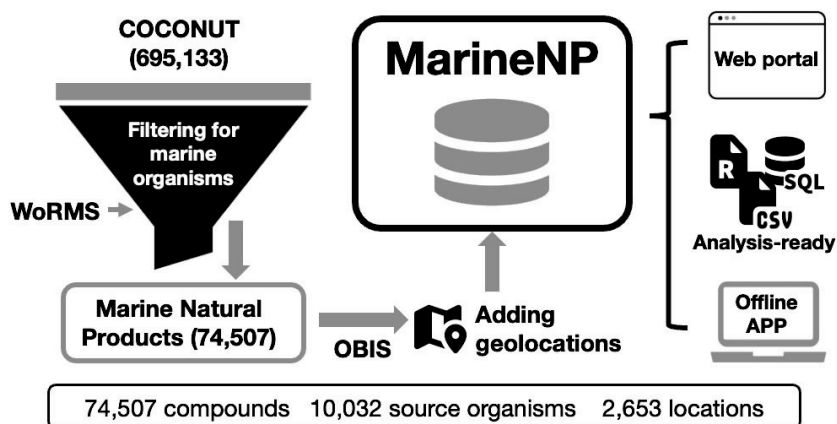
a: Blue Growth Research Lab, Ghent University, Wetenschapspark 1, Bluebridge, 8400 Oostende, Belgium.

*: Corresponding author: Zixia Liu (zixia.liu@ugent.be)

CRedit authorship contribution statement

ZL: Conceptualization, Data curation, Methodology, Software, Writing – original draft. **ED:** Data curation, Methodology, Writing – review & editing. **CJ:** Supervision, Writing – original draft. **JA:** Conceptualization, Project administration, Supervision, Writing – review & editing.

Graphical Abstract



Abstract

Marine natural products (MNPs) are a vital source of chemical diversity for drug discovery and chemical biology. However, researchers are often hindered by fragmented data sources, proprietary access models, and a lack of tools that support large-scale computational analysis or work in offline environments. We present MarineNP (<https://marinenp.scicloud.eu>), an open-access database designed to address these challenges. It offers three modes of access: (1) a powerful, web portal for interactive searching and data exportation; (2) complete database downloads in analysis-ready formats (SQL, CSV, RData); and (3) a cross-platform standalone application (Windows, macOS, Linux) that provides the full querying functionality in a fully offline environment. MarineNP improves existing resources by providing systematically processed and taxonomically data, including taxonomic validation against the World Register of Marine Species (WoRMS) and the inclusion of geolocation information for source organisms. In total, MarineNP currently contains 74,507 unique marine natural products, 10,032 unique source organisms, and 2,653 distinct collection locations.

Keywords

Marine Natural Products, Bioinformatics, Cheminformatics, Drug Discovery, Open-Source Software, Offline Database.

1. Introduction

The immense chemical diversity of marine natural products (MNPs) represents a crucial frontier for biomedical research, offering novel scaffolds for drug development and unique molecular probes to investigate biological processes. Effectively leveraging this chemical data is often constrained by the way it is stored and accessed. The landscape of MNP data is populated by several key resources, each with inherent limitations. Commercial databases, with MarinLit (Royal

Society of Chemistry, 2025) being a prominent forerunner, offer comprehensive and deeply curated datasets but are protected by a subscription paywall, limiting access for many academic researchers. On the open-access front, the landscape includes both specialized and generalized databases. Valuable marine-specific resources like the Comprehensive Marine Natural Products Database (CMNPD) have emerged, providing web-based portals to MNP data (Lyu et al., 2021). However, like many academic web services, their online availability can be inconsistent. At the same time, massive, generalized collections like COCONUT (Sorokina et al., 2021), the most comprehensive open collection of natural products to date, and other databases for example SuperNatural, NPAtlas, NPASS, LOTUS contain vast numbers of compounds but are not specific to the marine environment and thus lack marine-specific taxonomic validation and filtering relevant to marine biology (Gallo et al., 2023; Poynton et al., 2025; Rutz et al., 2022; Zeng et al., 2018). Other efforts exist as highly specialized, niche databases like the Seaweed Metabolite Database (SWMD), which are invaluable but focus only on a specific class of organisms (Davis and Vasanthi, 2011). Furthermore, a vast amount of data remains locked in static supplementary files accompanying publications in natural product-related journals, hindering its utility for systematic computational analysis.

A key bottleneck for bioinformaticians and researchers is the inability to easily integrate MNP data into local analysis pipelines. Web-only databases restrict large-scale querying and prevent offline use; their stability can be inconsistent and resource limitations often constrain complex user queries, creating significant barriers for researchers, especially those with limited internet access or handling sensitive workflows. While some databases offer data dumps, they rarely provide user-friendly tools to query this data locally.

To address this critical gap, we developed MarineNP, a comprehensive and locally deployable database of marine natural products. MarineNP

stands out by providing users with three access options: a user-friendly website, complete database downloads, and a new desktop app for offline analysis. It's designed to empower a wide range of research activities, from simple interactive queries to complex, high-throughput computational screening.

Table 1. Comparison with Existing Resources

#	Database	Reference	Number of Compounds	Version	Website Accessible*	Bulk Download	Offline Deploy	Marine Specialized
1	CMNPD	(Lyu et al. 2021)	31,561	1.0	No	Partial	No	Yes
2	COCONUT	(Sorokina et al. 2021)	695,133	10-2024	Yes	Yes	Yes	No
3	LOTUS	(Rutz et al. 2022)	276,518	44228	Yes	Partial	Yes	No
4	NPASS	(Zeng et al. 2018)	43,285	NPASS-2023	Yes	Yes	No	No
5	NPAtlas	(Poynton et al. 2025)	36,395	v2024_03	Yes	Yes	No	No
6	SuperNatural	(Gallo et al. 2023)	484,887	3.0	No	Yes	No	No
7	SWMD	(Davis and Vasanthi 2011)	1,191	2011	Yes	No	No	No
MarineNP			74,507	v1	Yes	Yes	Yes	Yes

*. Accessed on 2025-06-08

Feature comparison between MarineNP and publicly available natural product databases. MarineNP is distinguished by its comprehensive marine specialization, coupled with full website accessibility, bulk download, and offline deployment capabilities.

2. Materials and methods

2.1. Data collection

The MarineNP database integrates data from three primary open-access resources. The foundational chemical data was sourced from the COCONUT database (Sorokina et al., 2021), with the specific dataset (*coconut-dump-10-2024.sql.zip*) downloaded from its Zenodo repository with record number 13897048 and imported into a PostgreSQL database for further processing. Taxonomic validation and organism classification were performed using the World Register of Marine Species (WoRMS) (WoRMS Editorial Board, 2025), accessed via the Lifewatch Species Information Backbone (Vandepitte et al., 2019) for efficient batch matching. Geolocation information for the source organisms was retrieved from the Ocean Biodiversity Information System (OBIS) (Ocean Biodiversity Information System, 2025) using its API (<https://api.obis.org/v3/occurrence>) with the WoRMS AphiaID as the identifier.

2.2. Data processing

To create a specifically marine-focused dataset, we implemented a two-stage pipeline. First, an initial taxonomic validation was performed by matching all source organism names from the COCONUT dataset against WoRMS using the Lifewatch service. This initial pass resolved many names but left a subset with ambiguous or "fuzzy" matches. To resolve these ambiguities, we used the *worms* R package (v0.4.3) (Chamberlain and Vanhoorne, 2025) to programmatically re-query each fuzzy case against the WoRMS API and retained the top-ranked taxonomic match returned by WoRMS. This step was intended as a standardized taxonomic reconciliation rather than a manual taxonomic verification of every record. Organisms were assigned an *is_marine* flag based on their habitat environment recorded in WoRMS; organisms flagged as "Marine" or "Brackish" were retained for the final dataset. In the second stage, our Go backend systematically queried the OBIS API

for occurrence information using the official WoRMS AphiaID for each validated marine organism. The retrieved geolocation data is stored in a local cache table (*obis_cache*), which is updated regularly. This caching strategy significantly reduces the query load on the public OBIS API while ensuring the data remains current.

3. Data description

3.1. Database content

The entire process mentioned in 2.1 and 2.2 resulted in the current version of MarineNP. MarineNP is available online for interactive search and can also be downloaded in multiple file formats, including SQL, CSV, and RData. The database currently contains 74,507 unique marine natural products. Each molecular entry is defined by multiple standard chemical identifiers, including IUPAC name, Canonical SMILES, CAS number, standard InChI, and InChIKey, and is annotated with 38 computed physicochemical properties, such as molecular weight, AlogP, and topological polar surface area. The dataset also includes drug-likeness metrics like Lipinski's Rule of Five violations and QED scores, alongside natural product likeness values and detailed chemical and biosynthetic classifications via ClassyFire and NP-Classifier. The biological context for each molecule is provided through links to 10,032 unique source organisms, which are taxonomically validated with WoRMS AphiaID, and geographical context is established through connections to 2,653 distinct collection locations. The relational integrity of the database is maintained through linking tables that map over 130,000 molecule-organism and 39,000 molecule-location associations.

3.2. Implementation

MarineNP was built with a technology stack chosen for high performance and ease of local deployment. The entire database is stored in SQLite3, which provides a fast, file-based, and serverless solution that makes it trivial to deploy locally on all devices. The backend

server was written in Golang, selected for its high performance and its ability to compile into a single binary with no external dependencies. The web portal's user interface features a clean front end built with standard HTML and JavaScript, ensuring a responsive and user-friendly experience across all modern browsers, while facilitating simple local deployment. This architecture allows a user to run the entire MarineNP service locally with a single executable command.

The source code is available on GitHub:

<https://github.com/bluegrowthlab/MarineNP>.

4. Discussion

4.1. User interface and functionality

MarineNP provides three integrated access modes designed for diverse research workflows. The web portal offers intuitive data exploration through the following interfaces: 1. Keyword Search: Enables cross-field text matching (compound names, identifiers, SMILES, CAS numbers) with partial matching and case-insensitive functionality; 2. Advanced Search: Supports complex queries across 38 physicochemical properties (e.g., molecular weight ranges, Lipinski violations, QED scores) and biological metadata (taxonomic classifications, geographic origins). Filters are combinable for precision targeting compounds; 3. Details View: Presents comprehensive compound profiles with interactive navigation between chemical classifications, biological sources (linked to WoRMS), and geospatial distributions (sourced from OBIS). The cross-platform standalone application replicates full web functionality in offline environments. Offline deployment refers to local querying of a synchronized database release, not continuous real-time updating of WoRMS or OBIS-linked records. The analysis-ready file (available in SQL/CSV/RData) downloads facilitate pipeline integration. All interfaces share a unified schema ensuring consistency across access modes.

4.2. Comparative Advantages Over Existing Resources

MarineNP addresses critical gaps in MNP data accessibility by offering marine-specific taxonomic validation and filtering comparable to existing databases, for example CMNPD, yet surpassing them through integrated offline access, taxonomic validation via WoRMS, and geolocation integration. The detailed comparison with existing resources is listed in Table 1.

4.3. Conclusion

Despite the immense chemical diversity of marine natural products (MNPs), their potential for drug discovery is often constrained by data accessibility. In this context, integrative analyses of public data are crucial to uncover novel chemical scaffolds and biological activities. Today, many MNP datasets are publicly available, but significant cheminformatics knowledge is necessary to integrate and analyze them effectively. To bridge this gap between raw data and scientific discovery, interactive and locally deployable databases are valuable tools that enable exploratory analysis without requiring deep computational expertise. This type of resource has proven successful in other biological domains, enabling the reuse of public data and the generation of new hypotheses.

A resource of this nature is urgently needed in the field of MNP research to: (1) Compile data from fragmented sources into a single, marine-focused portal, (2) Provide access to taxonomically validated, marine-filtered data in analysis-ready formats, and (3) Empower researchers with full database functionality in an offline environment. We developed the MarineNP database to accomplish all of these objectives. To the best of our knowledge, MarineNP is the first resource to provide a comprehensive MNP database alongside a fully-featured, cross-platform, standalone application for offline use. We expect it to become a key reference in the field. In the coming years, MarineNP will be regularly updated to expand its data content and incorporate new

functionalities based on user feedback.

Part 2: A Quantitative Workflow for Assessing the Physicochemical Fate of Natural Compounds in Sea Spray Aerosol.

Zixia Liu^{a,*}, Yunmeng Li^b, Eveline Diopere^a, Maarten De Rijcke^b, Colin R. Janssen^a, Jana Asselman^a

a: Blue Growth Research Lab, Ghent University, Wetenschapspark 1, Bluebridge, 8400 Oostende, Belgium.

b: Flanders Marine Institute (VLIZ), Research Division, Ocean and Human Health, InnovOcean Campus, Jacobsenstraat 1, 8400, Oostende, Belgium

*: Corresponding author: Zixia Liu (zixia.liu@ugent.be)

CRedit authorship contribution statement

ZL: Conceptualization, Data curation, Methodology, Software, Writing – original draft. **YL:** Methodology, Data curation, Writing – review & editing. **ED:** Data curation, Methodology, Writing – review & editing. **MDR:** Methodology, Writing – review & editing. **CJ:** Supervision, Writing – original draft. **JA:** Conceptualization, Project administration, Supervision, Writing – review & editing.

Abstract

Sea spray aerosol (SSA) is a major natural particle source and a direct pathway for ocean-derived chemicals to enter coastal air, yet chemical attribution is hindered by complex mixtures and sparse compound identification. We present a reproducible, process-informed screening workflow to prioritize marine natural compounds that are physically plausible contributors to SSA organic composition. Using MarineNP (74,507 candidates) as a structure-resolved hypothesis space, we integrate physicochemical descriptors with screening-level EPI Suite estimates to represent three necessary steps for exposure relevance: enrichment at the sea surface microlayer (SSML), aerosolization and condensed-phase retention, and atmospheric persistence. Two mechanistic descriptor modules (enrichment; aerosolization) were

clustered into interpretable physicochemical families (16 and 18 clusters) using a stable two-stage k-means strategy. Atmospheric half-lives from AOPWIN provided an orthogonal fate filter, with OH lifetimes used as a conservative discriminator. We then anchored clusters to literature evidence (53 reference records) to summarize cluster-level enrichment and aerosolization factors and assess rank stability. Combining top-ranked clusters with an OH persistence threshold (>0.5 h) produced prioritized sets of 20,926 (enrichment) and 12,878 (aerosolization); their overlap contained 6,138 candidates, reduced to 4,364 in the triple intersection with stability. Forty literature references fell within this final set. The workflow yields an interpretable, updateable prioritization map and candidate list to guide targeted SSA chemical analysis and exposure-relevant hypothesis testing.

Keywords: Sea spray aerosol; sea surface microlayer; marine natural products; physicochemical screening; clustering; vapor pressure; atmospheric lifetime; OH oxidation; chemical prioritization; exposure relevance

1. Introduction

Sea spray aerosol (SSA) is a dominant natural source of particulate matter to the marine atmosphere and a key physical pathway linking ocean surface processes to coastal air composition (Grythe et al., 2014; O'Dowd and De Leeuw, 2007). The SSA is generated primarily by wave-driven bubble entrainment and bursting, producing film and jet droplets that rapidly evaporate to form particles spanning submicron to super micron sizes (Lewis and Schwartz, 2004; Prather et al., 2013). Although sea salt provides much of the mass in many SSA size fractions, decades of field and laboratory work demonstrate that SSA can be substantially enriched in organic material relative to bulk seawater, particularly in smaller particles, reflecting selective transfer of surfactants, biopolymers, and other marine-derived organics (Facchini et al., 2008; Quinn et al., 2014). This enrichment matters for exposure because SSA

represents a direct and recurrent coastal inhalation route for ocean-derived chemical and biological material, with concentrations and composition varying with wind forcing, sea state, and ecological conditions (Cochran et al., 2017).

Interest in SSA has expanded beyond climate-relevant aerosol–cloud interactions to include potential consequences for human health. Multiple studies have shown that SSA and SSA-derived extracts can elicit measurable, composition-dependent responses in human cell models. Under harmful algal bloom (HAB) conditions, SSA is more likely to carry algal toxins, and exposures have been associated with adverse endpoints such as reduced cell viability, increased cytotoxicity, oxidative stress, and pro-inflammatory signaling (Cheng et al., 2010; Vila et al., 2016). By contrast, under non-bloom or “ambient” marine conditions, SSA exposures have in several reports produced beneficial responses, including a transcriptomic signature consistent with mTOR-pathway inhibition and reduced PCSK9 (Proprotein Convertase Subtilisin/Kexin Type 9) expression in human lung cells, resembling the effect of a pharmacological mTOR inhibitor and interpreted as potentially health-promoting at environmentally relevant doses (Asselman et al., 2019; Van Acker et al., 2020). While the direction and magnitude of responses depend on dose, particle size, and source conditions, the collective evidence supports the idea that SSA is not uniformly benign and that its health relevance is context dependent, emerging from the interplay between aerosol composition, exposure intensity, and physiological targets. A persistent limitation, however, is chemical attribution: even when SSA-associated bioactivity is reproducible, the specific molecular drivers are rarely identified, and effects often reflect complex mixtures rather than single agents (Cochran et al., 2017; Zhang and Du, 2017).

This attribution challenge is amplified by the chemical complexity of the marine metabolome and of aerosol organic matter. High-resolution mass spectrometry of SSA frequently detects thousands of molecular features, yet only a minority can be annotated with high confidence, and

quantitative information is often limited by low collected mass, matrix interferences, and the scarcity of available reference standards and library compounds as many marine natural products remain to be structurally elucidated and described (Bertram et al., 2018; Schmitt-Kopplin et al., 2012). Consequently, a purely measurement-driven strategy struggles to move from “many unknown signals” to mechanistic hypotheses about which classes of marine natural compounds are physically likely to be transferred from the ocean into airborne particles and therefore plausibly contribute to exposure-relevant aerosol chemistry and bioactivity. This motivates complementary approaches that can prioritize compounds before targeted analytical investment, using physicochemical constraints based on the processes that govern ocean-to-air transfer and post-emission fate (Bertram et al., 2018).

Marine natural product (MNP) resources provide a tractable starting point for such prioritization because they define an explicit, structure-resolved candidate space that can be screened even when environmental measurements remain only partially annotated. In this work, we therefore leverage MarineNP, a marine natural product database developed as a companion resource that consolidates known MNP structures from publicly available records, harmonizes chemical identifiers, and provides a standardized set of computed physicochemical descriptors to support high-throughput, property-based analyses (Chapter 5 Part A). Although MarineNP cannot capture the full molecular diversity of dissolved and particulate organic matter in seawater, it offers a chemically explicit hypothesis space spanning major marine metabolite classes (e.g., lipids, terpenoids, alkaloids, polyketides, glycosides). Crucially, because each entry is linked to a defined molecular structure, MarineNP enables mechanistically motivated filtering, clustering, and prioritization in a way that remains informative even when SSA and SSML high-resolution mass spectrometry datasets contain many features that cannot yet be confidently identified. In the present study, MarineNP is therefore used not as a catalog of confirmed SSA constituents, but as a systematic

candidate universe in which compounds can be ranked and organized according to their physicochemical plausibility for SSML enrichment, aerosolization into SSA, and subsequent persistence in air.

We propose that marine natural compounds of interest for aerosolization into SSA can be prioritized using three necessary (though not sufficient) criteria: (i) enrichment in the SSML, and (ii) potential for aerosolization, (iii) resilience against atmospheric degradation. Together, these criteria reflect a minimal physical logic for exposure relevance: a compound must (a) be available at the interface where SSA droplets originate, (b) be transferable into aerosol particles and retained in the condensed phase to a meaningful extent, and (c) persist long enough against oxidative loss to support chronic, low-dose exposure in coastal populations, thereby creating potential for health-relevant impacts over time.

First, resilience against atmospheric degradation constrains whether an emitted compound can remain intact in air. Many organic species undergo rapid oxidation by OH radicals and ozone, leading to fragmentation, functionalization, and changes in volatility that alter gas–particle partitioning and lifetime (Atkinson and Arey, 2003; George and Abbatt, 2010). For coastal exposure, the aerosol mixture can shift from continental-dominated to surf-influenced essentially immediately once air crosses the coastline, because the surf zone acts as a very local and intense SSA source (Vignati et al., 2001). At larger near-field scales, coastal circulations such as sea-breeze trapping can return marine boundary-layer air toward land within ~2–3 hours after intrusion, keeping SSA-derived material within the exposure-relevant coastal zone (Pace et al., 2015). These short transport windows make atmospheric persistence a basic screening constraint. We therefore use screening-level estimates of oxidative loss to distinguish compounds likely to degrade rapidly from those more likely to persist; the implementation is described in the Methods section.

Second, enrichment in the sea surface microlayer (SSML) is a prerequisite for efficient transfer of many organic compounds into sea spray aerosol (SSA) because SSA production via bubble bursting preferentially samples the ocean's immediate surface film, including the bubble-cap layer that drains and ruptures to generate film droplets (Cravigan et al., 2020; Schiffer et al., 2018). The SSML—often operationally defined as the upper tens to hundreds of micrometers of the ocean boundary layer—is widely recognized as physicochemically and biologically distinct from underlying seawater, with the capacity to concentrate surface-active material and biogenic macromolecules (e.g., lipids, proteins, polysaccharides) and to form hydrated, gel-like interfacial matrices whose enrichment varies with ecosystem state and physical forcing (Cunliffe et al., 2013; Engel et al., 2017; Wurl and Holmes, 2008). Mechanistically, selective accumulation in the SSML reflects interfacial adsorption and retention governed by hydrophobic interactions, hydrogen-bonding capacity, and steric/packing constraints within surface films and gels. The descriptor set used to operationalize this SSML-enrichment module is described in the Methods section.

Third, potential for aerosolization constrains whether an organic compound present at the sea surface can be transferred into SSA and remain in the condensed phase after emission, rather than being rapidly lost through evaporation or unfavorable phase partitioning. Fresh SSA contains both volatile and non-volatile components, and volatility measurements of nascent SSA (e.g., thermodenuder-based approaches) have been used to demonstrate the presence of a low-volatility organic fraction whose magnitude varies with seawater conditions and biological state (Cravigan et al., 2020; Rasmussen et al., 2017). Because condensed-phase retention is fundamentally governed by volatility and gas-particle partitioning, aerosolization propensity is most directly approximated using Estimated Vapor Pressure, consistent with absorptive partitioning theory and volatility-based frameworks that treat (effective) saturation vapor pressure as a primary control on particle-phase abundance (Donahue et al., 2012; Pankow, 1994). The

descriptor set used to operationalize this aerosolization module is described in the Methods section.

Here, we present a reproducible, process-informed screening framework to prioritize marine natural compounds that are physically plausible contributors to SSA organic composition. The approach integrates structure-resolved descriptors with screening-level property estimates to evaluate three necessary steps for exposure relevance—persistence in air, enrichment at the sea surface, and transfer into aerosol—and organizes candidates into interpretable physicochemical families. By producing mechanism-grounded, ranked candidate sets linked to known exemplars where available, the framework is designed to focus targeted chemical analysis on the most plausible drivers of SSA composition and downstream bioactivity.

2. Methods

2.1. Data sources and initial preprocessing

The candidate universe was derived from MarineNP (version 1, r20250608; Zenodo distribution), using two tables: *molecules.csv* (identifiers and structures) and *properties.csv* (computed physicochemical descriptors). These tables were merged on the unique molecule identifier to create a unified structure–descriptor table. To ensure compatibility with downstream structure-based estimation tools, entries lacking valid structural identifiers (missing/empty SMILES) were removed. In addition, a small set of known problematic records (molecule IDs: 546623, 685189, 71730) was excluded to prevent parsing/estimation failures and to maintain a fully reproducible pipeline. Canonical SMILES strings were extracted and used as the primary structure input for external property estimation.

To augment the MarineNP descriptors with screening-level atmospheric and volatility properties, we used US EPA EPI Suite™ v4.1. Two EPI Suite modules were executed using canonical SMILES as input: (i) AOPWIN v1.93 to estimate gas-phase reaction rate constants for OH radicals and ozone (O₃), and to compute corresponding atmospheric half-lives under AOPWIN default assumptions. (ii) MPBPWIN v1.44 to estimate vapor pressure (VP) and supporting thermophysical predictors including boiling point (BP) and melting point (MP). EPI Suite produces text-based reports; these were programmatically parsed using dedicated import scripts (*scripts/03_1_import_aopwin_features.R* and *scripts/03_2_import_mpbpwin_features.R* in source code repository) to extract endpoint values, harmonize units/field names, and merge them back into the MarineNP feature table. The resulting integrated dataset is exported as *chemical_merged_features.csv*, which served as the common input for downstream processing and clustering.

2.2. Descriptor definition and module-specific feature sets

Feature selection was guided by mechanistic relevance to interfacial enrichment and aerosol partitioning processes and implemented as two clustered mechanistic modules (enrichment and aerosolization) that share a few common descriptors along with their own unique features, plus a non-clustered stability module based on half-life endpoints. The shared descriptors were included in both clustered modules to retain consistent information on molecular size/architecture, hydrophobicity, charge state, ring structure, and saccharide motifs, while module-specific descriptors were added to capture process-relevant variance unique to enrichment or aerosolization. The complete descriptor inventory used in the workflow with module membership, units, mechanistic rationale, and key citations is provided in Table 1.

For the enrichment module, hydrophobic partitioning and amphiphilicity are pragmatically approximated by ALogP, while polarity and hydration constraints at the interface are represented by Topological Polar

Surface Area, and specific polar interaction capacity is captured by Hydrogen Bond Donors and Hydrogen Bond Acceptors (Ertl et al., 2000; Kelly et al., 2004). Because interfacial films can impose structural selectivity beyond bulk solubility—favoring molecules whose conformational freedom and rigidity permit accommodation within condensed organic layers—molecular flexibility is represented by Rotatable Bond Count (Veber et al., 2002), while ring-driven rigidity is complemented by Number of Minimal Rings (Braun et al., 2024). Aromaticity is explicitly tracked using Aromatic Rings Count, motivated by both mechanistic and observational evidence that aromatic species exhibit strong surface enhancement at the air–water interface and that polycyclic aromatic hydrocarbons and related hydrophobic organics can be enriched in the SSML relative to subsurface waters (Cincinelli et al., 2005; Vácha et al., 2006). To distinguish more saturated, lipid-like architectures from flatter aromatic scaffolds—relevant to association with surfactant-rich SSML material—structural saturation is represented by Fraction Csp3 (Lovering et al., 2009). Finally, molecular size constraints on diffusion and interfacial accommodation are represented by Molecular Weight and Total Atom Count, consistent with adsorption frameworks in which interfacial partitioning scales with molecular surface area/size terms (Kelly et al., 2004). Given the documented importance of carbohydrate-rich material in the SSML/SSA continuum, we additionally include a motif indicator, Contains Ring Sugars, to flag candidates plausibly associated with SSML gels/foams and with saccharide enrichment pathways observed in nascent SSA (Jayarathne et al., 2016).

For the aerosolization module, to emphasize this distinct step—SSML-to-SSA transfer and post-emission retention—we intentionally restrict the aerosolization mechanistic module to a compact set of descriptors that capture volatility-relevant structure while minimizing overlap with SSML-enrichment proxies. In addition to Estimated Vapor Pressure, molecular size is represented by Molecular Weight and Total Atom Count, supported by evidence that evaporation behavior correlates with

both vapor pressure and molar mass (Li et al., 2018). We retain ALogP to reflect affinity for organic-rich condensed phases that can stabilize partitioning within mixed SSA organic material (Pankow, 1994). Structural effects on volatility are further represented by Number of Minimal Rings, consistent with findings that ring structure and isomerism can materially alter vapor pressures (Dang et al., 2019). Finally, we include Contains Ring Sugars the same reason as discussed in the enrichment module (Jayarathne et al., 2016).

Table 1. Descriptor set used in the enrichment, aerosolization, and stability mechanistic modules

Variable	Full Name	Unit	Module (s)*	Process relevance / rationale	References
total_atom_count	Total Atom Count	count	EA	Size proxy; relates to interfacial accommodation/packing and volatility/evaporation tendencies	Kelly et al., 2004; Li et al., 2018
heavy_atom_count	Heavy Atom Count	count	EA	Size proxy excluding H; complements total atoms for structural scaling	Kelly et al., 2004; Li et al., 2018
molecular_weight	Molecular Weight	g/mol	EA	Size proxy; affects diffusion/adsorption and volatility-driven partitioning	Kelly et al., 2004; Li et al., 2018
exact_molecular_weight	Exact Molecular Weight	g/mol	EA	High-precision MW; retained as complementary size descriptor	Kelly et al., 2004; Li et al., 2018
alogp	ALogP	log unit	EA	Hydrophobicity/amphiphilicity proxy for interfacial adsorption and organic-phase affinity in mixed SSA	Kelly et al., 2004; Pankow, 1994
formal_charge	Formal Charge	e	EA	Charge-state proxy; influences solubility and ion-organic interactions in saline interfacial films/SSA	Burrows et al., 2016; Jayarathne et al., 2016
number_of_minimal_rings	Number of Minimal Rings	count	EA	Rigidity/packing proxy; ring structure can affect accommodation and volatility	Braun et al., 2024; Dang et al., 2019

contains_sugar	Contains Sugar	—	EA	Saccharide motif indicator; consistent with saccharide enrichment/transfer pathways into nascent SSA	Jayarathne et al., 2016
contains_ring_sugars	Contains Ring Sugars	—	EA	Cyclic saccharide motif; flags common monosaccharide-like substructures	Jayarathne et al., 2016
contains_linear_sugars	Contains Linear Sugars	—	EA	Linear saccharide motif; flags (poly)saccharide-like architectures	Jayarathne et al., 2016
topological_polar_surface_area	Topological Polar Surface Area	A ²	E	Polarity/hydration proxy; constrains interfacial adsorption and retention in SSML films/gels	Ertl et al., 2000; Kelly et al., 2004
rotatable_bond_count	Rotatable Bond Count	count	E	Flexibility proxy; influences packing/ordering in condensed interfacial layers	Veber et al., 2002
hydrogen_bond_acceptors	Hydrogen Bond Acceptors	count	E	Polar interaction capacity; supports hydration and specific interactions within SSML matrices	Kelly et al., 2004
hydrogen_bond_donors	Hydrogen Bond Donors	count	E	Polar interaction capacity; as above	Kelly et al., 2004
aromatic_rings_count	Aromatic Rings Count	count	E	Aromaticity proxy; aromatic species often show strong air–water interfacial enhancement	Cincinelli et al., 2005; Vácha et al., 2006
fractioncsp3	Fraction Csp3	ratio	E	Saturation proxy; distinguishes lipid-like architectures from flatter aromatic scaffolds	Lovering et al., 2009
van_der_walls_volume	Van der Waals Volume	—	E	Steric/packing proxy; contributes to interfacial accommodation and adsorption scaling	Kelly et al., 2004

bp_est	Boiling Point (Estimated)	°C	A	Volatility-related proxy supporting retention/evaporation behavior	Donahue et al., 2012; Pankow, 1994
mp_est	Melting Point (Estimated)	°C	A	Phase-state proxy (solid vs. liquid); influences condensed-phase persistence in particles	Koop et al., 2011
vp_est	Vapor Pressure (Estimated)	mm Hg	A	Primary volatility driver for gas-particle partitioning and condensed-phase retention after emission	Donahue et al., 2012; Pankow, 1994
aop_oh_half_life_hours	AOP OH Half-life	h	S	Screening-level atmospheric persistence vs OH oxidation; used directly (not clustered)	Atkinson and Arey, 2003; US EPA, 2015
aop_ozone_half_life_hours	AOP O ₃ Half-life	h	S	Screening-level atmospheric persistence vs ozonolysis; used directly (not clustered)	Atkinson and Arey, 2003; US EPA, 2015

*: E = SSML enrichment mechanistic module (clustered); A = aerosolization/condensed-phase retention mechanistic module (clustered); EA = included in both clustered modules; S = atmospheric stability module (not clustered; assessed directly using AOPWIN-derived half-life estimates).

2.3. Statistical feature cleaning and transformation

For the enrichment and aerosolization clustered modules, descriptors underwent a standardized cleaning pipeline prior to clustering: (1) Near-zero variance (NZV) filtering: descriptors with >95% identical values were removed. (2) Missingness filtering: variables with >40% missing values were removed, except for theoretically central endpoints (notably vapor pressure) which were retained and accompanied by a Boolean missingness indicator to preserve coverage information without imposing ungrounded numeric imputation. In practice, estimated vapor pressure (`vp_est`) was treated as a critical variable and retained despite elevated missingness because it represented the primary volatility descriptor in the aerosolization module. (3) Collinearity pruning: highly correlated features were identified using Spearman correlation ($\rho > 0.95$). This pruning step was intentionally conservative and aimed to reduce redundancy rather than enforce strict orthogonality among descriptors. For each correlated pair, the feature with lower missingness and/or higher mechanistic priority was retained. Mechanistic priority was assigned explicitly for a small set of central descriptors, including `vp_est`, `molecular_weight`, `exact_molecular_weight`, `heavy_atom_count`, and `total_atom_count`. As a result, some strongly correlated but sub-threshold descriptor pairs were intentionally retained when they preserved complementary mechanistic information. Feature distributions were then assessed for skewness. Variables with absolute skewness > 1.0 were log-transformed using $\log_{10}(x)$ for strictly positive values and $\log_{10}(x + 1)$ for non-negative, count-like variables. Because estimated vapor pressures were frequently extremely small and occasionally reported as zeros, vapor pressure was treated as left-censored at the lower bound and transformed as $\log_{10}(\max(\text{vp\_est}, \epsilon))$, where ϵ was set to one order of magnitude below the smallest positive `vp_est` in the dataset (with a conservative minimum floor). Finally, all continuous variables were standardized (z-score normalization) to ensure equal weighting in distance-based clustering.

2.4. Assessment across mechanistic modules

Candidate compounds were assessed across three mechanistic modules representing atmospheric stability, SSML enrichment, and aerosolization/condensed-phase retention. For the enrichment and aerosolization modules, we applied unsupervised learning independently to identify physicochemical families aligned with each process step. Candidate clustering algorithms (k-means, hierarchical clustering with Ward's linkage, and CLARA) were compared using average silhouette width and clustering stability across 25 bootstrap iterations, and the final clustering used a two-stage hierarchical k-means procedure consisting of (i) global k-means partitioning with $k = 10$, followed by (ii) iterative sub-clustering of high-variance clusters (top 25% by within-cluster variance) into 2–5 sub-clusters when the split improved silhouette score by >0.1 , enabling coarse-to-fine resolution while limiting over-fragmentation and preserving interpretability. In contrast, the stability module was evaluated without clustering: atmospheric persistence was quantified directly using AOPWIN-derived half-life estimates for OH and O₃ reactions (`aop_oh_half_life_hours`, `aop_ozone_half_life_hours`), which were carried forward as screening-level fate metrics to rank candidates within and across enrichment/aerosolization clusters and to support downstream filtering consistent with exposure-relevant residence times; compounds lacking half-life estimates were retained but flagged as missing to maintain transparent interpretation.

2.5. Literature-based cluster annotation using reference compounds

To connect the unsupervised enrichment and aerosolization clusters to empirical evidence, we performed literature-based cluster annotation using a curated set of reference compounds for which published measurements (or reported enrichment/aerosolization factors) exist. This step is intended to contextualize and rank clusters by observed

behavior, not to train a predictive model or to redefine cluster boundaries. Reference information was entered through a lightweight data-entry interface and stored as a tidy table with the following fields: (i) identifiers and provenance (molecule_id, source_type (e.g., Field Sample, Wetlab), reference_doi, remarks, timestamp), (ii) input mode (input_type $\in$ {Concentrations, Direct Factor}), (iii) concentration inputs where available (conc_bsw, conc_sml, conc_ssa for bulk seawater, sea surface microlayer, and sea spray aerosol, respectively; missing allowed), and (iv) directly reported factors (direct_factor_value, direct_factor_type). For records entered as concentrations, enrichment and aerosolization factors were derived according to the definition used in the original source; where Na^+ -normalized enrichment factors were reported or could be computed (i.e., sufficient Na^+ information was available), we used the Na^+ -normalized form rather than raw concentration ratios. For records entered as Direct Factor, published ratios were captured verbatim in direct_factor_value and interpreted using direct_factor_type, where the “/” denotes the compartment-to-compartment ratio as reported (e.g., SSA relative to BSW) rather than a formula to be re-derived. Each reference compound was then mapped to its assigned enrichment and/or aerosolization cluster, and cluster-level factor distributions (e.g., range and central tendency on an appropriate scale) were used to provide an empirical basis for ranking clusters by relative likelihood of SSML enrichment or SSA aerosolization potential. This reference-compound-based calibration is conceptually aligned with established metabolomics practices in molecular networking, where sparse library-identified compounds are used to annotate and interpret (and in some workflows propagate annotations across) unsupervised molecular families (da Silva et al., 2018). The current labeled set was assembled opportunistically and is not the result of a systematic literature search, serving as a proof-of-concept demonstrating how sparse empirical anchor data can be integrated to calibrate and interpret physicochemical clusters.

2.6. Software and reproducibility

All analyses were implemented in R (≥ 4.0) using tidyverse, cluster, factoextra, fastcluster, ClusterR, shiny, plotly, DT, and shinyjs. The workflow is organized as script modules from data acquisition through estimation parsing, feature cleaning/transformation, clustering, visualization, and labeling, with intermediate and final outputs written to structured data/ and reports/ directories of the project repository (<https://github.com/bluegrowthlab/ssa-natural-compounds-fate>) to support complete reruns.

3. Results

3.1. Data integration and final descriptor spaces

MarineNP (v1, r20250608) contained 74,507 candidate compounds. After filtering for valid SMILES and excluding three records that failed SMILES parsing due to nonstandard/unsupported stereochemical annotations (IDs 546623, 685189, 71730),

74,504 compounds remained and were carried forward to the integrated feature table. Merging MarineNP descriptors with EPI Suite outputs (AOPWIN, MPBPWIN) produced an integrated dataset with 73 descriptors, of which 22 were retained as process-relevant candidates based on the mechanistic rationale in Table 1. Module-specific cleaning produced compact descriptor spaces for clustering. The SSML enrichment module retained 11 descriptors, and the aerosolization module retained 6 descriptors. All retained descriptors were complete after filtering, except estimated vapor pressure, which still contained 838 missing values. Collinearity pruning reduced redundancy while preserving mechanistic coverage. After pruning, the maximum retained pairwise Spearman correlation was $|\rho|_{\max} = 0.942$ for the enrichment module and $|\rho|_{\max} = 0.918$ for the aerosolization module. Some highly correlated descriptor pairs were intentionally retained below the predefined pruning threshold. Specifically, the maximum retained correlation occurred between `topological_polar_surface_area` and

hydrogen_bond_acceptors in the enrichment module and between total_atom_count and molecular_weight in the aerosolization module. Skewed variables were log-transformed and all continuous descriptors were z-score standardized to support distance-based clustering. Figure 1 summarizes the final descriptor spaces used for clustering.

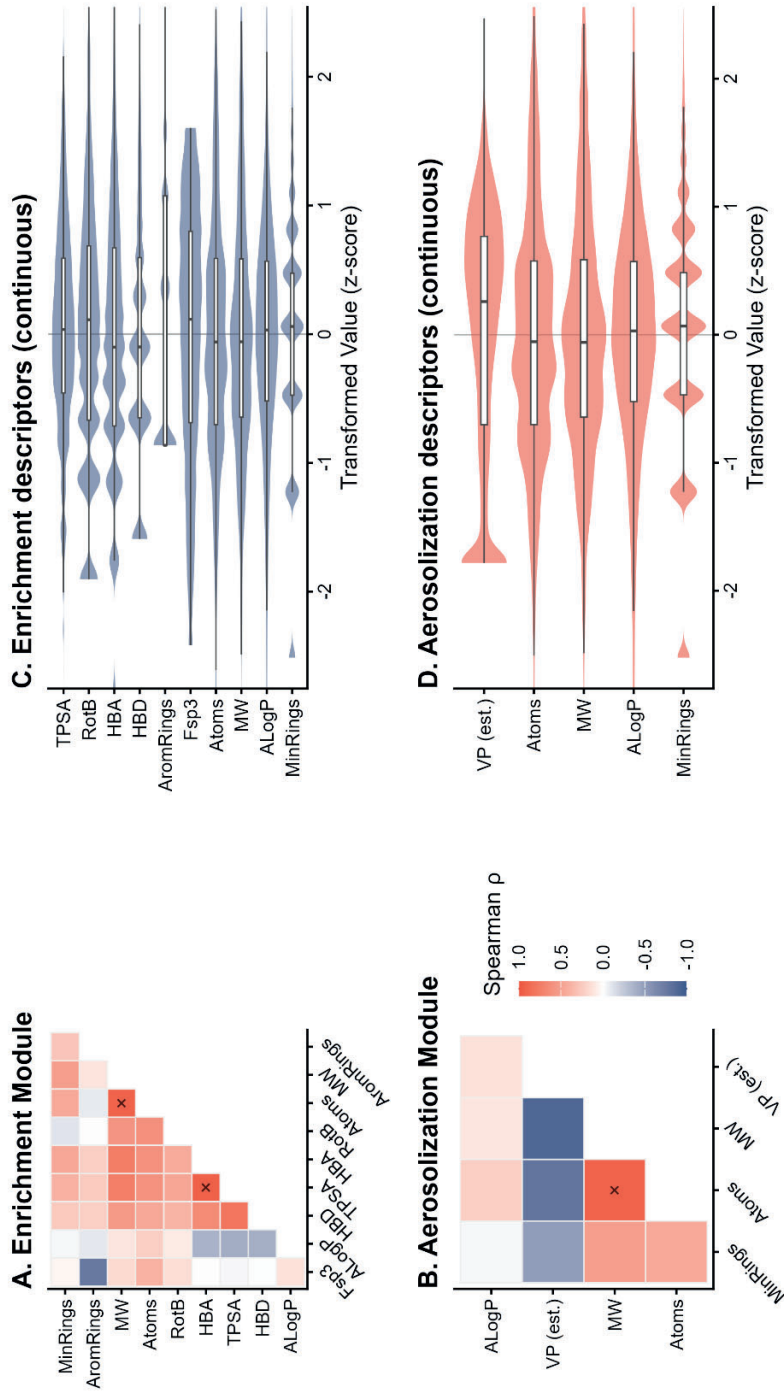


Figure 1 | Final descriptors used for module-specific clustering. (A–B) Upper-triangular heatmaps of pairwise

Spearman correlations among retained descriptors in the enrichment (A) and aerosolization (B) modules after collinearity pruning, illustrating residual correlation structure in the final feature sets (color scale shows Spearman ρ ; $\times$ marks indicate highly correlated pairs). (C–D) Distributions of the continuous descriptors used for clustering in the enrichment (C) and aerosolization (D) modules after transformation (log-transformation where needed) and z-score standardization. Violin shapes represent the full distribution and embedded boxplots summarize median and interquartile range; values are shown on the transformed (z-score) scale.

3.2. Module results: enrichment, aerosolization, and atmospheric stability

To select a robust clustering approach (i.e., both the clustering method and the cluster number, k) for each mechanistic module, we compared k -means, Ward hierarchical clustering, and CLARA across a range of candidate cluster numbers using average silhouette width with 25 bootstrap iterations (Figure 2). In both descriptor spaces, k -means consistently achieved the highest mean silhouette values and exhibited minimal bootstrap variability. For the enrichment module, k -means peaked at $k = 6$ (mean silhouette = 0.195) and remained relatively stable across larger k values (minimum mean silhouette = 0.167 across the evaluated range), outperforming CLARA (0.127–0.173) and Ward clustering (0.114–0.137). For the aerosolization module, k -means showed the strongest separation at low k (mean silhouette = 0.275 at $k = 3$) and retained comparatively high silhouettes across higher-resolution solutions (minimum mean silhouette = 0.213), again exceeding CLARA (0.164–0.270) and Ward clustering (0.157–0.263). Ward clustering showed substantially higher bootstrap variability than k -means or CLARA, consistent with greater sensitivity to resampling in these standardized descriptor spaces. Based on these comparisons, k -means was selected as the primary clustering algorithm for both modules, and a common base resolution of $k = 10$ was carried forward as the first-stage partition in the subsequent two-stage refinement strategy.

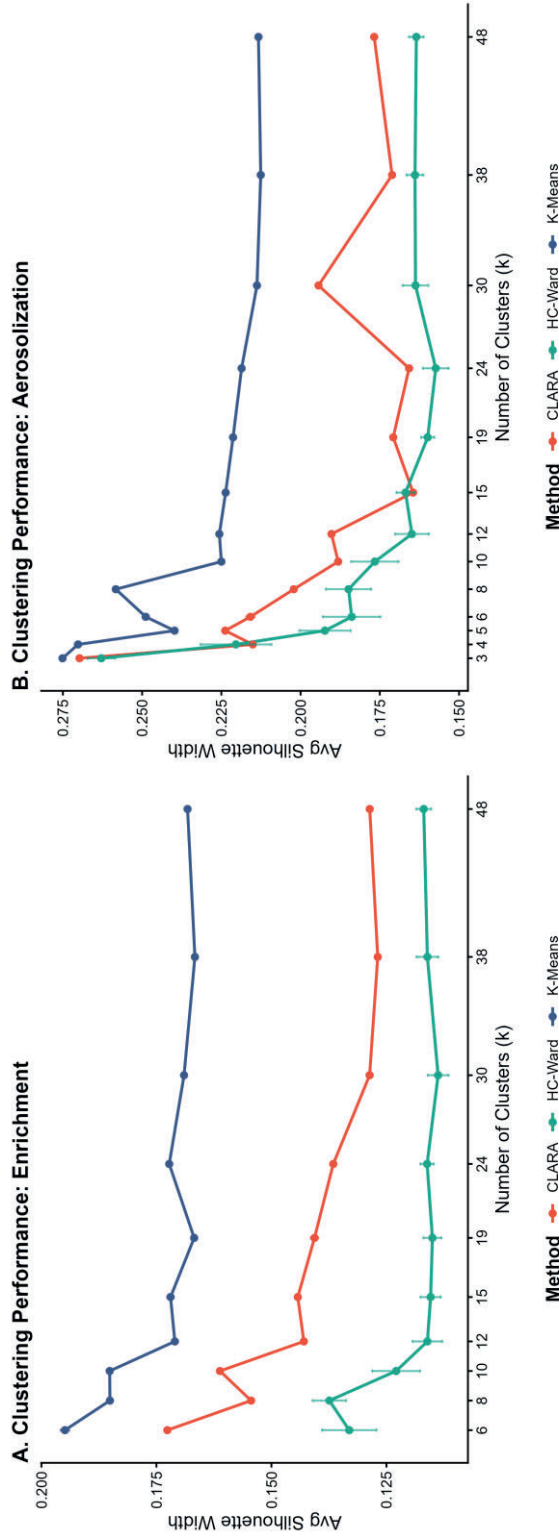


Figure 2 | Clustering performance across candidate methods. Mean silhouette width (25 bootstrap iterations) is shown as a function of the number of clusters (k) for the enrichment (A) and aerosolization (B) modules. K-means consistently achieved the highest silhouette values across k , while Ward hierarchical clustering showed greater variability across bootstrap iterations.

Following method selection (Figure 2), we applied a two-stage k-means strategy to obtain a compact global partition while increasing resolution in heterogeneous regions of descriptor space. For the enrichment module, an initial solution with $k = 10$ was refined by sub-clustering three high-variance parent clusters: cluster 1 was split into two sub-clusters (1_1–1_2), cluster 2 into two sub-clusters (2_1–2_2), and cluster 5 into five sub-clusters (5_1–5_5), yielding 16 final enrichment clusters in total. All 74,504 candidates were retained in enrichment clustering, with final cluster sizes ranging from 344 to 12,726 compounds (median = 2,827). The four largest clusters (4, 7, 10, and 9) together accounted for 55.7% of candidates, while the refined sub-clusters (IDs containing underscores) comprised 15.2%, indicating that refinement was focused on a minority of structurally diverse regions rather than broadly fragmenting the solution.

For the aerosolization module, clustering was performed on the subset with non-missing estimated vapor pressure (73,666 compounds), consistent with `vp_est` being the primary volatility driver for this module (838 compounds lacked `vp_est` and were retained in the integrated dataset but excluded from aerosolization clustering). The initial $k = 10$ partition was refined by splitting three high-variance parent clusters: cluster 10 into five sub-clusters (10_1–10_5), cluster 2 into three sub-clusters (2_1–2_3), and cluster 5 into three sub-clusters (5_1–5_3), yielding 18 final aerosolization clusters. Cluster sizes ranged from 254 to 15,776 compounds (median = 2,195). The three largest clusters (3, 1, and 8) accounted for 49.2% of candidates, while refined sub-clusters comprised 18.6%, reflecting targeted resolution of heterogeneous volatility/structure regimes.

Cluster fingerprints (median z-scores of retained descriptors within each cluster) revealed coherent physicochemical families aligned with the mechanistic intent of each module (Figure 3). In the enrichment module (Figure 3A), separation was driven primarily by a polarity/hydrogen-bonding axis (TPSA, H-bond donors/acceptors) contrasted with

hydrophobicity and structural architecture (ALogP, aromatic rings, fraction Csp³, flexibility and size proxies). A distinct high-polarity, H-bond-rich family emerged in the split cluster-5 group (5_1–5_5), characterized by elevated TPSA and hydrogen-bonding capacity, increased size (molecular weight/total atom count), and strongly reduced ALogP; sub-clusters within this family further differentiated flexibility/size and ring-sugar motifs. At the opposite extreme, clusters 2_1 and 2_2 represented very low-polarity profiles (strongly depleted TPSA and hydrogen-bonding capacity), while aromatic/low-saturation families (notably clusters 4 and 6) exhibited elevated aromatic ring counts and reduced fraction Csp³, consistent with more rigid, aromatic scaffolds.

In the aerosolization module (Figure 3B), fingerprints were dominated by a volatility–size gradient, separating higher vapor-pressure, smaller candidates (e.g., cluster 7 and cluster 1) from low-vapor-pressure families with increased molecular weight and total atom count. The split cluster-2 and cluster-5 families (2_1–2_3; 5_1–5_3) occupied an extreme low-vapor-pressure, high-size regime but differed strongly in hydrophobicity: cluster-5 subgroups were markedly hydrophilic (strongly reduced ALogP) and included sugar-motif-positive profiles, whereas cluster-2 subgroups spanned more intermediate hydrophobicity. In contrast, the split cluster-10 family (10_1–10_5) captured low-vapor-pressure candidates with comparatively higher ALogP (more hydrophobic retention-favored profiles), including the most hydrophobic subcluster (10_4). Together, these fingerprints indicate that the clustering resolved interpretable families reflecting distinct combinations of volatility-relevant structure and partitioning proxies, setting up the subsequent stability overlay and literature anchoring.

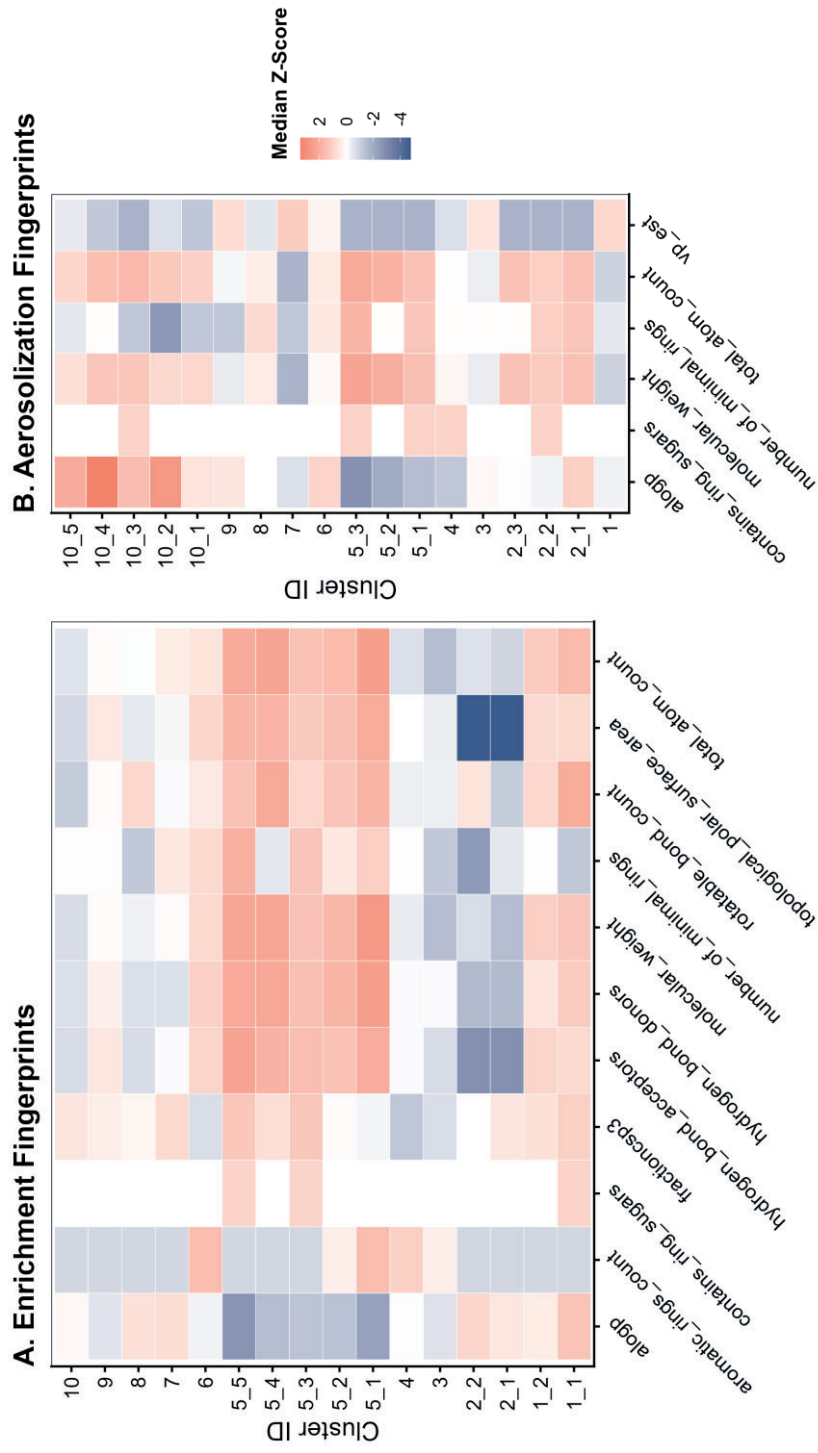


Figure 3 | Cluster fingerprints for enrichment and aerosolization modules. Heatmaps show the median z-score

of each retained descriptor within each final cluster after transformation and standardization. (A) Enrichment module fingerprints across 16 clusters derived from two-stage k-means refinement (underscored IDs denote second-stage sub-clusters), summarizing variation across polarity/hydrogen-bonding capacity, hydrophobicity, aromaticity/saturation, flexibility, and size proxies. (B) Aerosolization module fingerprints across 18 clusters (two-stage refinement), dominated by gradients in estimated vapor pressure (*vp_est*) and size (molecular weight, total atom count), with additional separation by hydrophobicity (*ALogP*), ring structure, and ring-sugar motifs. Aerosolization clustering was performed on compounds with non-missing *vp_est* (*n* = 73,666); compounds lacking *vp_est* were retained in the integrated dataset but excluded from aerosolization clustering.

Atmospheric persistence estimates derived from AOPWIN showed distinct stability regimes for OH-initiated oxidation and ozonolysis (Figure 4). OH half-life estimates were available for 73,490 of 74,504 candidates (98.6% coverage), whereas O₃ half-lives were available for 50,187 candidates (67.4% coverage), reflecting substantially higher missingness for the ozone endpoint. Using a conservative 30-min persistence threshold (Figure 4, dashed line), 69.8% of compounds with OH estimates (51,280/73,490) exceeded the cutoff, while 80.2% of compounds with O₃ estimates (40,252/50,187) exceeded the cutoff. Consistent with the broader distribution in Figure 4, ozone half-lives were often longer than OH half-lives when estimates were available, while OH provided a more stringent discriminator for rapid oxidative loss.

Overlaying stability onto the clustering solutions highlighted strong between-cluster differences in persistence that were not encoded in the descriptor spaces used for clustering. In the enrichment module, the fraction exceeding the 30-min cutoff spanned 5.98%–94.81% for OH and 60.30%–97.47% for O₃ across clusters. For example, cluster 3 exhibited high persistence against both oxidants (OH-stable 94.8%, O₃-stable 91.5%; median OH 1.60 h, median O₃ 4.03 h), whereas several sub-clusters within the refined cluster-5 family were strongly OH-limited despite relatively high O₃ stability (e.g., cluster 5_1: OH-stable 6.0% vs O₃-stable 84.7%; median OH 0.276 h, median O₃ 12.10 h). In the aerosolization module (vp_est non-missing subset; n = 73,666), cluster-level stability likewise varied widely (OH-stable 19.36%–97.76%, O₃-stable 58.17%–95.84%), with a high-persistence family represented by cluster 7 (OH-stable 97.8%, O₃-stable 94.4%; median OH 2.00 h, median O₃ 4.37 h) and a comparatively short-lived subgroup represented by cluster 10_5 (O₃-stable 58.2%; median OH 0.228 h, median O₃ 0.544 h). Together, these results support using half-life thresholds as an independent fate filter applied on top of enrichment/aerosolization families, while making endpoint-specific missingness explicit.

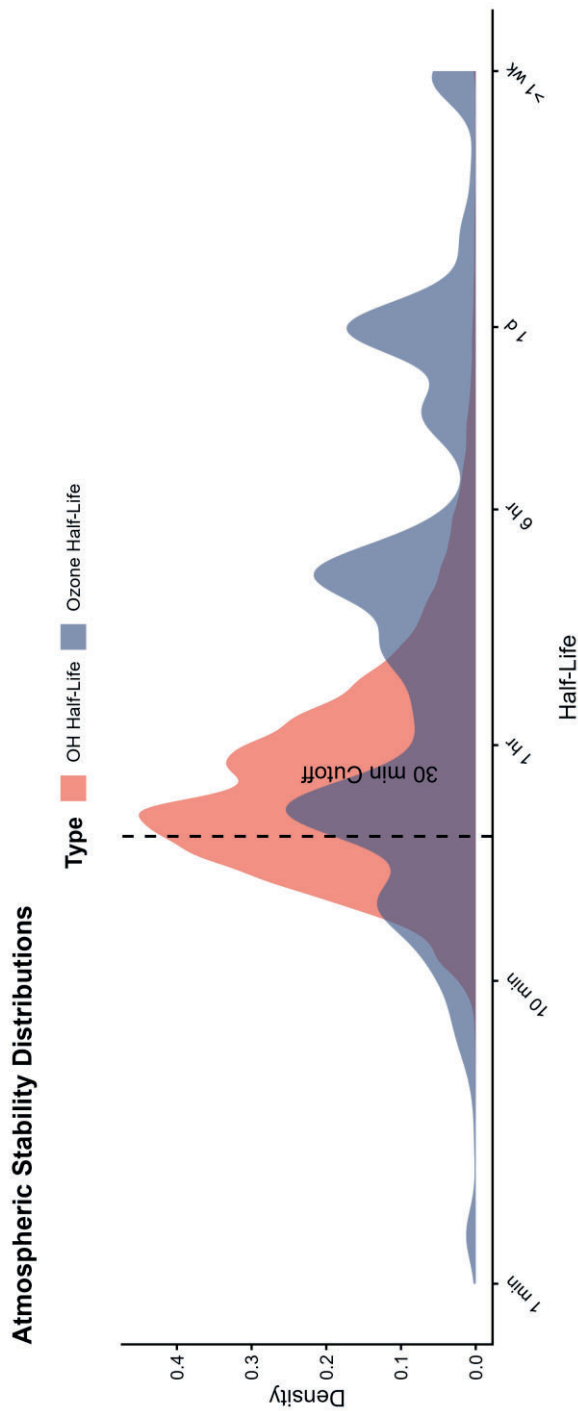


Figure 4 | Atmospheric stability distributions and stability-by-cluster overlay. (A) Kernel density distributions of AOPWIN-derived atmospheric half-lives against OH oxidation and O₃ reactions (log-scaled x-axis). The dashed

line indicates the 30-min persistence cutoff used for screening. OH half-lives were available for 73,490/74,504 candidates (1,014 missing), whereas O_3 half-lives were available for 50,187/74,504 candidates (24,317 missing).

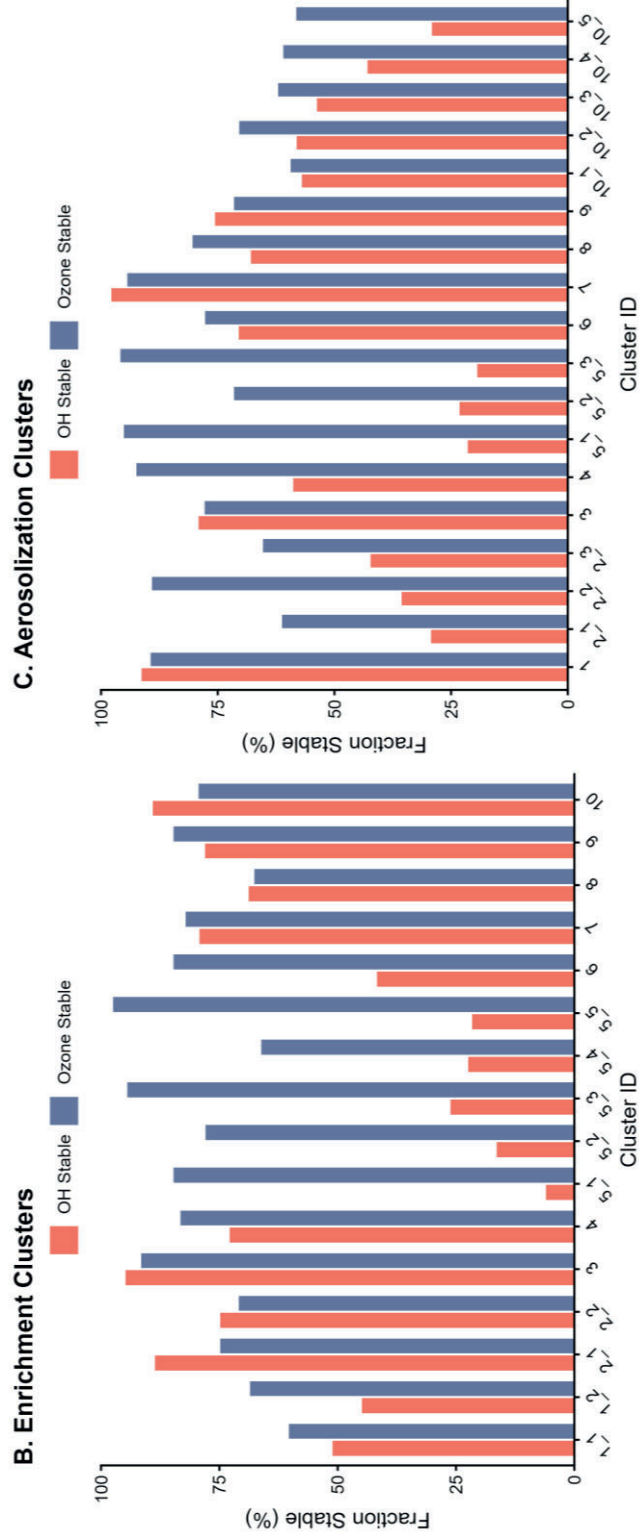


Figure 4 | Atmospheric stability distributions and stability-by-cluster overlay. (B–C) Fraction of compounds exceeding the 30-min cutoff within each enrichment cluster (B) and aerosolization cluster (C), reported separately for OH and O₃ among compounds with available estimates.

3.3. Empirical anchoring of clusters using literature reference compounds

To connect the unsupervised physicochemical families to empirical evidence, we curated a literature-derived reference set and mapped each reference compound to its assigned enrichment and/or aerosolization cluster. The reference table (SI Table S1) comprised 53 records of chemicals that have experimental or field evidence of enrichment or aerosolization. It included 42 with finite enrichment evidence ($\log_{10}$ EF) and 31 with finite aerosolization evidence ($\log_{10}$ AF). Enrichment references mapped to 6/16 enrichment clusters, while aerosolization references mapped to 4/18 aerosolization clusters (Figure 5). Because aerosolization clustering was restricted to candidates with non-missing vapor pressure, two aerosolization-evidence records were not cluster-mapped (mapped aerosolization evidence: $n = 29$).

Projecting the candidate space into two dimensions (PCA) and overlaying mapped references provides an interpretable view of where empirical anchors fall within each descriptor space (Figure 5A). Reference points occur both within high-density candidate regions and in distinct subregions, and cluster-level reference centroids indicate which physicochemical neighborhoods are supported by published evidence.

Cluster-level factor summaries showed clear differences in empirical behavior among the labeled families (Figure 5B). In the enrichment module, references were concentrated in cluster 3 ($n = 26$), which showed modest enrichment (median $\log_{10}$ EF ≈ 0.15 , $\sim 1.4\times$). Several other clusters exhibited higher median enrichment, including cluster 8 (median $\log_{10}$ EF ≈ 1.2 , $\sim 16\times$) and clusters 5_3, 2_1, and 1_2 (medians ≈ 0.6 – 0.7 , ~ 4 – $5\times$), while cluster 5_2 remained near unity enrichment (median $\log_{10}$ EF ≈ 0.0 , $\sim 1\times$). These distributions provide an empirical basis for ordering enrichment clusters by relative likelihood of SSML

enrichment while retaining within-cluster variability via bootstrap uncertainty.

In the aerosolization module, most mapped references fell in cluster 7 ($n = 24$), which showed a moderate median aerosolization signal (median $\log_{10} AF \approx 0.25$, $\sim 1.8\times$). Two clusters exhibited orders-of-magnitude higher aerosolization factors: cluster 2_1 (median $\log_{10} AF \approx 3.0$, $\sim 10^3\times$) and cluster 5_1 (median $\log_{10} AF \approx 2.7$, $\sim 5\times 10^2\times$). By contrast, cluster 4 showed slightly sub-unity aerosolization (median $\log_{10} AF \approx -0.13$, $\sim 0.7\times$). Together, these cluster-resolved summaries indicate that reference evidence differentiates aerosolization families spanning from near-unity behavior to strong transfer/retention regimes.

To assess sensitivity of cluster ordering to individual reference points, we evaluated rank stability using leave-one-out (LOO) recalculation of cluster median-factor ranks (Figure 5C). In the enrichment module, clusters with stronger internal support showed stable ordering (e.g., cluster 3 remained rank 5, cluster 5_2 remained rank 6), whereas the highest-ranked clusters with fewer anchors showed wider rank intervals (e.g., cluster 8 ranged 1–4 across LOO). In the aerosolization module, the top-ranked clusters were fully stable (cluster 2_1 always rank 1; cluster 5_1 always rank 2), while cluster 7 shifted modestly (ranks 2–3). These diagnostics support using cluster-level medians (with uncertainty) as a transparent calibration statistic for defining high-likelihood cluster sets used in downstream prioritization.

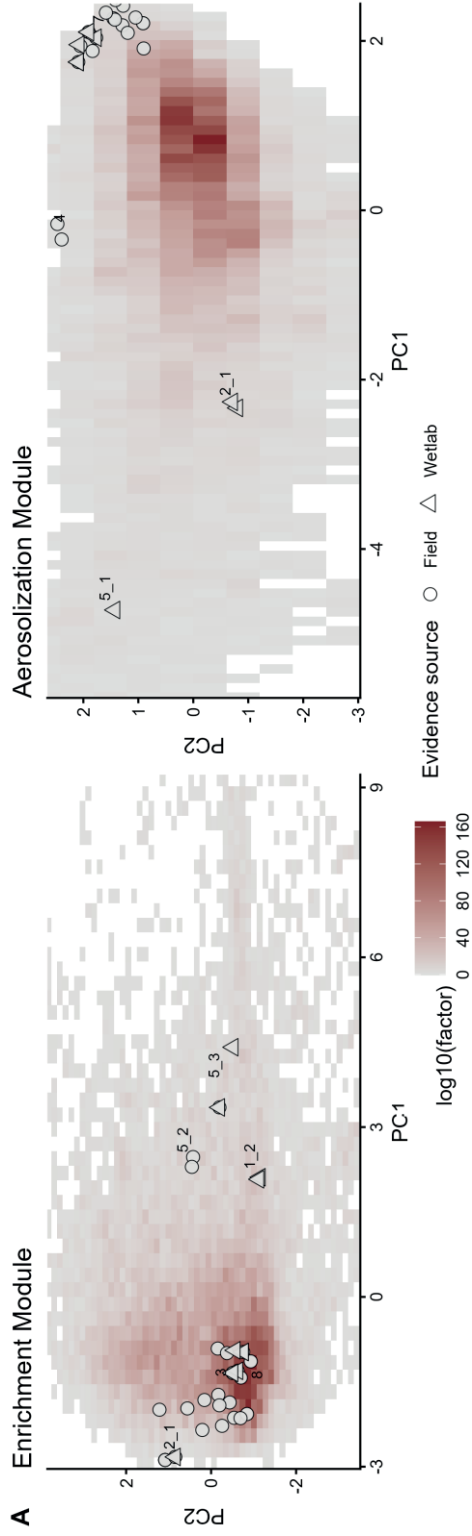


Figure 5 (A) | Literature-anchored calibration of enrichment and aerosolization clusters. Reference compounds overlaid on the module-specific PCA descriptor spaces for enrichment (left) and aerosolization (right). Background shading shows binned candidate density (computed from a representative candidate sample), while reference points are shown with shapes denoting evidence source (field vs wet-lab); labels indicate cluster-level reference centroids.

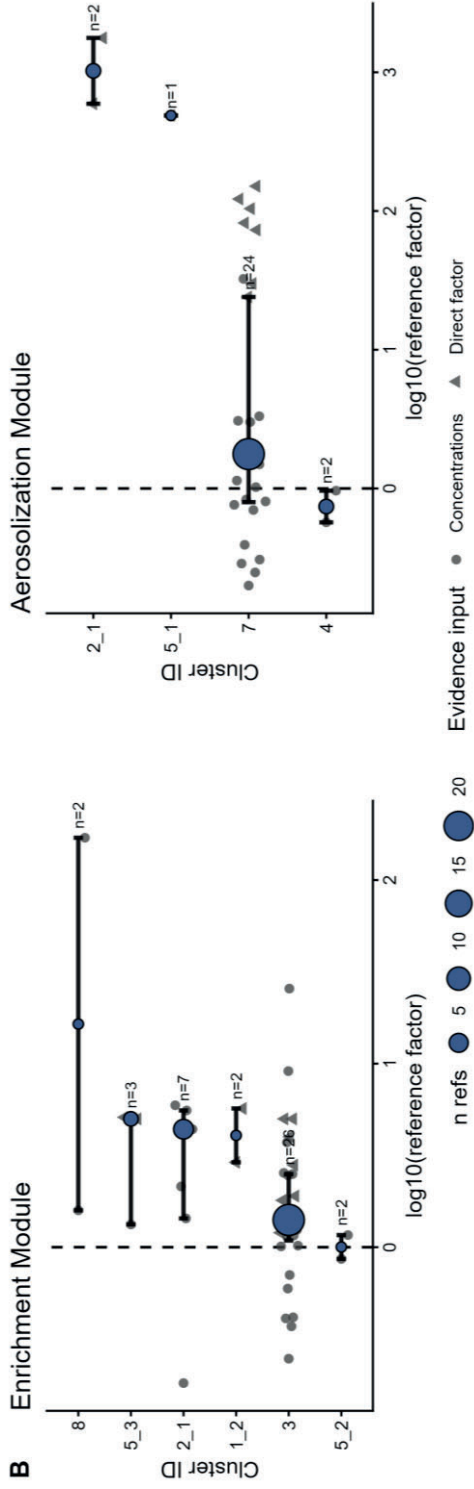


Figure 5 (B) | Literature-anchored calibration of enrichment and aerosolization clusters. Reference factor distributions by cluster for enrichment (left; log10 EF) and aerosolization (right; log10 AF). Points represent individual reference records (shape denotes evidence input: concentration-derived vs direct factor). Blue dots show cluster medians and whiskers show bootstrap 95% confidence intervals for the median; dot size scales with the number of reference records mapped to the cluster.

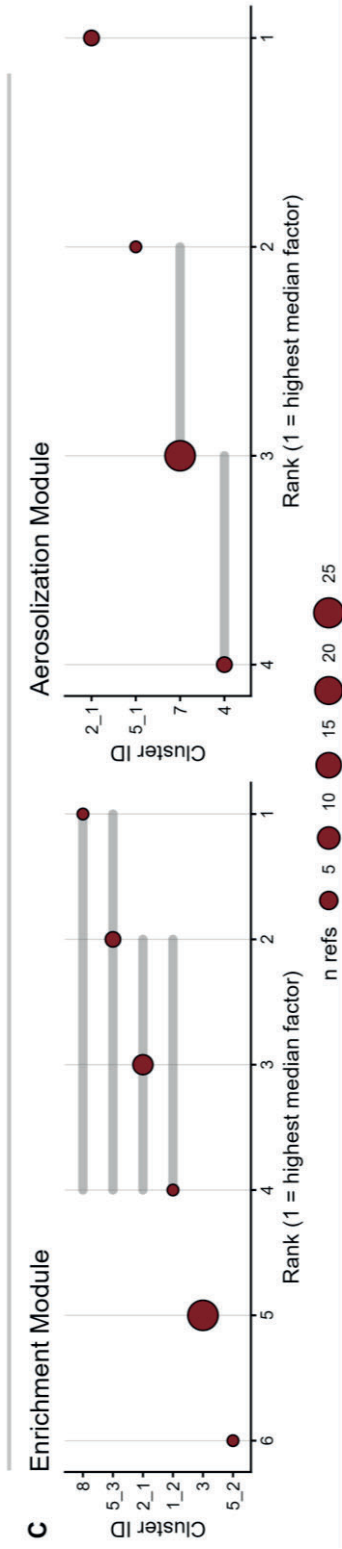


Figure 5 (C) | Literature-anchored calibration of enrichment and aerosolization clusters. Rank stability under leave-one-out (LOO) recalculation of cluster median-factor ranks (1 = highest median factor). Dots show the median LOO rank per cluster and horizontal segments show the min-max rank range across LOO iterations; point size scales with the number of reference records

3.4. Prioritized candidate sets and overlaps

We next combined the reference-calibrated cluster rankings (Section 3.3) with an atmospheric persistence filter to derive a focused candidate set that is simultaneously (i) likely to enrich at the sea surface, (ii) likely to transfer/retain in SSA, and (iii) sufficiently persistent for near-field coastal exposure. Prioritized sets were defined as all candidates assigned to the top-ranked enrichment clusters (Enrichment set; $n = 20,926$) and the top-ranked aerosolization clusters (Aerosolization set; $n = 12,878$). Atmospheric stability was defined using OH reactivity as a conservative discriminator, with Stable candidates requiring OH half-life > 0.5 h ($n = 51,280$).

Intersecting the two process-based sets produced 6,138 candidates that were simultaneously prioritized for enrichment and aerosolization ($E \cap A$; Figure 6A). Applying the stability criterion reduced this overlap to 4,364 candidates in the triple intersection ($E \cap A \cap \text{Stable}$), defining the final prioritized set for downstream investigation. Importantly, 40 literature reference records mapped into this triple intersection, indicating that the combined selection recovers a substantial fraction of the empirically anchored compounds while expanding the candidate space beyond the reference list via cluster-based generalization.

The overlap was driven by a small number of high-contributing clusters (Figure 6B). On the enrichment side, cluster 3 accounted for the majority of overlap candidates (3,694/6,138), and this contribution was largely retained after stability filtering (3,611 stable; $\sim 98\%$). The second-largest enrichment contributor was cluster 5_3 (1,078 total), but only a minority of these remained stable (288; $\sim 27\%$), illustrating how the OH half-life threshold selectively downweights certain physicochemical families within the $E \cap A$ set. On the aerosolization side, cluster 7 dominated the overlap (3,979/6,138), again with a high stable fraction (3,870; $\sim 97\%$). Additional overlap contributions came from cluster 5_1 (1,288 total; 277 stable; $\sim 22\%$) and cluster 2_1 (664 total; 99 stable; $\sim 15\%$). Together,

these patterns show that the final candidate list is shaped both by the evidence-guided cluster selection and by stability acting as an orthogonal fate constraint, preferentially retaining overlap candidates from the most persistent cluster families.

The full list of the candidates can be found in file
reports\07_module_results\prioritized_molecules_triple_intersection.csv of the repository.

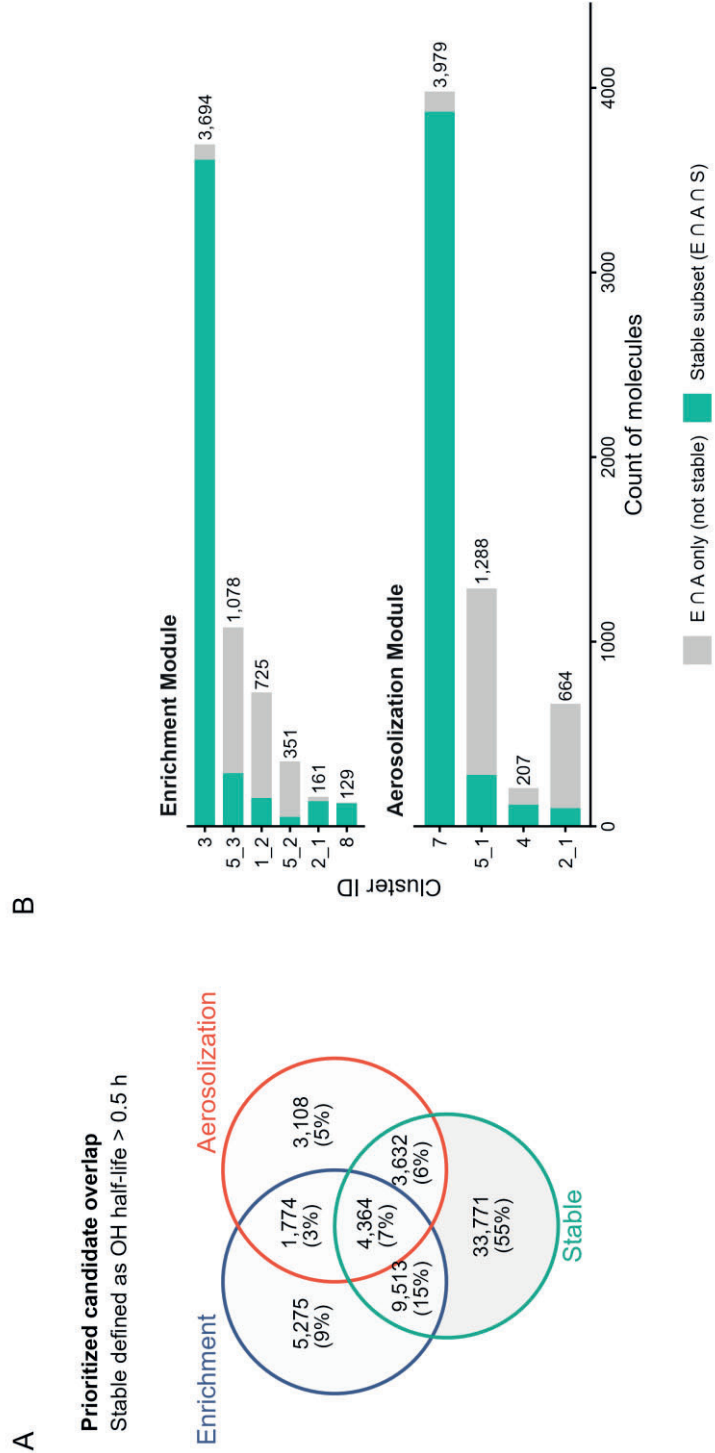


Figure 6 | Prioritized candidate sets and overlaps across mechanistic modules. (A) Overlap of prioritized candidate sets derived from the enrichment module (top-ranked enrichment clusters), aerosolization module

(top-ranked aerosolization clusters), and atmospheric stability screening (Stable: OH half-life > 0.5 h). Counts within each region show the number (and percentage) of candidates in each unique or shared set, highlighting the final triple-intersection candidate set ($E \cap A \cap \text{Stable}$). (B) Cluster drivers of the enrichment–aerosolization overlap. Bars show, for each contributing cluster, the number of candidates in the enrichment–aerosolization intersection (light) and the subset that also meets the stability criterion (green). The top panel summarizes contributions by enrichment cluster ID and the bottom panel summarizes contributions by aerosolization cluster ID.

4. Discussion

4.1. Module spaces, clustering choices, and fate structure

This workflow rests on a simple but important premise: for a marine natural compound to be exposure-relevant via sea spray aerosol, it must pass through multiple distinct bottlenecks—accumulation at the ocean–air interface, transfer and retention in particles after emission, and persistence against atmospheric loss. Treating these steps as separable is more than a conceptual convenience; it reflects the process reality of SSA production from bubble bursting and rapid droplet evaporation, followed by oxidative aging in the marine boundary layer (Atkinson and Arey, 2003; George and Abbatt, 2010). Rather than collapsing all “SSA relevance” into a single score, we intentionally organize candidates into process-informed physicochemical families and then apply an independent fate constraint to identify combinations that remain plausible across steps.

Mechanistic meaning of the two process spaces

Interfacial enrichment and aerosol-phase retention are governed by different physical controls. Interfacial accumulation depends strongly on amphiphilicity and adsorption energetics, but also on hydration constraints and the ability of compounds to reside within heterogeneous, often gel-like surface films that are chemically and biologically distinct from underlying seawater (Cunliffe et al., 2013; Engel et al., 2017). For that reason, the enrichment step is represented by descriptors that capture polarity and hydrogen-bonding capacity alongside hydrophobicity, size, and architectural constraints. These proxies align with established views that SSML enrichment reflects a combination of hydrophobic partitioning and specific polar interactions within complex interfacial matrices, rather than simple bulk solubility alone (Ertl et al., 2000). Structural degrees of freedom and rigidity further matter because condensed surface films impose packing constraints, and surface enhancement has been observed for aromatic and hydrophobic organics at the air–water interface (Cincinelli et al., 2005; Vácha et al.,

2006).

In contrast, post-emission retention in SSA is fundamentally constrained by volatility and gas–particle partitioning as droplets evaporate and particles equilibrate (Donahue et al., 2012; Pankow, 1994). A compact volatility-centered representation is therefore appropriate: vapor pressure and size proxies define broad retention regimes, while hydrophobicity and ring/sugar motifs add structure reflecting affinity for organic phases and the possibility of stabilization in salt- and water-rich particle matrices. This is consistent with volatility-based frameworks and with empirical SSA observations showing that the magnitude of a low-volatility organic fraction can vary with seawater conditions and biological state (Cravigan et al., 2020; Rasmussen et al., 2017). In this sense, the enrichment and aerosolization modules are not competing definitions; they are complementary “views” of the same candidate universe under different physical filters operating at different stages of ocean-to-air transfer.

A key benefit of these mechanistic representations is interpretability. They allow “plausibility” to be narrated in process terms—how a candidate could be enriched at the interface, why it might remain condensed after emission—rather than as an opaque statistical label. This is particularly valuable in SSA chemistry, where molecular diversity is large and where many environmentally detected features remain unidentified (Bertram et al., 2018). In such settings, family-level hypotheses can guide analytical choices more effectively than attempting to predict specific compounds from first principles with limited observational constraints.

Why unsupervised families are a reasonable organizing principle

The chemical space defined by a large structure-resolved database is heterogeneous and high-dimensional, but it is not arbitrary. Many descriptors co-vary because they reflect coupled physical constraints: larger molecules often carry more functional groups (Ertl, 2017); polarity

and hydrogen-bonding capacity are linked to solvation and partitioning (el Tayar et al., 1991); ring structure influences rigidity and volatility (Brandt et al., 2023). These couplings imply that candidates occupy recurrent “neighborhoods” corresponding to recognizable physicochemical families. Unsupervised clustering is therefore not used here to discover a hidden truth about the environment; it is used to compress a vast candidate universe into an interpretable taxonomy of neighborhoods that can be compared, ranked, and updated as evidence accumulates (Quinn et al., 2017; Wang et al., 2016), it is an approach conceptually aligned with how metabolomics uses sparse identifications to interpret broader molecular families (da Silva et al., 2018).

Methodologically, the clustering strategy is defensible because it is tuned for stability and interpretability rather than maximal granularity. A single high-resolution partition can over-fragment families and obscure process narratives, while an overly coarse partition can merge distinct regimes that likely behave differently during enrichment or aerosolization (Li et al., 2022). A coarse-to-fine approach preserves a readable global map while increasing resolution where internal heterogeneity is high (Li et al., 2022). Critically, none of these steps uses enrichment/aerosolization outcomes as targets; the families are defined purely by descriptors, and empirical behavior is incorporated later through separate anchoring and filtering steps. This separation reduces circularity and keeps the workflow transparent about what is learned from structure versus what is learned from external evidence.

Atmospheric stability as an independent gate

Persistence against oxidation is treated as an orthogonal constraint rather than something inferred indirectly from enrichment or aerosolization proxies. This is deliberate: atmospheric degradation depends on oxidant-specific chemistry and reactive functional groups, and many organic species undergo rapid oxidation by OH radicals and ozone, altering lifetimes and partitioning behavior (Atkinson and Arey, 2003; George and Abbatt, 2010). By evaluating oxidative half-lives

independently and applying a persistence threshold as a separate filter, the workflow ensures that candidates are not retained simply because they appear surface-active or low-volatility. Instead, they must also satisfy a minimal residence-time criterion consistent with coastal exposure-relevant time scales discussed for near-field transport and recirculation (Pace et al., 2015; Vignati et al., 2001).

This separation clarifies interpretation: a candidate can be plausible for enrichment and aerosol retention yet still be downweighted because it is short-lived; conversely, a stable candidate may remain low priority if it lacks plausible transfer pathways. In practice, that is where the workflow becomes most useful—by making trade-offs explicit and allowing prioritization to reflect the intersection of mechanistically distinct requirements rather than any single proxy.

4.2. Limits and uncertainty

This workflow is deliberately screening-level, and its value depends on being explicit about what it does and does not claim.

Candidate universe bias. MarineNP is a structure-resolved hypothesis space, not a census of SSML/SSA constituents. It is shaped by reporting, isolation, and annotation biases in marine natural products research. Therefore, the workflow prioritizes plausibility within MarineNP, not prevalence in seawater or aerosols.

Screening-property uncertainty. AOPWIN and MPBPWIN provide rapid estimates that enable scale, but their uncertainties can be class-dependent, particularly for large multifunctional molecules, ionic species, and compounds whose behavior is dominated by phase state or specific interactions (US EPA, 2015). Vapor pressure in particular is central to the aerosolization module, so uncertainties in `vp_est` will propagate into cluster assignments and family boundaries. The workflow partially addresses this by retaining missingness as explicit information rather

than silently imputing.

Ionization, salts, and condensed-phase complexity. Descriptors such as ALogP and TPSA are useful proxies, but SSML and SSA are saline, dynamic matrices where ion pairing, complexation, pH-dependent speciation, and phase separation can materially alter partitioning and persistence (Burrows et al., 2014). The present analysis does not explicitly represent pKa, ionic strength effects, or inorganic–organic interactions beyond what is indirectly encoded through structure.

Reference anchoring sparsity and heterogeneity. Literature anchors were assembled opportunistically rather than via systematic review, and reference data reflect diverse measurement contexts (field vs lab), definitions (Na⁺ normalization vs raw ratios), and uncertainty structures. The calibration step is therefore best viewed as a proof-of-concept showing how sparse empirical evidence can be integrated to rank physicochemical families—not as a definitive mapping from cluster to real-world enrichment/aerosolization factors.

Process incompleteness. The workflow encodes necessary constraints but not the full physics and biology of SSA production. Bubble-mediated transport, biological state, particle size fractionation, aging, and heterogeneous chemistry can all determine whether a plausible candidate becomes an actual SSA constituent in meaningful abundance (Cravigan et al., 2020; George and Abbatt, 2010; Schiffer et al., 2018). As such, the candidate lists are best used to guide targeted measurement and validation, not to replace it.

4.3. Contribution, implications, and outlook

This study contributes a reproducible way to translate a structure-resolved marine natural products database into process-grounded hypotheses about SSA-relevant chemistry. The key advance is not any single descriptor or clustering method, but the workflow logic:

mechanistic module design separates enrichment from aerosol retention; unsupervised clustering converts high-dimensional descriptors into interpretable families; literature anchors provide empirical ordering without turning the problem into supervised prediction; and atmospheric stability acts as an explicit, orthogonal fate gate. Together, these elements create a transparent prioritization pipeline that can be updated as new data arrive and that makes uncertainty visible rather than implicit.

A persistent obstacle in SSA health research is chemical attribution: exposures can produce reproducible, composition-dependent responses, yet the responsible molecules are rarely identified in complex mixtures (Asselman et al., 2019; Cochran et al., 2017; Van Acker et al., 2020). The present framework supports attribution in a practical way: it proposes a tractable set of chemically explicit candidates and families that can be targeted for MS/MS confirmation, standards acquisition, and bioactivity-guided fractionation. Even when most SSA HRMS features remain unannotated, family-level hypotheses can help interpret unknowns (e.g., “low-volatility hydrophobic retained family” versus “high-polarity sugar-associated family”), guiding what to prioritize analytically. More broadly, the workflow offers a common language for connecting ocean surface composition, SSA particle-phase retention, and atmospheric persistence—an essential foundation for mechanistic exposure hypotheses.

Three extensions would most directly strengthen predictive and interpretive power:

1. Systematic reference expansion. A systematic literature review and community-driven reference contributions would broaden anchor coverage across clusters, improve ranking stability, and enable stratification by measurement context (field vs lab; size fraction; normalization choices).
2. Speciation- and matrix-aware descriptors. Incorporating

pKa/ionization state, salt effects, and phase-state proxies would better represent SSML/SSA chemistry beyond neutral-form logP and TPSA. Where feasible, pairing structure-based descriptors with seawater-relevant partitioning frameworks could improve mechanistic fidelity.

3. Validation loops. Targeted measurements in SSML and nascent SSA—focused on representatives from dominant overlap-driving families—could test whether the workflow’s prioritized families are enriched/retained as expected. Iterative recalibration using those results would turn the current screening pipeline into a living framework.

Supplementary Information

Table S1 Reference for literature-anchored calibration of enrichment and aerosolization clusters

Molecule ID	Source Type	Reference DOI	Input Type	Conc. BSW	Conc. SML	Conc. SSA	Direct Factor Value	Direct Factor Type
75317	Field Sample	10.1016/S0304-4203(01)00049-4	Concentrations	0.028	0.123	NA	NA	NA
4125	Field Sample	10.5194/acp-23-6571-2023	Concentrations	2.15	0.88	0.66	NA	NA
5708	Field Sample	10.1016/S0304-4203(01)00049-4	Concentrations	0.011	0.002	NA	NA	NA
5708	Field Sample	10.1016/S0304-4203(01)00049-4	Concentrations	0.029	0.172	NA	NA	NA
4775	Field Sample	10.5194/acp-23-6571-2023	Concentrations	1.65	4.19	1.33	NA	NA
1312	Field Sample	10.5194/acp-23-6571-2023	Concentrations	15.39	17.56	6.04	NA	NA
23494	Field Sample	10.5194/acp-23-6571-2023	Concentrations	3.23	3.76	1.84	NA	NA
6323	Field Sample	10.5194/acp-23-6571-2023	Concentrations	0.19	1.73	0.63	NA	NA

425158	Field Sample	10.5194/acp-23-6571-2023	Concentrations	7.24	9.63	5.52	NA	NA
695076	Field Sample	10.5194/acp-23-6571-2023	Concentrations	30.07	25.96	28.98	NA	NA
36050	Field Sample	10.5194/acp-23-6571-2023	Concentrations	8.82	8.91	1.76	NA	NA
5876	Field Sample	10.5194/acp-23-6571-2023	Concentrations	8.59	6.06	7.15	NA	NA
34818	Field Sample	10.5194/acp-23-6571-2023	Concentrations	3.21	1.91	9.87	NA	NA
6086	Field Sample	10.5194/acp-23-6571-2023	Concentrations	0.19	4.87	6.16	NA	NA
6202	Field Sample	10.5194/acp-23-6571-2023	Concentrations	1.59	3.98	4.79	NA	NA
6179	Field Sample	10.5194/acp-23-6571-2023	Concentrations	12.87	4.77	3.7	NA	NA
6738	Field Sample	10.5194/acp-23-6571-2023	Concentrations	7.24	9.63	5.52	NA	NA
5994	Field Sample	10.5194/acp-23-6571-2023	Concentrations	8.56	8.73	8.75	NA	NA
72087	Field Sample	10.5194/acp-23-6571-2023	Concentrations	17.94	4.43	4.45	NA	NA

5343	Field Sample	10.5194/acp-23-6571-2023	Concentrations	17.66	7.33	12.35	NA	NA
38729	Field Sample	10.5194/acp-23-6571-2023	Concentrations	16.39	24.42	18.68	NA	NA
5873	Field Sample	10.5194/acp-23-6571-2023	Concentrations	5.56	13.67	11.67	NA	NA
6712	Field Sample	10.5194/acp-23-6571-2023	Concentrations	23.68	29.37	35.33	NA	NA
15203	Field Sample	10.1016/S0304-4203 (01)00049-4	Concentrations	0.019	0.022	NA	NA	NA
5708	Field Sample	10.1016/S0304-4203 (01)00049-4	Concentrations	0.023	0.033	NA	NA	NA
5708	Field Sample	10.1016/S0304-4203 (01)00049-4	Concentrations	0.034	0.189	NA	NA	NA
566961	Field Sample	10.1016/S0304-4203 (01)00049-4	Concentrations	0.36	1.34	NA	NA	NA
75317	Field Sample	10.1016/S0304-4203 (01)00049-4	Concentrations	0.007	0.015	NA	NA	NA
10122	Field Sample	10.1016/S0304-4203 (01)00049-4	Concentrations	0.097	0.154	NA	NA	NA
10122	Field Sample	10.1016/S0304-4203 (01)00049-4	Concentrations	0.194	33	NA	NA	NA

75317	Wetlab	10.1016/S0304-4203 (01)00049-4	Concentrations	0.028	0.123	NA	NA	NA
530883	Wetlab	10.1021/acs.est.1c00995	Direct Factor	NA	NA	NA	5.7	Enrichment Factor (SML/BSW)
530883	Wetlab	10.1021/acs.est.1c00995	Direct Factor	NA	NA	NA	1769	Aerosolization Factor (SSA/SML)
529907	Wetlab	10.1021/acs.est.1c00995	Direct Factor	NA	NA	NA	488	Aerosolization Factor (SSA/SML)
529907	Wetlab	10.1021/acs.est.1c00995	Direct Factor	NA	NA	NA	5.1	Enrichment Factor (SML/BSW)
21351	Wetlab	10.1021/acs.est.1c00995	Direct Factor	NA	NA	NA	2.9	Enrichment Factor (SML/BSW)
21351	Wetlab	10.1021/acs.est.1c00995	Direct Factor	NA	NA	NA	593	Aerosolization Factor (SSA/SML)
1312	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	2.8	Enrichment Factor (SML/BSW)
1312	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	73	Aerosolization Factor (SSA/SML)
1396	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	151	Aerosolization Factor (SSA/SML)
1396	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	1.9	Enrichment Factor (SML/BSW)

425158	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	5	Enrichment Factor (SML/BSW)
425158	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	321	Aerosolization Factor (SSA/SML)
6738	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	82	Aerosolization Factor (SSA/SML)
6738	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	5	Enrichment Factor (SML/BSW)
6983	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	4	Enrichment Factor (SML/BSW)
6983	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	122	Aerosolization Factor (SSA/SML)
5994	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	104	Aerosolization Factor (SSA/SML)
5994	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	5	Enrichment Factor (SML/BSW)
36050	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	1.2	Enrichment Factor (SML/BSW)
36050	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	24	Aerosolization Factor (SSA/BSW)
15234	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	30	Aerosolization Factor (SSA/BSW)

15234	Wetlab	10.1021/acs.est.6b02988	Direct Factor	NA	NA	NA	1.8	Enrichment Factor (SML/BSW)

The column input_type denotes either measured concentrations in bulk seawater (conc_bsw), sea-surface microlayer (conc_sml), and/or sea-spray aerosol (conc_ssa), or a reported direct factor (direct_factor_value, direct_factor_type; enrichment factor SSM/BSW; aerosolization factor SSA/SML or SSA/BSW). NA = not available in the source.

Chapter 6 Conclusions and Perspectives

The aim of this PhD thesis is to investigate how natural sea spray aerosols (SSAs) can act as a mechanistic link between ocean ecology and human health. The work is based on measurements taken at the coast of Ostend, Belgium, in the Southern Bight of the North Sea. This study area is a well characterized temperate coastal system with recurrent spring blooms of *Phaeocystis* and diatoms, making it an ideal natural laboratory to examine how seasonal ecosystem dynamics shape the composition and emission of SSAs. However, the conclusions are expected to be relevant for other coastal regions and, in some cases, to inland areas that are intermittently influenced by transported SSAs.

To fulfil this aim, the thesis brings together four interdisciplinary studies that span field observations, laboratory experiments, ecological analysis, and bio- and cheminformatics:

- 1) A field and modelling study that characterizes the occurrence of phytoplankton-derived surfactant lipids, in particular dipalmitoylphosphatidylcholine (DPPC), in ambient coastal SSAs and identifies the physical and ecological drivers of their aerosolization (Chapter 2).
- 2) An *in vitro* transcriptomic study that links seasonal variability in natural SSA mixtures to pathway-level responses in human bronchial epithelial cells (Chapter 3).
- 3) A three-year analysis of phytoplankton bloom timing and community composition in the Southern Bight of the North Sea that constrains the ecological background for changes in SSA composition and exposure (Chapter 4).
- 4) The development of a comprehensive marine natural products database that is used to identify classes of bioactive compounds with physicochemical properties favoring aerosolization and potential respiratory relevance (Chapter 5).

Together, these four studies trace the pathway from ocean ecological dynamics, through SSA production and composition, to cellular and molecular responses and the wider chemical space of marine natural products, thereby operationalizing the central theme of this thesis that SSAs link ocean processes to human health.

In Chapter 1, I formulated four central research questions about this ocean–air–health connection, and in Chapters 2 to 5 I addressed these questions using the complementary approaches outlined above. In the following sections, I summarize how the main findings of each chapter collectively answer these questions and what they imply for future research on marine aerosols and human health.

1. Synthesis of the main findings

1.1. Q1: Phytoplankton derived surfactant lipids in natural SSAs

To what extent can phytoplankton derived surfactant lipids, in particular DPPC, produced during ambient, non HAB conditions be detected in natural coastal SSAs, and can their occurrence be quantitatively linked to ocean surface ecological state, for example through in situ indicators and satellite derived proxies of phytoplankton biomass and community composition?

Short answer: DPPC is a recurrent component of ambient coastal SSAs at Ostend, not restricted to bloom events, and its enrichment is jointly controlled by physical forcing and phytoplankton dynamics in the source waters.

In Chapter 2, I showed that DPPC is routinely present in natural SSAs at picogram per cubic meter concentrations. It was detected in the majority of weekly SSA samples from Ostend, with concentrations and

DPPC to Na^+ ratios summarized over the full sampling period (Figure 2 in Chapter 2). These data demonstrate that DPPC is not an exceptional or rare constituent, but a persistent feature of the coastal marine aerosol mixture in this temperate system.

The marine origin of DPPC in the aerosol phase was supported by its consistent association with sea salt tracers. Co-variation of DPPC with Na^+ in coastal air and the stability of the DPPC to Na^+ ratio across samples indicate that the main source is sea spray rather than non-marine emissions (Figure 2B in Chapter 2). Enrichment factors for Mg^{2+} and Ca^{2+} further increased in parallel with DPPC enrichment (Figure 3 in Chapter 2), which is consistent with interfacial complexation between biogenic surfactants and divalent cations in the sea surface microlayer.

To identify the drivers of this enrichment, I used a generalized linear modelling framework that combined wave climate and phytoplankton related predictors. The model results (Table 1 in Chapter 2) and the corresponding time series of predictors and DPPC concentrations (Figure 4 in Chapter 2) show that significant wave height has a strong negative effect on DPPC enrichment, while phytoplankton biomass and its recent changes contribute positively. The best fit was obtained with a 6-day lag between changes in phytoplankton biomass and DPPC in the aerosol, suggesting that surfactant supply through exudation and cell lysis feeds into the aerosolizable pool with a characteristic delay.

Together, these lines of evidence answer Q1 by demonstrating that phytoplankton derived surfactant lipids, exemplified by DPPC, are recurrently aerosolized under ambient, non HAB conditions, and that their abundance in SSAs reflects a combination of physical conditions at the sea surface and the ecological state of the upper ocean.

1.2 Q2: Seasonal variability in SSA composition and bioactivity

How does the composition and bioactivity of natural SSAs vary over seasonal and environmental gradients in a temperate coastal system, and how do these changes translate into pathway level responses in human bronchial epithelial cells exposed to environmentally scaled doses of SSA derived material?

Short answer: SSA composition and bioactivity at Ostend vary strongly with season and ecological state, and natural SSA mixtures at environmentally scaled doses modulate central growth, inflammatory, and metabolic pathways in human bronchial epithelial cells.

In Chapter 3, I linked a one-year SSA sampling campaign to *in vitro* exposures of BEAS-2B cells and RNA sequencing. SSA extracts collected at Ostend and dosed at environmentally scaled concentrations induced clear differential gene expression, with the number of differentially expressed genes varying widely between sampling dates (Figure 1 in Chapter 3). This demonstrates that realistic SSA mixtures can measurably alter the transcriptome of human lung cells.

Pathway analyses showed that these transcriptional changes converge on key signaling axes. SSA exposures repeatedly affected PI3K, Akt, mTOR, NF- κ B, and AMPK, with a consistent pattern of downregulated mTOR-related and pro-inflammatory signaling and activation of AMPK-linked responses (Figure 2 in Chapter 3). This profile is compatible with the biogenics hypothesis (Moore, 2015), which proposes that low dose exposure to natural marine compounds can shift cellular set points toward more restrained growth and enhanced stress regulation.

The magnitude and character of SSA-induced responses were strongly seasonal. SSA samples collected during periods of higher phytoplankton biomass and greater chemical feature richness elicited more extensive transcriptional changes than samples from colder, less productive periods. Ordinations and clustering of chemical fingerprints and exposure effects show that spring and summer SSA are both chemically richer and more bioactive than winter SSA (Figure 3a & b in Chapter 3). These patterns link variability in SSA composition and diversity to variability in pathway level responses.

Comparisons with ambient coastal aerosols and inland city aerosols indicate that the strongest pathway modulation is associated with marine influenced mixtures, although not all variability can be attributed to SSA alone (Figure 1 and Figure 3 in Chapter 3). Overall, these results answer Q2 by showing that natural SSA mixtures in this coastal system are bioactive at environmentally scaled doses and that their molecular effects on human bronchial epithelial cells vary systematically with season and source region ecology

1.3 Q3: Changing phytoplankton phenology and implications for SSA mediated exposure

How do the timing, magnitude, and taxonomic composition of phytoplankton blooms in the Southern Bight of the North Sea vary on interannual timescales, and what are the implications of these shifts for the seasonality and intensity of SSA-mediated human exposure to marine biogenics?

Short answer: Over the three-year study period, spring and early summer blooms in the Southern Bight showed clear interannual variability in bloom timing, as well as in magnitude and dominant taxa. These shifts in bloom timing are expected to alter the seasonal windows

during which coastal SSAs are most enriched in phytoplankton derived biogenics.

In Chapter 4, I analyzed a three-year monthly dataset of phytoplankton community composition and environmental drivers at a coastal station in the Southern Bight. Metabarcoding data revealed that Heterokontophyta, including *Phaeocystis* and diverse diatoms, dominate the community, with temporal dynamics of major phyla summarized in Figure 1b & c in Chapter 4. Time series of phytoplankton carbon concentrations and normalized densities documented recurring spring and early summer blooms, in which *Phaeocystis* dominated events were followed by diatom dominated phases (Figure 2 in Chapter 4).

Within this recurring pattern, there were clear interannual differences in bloom timing. In 2021, both the *Phaeocystis* and *Thalassiosira* blooms started earlier than in 2018 and 2019, as shown by the earlier peaks in normalized densities for these taxa (Figure 2c in Chapter 4). These observations document shifts in bloom phenology, in the strict sense of changes in the timing of bloom onset and peak. At the same time, the relative contribution of individual taxa to total biomass and the overall bloom magnitude also differed between years, indicating interannual variability in community structure and intensity superimposed on the phenological shifts. Trends in biodiversity indices and environmental variables, including temperature, salinity, and suspended particulate matter, provide additional context for these contrasts between years (Figure 4 in Chapter 4).

Correlations between phytoplankton metrics and environmental variables, together with redundancy analysis, point to nutrients and salinity as key drivers of these interannual differences. Summary

statistics and correlation patterns (Table 1 and Figure 3a in Chapter 4) show strong positive associations between phytoplankton abundance and nitrate, phosphate, and silicate, as well as marked relationships with salinity. The ordination of taxa and environmental variables (Figure 3C in Chapter 4) further indicates that different groups align with distinct combinations of nutrients and salinity, supporting the interpretation that interannual changes in these conditions affect both bloom timing and community structure.

From the perspective of SSA mediated exposure, these results imply that the timing of periods when the sea surface microlayer is richest in aerosolizable biogenics is not fixed from year to year. Earlier blooms, as observed in 2021, mean that surface waters become enriched in *Phaeocystis* mucilage, diatom derived lipids, and other phytoplankton products earlier in the season, which is expected to shift the seasonal window of enhanced SSA enrichment in marine biogenics. In this way, the results of Chapter 4 answer Q3 by showing that interannual variability in bloom timing, together with associated changes in magnitude and composition, is likely to translate into interannual variability in the seasonality and intensity of SSA mediated exposure to marine biogenics.

1.4 Q4: Marine natural products as candidate aerosolizable biogenics

Which classes of marine natural products are both likely to be produced in surface ocean ecosystems and have physicochemical properties that favor aerosolization and inhalation, and what health relevant molecular pathways do these candidate aerosolizable compounds target according to current bioactivity knowledge and cheminformatics analyses?

Short answer: Using MarineNP and a property-based clustering approach, I identified groups of marine natural products that are produced in surface ocean ecosystems and have high Enrichment and Aerosolization Potential, and within these groups many compounds are known or predicted to target signaling pathways that are also modulated by natural SSA exposure, such as PI3K, Akt, mTOR, NF- κ B, and AMPK.

In Chapter 5, I first addressed Q4 by developing MarineNP, a comprehensive and locally deployable database of marine natural products. MarineNP integrates chemical structures, physicochemical descriptors, producing taxa, and georeferenced occurrence data into a single resource tailored to marine systems. Its scope and distinguishing features relative to existing natural product databases are summarized in Table 1 in Chapter 5, which highlights the marine specific coverage and the ability to work fully offline, both of which are important for linking chemical information to oceanographic and aerosol datasets.

Building on this database, I then used cheminformatic filtering and clustering to identify marine compounds that are plausible candidates for transfer into SSAs. A key step was to organize molecular descriptors into four conceptually distinct property sets: Enrichment Potential, Aerosolization Potential, Health Effect Potential, and Stability in Air. Enrichment Potential was defined using aLogP, topological polar surface area (TPSA), molecular weight, number of rotatable bonds, and hydrogen bond acceptors, in order to capture the balance of hydrophobicity, polarity, size, and flexibility that governs accumulation in the sea surface microlayer. Aerosolization Potential was described using aLogP, TPSA, logVP, number of aromatic rings, and number of heavy atoms, which together reflect a compound's tendency to partition into and persist in the aerosol phase. Health Effect Potential was proxied by the quantitative estimate of drug likeness (QED), and Stability in Air

was represented by logOH, a descriptor of reactivity with hydroxyl radicals.

Using these property sets, I clustered MarineNP compounds in the multidimensional descriptor space and then used chemicals with known enrichment factors from the literature to label clusters with relative ranks of Enrichment Potential and Aerosolization Potential. In other words, clusters that contained many compounds already known to be strongly enriched in the sea surface microlayer or in SSAs were assigned higher enrichment or aerosolization ranks. This procedure yielded a subset of clusters, and therefore a concrete list of compounds, that are both produced in surface ocean ecosystems and have chemical profiles that favor accumulation at the sea surface and transfer into sea spray.

Finally, I intersected these high-ranking clusters with bioactivity and target information available in MarineNP and in external resources. Many of the compounds in the top Enrichment and Aerosolization clusters are reported to modulate kinases in the PI3K, Akt, and mTOR cascades, to influence NF- κ B and other inflammation related transcription factors, or to affect regulators of oxidative stress and energy metabolism such as AMPK. These are the same signaling axes that respond to natural SSA exposure in Chapter 3. Although the current work does not yet demonstrate the presence of specific MarineNP compounds in Ostend SSA samples, it shows that the marine natural product space contains well defined groups of molecules that are both physically likely to reach the aerosol phase and mechanistically capable of affecting pathways relevant to respiratory and systemic health.

In this way, Chapter 5 answers Q4 by combining a dedicated marine natural products database with a structured property-based clustering and ranking approach to produce a prioritized list of candidate

aerosolizable biogenics. This list provides a rational starting point for targeted chemical analyses of SSA and for future mechanistic studies that seek to link individual marine metabolites to the cellular responses observed in SSA exposure experiments.

1.5 Conceptual and methodological contributions

Taken together, the four research chapters advance both the conceptual framing and the methodological toolkit needed to understand sea spray aerosols as a mechanistic link between ocean ecology and human health. Conceptually, the thesis moves beyond the traditional view of SSAs as either benign or harmful agents and instead treats them as context dependent carriers of marine biogenics whose effects emerge from the interaction between composition, dose, and ecological state. Methodologically, it combines long term field observations, environmentally scaled in vitro exposures, high resolution molecular profiling, bio- and cheminformatics in a way that can be generalized to other coastal systems and compound classes.

A first key conceptual contribution is the explicit embedding of the biogenics hypothesis within a physically and ecologically constrained exposure framework. By connecting surfactant lipids and broader classes of marine natural products to specific ocean surface processes, such as changes in wind speed and sea state that modulate SSA production (Chapter 2), the development of coastal phytoplankton blooms and associated enrichment of the sea surface microlayer (Chapters 2 and 4), the seasonal evolution of SSA composition and cellular bioactivity in coastal air (Chapter 3), and shifts in bloom phenology driven by climate variability and eutrophication (Chapter 4), the thesis reframes SSAs as ecosystem-dependent mixtures whose bioactivity reflects the timing, magnitude, and taxonomic structure of phytoplankton blooms rather than a static background signal, in line with emerging views of the sea surface microlayer as a biologically structured, surfactant rich interface (Cunliffe et al., 2013; Engel et al.,

2017). The notion of exposure windows, in which the sea surface microlayer is transiently enriched in aerosolizable biogenics due to bloom development and lysis, provides a concrete way to link bloom phenology to periods of enhanced coastal aerosol bioactivity. In parallel, the work clarifies that, at the environmentally scaled doses applied in Chapter 3 (corresponding to 10–80-fold ambient exposure), marine aerosols may exert their influence primarily through low dose modulation of conserved signaling pathways, such as PI3K, Akt, mTOR, NF- κ B, and AMPK, rather than through overt toxicity, thereby aligning SSA exposure more closely with nutritional or pharmacological cues than with classical air pollutants, consistent with the biogenics (Moore, 2015) and old friends (Rook, 2023) hypothesis,

A second conceptual advance is introducing a structured way of thinking about candidate aerosolizable biogenics in terms of four property domains, Enrichment Potential, Aerosolization Potential, Health Effect Potential, and Stability in Air. This framework allows the vast chemical space of marine metabolites to be organized according to their estimated potential to become enriched at the ocean surface and transferred into sea spray aerosols, and according to their potential to engage health relevant pathways. By calibrating the property space with compounds that are already known to be enriched in the sea surface microlayer or in SSAs, the framework provides a first pass, compound specific estimate of which marine natural products are most likely to end up in the aerosol phase, complementing physically based descriptions of organic fractionation and surfactant enrichment in SSAs (Bertram et al., 2018; Burrows et al., 2014). Because the underlying property sets are not specific to phytoplankton derived metabolites, a similar approach can be adapted to other marine organism groups beyond phytoplankton, and to exposure media, for example freshwater spray aerosols.

On the methodological side, an important contribution is the integration of long-term coastal aerosol sampling with quantitative environmental

modelling. The one year and multi-year datasets combine measurements of SSA composition with wave climate, meteorological variables, and phytoplankton biomass and community composition. The generalized linear models that incorporate time lags between ocean surface signals and aerosol composition provide a template for attributing aerosol constituents to ecological and physical drivers while explicitly accounting for transport and processing times, extending previous efforts (Callaghan et al., 2014; Christiansen et al., 2019) to link bubble entrainment, water temperature, and biological productivity to SSA production in laboratory and field studies. This approach is broadly applicable to other surfactant classes and coastal settings where joint physical and biological control of aerosol composition is expected.

A second methodological innovation is the use of environmentally scaled *in vitro* exposure schemes coupled to transcriptomics for complex natural aerosol mixtures. By basing doses on alveolar surface area scaling and realistic inhalation scenarios, the thesis narrows the gap between *in vitro* design and plausible human exposure. The combination of RNA sequencing with untargeted chemical fingerprinting allows the extent and direction of pathway modulation to be related to both the number and diversity of chemical features in SSA samples, rather than to single marker compounds. This mixture oriented, pathway level readout is in line with recent work that has used human airway epithelium to probe inflammatory and transcriptomic responses to cyanobacterial and HAB related aerosols (Breidenbach et al., 2022; Labohá et al., 2023) and is particularly well suited for characterizing subtle regulatory effects of natural aerosols and can be extended to other cell types or organ systems.

Finally, the thesis contributes data infrastructure and informatics methods that support future work at the interface of marine chemistry, aerosol science, and health. The MarineNP database consolidates structural, taxonomic, geographic, and bioactivity information for marine natural products in a form that is readily linkable to oceanographic and

atmospheric datasets, complementing existing compilations of marine bioactive compounds developed for drug discovery and chemical ecology (Lyu et al., 2021). The property-based clustering and ranking procedure, which uses compounds with known enrichment characteristics to calibrate cluster level Enrichment and Aerosolization Potential, provides a reproducible pipeline for generating prioritized candidate lists for targeted analyses. In combination with the ecological phenology analyses and exposure experiments, these tools demonstrate a full cross scale workflow that runs from ocean community dynamics, through aerosol formation and composition, to molecular responses in human cells. This integrative framework constitutes a central methodological advance of the thesis and offers a foundation for more predictive assessments of how changes in coastal ecosystems may propagate to human exposure and health.

2. Limitations and perspectives

2.1 Limitations of the present work

The topic of sea spray aerosols and human health is intrinsically broad and interdisciplinary, spanning physical oceanography, atmospheric chemistry, microbiology, toxicology, epidemiology, and public health. This thesis could only address a small, though coherent, subset of this landscape. While chapter specific limitations are discussed in detail elsewhere, several general constraints cut across the work as a whole and should be acknowledged explicitly.

A first limitation is spatial and environmental representativeness. All empirical work was anchored in a single temperate coastal system, the Southern Bight of the North Sea, with intensive sampling at or near Ostend. This region is shallow, tidally energetic, and influenced by strong anthropogenic nutrient inputs, which shape both bloom dynamics and the pool of surface-active material available for aerosolization (Desmit et al., 2020; Rousseau et al., 2013). As SSA production, size distributions, and organic fractionation are known to vary strongly

across oceanographic regimes and biogeochemical states (Grythe et al., 2014; O'Dowd et al., 2015), the patterns described here, including DPPC enrichment, seasonal variability in SSA bioactivity, and phytoplankton phenology, may not be directly transferable to oligotrophic subtropical gyres, upwelling systems, polar oceans, or highly urbanized coasts. Similarly, the focus on near-shore conditions provides limited insight into open ocean or long-range transported marine aerosols, which can differ in both chemical makeup and particle size, and therefore in potential deposition and health relevance (Lewis and Schwartz, 2004).

Second, the thesis is temporally restricted. The aerosol and cell exposure work covers approximately one year of SSA variability, and the phenological analysis spans three years of phytoplankton observations. These windows are sufficient to reveal seasonal and interannual differences but are too short to robustly quantify long term trends or to separate climate driven changes from internal variability, particularly in coastal systems where phenology responds to both warming and changing nutrient regimes (Alvarez-Fernandez et al., 2012; Desmit et al., 2020). As a result, any interpretation of shifting bloom timing or changing exposure windows must be regarded as suggestive rather than definitive with respect to decadal scale climate change (Hjerne et al., 2019).

Third, the links between ocean ecology, SSA composition, and cellular responses are primarily correlative. DPPC enrichment, SSA chemical fingerprints, phytoplankton metrics, and transcriptomic responses were analyzed jointly, yet the molecular composition of the bioactive fraction remains only partially resolved, consistent with the high chemical complexity of organic-enriched SSAs and the difficulty of attributing mixture-level responses to specific constituents (Bertram et al., 2018; Quinn et al., 2015). In the exposure experiments, SSA extracts behave as a chemical black box: they elicit clear pathway level responses, but the specific compounds or mixtures responsible are not identified, which

limits causal inference and mechanism assignment. Likewise, the property-based clustering in MarineNP yields mechanistically plausible candidates, but their actual presence in the aerosol samples was not confirmed and their enrichment behavior remains uncertain without targeted validation. Together, these constraints mean that the thesis advances a mechanistic framework but stops short of establishing causal chains from individual marine metabolites to defined health outcomes.

Finally, the thesis does not address several important dimensions of vulnerability and co exposure. All cellular experiments used a single genetic background and did not consider age, sex, pre-existing disease, or co-morbid exposures such as urban air pollution, tobacco smoke, or occupational aerosols, all of which can modify inhaled dose and epithelial responses (Ching and Kajino, 2018). Interactions between marine aerosols and anthropogenic pollutants, including potential modification of toxicity by marine surfactants (Finlayson-Pitts and Pitts, 2000; Hidalgo et al., 2015), were beyond the scope of this work. Likewise, the step from cellular responses to population level health, for example through integration with epidemiological data on blue space exposure, remains to be fully developed.

Taken together, these limitations mean that the conclusions of this thesis should be understood as hypothesis generating and mechanistic in scope, rather than as definitive statements about the net health impact of SSA exposure in coastal populations. They also point directly toward the perspectives outlined in the next subsection, where I highlight how future work can build on the present findings by broadening spatial and temporal coverage, resolving chemical identities, moving from correlations toward causality, and bridging the remaining gaps between ocean processes, aerosols, and human health.

2.2 Perspectives

The limitations outlined above suggest several directions in which future work on SSAs and human health could meaningfully extend the present thesis, particularly by integrating atmospheric processing, chemical speciation, and health endpoints across scales.

A first priority is to complement mechanistic *in vitro* work with epidemiological studies that explicitly incorporate marine aerosol exposure, building on lessons from HAB events where aerosolized marine biotoxins have been linked to respiratory symptoms in both healthy adults and people with asthma (Fleming et al., 2011). Most blue-space health studies still rely on coarse proxies for coastal exposure, such as residence distance or self-reported visits, and rarely distinguish between shoreline environments or isolate the sea-spray aerosol contribution (Gascon et al., 2017). Prospective or retrospective analyses that combine high-resolution coastal air-quality measurements with SSA tracers and chemical markers, alongside routinely collected health data, could test whether episodes of enhanced marine aerosol influence are associated with short-term changes in respiratory outcomes, and whether risks differ across susceptible groups (Bertram et al., 2018; Grythe et al., 2014). Natural experiments comparing seasons with distinct SSA composition or sites with contrasting phytoplankton regimes could strengthen causal inference by providing quasi-controlled contrasts in exposure mixtures. Together, such designs would help bridge the gap between cellular pathway modulation and endpoints that matter clinically and for public health.

A second key avenue concerns the transport and aging of bioactive SSA components. Although this thesis focused on fresh or near-coastal aerosols, marine particles can contribute to particulate matter far inland, and their composition and mixing state can shift substantially during transport (Ching and Kajino, 2018; May et al., 2018). Future studies should therefore follow how size distributions, hygroscopicity, surfactant content, and specific biogenic compounds evolve under realistic

meteorological conditions, and how oxidation, heterogeneous reactions, and mixing with anthropogenic pollutants alter bioactivity (Finlayson-Pitts and Pitts, 2000; Jayarathne et al., 2016; Sievering et al., 2004). Controlled aging experiments using sea-spray generation and environmental chambers, coupled to repeated chemical characterization and bioassays (Su et al., 2022), would help determine whether initially benign mixtures become more or less bioactive with time and distance.

A third major challenge is to identify the specific compound or mixtures responsible for the bioactivities observed *in vitro*. As shown in Chapter 3, natural SSA extracts can modulate key signaling pathways, but in the present work the extract largely remains a black box. The present work shows that natural SSA extracts modulate key signaling pathways, but treats the extract as a black box. This black-box treatment is not only conceptual, but also analytical: for many aerosol-derived compounds the absolute collected mass is extremely low, meaning that even UHPLC-HRMS signals frequently fall below LOD/LOQ and do not yield sufficient material for confident confirmation, fractionation, or isolation (e.g., targeted SSA measurements where LOQs still correspond to pg m^{-3} air concentrations despite high-volume sampling) (Van Acker et al., 2021b). In parallel, untargeted HRMS typically produces hundreds to thousands of molecular features, yet only a small fraction can be annotated because public MS/MS libraries and natural-product structure databases incompletely cover marine chemical space; consequently, a large portion of spectra remains unmatched or only tentatively assigned (da Silva et al., 2015). This database gap links directly to the database chapter: while MarineNP substantially improves access and computational querying of known marine natural products, it cannot represent compounds that are unreported, structurally unresolved, or absent from source datasets, leaving many SSA features outside current reference space. Future research should combine untargeted high resolution mass spectrometry with bioassay guided fractionation and chemical isolation, while also scaling

sampling/production to obtain enough material above LOQ (e.g., higher-volume sampling, pooled filters/time windows, or controlled SSA generation) to pinpoint active fractions and, ideally, individual structures. Once candidates are identified, targeted quantification in SSA samples, alone or in combination with MarineNP-driven predictions, would allow dose response relationships to be established for defined compounds. Parallel experiments in which particular chemical classes are selectively removed, enriched, or synthetically reconstituted would help to disentangle the relative contributions of surfactant lipids, small polar metabolites, polysaccharides, and microbial fragments to the overall cellular response.

At the same time, exposure models themselves should be diversified and made more physiologically realistic. Extending from submerged monocultures of BEAS-2B cells to air liquid interface systems, co cultures that include epithelial and immune cells, organoids, or *ex vivo* tissue slices would provide a more faithful representation of the lung environment. Incorporating repeated low dose exposures that mimic chronic coastal residence, together with functional endpoints such as barrier integrity, mucus production, immune signaling, and repair, would help to assess whether the transcriptional patterns observed here translate into durable changes in tissue behavior.

From an ecosystem and climate perspective, it will be important to broaden the spatial and temporal scope of ocean and aerosol observations. Comparative studies across different coastal systems, including oligotrophic, upwelling, polar, and tropical environments, would reveal how general the Ostend patterns are and whether particular biogeographic provinces or phytoplankton assemblages are associated with stronger SSA enrichment in biogenics. Extending phenological and aerosol measurements over longer periods, and embedding them in coupled physical biogeochemical models, would allow scenarios to be constructed in which future changes in nutrient regimes, stratification, and bloom timing are translated into altered SSA

exposure windows for nearby populations.

Finally, there is substantial scope to further develop MarineNP as a general-purpose tool for marine biogenics research. Priorities include expanding compound coverage, especially for bacterial and viral metabolites, improving the completeness and quality of bioactivity annotations, and incorporating descriptors that better capture aerosol related behavior, for example interfacial tension effects or propensity to form aggregates. Providing programmatic interfaces, harmonizing MarineNP identifiers with those used in atmospheric chemistry databases, and linking directly to mass spectral libraries would facilitate integration into SSA studies and beyond. In the longer term, MarineNP could serve as a hub that connects ocean observatories, aerosol measurement networks, and health datasets through shared chemical entities, enabling more systematic, cross disciplinary analyses of how marine metabolites move through the ocean atmosphere human system.

Together, these perspectives point toward a research agenda in which SSAs are studied not only as carriers of inorganic sea salt and climate relevant particles, but also as dynamic, ecologically structured mixtures of bioactive compounds. Progress along these lines will require sustained collaboration between oceanographers, atmospheric scientists, chemists, toxicologists, epidemiologists, and data scientists. The framework developed in this thesis provides one possible foundation for such efforts, but realizing its full potential will depend on extending it across scales, regions, and disciplines.

3. Open science, data stewardship, and tool accessibility

The work presented in this thesis also highlights a more structural issue that affects the entire field of marine biogenics and, more broadly, environmental health research. Many of the scientific questions addressed here are limited not only by sampling and experimental constraints, but also by the availability, persistence, and interoperability

of the underlying data and tools. Progress along the ocean–air–health axis will therefore depend increasingly on whether the community embraces open science practices that strengthen transparency and reproducibility (Nosek et al., 2015) and invests in robust, well-curated data infrastructures guided by **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse (FAIR) principles (Munafò et al., 2017). Complementary governance frameworks such as **C**ollective benefit, **A**uthority to control, **R**esponsibility, and **E**thics (CARE) further emphasize that stewardship is not only technical, but also social and ethical, including questions of authority to control and benefit-sharing (Carroll et al., 2020).

A central motivation for developing MarineNP was the temporary unavailability of the Comprehensive Marine Natural Products Database (CMNPD). For a period of time, CMNPD, which represented the most extensive resource on marine natural products, was simply inaccessible because its website was down. Since the data were hosted exclusively on that site, with no mirrored copy in a public repository, users had no alternative access. This experience underscored how vulnerable research workflows become when key datasets reside behind a single institutional front end. In contrast, open repositories such as Zenodo, GitHub, and related platforms provide long term, versioned storage that is independent of individual institutional websites and can be mirrored or forked if necessary (Lin et al., 2020; Whitaker, 2019). Archiving database snapshots and code in such repositories would greatly reduce the risk that critical resources disappear due to technical, administrative, or funding-related disruptions. This necessity is underscored by research showing that the availability of research data declines at an alarming rate; for every year that passes after publication, the odds of finding an extant dataset fall by 17% (Vines et al., 2014).

The same tension arises for important online tools. Resources such as ClassyFire for chemical classification and WoRMS for taxonomic matching are extremely valuable, and their curators have strong and legitimate reasons to maintain central control for purposes of data

integrity, authority, and intellectual property (WoRMS Editorial Board, 2025). At the same time, restricting access to web-based interfaces creates practical bottlenecks for large scale analyses that need to be run repeatedly, integrated into automated pipelines, or executed on local computing clusters. Where possible, providing documented application programming interfaces, downloadable reference files, or containerized local instances that can be run on institutional servers would significantly lower barriers for researchers while preserving attribution and governance by the original developers (Boettiger, 2015; Sandve et al., 2013). Such hybrid models, in which authoritative central services coexist with controlled local deployments, would be particularly beneficial for interdisciplinary fields that need to connect chemistry, taxonomy, and environmental metadata at scale. Recent proposals extend these ideas to the software itself through FAIR4RS principles, which adapt FAIR to the specific needs of research code and software (Barker et al., 2022), by promoting clear metadata and citation, archived/versioned releases, transparent dependencies/environments, and licensing/documentation so tools remain runnable and reusable beyond their original publication.

The meta-analytical work attempted in Chapter 5 also illustrates the challenges that arise when data are not managed according to FAIR principles. Searching for well described, machine readable datasets on marine natural products, enrichment factors, and SSA composition proved time consuming and often frustrating. Key obstacles included incomplete or inconsistent metadata, lack of standardized identifiers for compounds and taxa, and data that are embedded in figures or supplementary PDFs rather than provided in reusable tables or repositories. As a result, much effort had to be devoted to reconstructing or approximating information that could, in principle, have been shared in a ready to use form. Resources such as FAIRsharing, which map and connect reporting standards, repositories, and data policies across disciplines, provide one practical way to navigate this landscape and encourage convergence on interoperable practice (the FAIRsharing

Community et al., 2019). Adopting FAIR practices, including deposition of raw and processed data, code, and detailed metadata in open repositories at the time of publication, would greatly facilitate secondary use and synthesis across studies.

There is therefore a strong case for journals, funders, and research institutions to treat open data and open tools as core infrastructure for marine biogenics and ocean–air–health research. Requirements for repository deposition of data underlying figures, use of persistent identifiers for samples, compounds, and taxa, and publication of analysis scripts would not only increase transparency and reproducibility, they would also allow cross study integration that is essential for understanding variability across regions, timescales, and experimental systems. MarineNP is one concrete attempt to contribute to this landscape by providing an open, locally deployable database tailored to marine natural products, but it can only be as complete and accurate as the underlying published data allow.

The experiences from this thesis reflect a broader reality that extends well beyond marine and environmental sciences. Most contemporary research fields, from environmental change and ecology to biomedical science, social sciences, and data intensive physics, increasingly rely on long term, interoperable, and openly maintained data and tools. When core resources are fragile, fragmented, or confined to single institutions, the continuity and cumulative nature of science are put at risk. Investing in open, well curated, and locally deployable databases, transparent analysis workflows, and FAIR data practices is therefore not a niche concern, but a general requirement for modern science. By committing to such open infrastructures, the research community improves the chances that knowledge generated today remains accessible, reusable, and genuinely useful for future researchers, decision makers, and societies.

References

- Abraham, W.M., Bourdelais, A.J., Sabater, J.R., Ahmed, A., Lee, T.A., Serebriakov, I., Baden, D.G., 2005. Airway Responses to Aerosolized Brevetoxins in an Animal Model of Asthma. *Am. J. Respir. Crit. Care Med.* 171, 26–34. <https://doi.org/10.1164/rccm.200406-735OC>
- Adams, E.M., Casper, C.B., Allen, H.C., 2016. Effect of cation enrichment on dipalmitoylphosphatidylcholine (DPPC) monolayers at the air-water interface. *J. Colloid Interface Sci.* 478, 353–364. <https://doi.org/10.1016/j.jcis.2016.06.016>
- Aller, J.Y., Radway, J.C., Kiltbau, W.P., Bothe, D.W., Wilson, T.W., Vaillancourt, R.D., Quinn, P.K., Coffman, D.J., Murray, B.J., Knopf, D.A., 2017. Size-resolved characterization of the polysaccharidic and proteinaceous components of sea spray aerosol. *Atmos. Environ.* 154, 331–347. <https://doi.org/10.1016/j.atmosenv.2017.01.053>
- Alvarez-Fernandez, S., Lindeboom, H., Meesters, E., 2012. Temporal changes in plankton of the North Sea: community shifts and environmental drivers. *Mar. Ecol. Prog. Ser.* 462, 21–38. <https://doi.org/10.3354/meps09817>
- Anderson, S.I., Barton, A.D., Clayton, S., Dutkiewicz, S., Rynearson, T.A., 2021. Marine phytoplankton functional types exhibit diverse responses to thermal change. *Nat. Commun.* 12, 6413. <https://doi.org/10.1038/s41467-021-26651-8>
- Arrieta, J.M., Arnaud-Haond, S., Duarte, C.M., 2010. What lies underneath: Conserving the oceans' genetic resources. *Proc. Natl. Acad. Sci.* 107, 18318–18324. <https://doi.org/10.1073/pnas.0911897107>
- Asselman, J., Van Acker, E., De Rijcke, M., Tilleman, L., Van Nieuwerburgh, F., Mees, J., De Schamphelaere, K.A.C., Janssen, C.R., 2019. Marine biogenics in sea spray aerosols interact with the mTOR signaling pathway. *Sci. Rep.* 9, 675. <https://doi.org/10.1038/s41598-018-36866-3>
- Ault, A.P., Zhao, D., Ebben, C.J., Tauber, M.J., Geiger, F.M., Prather,

- K.A., Grassian, V.H., 2013. Raman microspectroscopy and vibrational sum frequency generation spectroscopy as probes of the bulk and surface compositions of size-resolved sea spray aerosol particles. *Phys. Chem. Chem. Phys.* 15, 6206. <https://doi.org/10.1039/c3cp43899f>
- Ávila-Román, J., García-Gil, S., Rodríguez-Luna, A., Motilva, V., Talero, E., 2021. Anti-Inflammatory and Anticancer Effects of Microalgal Carotenoids. *Mar. Drugs* 19, 531. <https://doi.org/10.3390/md19100531>
- Azetsu-Scott, K., Passow, U., 2004. Ascending marine particles: Significance of transparent exopolymer particles (TEP) in the upper ocean. *Limnol. Oceanogr.* 49, 741–748. <https://doi.org/10.4319/lo.2004.49.3.0741>
- Backer, L.C., Kirkpatrick, B., Fleming, L.E., Cheng, Y.S., Pierce, R., Bean, J.A., Clark, R., Johnson, D., Wanner, A., Tamer, R., Zhou, Y., Baden, D.G., 2005. Occupational Exposure to Aerosolized Brevetoxins during Florida Red Tide Events: Effects on a Healthy Worker Population. *Environ. Health Perspect.* 113, 644–649. <https://doi.org/10.1289/ehp.7502>
- Barker, D.R., Zeitlin, H., 1972. Metal-ion concentrations in sea-surface microlayer and size-separated atmospheric aerosol samples in Hawaii. *J. Geophys. Res.* 77, 5076–5086. <https://doi.org/10.1029/JC077i027p05076>
- Behrenfeld, M.J., Boss, E.S., 2014. Resurrecting the Ecological Underpinnings of Ocean Plankton Blooms. *Annu. Rev. Mar. Sci.* 6, 167–194. <https://doi.org/10.1146/annurev-marine-052913-021325>
- Behrenfeld, M.J., O'Malley, R.T., Siegel, D.A., McClain, C.R., Sarmiento, J.L., Feldman, G.C., Milligan, A.J., Falkowski, P.G., Letelier, R.M., Boss, E.S., 2006. Climate-driven trends in contemporary ocean productivity. *Nature* 444, 752–755. <https://doi.org/10.1038/nature05317>
- Berdalet, E., Pavaux, A.-S., Abós-Herrándiz, R., Travers, M., Appéré, G., Vila, M., Thomas, J., de Haro, L., Estrada, M., Medina-Pérez,

- N.I., Viure, L., Karlson, B., Lemée, R., 2022. Environmental, human health and socioeconomic impacts of *Ostreopsis* spp. Blooms in the NW Mediterranean. *Harmful Algae* 119, 102320. <https://doi.org/10.1016/j.hal.2022.102320>
- Bertram, T.H., Cochran, R.E., Grassian, V.H., Stone, E.A., 2018. Sea spray aerosol chemical composition: elemental and molecular mimics for laboratory studies of heterogeneous and multiphase reactions. *Chem. Soc. Rev.* 47, 2374–2400. <https://doi.org/10.1039/C7CS00008A>
- Blanchard, D.C., 1964. Sea-to-Air Transport of Surface Active Material. *Science* 146, 396–397. <https://doi.org/10.1126/science.146.3642.396>
- Blanchard, D.C., 1963a. The electrification of the atmosphere by particles from bubbles in the sea. *Prog. Oceanogr.* 1, 73–202. [https://doi.org/10.1016/0079-6611\(63\)90004-1](https://doi.org/10.1016/0079-6611(63)90004-1)
- Blanchard, D.C., 1963b. The electrification of the atmosphere by particles from bubbles in the sea. *Prog. Oceanogr.* 1, 73–202. [https://doi.org/10.1016/0079-6611\(63\)90004-1](https://doi.org/10.1016/0079-6611(63)90004-1)
- Boyce, S.G., 1951. Source of Atmospheric Salts. *Science* 113, 620–621. <https://doi.org/10.1126/science.113.2944.620>
- Breidenbach, J.D., French, B.W., Gordon, T.T., Kleinhenz, A.L., Khalaf, F.K., Willey, J.C., Hammersley, J.R., Mark Wooten, R., Crawford, E.L., Modyanov, N.N., Malhotra, D., Teeguarden, J.G., Haller, S.T., Kennedy, D.J., 2022. Microcystin-LR aerosol induces inflammatory responses in healthy human primary airway epithelium. *Environ. Int.* 169, 107531. <https://doi.org/10.1016/j.envint.2022.107531>
- Brereton, F., Clinch, J.P., Ferreira, S., 2008. Happiness, geography and the environment. *Ecol. Econ.* 65, 386–396. <https://doi.org/10.1016/j.ecolecon.2007.07.008>
- Burrows, S.M., Ogunro, O., Frossard, A.A., Russell, L.M., Rasch, P.J., Elliott, S.M., 2014. A physically based framework for modeling the organic fractionation of sea spray aerosol from bubble film Langmuir equilibria. *Atmospheric Chem. Phys.* 14, 13601–

13629. <https://doi.org/10.5194/acp-14-13601-2014>
- Calder, P.C., 2017. Omega-3 fatty acids and inflammatory processes: from molecules to man. *Biochem. Soc. Trans.* 45, 1105–1115. <https://doi.org/10.1042/BST20160474>
- Callaghan, A.H., Stokes, M.D., Deane, G.B., 2014. The effect of water temperature on air entrainment, bubble plumes, and surface foam in a laboratory breaking-wave analog. *J. Geophys. Res. Oceans* 119, 7463–7482. <https://doi.org/10.1002/2014JC010351>
- Casper, C.B., Verreault, D., Adams, E.M., Hua, W., Allen, H.C., 2016. Surface Potential of DPPC Monolayers on Concentrated Aqueous Salt Solutions. *J. Phys. Chem. B* 120, 2043–2052. <https://doi.org/10.1021/acs.jpcc.5b10483>
- Chan, M.N., Choi, M.Y., Ng, N.L., Chan, C.K., 2005. Hygroscopicity of Water-Soluble Organic Compounds in Atmospheric Aerosols: Amino Acids and Biomass Burning Derived Organic Species. *Environ. Sci. Technol.* 39, 1555–1562. <https://doi.org/10.1021/es049584l>
- Ching, J., Kajino, M., 2018. Aerosol mixing state matters for particles deposition in human respiratory system. *Sci. Rep.* 8, 8864. <https://doi.org/10.1038/s41598-018-27156-z>
- Chono, S., Fukuchi, R., Seki, T., Morimoto, K., 2009. Aerosolized liposomes with dipalmitoyl phosphatidylcholine enhance pulmonary insulin delivery. *J. Controlled Release* 137, 104–109. <https://doi.org/10.1016/j.jconrel.2009.03.019>
- Christiansen, S., Salter, M.E., Gorokhova, E., Nguyen, Q.T., Bilde, M., 2019. Sea Spray Aerosol Formation: Laboratory Results on the Role of Air Entrainment, Water Temperature, and Phytoplankton Biomass. *Environ. Sci. Technol.* 53, 13107–13116. <https://doi.org/10.1021/acs.est.9b04078>
- Ciavatta, S., Brewin, R.J.W., Skákala, J., Polimene, L., De Mora, L., Artioli, Y., Allen, J.I., 2018. Assimilation of Ocean-Color Plankton Functional Types to Improve Marine Ecosystem Simulations. *J. Geophys. Res. Oceans* 123, 834–854.

<https://doi.org/10.1002/2017JC013490>

- Ciminiello, P., Dell'Aversano, C., Iacovo, E.D., Fattorusso, E., Forino, M., Tartaglione, L., Benedettini, G., Onorari, M., Serena, F., Battocchi, C., Casabianca, S., Penna, A., 2014. First Finding of *Ostreopsis* cf. *ovata* Toxins in Marine Aerosols. *Environ. Sci. Technol.* 48, 3532–3540. <https://doi.org/10.1021/es405617d>
- Cipriano, R.J., Monahan, E.C., Bowyer, P.A., Woolf, D.K., 1987. Marine condensation nucleus generation inferred from whitecap simulation tank results. *J. Geophys. Res. Oceans* 92, 6569–6576. <https://doi.org/10.1029/JC092iC06p06569>
- Clarenbach, C.F., Senn, O., Brack, T., Kohler, M., Bloch, K.E., 2005. Monitoring of Ventilation During Exercise by a Portable Respiratory Inductive Plethysmograph. *Chest* 128, 1282–1290. <https://doi.org/10.1378/chest.128.3.1282>
- Cochran, R.E., Laskina, O., Jayarathne, T., Laskin, A., Laskin, J., Lin, P., Sultana, C., Lee, C., Moore, K.A., Cappa, C.D., Bertram, T.H., Prather, K.A., Grassian, V.H., Stone, E.A., 2016. Analysis of Organic Anionic Surfactants in Fine and Coarse Fractions of Freshly Emitted Sea Spray Aerosol. *Environ. Sci. Technol.* 50, 2477–2486. <https://doi.org/10.1021/acs.est.5b04053>
- Colantonio, S., Simpson, J.T., Fisher, R.J., Yavlovich, A., Belanger, J.M., Puri, A., Blumenthal, R., 2011. Quantitative Analysis of Phospholipids Using Nanostructured Laser Desorption Ionization Targets. *Lipids* 46, 469–477. <https://doi.org/10.1007/s11745-010-3493-1>
- Cosby, A.G., Lebakula, V., Smith, C.N., Wanik, D.W., Bergene, K., Rose, A.N., Swanson, D., Bloom, D.E., 2024. Accelerating growth of human coastal populations at the global and continent levels: 2000–2018. *Sci. Rep.* 14, 22489. <https://doi.org/10.1038/s41598-024-73287-x>
- Cravigan, L.T., Mallet, M.D., Vaattovaara, P., Harvey, M.J., Law, C.S., Modini, R.L., Russell, L.M., Stelcer, E., Cohen, D.D., Olsen, G., Safi, K., Burrell, T.J., Ristovski, Z., 2020. Sea spray aerosol organic enrichment, water uptake and surface tension effects.

- Atmospheric Chem. Phys. 20, 7955–7977.
<https://doi.org/10.5194/acp-20-7955-2020>
- Crocker, D.R., Kaluarachchi, C.P., Cao, R., Dinasquet, J., Franklin, E.B., Morris, C.K., Amiri, S., Petras, D., Nguyen, T., Torres, R.R., Martz, T.R., Malfatti, F., Goldstein, A.H., Tivanski, A.V., Prather, K.A., Thiemens, M.H., 2022. Isotopic Insights into Organic Composition Differences between Supramicron and Submicron Sea Spray Aerosol. *Environ. Sci. Technol.* 56, 9947–9958.
<https://doi.org/10.1021/acs.est.2c02154>
- Cunliffe, M., Engel, A., Frka, S., Gašparović, B., Guitart, C., Murrell, J.C., Salter, M., Stolle, C., Upstill-Goddard, R., Wurl, O., 2013. Sea surface microlayers: A unified physicochemical and biological perspective of the air–ocean interface. *Prog. Oceanogr.* 109, 104–116. <https://doi.org/10.1016/j.pocean.2012.08.004>
- Cunliffe, M., Upstill-Goddard, R.C., Murrell, J.C., 2011. Microbiology of aquatic surface microlayers. *FEMS Microbiol. Rev.* 35, 233–246.
<https://doi.org/10.1111/j.1574-6976.2010.00246.x>
- Daewel, U., Schrum, C., 2013. Simulating long-term dynamics of the coupled North Sea and Baltic Sea ecosystem with ECOSMO II: Model description and validation. *J. Mar. Syst.* 119–120, 30–49.
<https://doi.org/10.1016/j.jmarsys.2013.03.008>
- Daewel, U., Schrum, C., Macdonald, J.I., 2019. Towards end-to-end (E2E) modelling in a consistent NPZD-F modelling framework (ECOSMO E2E_v1.0): application to the North Sea and Baltic Sea. *Geosci. Model Dev.* 12, 1765–1789.
<https://doi.org/10.5194/gmd-12-1765-2019>
- Dahlin, M.E., O'Connor, T.E., Martinasek, M.P., 2024. A Scoping Review of the Respiratory Effects of Red Tide. *Respir. Care* 69, 1555–1563. <https://doi.org/10.4187/respcare.11955>
- Daniels, C.B., Orgeig, S., 2003. Pulmonary Surfactant: The Key to the Evolution of Air Breathing. *Physiology* 18, 151–157.
<https://doi.org/10.1152/nips.01438.2003>
- Desmit, X., Nohe, A., Borges, A.V., Prins, T., De Cauwer, K., Lagring, R., Van der Zande, D., Sabbe, K., 2020. Changes in chlorophyll

- concentration and phenology in the North Sea in relation to de-eutrophication and sea surface warming. *Limnol. Oceanogr.* 65, 828–847. <https://doi.org/10.1002/lno.11351>
- Dickson, A.G., Goyet, C., 1994. Handbook of methods for the analysis of the various parameters of the carbon dioxide system in sea water. Version 2 (No. ORNL/CDIAC--74, 10107773). <https://doi.org/10.2172/10107773>
- Dyshlovoy, S.A., 2021. Recent Updates on Marine Cancer-Preventive Compounds. *Mar. Drugs* 19, 558. <https://doi.org/10.3390/md19100558>
- Engel, A., Bange, H.W., Cunliffe, M., Burrows, S.M., Friedrichs, G., Galgani, L., Herrmann, H., Hertkorn, N., Johnson, M., Liss, P.S., Quinn, P.K., Schartau, M., Soloviev, A., Stolle, C., Upstill-Goddard, R.C., Van Pinxteren, M., Zäncker, B., 2017. The Ocean's Vital Skin: Toward an Integrated Understanding of the Sea Surface Microlayer. *Front. Mar. Sci.* 4, 165. <https://doi.org/10.3389/fmars.2017.00165>
- Facchini, M.C., Rinaldi, M., Decesari, S., Carbone, C., Finessi, E., Mircea, M., Fuzzi, S., Ceburnis, D., Flanagan, R., Nilsson, E.D., De Leeuw, G., Martino, M., Woeltjen, J., O'Dowd, C.D., 2008. Primary submicron marine aerosol dominated by insoluble organic colloids and aggregates. *Geophys. Res. Lett.* 35, 2008GL034210. <https://doi.org/10.1029/2008GL034210>
- Figgatt, M., Muscatiello, N., Wilson, L., Dziewulski, D., 2016. Harmful Algal Bloom–Associated Illness Surveillance: Lessons From Reported Hospital Visits in New York, 2008–2014. *Am. J. Public Health* 106, 440–442. <https://doi.org/10.2105/AJPH.2015.302988>
- Finlayson-Pitts, B.J., Pitts, J.N., 2000. Overview of the Chemistry of Polluted and Remote Atmospheres, in: *Chemistry of the Upper and Lower Atmosphere*. Elsevier, pp. 1–14. <https://doi.org/10.1016/B978-012257060-5/50003-4>
- Finney, D.J., 1948. Main Effects and Interactions. *J. Am. Stat. Assoc.* 43, 566–571. <https://doi.org/10.1080/01621459.1948.10483283>

- Fleming, L.E., Kirkpatrick, B., Backer, L.C., Bean, J.A., Wanner, A., Dalpra, D., Tamer, R., Zaias, J., Cheng, Y.S., Pierce, R., Naar, J., Abraham, W., Clark, R., Zhou, Y., Henry, M.S., Johnson, D., Van De Bogart, G., Bossart, G.D., Harrington, M., Baden, D.G., 2005. Initial Evaluation of the Effects of Aerosolized Florida Red Tide Toxins (Brevetoxins) in Persons with Asthma. *Environ. Health Perspect.* 113, 650–657. <https://doi.org/10.1289/ehp.7500>
- Fleming, L.E., Kirkpatrick, B., Backer, L.C., Bean, J.A., Wanner, A., Reich, A., Zaias, J., Cheng, Y.S., Pierce, R., Naar, J., Abraham, W.M., Baden, D.G., 2007. Aerosolized Red-Tide Toxins (Brevetoxins) and Asthma. *Chest* 131, 187–194. <https://doi.org/10.1378/chest.06-1830>
- Fleming, L.E., Kirkpatrick, B., Backer, L.C., Walsh, C.J., Nierenberg, K., Clark, J., Reich, A., Hollenbeck, J., Benson, J., Cheng, Y.S., Naar, J., Pierce, R., Bourdelais, A.J., Abraham, W.M., Kirkpatrick, G., Zaias, J., Wanner, A., Mendes, E., Shalat, S., Hoagland, P., Stephan, W., Bean, J., Watkins, S., Clarke, T., Byrne, M., Baden, D.G., 2011. Review of Florida red tide and human health effects. *Harmful Algae* 10, 224–233. <https://doi.org/10.1016/j.hal.2010.08.006>
- Flemisch Government, 2021. Meetnet Vlaamse Banken [WWW Document]. URL <https://meetnetvlaamsebanken.be/> (accessed 5.5.21).
- Fox, J., Monette, G., 1992. Generalized Collinearity Diagnostics. *J. Am. Stat. Assoc.* 87, 178–183. <https://doi.org/10.1080/01621459.1992.10475190>
- Fox, J., Weisberg, S., 2018. An R companion to applied regression, Third edition. ed. SAGE, Los Angeles.
- Frossard, A.A., Li, W., Gérard, V., Nozière, B., Cohen, R.C., 2018. Influence of surfactants on growth of individual aqueous coarse mode aerosol particles. *Aerosol Sci. Technol.* 52, 459–469. <https://doi.org/10.1080/02786826.2018.1424315>
- Gantt, B., Meskhidze, N., 2013. The physical and chemical

- characteristics of marine primary organic aerosol: a review. *Atmospheric Chem. Phys.* 13, 3979–3996. <https://doi.org/10.5194/acp-13-3979-2013>
- Gascon, M., Zijlema, W., Vert, C., White, M.P., Nieuwenhuijsen, M.J., 2017. Outdoor blue spaces, human health and well-being: A systematic review of quantitative studies. *Int. J. Hyg. Environ. Health* 220, 1207–1221. <https://doi.org/10.1016/j.ijheh.2017.08.004>
- Gašparović, B., Frka, S., Kozarac, Z., Nelson, A., 2008. A method for characterization of sea surface microlayer based on monolayer properties in presence and absence of phospholipids. *Anal. Chim. Acta* 620, 64–72. <https://doi.org/10.1016/j.aca.2008.05.043>
- Ghadiri, M., Young, P.M., Traini, D., 2019. Strategies to Enhance Drug Absorption via Nasal and Pulmonary Routes. *Pharmaceutics* 11, 113. <https://doi.org/10.3390/pharmaceutics11030113>
- Ghallab, D.S., Ibrahim, R.S., Mohyeldin, M.M., Shawky, E., 2024. Marine algae: A treasure trove of bioactive anti-inflammatory compounds. *Mar. Pollut. Bull.* 199, 116023. <https://doi.org/10.1016/j.marpolbul.2023.116023>
- González-Gil, R., Banas, N.S., Bresnan, E., Heath, M.R., 2022. The onset of the spring phytoplankton bloom in the coastal North Sea supports the Disturbance Recovery Hypothesis. *Biogeosciences* 19, 2417–2426. <https://doi.org/10.5194/bg-19-2417-2022>
- González-Gil, R., González Taboada, F., Cáceres, C., Largier, J.L., Anadón, R., 2018. Winter-mixing preconditioning of the spring phytoplankton bloom in the Bay of Biscay. *Limnol. Oceanogr.* 63, 1264–1282. <https://doi.org/10.1002/lno.10769>
- Griese, M., 1999. Pulmonary surfactant in health and human lung diseases: state of the art. *Eur. Respir. J.* 13, 1455–1476. <https://doi.org/10.1183/09031936.99.13614779>
- Groß, E., Di Pane, J., Boersma, M., Meunier, C.L., 2022. River discharge-related nutrient effects on North Sea coastal and offshore phytoplankton communities. *J. Plankton Res.* 44, 947–

960. <https://doi.org/10.1093/plankt/fbac049>
- Grythe, H., Ström, J., Krejci, R., Quinn, P., Stohl, A., 2014. A review of sea-spray aerosol source functions using a large global set of sea salt aerosol concentration measurements. *Atmospheric Chem. Phys.* 14, 1277–1297. <https://doi.org/10.5194/acp-14-1277-2014>
- Hadinoto, K., Phanapavudhikul, P., Kewu, Z., Tan, R.B.H., 2007. Dry powder aerosol delivery of large hollow nanoparticulate aggregates as prospective carriers of nanoparticulate drugs: Effects of phospholipids. *Int. J. Pharm.* 333, 187–198. <https://doi.org/10.1016/j.ijpharm.2006.10.009>
- Hamed, I., Özogul, F., Özogul, Y., Regenstein, J.M., 2015. Marine Bioactive Compounds and Their Health Benefits: A Review. *Compr. Rev. Food Sci. Food Saf.* 14, 446–465. <https://doi.org/10.1111/1541-4337.12136>
- Hannan, Md.A., Sohag, A.A.M., Dash, R., Haque, Md.N., Mohibbullah, Md., Oktaviani, D.F., Hossain, Md.T., Choi, H.J., Moon, I.S., 2020. Phytosterols of marine algae: Insights into the potential health benefits and molecular pharmacology. *Phytomedicine* 69, 153201. <https://doi.org/10.1016/j.phymed.2020.153201>
- Hardy, J.T., 1982. The sea surface microlayer: Biology, chemistry and anthropogenic enrichment. *Prog. Oceanogr.* 11, 307–328. [https://doi.org/10.1016/0079-6611\(82\)90001-5](https://doi.org/10.1016/0079-6611(82)90001-5)
- Hayakawa, K., Handa, N., Kawanobe, K., Wong, C.S., 1996. Factors controlling the temporal variation of fatty acids in piculate matter during a phytoplankton bloom in a marine mesocosm. *Mar. Chem.* 52, 233–244. [https://doi.org/10.1016/0304-4203\(95\)00087-9](https://doi.org/10.1016/0304-4203(95)00087-9)
- Heching, M., Lev, S., Shitenberg, D., Dicker, D., Kramer, M.R., 2021. Surfactant for the Treatment of ARDS in a Patient With COVID-19. *Chest* 160, e9–e12. <https://doi.org/10.1016/j.chest.2021.01.028>
- Hentschel, R., Bohlin, K., Van Kaam, A., Fuchs, H., Danhaive, O., 2020. Surfactant replacement therapy: from biological basis to current

- clinical practice. *Pediatr. Res.* 88, 176–183.
<https://doi.org/10.1038/s41390-020-0750-8>
- Hidalgo, A., Cruz, A., Pérez-Gil, J., 2015. Barrier or carrier? Pulmonary surfactant and drug delivery. *Eur. J. Pharm. Biopharm.* 95, 117–127. <https://doi.org/10.1016/j.ejpb.2015.02.014>
- Hjerne, O., Hajdu, S., Larsson, U., Downing, A.S., Winder, M., 2019. Climate Driven Changes in Timing, Composition and Magnitude of the Baltic Sea Phytoplankton Spring Bloom. *Front. Mar. Sci.* 6, 482. <https://doi.org/10.3389/fmars.2019.00482>
- Höfer, J., Giesecke, R., Hopwood, M.J., Carrera, V., Alarcón, E., González, H.E., 2019. The role of water column stability and wind mixing in the production/export dynamics of two bays in the Western Antarctic Peninsula. *Prog. Oceanogr., Subantarctic and Antarctic Marine Ecosystems: outlining patterns and processes in a changing ocean* 174, 105–116. <https://doi.org/10.1016/j.pocean.2019.01.005>
- Hoffman, E.J., Duce, R.A., 1976. Factors influencing the organic carbon content of marine aerosols: A laboratory study. *J. Geophys. Res.* 81, 3667–3670. <https://doi.org/10.1029/JC081i021p03667>
- Hooyberg, A., Roose, H., Grellier, J., Elliott, L.R., Lonneville, B., White, M.P., Michels, N., De Henauw, S., Vandegehuchte, M., Everaert, G., 2020. General health and residential proximity to the coast in Belgium: Results from a cross-sectional health survey. *Environ. Res.* 184, 109225. <https://doi.org/10.1016/j.envres.2020.109225>
- Iasella, S.V., Stetten, A.Z., Corcoran, T.E., Garoff, S., Przybycien, T.M., Tilton, R.D., 2018. Aerosolizing Lipid Dispersions Enables Antibiotic Transport Across Mimics of the Lung Airway Surface Even in the Presence of Pre-existing Lipid Monolayers. *J. Aerosol Med. Pulm. Drug Deliv.* 31, 212–220. <https://doi.org/10.1089/jamp.2017.1412>
- ICH, 2005. Harmonized Tripartite Guideline: Validation of Analytical Procedures: Text and Methodology.
- Jayarathne, T., Sultana, C.M., Lee, C., Malfatti, F., Cox, J.L., Pendergraft,

- M.A., Moore, K.A., Azam, F., Tivanski, A.V., Cappa, C.D., Bertram, T.H., Grassian, V.H., Prather, K.A., Stone, E.A., 2016. Enrichment of Saccharides and Divalent Cations in Sea Spray Aerosol During Two Phytoplankton Blooms. *Environ. Sci. Technol.* 50, 11511–11520. <https://doi.org/10.1021/acs.est.6b02988>
- Jungwirth, P., Rosenfeld, D., Buch, V., 2005. A possible new molecular mechanism of thundercloud electrification. *Atmospheric Res.* 76, 190–205. <https://doi.org/10.1016/j.atmosres.2004.11.016>
- Keene, W.C., Maring, H., Maben, J.R., Kieber, D.J., Pszenny, A.A.P., Dahl, E.E., Izaguirre, M.A., Davis, A.J., Long, M.S., Zhou, X., Smoydzin, L., Sander, R., 2007. Chemical and physical characteristics of nascent aerosols produced by bursting bubbles at a model air-sea interface. *J. Geophys. Res. Atmospheres* 112, 2007JD008464. <https://doi.org/10.1029/2007JD008464>
- Kientzler, C.F., Arons, A.B., Blanchard, D.C., Woodcock, A.H., 1954. Photographic Investigation of the Projection of Droplets by Bubbles Bursting at a Water Surface. *Tellus* 6, 1–7. <https://doi.org/10.1111/j.2153-3490.1954.tb01085.x>
- Kirkpatrick, B., Bean, J.A., Fleming, L.E., Kirkpatrick, G., Grief, L., Nierenberg, K., Reich, A., Watkins, S., Naar, J., 2010. Gastrointestinal emergency room admissions and Florida red tide blooms. *Harmful Algae* 9, 82–86. <https://doi.org/10.1016/j.hal.2009.08.005>
- Kirkpatrick, B., Fleming, L.E., Backer, L.C., Bean, J.A., Tamer, R., Kirkpatrick, G., Kane, T., Wanner, A., Dalpra, D., Reich, A., Baden, D.G., 2006. Environmental exposures to Florida red tides: Effects on emergency room respiratory diagnoses admissions. *Harmful Algae* 5, 526–533. <https://doi.org/10.1016/j.hal.2005.09.004>
- Koop, T., Bookhold, J., Shiraiwa, M., Pöschl, U., 2011. Glass transition and phase state of organic compounds: dependency on molecular properties and implications for secondary organic

- aerosols in the atmosphere. *Phys. Chem. Chem. Phys.* 13, 19238–19255. <https://doi.org/10.1039/C1CP22617G>
- Kruve, A., Rebane, R., Kipper, K., Oldekop, M.-L., Evard, H., Herodes, K., Ravio, P., Leito, I., 2015. Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part I. *Anal. Chim. Acta* 870, 29–44. <https://doi.org/10.1016/j.aca.2015.02.017>
- Labohá, P., Sychrová, E., Brózman, O., Sovadinová, I., Bláhová, L., Prokeš, R., Ondráček, J., Babica, P., 2023. Cyanobacteria, cyanotoxins and lipopolysaccharides in aerosols from inland freshwater bodies and their effects on human bronchial cells. *Environ. Toxicol. Pharmacol.* 98, 104073. <https://doi.org/10.1016/j.etap.2023.104073>
- Lacroix, G., Ruddick, K., Gypens, N., Lancelot, C., 2007. Modelling the relative impact of rivers (Scheldt/Rhine/Seine) and Western Channel waters on the nutrient and diatoms/*Phaeocystis* distributions in Belgian waters (Southern North Sea). *Cont. Shelf Res.* 27, 1422–1446. <https://doi.org/10.1016/j.csr.2007.01.013>
- Lancelot, C., 1983. Factors affecting phytoplankton extracellular release in the Southern Bight of the North Sea. *Mar. Ecol. Prog. Ser.* 12, 115–121.
- Lavery, A.M., Kieszak, S.M., Law, R., Bronstein, A.C., Funk, A.R., Banerji, S., Brown, K., Sollee, D.R., Backer, L.C., 2023. Harmful Algal Bloom Exposures Self-reported to Poison Centers in the United States, May–October 2019. *Public Health Reports®* 138, 865–869. <https://doi.org/10.1177/00333549221146654>
- Leck, C., Svensson, E., 2015. Importance of aerosol composition and mixing state for cloud droplet activation over the Arctic pack ice in summer. *Atmospheric Chem. Phys.* 15, 2545–2568. <https://doi.org/10.5194/acp-15-2545-2015>
- Lewis, R., Schwartz, E., 2004. Sea Salt Aerosol Production: Mechanisms, Methods, Measurements and Models—A Critical Review, Geophysical Monograph Series. American Geophysical Union, Washington, D. C. <https://doi.org/10.1029/GM152>

- Li, Y., Ezell, M.J., Finlayson-Pitts, B.J., 2011. The impact of organic coatings on light scattering by sodium chloride particles. *Atmos. Environ.* 45, 4123–4132. <https://doi.org/10.1016/j.atmosenv.2011.05.031>
- Li, Y., Schütte, W., Dekeukeleire, M., Janssen, C., Boon, N., Asselman, J., Lebeer, S., Spacova, I., De Rijcke, M., 2024. The immunostimulatory activity of sea spray aerosols: bacteria and endotoxins activate TLR4, TLR2/6, NF- κ B and IRF in human cells. *Sci. Total Environ.* 927, 171969. <https://doi.org/10.1016/j.scitotenv.2024.171969>
- Liénart, C., Savoye, N., David, V., Ramond, P., Rodriguez Tress, P., Hanquiez, V., Marieu, V., Aubert, F., Aubin, S., Bichon, S., Boinet, C., Bourasseau, L., Bozec, Y., Bréret, M., Breton, E., Caparros, J., Cariou, T., Claquin, P., Conan, P., Corre, A.-M., Costes, L., Crouvoisier, M., Del Amo, Y., Derriennic, H., Dindinaud, F., Duran, R., Durozier, M., Devesa, J., Ferreira, S., Feunteun, E., Garcia, N., Geslin, S., Grossteffan, E., Gueux, A., Guillaudeau, J., Guillou, G., Jolly, O., Lachaussée, N., Lafont, M., Lagadec, V., Lamoureux, J., Lauga, B., Lebreton, B., Lecuyer, E., Lehodey, J.-P., Leroux, C., L’Helguen, S., Macé, E., Maria, E., Mousseau, L., Nowaczyk, A., Pineau, P., Petit, F., Pujo-Pay, M., Raimbault, P., Rimmelin-Maury, P., Rouaud, V., Sauriau, P.-G., Sultan, E., Susperregui, N., 2018. Dynamics of particulate organic matter composition in coastal systems: Forcing of spatio-temporal variability at multi-systems scale. *Prog. Oceanogr.* 162, 271–289. <https://doi.org/10.1016/j.pocean.2018.02.026>
- Lim, C.C., Yoon, J., Reynolds, K., Gerald, L.B., Ault, A.P., Heo, S., Bell, M.L., 2023a. Harmful algal bloom aerosols and human health. *eBioMedicine* 93, 104604. <https://doi.org/10.1016/j.ebiom.2023.104604>
- Lim, C.C., Yoon, J., Reynolds, K., Gerald, L.B., Ault, A.P., Heo, S., Bell, M.L., 2023b. Harmful algal bloom aerosols and human health. *eBioMedicine* 93, 104604. <https://doi.org/10.1016/j.ebiom.2023.104604>

- Lipps, W.C., Braun-Howland, E.B., Baxter, T.E., American Public Health Association, American Water Works Association, Water Environment Federation (Eds.), 2023. Standard methods for the examination of water and wastewater, 24th edition. ed. American Public Health Association, Washington.
- Long, M.S., Keene, W.C., Kieber, D.J., Frossard, A.A., Russell, L.M., Maben, J.R., Kinsey, J.D., Quinn, P.K., Bates, T.S., 2014. Light-enhanced primary marine aerosol production from biologically productive seawater. *Geophys. Res. Lett.* 41, 2661–2670. <https://doi.org/10.1002/2014GL059436>
- Marty, J.C., Saliot, A., Buat-Ménard, P., Chesselet, R., Hunter, K.A., 1979. Relationship between the lipid compositions of marine aerosols, the sea surface microlayer, and subsurface water. *J. Geophys. Res. Oceans* 84, 5707–5716. <https://doi.org/10.1029/JC084iC09p05707>
- May, N.W., Gunsch, M.J., Olson, N.E., Bondy, A.L., Kirpes, R.M., Bertman, S.B., China, S., Laskin, A., Hopke, P.K., Ault, A.P., Pratt, K.A., 2018. Unexpected Contributions of Sea Spray and Lake Spray Aerosol to Inland Particulate Matter. *Environ. Sci. Technol. Lett.* 5, 405–412. <https://doi.org/10.1021/acs.estlett.8b00254>
- Meng, S.-S., Chang, W., Lu, Z.-H., Xie, J.-F., Qiu, H.-B., Yang, Y., Guo, F.-M., 2019. Effect of surfactant administration on outcomes of adult patients in acute respiratory distress syndrome: a meta-analysis of randomized controlled trials. *BMC Pulm. Med.* 19, 9. <https://doi.org/10.1186/s12890-018-0761-y>
- Modini, R.L., Russell, L.M., Deane, G.B., Stokes, M.D., 2013. Effect of soluble surfactant on bubble persistence and bubble-produced aerosol particles. *J. Geophys. Res. Atmospheres* 118, 1388–1400. <https://doi.org/10.1002/jgrd.50186>
- Moore, M.N., 2015. Do airborne biogenic chemicals interact with the PI3K/Akt/mTOR cell signalling pathway to benefit human health and wellbeing in rural and coastal environments? *Environ. Res.* 140, 65–75. <https://doi.org/10.1016/j.envres.2015.03.015>
- Mueller, W., Milner, J., Loh, M., Vardoulakis, S., Wilkinson, P., 2022.

- Exposure to urban greenspace and pathways to respiratory health: An exploratory systematic review. *Sci. Total Environ.* 829, 154447. <https://doi.org/10.1016/j.scitotenv.2022.154447>
- Muralidharan, P., Malapit, M., Mallory, E., Hayes, D., Mansour, H.M., 2015. Inhalable nanoparticulate powders for respiratory delivery. *Nanomedicine Nanotechnol. Biol. Med.* 11, 1189–1199. <https://doi.org/10.1016/j.nano.2015.01.007>
- Naselli-Flores, L., Padisák, J., 2023. Ecosystem services provided by marine and freshwater phytoplankton. *Hydrobiologia* 850, 2691–2706. <https://doi.org/10.1007/s10750-022-04795-y>
- Nielsen, L.S., Bilde, M., 2020. Exploring controlling factors for sea spray aerosol production: temperature, inorganic ions and organic surfactants. *Tellus B Chem. Phys. Meteorol.* 72, 1801305. <https://doi.org/10.1080/16000889.2020.1801305>
- O'Dowd, C., Ceburnis, D., Ovadnevaite, J., Bialek, J., Stengel, D.B., Zacharias, M., Nitschke, U., Connan, S., Rinaldi, M., Fuzzi, S., Decesari, S., Cristina Facchini, M., Marullo, S., Santoleri, R., Dell'Anno, A., Corinaldesi, C., Tangherlini, M., Danovaro, R., 2015. Connecting marine productivity to sea-spray via nanoscale biological processes: Phytoplankton Dance or Death Disco? *Sci. Rep.* 5, 14883. <https://doi.org/10.1038/srep14883>
- O'Dowd, C.D., De Leeuw, G., 2007. Marine aerosol production: a review of the current knowledge. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 365, 1753–1774. <https://doi.org/10.1098/rsta.2007.2043>
- O'Dowd, C.D., Smith, M.H., Consterdine, I.E., Lowe, J.A., 1997. Marine aerosol, sea-salt, and the marine sulphur cycle: a short review. *Atmos. Environ.* 31, 73–80. [https://doi.org/10.1016/S1352-2310\(96\)00106-9](https://doi.org/10.1016/S1352-2310(96)00106-9)
- Orgeig, S., Daniels, C.B., 2001. The roles of cholesterol in pulmonary surfactant: insights from comparative and evolutionary studies. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* 129, 75–89. [https://doi.org/10.1016/S1095-6433\(01\)00307-5](https://doi.org/10.1016/S1095-6433(01)00307-5)
- Otero, V., Pint, S., Deneudt, K., De Rijcke, M., Mortelmans, J., Schepers, L., Cabrera, P., Sabbe, K., Vyverman, W., Vandegheuchte, M.,

- Everaert, G., 2022. Pronounced seasonal and spatial variability in determinants of phytoplankton biomass dynamics along a near–offshore gradient in the southern North Sea. *Biogeosciences Discuss.* 1–45. <https://doi.org/10.5194/bg-2022-11>
- Panda, A., Nag, K., Harbottle, R., Possmayer, F., Petersen, N., 2004. Thermodynamic studies on mixed molecular langmuir filmsPart 2. Mutual mixing of DPPC and bovine lung surfactant extract with long-chain fatty acids. *Colloids Surf. Physicochem. Eng. Asp.* 247, 9–17. [https://doi.org/10.1016/S0927-7757\(04\)00418-2](https://doi.org/10.1016/S0927-7757(04)00418-2)
- Peña, E.A., Slate, E.H., 2006. Global Validation of Linear Model Assumptions. *J. Am. Stat. Assoc.* 101, 341–354. <https://doi.org/10.1198/016214505000000637>
- Pendergraft, M.A., Belda-Ferre, P., Petras, D., Morris, C.K., Mitts, B.A., Aron, A.T., Bryant, M., Schwartz, T., Ackermann, G., Humphrey, G., Kaandorp, E., Dorrestein, P.C., Knight, R., Prather, K.A., 2023. Bacterial and Chemical Evidence of Coastal Water Pollution from the Tijuana River in Sea Spray Aerosol. *Environ. Sci. Technol.* 57, 4071–4081. <https://doi.org/10.1021/acs.est.2c02312>
- Peng, J., Yuan, J.-P., Wu, C.-F., Wang, J.-H., 2011. Fucoxanthin, a Marine Carotenoid Present in Brown Seaweeds and Diatoms: Metabolism and Bioactivities Relevant to Human Health. *Mar. Drugs* 9, 1806–1828. <https://doi.org/10.3390/md9101806>
- Petersen, L.-E., Kellermann, M.Y., Schupp, P.J., 2020. Secondary Metabolites of Marine Microbes: From Natural Products Chemistry to Chemical Ecology, in: Jungblut, S., Liebich, V., Bode-Dalby, M. (Eds.), *YOUMARES 9 - The Oceans: Our Research, Our Future*. Springer International Publishing, Cham, pp. 159–180. https://doi.org/10.1007/978-3-030-20389-4_8
- Pierce, R.H., Henry, M.S., Blum, P.C., Hamel, S.L., Kirkpatrick, B., Cheng, Y.S., Zhou, Y., Irvin, C.M., Naar, J., Weidner, A., Fleming, L.E., Backer, L.C., Baden, D.G., 2005. Brevetoxin composition

- in water and marine aerosol along a Florida beach: Assessing potential human exposure to marine biotoxins. *Harmful Algae* 4, 965–972. <https://doi.org/10.1016/j.hal.2004.11.004>
- Pilson, M.E.Q. (Ed.), 2012. Major constituents of seawater, in: *An Introduction to the Chemistry of the Sea*. Cambridge University Press, Cambridge, pp. 66–73. <https://doi.org/10.1017/CBO9781139047203.005>
- Piotrowicz, S.R., Ray, B.J., Hoffman, G.L., Duce, R.A., 1972. Trace metal enrichment in the sea-surface microlayer. *J. Geophys. Res.* 77, 5243–5254. <https://doi.org/10.1029/JC077i027p05243>
- Poelma, D.L., Ju, M.R., Bakker, S.C., Zimmermann, L.J., Lachmann, B.F., Van Iwaarden, J.F., 2004. A Common Pathway for the Uptake of Surfactant Lipids by Alveolar Cells. *Am. J. Respir. Cell Mol. Biol.* 30, 751–758. <https://doi.org/10.1165/rcmb.2003-0127OC>
- Poli, M., Ruiz-Olvera, P., Nalca, A., Ruiz, S., Livingston, V., Frick, O., Dyer, D., Schellhase, C., Raymond, J., Kulis, D., Anderson, D., McGrath, S., Deeds, J., 2018. Toxicity and pathophysiology of palytoxin congeners after intraperitoneal and aerosol administration in rats. *Toxicon* 150, 235–250. <https://doi.org/10.1016/j.toxicon.2018.06.067>
- Prather, K.A., Bertram, T.H., Grassian, V.H., Deane, G.B., Stokes, M.D., DeMott, P.J., Aluwihare, L.I., Palenik, B.P., Azam, F., Seinfeld, J.H., Moffet, R.C., Molina, M.J., Cappa, C.D., Geiger, F.M., Roberts, G.C., Russell, L.M., Ault, A.P., Baltrusaitis, J., Collins, D.B., Corrigan, C.E., Cuadra-Rodriguez, L.A., Ebben, C.J., Forestieri, S.D., Guasco, T.L., Hersey, S.P., Kim, M.J., Lambert, W.F., Modini, R.L., Mui, W., Pedler, B.E., Ruppel, M.J., Ryder, O.S., Schoepp, N.G., Sullivan, R.C., Zhao, D., 2013. Bringing the ocean into the laboratory to probe the chemical complexity of sea spray aerosol. *Proc. Natl. Acad. Sci.* 110, 7550–7555. <https://doi.org/10.1073/pnas.1300262110>
- Quinn, P.K., Bates, T.S., Schulz, K.S., Coffman, D.J., Frossard, A.A., Russell, L.M., Keene, W.C., Kieber, D.J., 2014. Contribution of

- sea surface carbon pool to organic matter enrichment in sea spray aerosol. *Nat. Geosci.* 7, 228–232. <https://doi.org/10.1038/ngeo2092>
- Quinn, P.K., Collins, D.B., Grassian, V.H., Prather, K.A., Bates, T.S., 2015. Chemistry and Related Properties of Freshly Emitted Sea Spray Aerosol. *Chem. Rev.* 115, 4383–4399. <https://doi.org/10.1021/cr500713g>
- Rastelli, E., Corinaldesi, C., Dell’Anno, A., Lo Martire, M., Greco, S., Cristina Facchini, M., Rinaldi, M., O’Dowd, C., Ceburnis, D., Danovaro, R., 2017. Transfer of labile organic matter and microbes from the ocean surface to the marine aerosol: an experimental approach. *Sci. Rep.* 7, 11475. <https://doi.org/10.1038/s41598-017-10563-z>
- Rinaldi, M., Fuzzi, S., Decesari, S., Marullo, S., Santoleri, R., Provenzale, A., Von Hardenberg, J., Ceburnis, D., Vaishya, A., O’Dowd, C.D., Facchini, M.C., 2013. Is chlorophyll- *a* the best surrogate for organic matter enrichment in submicron primary marine aerosol? *J. Geophys. Res. Atmospheres* 118, 4964–4973. <https://doi.org/10.1002/jgrd.50417>
- Rook, G.A.W., Martinelli, R., Brunet, L.R., 2003. Innate immune responses to mycobacteria and the downregulation of atopic responses. *Curr. Opin. Allergy Clin. Immunol.* 3, 337.
- Rousseau, V., Lantoine, F., Rodriguez, F., LeGall, F., Chrétiennot-Dinet, M.-J., Lancelot, C., 2013. Characterization of *Phaeocystis globosa* (Prymnesiophyceae), the blooming species in the Southern North Sea. *J. Sea Res.* 76, 105–113. <https://doi.org/10.1016/j.seares.2012.07.011>
- Salter, M., Hamacher-Barth, E., Leck, C., Werner, J., Johnson, C., Riipinen, I., Nilsson, E., Zieger, P., 2016. Calcium enrichment in sea spray aerosol particles. *Geophys. Res. Lett.* 43, n/a-n/a. <https://doi.org/10.1002/2016GL070275>
- Saltzman, E.S., 2009. Marine aerosols, in: Le Quéré, C., Saltzman, E.S. (Eds.), *Geophysical Monograph Series*. American Geophysical Union, Washington, D. C., pp. 17–35.

- <https://doi.org/10.1029/2008GM000769>
- Samsudin, D.D., 2013. Current Issues and Challenges in the Use of Aerosolized Surfactant for Respiratory Distress Syndrome in the Newborns. *Indones. Biomed. J.* 5, 91. <https://doi.org/10.18585/inabj.v5i2.57>
- Sandifer, P. A. (2023). Linking coastal environmental and health observations for human wellbeing. *Frontiers in Public Health*, 11. <https://doi.org/10.3389/fpubh.2023.1202118>
- Schiffer, J.M., Mael, L.E., Prather, K.A., Amaro, R.E., Grassian, V.H., 2018. Sea Spray Aerosol: Where Marine Biology Meets Atmospheric Chemistry. *ACS Cent. Sci.* 4, 1617–1623. <https://doi.org/10.1021/acscentsci.8b00674>
- Schultze, S.M., Hemmings, B.A., Niessen, M., Tschopp, O., 2012. PI3K/AKT, MAPK and AMPK signalling: protein kinases in glucose homeostasis. *Expert Rev. Mol. Med.* 14, e1. <https://doi.org/10.1017/S1462399411002109>
- Schulz, M., De Leeuw, G., Balkanski, Y., 2004. Sea-salt aerosol source functions and emissions, in: Granier, C., Artaxo, P., Reeves, C.E. (Eds.), *Emissions of Atmospheric Trace Compounds, Advances in Global Change Research*. Springer Netherlands, Dordrecht, pp. 333–359. https://doi.org/10.1007/978-1-4020-2167-1_9
- Schwier, A.N., Sellegri, K., Mas, S., Charrière, B., Pey, J., Rose, C., Temime-Roussel, B., Jaffrezo, J.-L., Parin, D., Picard, D., Ribeiro, M., Roberts, G., Sempéré, R., Marchand, N., D'Anna, B., 2017. Primary marine aerosol physical flux and chemical composition during a nutrient enrichment experiment in mesocosms in the Mediterranean Sea. *Atmospheric Chem. Phys.* 17, 14645–14660. <https://doi.org/10.5194/acp-17-14645-2017>
- Sellegri, K., O'Dowd, C.D., Yoon, Y.J., Jennings, S.G., de Leeuw, G., 2006. Surfactants and submicron sea spray generation. *J. Geophys. Res. Atmospheres* 111. <https://doi.org/10.1029/2005JD006658>
- Sievering, H., Caine, J., Harvey, M., McGregor, J., Nichol, S., Quinn,

- P., 2004. Aerosol non-sea-salt sulfate in the remote marine boundary layer under clear-sky and normal cloudiness conditions: Ocean-derived biogenic alkalinity enhances sea-salt sulfate production by ozone oxidation. *J. Geophys. Res. Atmospheres* 109, 2003JD004315. <https://doi.org/10.1029/2003JD004315>
- Silva, R. A., Rogers, K., & Buckley, T. J. (2018). Advancing Environmental Epidemiology to Assess the Beneficial Influence of the Natural Environment on Human Health and Well-Being. *Environmental Science & Technology*, 52(17), 9545–9555. <https://doi.org/10.1021/acs.est.8b01781>
- Song, X., Xiong, Y., Qi, X., Tang, W., Dai, J., Gu, Q., Li, J., 2018. Molecular Targets of Active Anticancer Compounds Derived from Marine Sources. *Mar. Drugs* 16, 175. <https://doi.org/10.3390/md16050175>
- Stuhlman, O., 1932. The Mechanics of Effervescence. *Physics* 2, 457–466. <https://doi.org/10.1063/1.1745071>
- Su, B., Wang, T., Zhang, G., Liang, Y., Lv, C., Hu, Y., Li, L., Zhou, Z., Wang, X., & Bi, X., 2022. A review of atmospheric aging of sea spray aerosols: Potential factors affecting chloride depletion. *Atmospheric Environment*, 290, 119365. <https://doi.org/10.1016/j.atmosenv.2022.119365>
- Suzumura, M., 2005. Phospholipids in marine environments: a review. *Talanta, Analysis of Phosphorus in Environmental and Agricultural Samples* 66, 422–434. <https://doi.org/10.1016/j.talanta.2004.12.008>
- Syedain, Z.H., Naqwi, A.A., Dolovich, M., Somani, A., 2015. *In Vitro* Evaluation of a Device for Intra-Pulmonary Aerosol Generation and Delivery. *Aerosol Sci. Technol.* 49, 747–752. <https://doi.org/10.1080/02786826.2015.1067670>
- T, C., Chandran P, A., Kurup G, M., 2015. Omega-3 fatty acid concentrate from *Dunaliella salina* possesses anti-inflammatory properties including blockade of NF- κ B nuclear translocation. *Immunopharmacol. Immunotoxicol.* 37, 81–89.

- <https://doi.org/10.3109/08923973.2014.981639>
- Talero, E., García-Mauriño, S., Ávila-Román, J., Rodríguez-Luna, A., Alcaide, A., Motilva, V., 2015. Bioactive Compounds Isolated from Microalgae in Chronic Inflammation and Cancer. *Mar. Drugs* 13, 6152–6209. <https://doi.org/10.3390/md13106152>
- Thorpe, S.A., Humphries, P.N., 1980. Bubbles and breaking waves. *Nature* 283, 463–465. <https://doi.org/10.1038/283463a0>
- Triesch, N., van Pinxteren, M., Frka, S., Stolle, C., Spranger, T., Hoffmann, E.H., Gong, X., Wex, H., Schulz-Bull, D., Gašparović, B., Herrmann, H., 2021. Concerted measurements of lipids in seawater and on submicrometer aerosol particles at the Cabo Verde islands: biogenic sources, selective transfer and high enrichments. *Atmospheric Chem. Phys.* 21, 4267–4283. <https://doi.org/10.5194/acp-21-4267-2021>
- Ullah, S., Sandqvist, S., Beck, O., 2015. Measurement of Lung Phosphatidylcholines in Exhaled Breath Particles by a Convenient Collection Procedure. *Anal. Chem.* 87, 11553–11560. <https://doi.org/10.1021/acs.analchem.5b03433>
- Van Acker, E., De Rijcke, M., Asselman, J., Beck, I.M., Huysman, S., Vanhaecke, L., De Schamphelaere, K.A.C., Janssen, C.R., 2020. Aerosolizable Marine Phycotoxins and Human Health Effects: In Vitro Support for the Biogenics Hypothesis. *Mar. Drugs* 18, 46. <https://doi.org/10.3390/md18010046>
- Van Acker, E., Huysman, S., De Rijcke, M., Asselman, J., De Schamphelaere, K.A.C., Vanhaecke, L., Janssen, C.R., 2021. Phycotoxin-Enriched Sea Spray Aerosols: Methods, Mechanisms, and Human Exposure. *Environ. Sci. Technol.* 55, 6184–6196. <https://doi.org/10.1021/acs.est.1c00995>
- Van Meulebroek, L., De Paepe, E., Vercruysse, V., Pomian, B., Bos, S., Lapauw, B., Vanhaecke, L., 2017. Holistic Lipidomics of the Human Gut Phenotype Using Validated Ultra-High-Performance Liquid Chromatography Coupled to Hybrid Orbitrap Mass Spectrometry. *Anal. Chem.* 89, 12502–12510. <https://doi.org/10.1021/acs.analchem.7b03606>

- Velino, C., Carella, F., Adamiano, A., Sanguinetti, M., Vitali, A., Catalucci, D., Bugli, F., Iafisco, M., 2019. Nanomedicine Approaches for the Pulmonary Treatment of Cystic Fibrosis. *Front. Bioeng. Biotechnol.* 7, 406. <https://doi.org/10.3389/fbioe.2019.00406>
- Vila, M., Abós-Herrándiz, R., Isern-Fontanet, J., Álvarez, J., Berdalet, E., 2016. Establishing the link between *Ostreopsis cf. ovata* blooms and human health impacts using ecology and epidemiology. *Sci. Mar.* 80, 107–115. <https://doi.org/10.3989/scimar.04395.08A>
- Vu, T.V., Delgado-Saborit, J.M., Harrison, R.M., 2015. A review of hygroscopic growth factors of submicron aerosols from different sources and its implication for calculation of lung deposition efficiency of ambient aerosols. *Air Qual. Atmosphere Health* 8, 429–440. <https://doi.org/10.1007/s11869-015-0365-0>
- Wang, X., Sultana, C.M., Trueblood, J., Hill, T.C.J., Malfatti, F., Lee, C., Laskina, O., Moore, K.A., Beall, C.M., McCluskey, C.S., Cornwell, G.C., Zhou, Y., Cox, J.L., Pendergraft, M.A., Santander, M.V., Bertram, T.H., Cappa, C.D., Azam, F., DeMott, P.J., Grassian, V.H., Prather, K.A., 2015. Microbial Control of Sea Spray Aerosol Composition: A Tale of Two Blooms. *ACS Cent. Sci.* 1, 124–131. <https://doi.org/10.1021/acscentsci.5b00148>
- Wauthoz, N., Amighi, K., 2014. Phospholipids in pulmonary drug delivery. *Eur. J. Lipid Sci. Technol.* 116, 1114–1128. <https://doi.org/10.1002/ejlt.201300368>
- Wheeler, B.W., White, M., Stahl-Timmins, W., Depledge, M.H., 2012. Does living by the coast improve health and wellbeing? *Health Place* 18, 1198–1201. <https://doi.org/10.1016/j.healthplace.2012.06.015>
- White, M.P., Alcock, I., Wheeler, B.W., Depledge, M.H., 2013. Coastal proximity, health and well-being: Results from a longitudinal panel survey. *Health Place* 23, 97–103. <https://doi.org/10.1016/j.healthplace.2013.05.006>
- Wilson, T.W., Ladino, L.A., Alpert, P.A., Breckels, M.N., Brooks, I.M., Browse, J., Burrows, S.M., Carslaw, K.S., Huffman, J.A., Judd,

- C., Kilthau, W.P., Mason, R.H., McFiggans, G., Miller, L.A., Nájera, J.J., Polishchuk, E., Rae, S., Schiller, C.L., Si, M., Temprado, J.V., Whale, T.F., Wong, J.P.S., Wurl, O., Yakobi-Hancock, J.D., Abbatt, J.P.D., Aller, J.Y., Bertram, A.K., Knopf, D.A., Murray, B.J., 2015. A marine biogenic source of atmospheric ice-nucleating particles. *Nature* 525, 234–238. <https://doi.org/10.1038/nature14986>
- Witt, M.L.I., Mather, T.A., Baker, A.R., De Hoog, J.C.M., Pyle, D.M., 2010. Atmospheric trace metals over the south-west Indian Ocean: Total gaseous mercury, aerosol trace metal concentrations and lead isotope ratios. *Mar. Chem.* 121, 2–16. <https://doi.org/10.1016/j.marchem.2010.02.005>
- Woodcock, A., 1948. Note concerning human respiratory irritation associated with high concentrations of plankton and mass mortality of marine organisms. *J. Mar. Res.* 7.
- Zhang, R., Ren, Y., Ren, T., Yu, Y., Li, B., Zhou, X., 2025. Marine-Derived Antioxidants: A Comprehensive Review of Their Therapeutic Potential in Oxidative Stress-Associated Diseases. *Mar. Drugs* 23, 223. <https://doi.org/10.3390/md23060223>