MAREL Carnot data and metadata from Coriolis Data Center

Raed Halawi Ghosn[1,~~2~~], Émilie Poisson-Caillault[2], Guillaume Charria[3], Armel Bonnat[4], Michel Repecaud[5], Jean-Valery Facq[6], Loïc Quéméner[5], Vincent Duquesne[1], Camille Blondel[1], Alain Lefebvre[1]

[1]Ifremer, Unité Littoral, Laboratoire Environnement et Ressources, 150 quai Gambetta, 62321 Boulogne-sur-mer, France
[2]LISIC EA 4491 ULCO/University Littoral, 62228 Calais, France
[3]Ifremer, Univ. Brest, Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, F-29280, Brest, France
[4]Ifremer, Univ. Brest, Service des Systèmes d'Informations Scientifiques pour la MER, F-29280, Brest, France
[5]Ifremer , REM/RDT/DCM, Centre de Brest, Plouzané, France
[6]Ifremer, Marine Structure Laboratory, 150 Quai Gambetta, 62200 Boulogne-sur-mer, France

Correspondence to: Alain Lefebvre (alain.lefebvre@ifremer.fr)

Abstract. The French coast of the Eastern English Channel (ECC) is classified as a potential eutrophication zone by the Paris and Oslo Convention (OSPAR)~~),~~ and as moderate to poor according to the phytoplankton quality element of the Water Framework Directive (WFD). It~~Indeed, the French part of the EEC~~ is regularly affected by *Phaeocystis globosa* bloom events, which have detrimental effects on the marine ecosystem, economy, and ~~as well as~~ public health. In this context and to improve our observation strategy, ~~Since phytoplankton is an important indicator of water quality, the~~ MAREL Carnot, a multi-sensor oceanographic ~~multi-sensor~~ station, was installed in the Eastern English Channel in 2004 at the Carnot wall in Boulogne sur Mer~~ in 2004~~. The aim of this station was to collect high frequency measurements of several~~monitor~~ water quality parameters~~and phytoplankton in order~~ to complement ~~results from existing more~~ conventional low-~~-~~resolution monitoring programs.~~, with high frequency data (sampling every 20 minutes).~~ The purpose of this paper is to describe~~introduce~~ the MAREL Carnot dataset and show how it can be used for several research objectives. MAREL Carnot collects high-~~-~~frequency, multi-parameter observations from surface water~~,~~ as well as meteorological measurements~~,~~ and sends the~~send~~ data in near real-time~~almost immediately~~ to an inshore data center. In this paper, we present several physical, chemical,~~physiochemical~~ and biological parameters measured by this station. We also demonstrate~~In addition, we demonstrated, based on previous research activities,~~ that the MAREL Carnot dataset can be used to assess~~is useful for evaluating~~ environmental or ecological statuses and conduct research in the field of~~,~~ marine phytoplankton ecology, and~~physical~~ oceanography. In addition, we show that this dataset may indirectly aid in improving European environmental management strategies~~, turbulence, as well as public policy. Most importantly, we showed its contribution to Marine Strategy Framework Directive (MSFD) and other regional or universal conventions~~.

1 Introduction

For millennia, the marine environment has been subjected to various sources of pollution. Major inputs of nitrate, phosphate, and other pollutants have been causing detrimental effects on the marine environment, including harmful algal blooms (HAB) and eutrophication (Le Moal et al., 2019). Since phytoplankton are at the base of the food web, their blooms can affect higher trophic levels and cause serious changes in marine biodiversity and water quality (e.g., oxygen deficiency) (Kazmi et al., 2022; Young et al., 2020). HABs can produce toxins that degrade water quality and cause health problems in humans and marine animals (Ross Brown et al., 2022; Young et al., 2020). They can also be associated with mass foam accumulations on beaches, as with *Phaeocystis globosa* blooms (Blauw et al., 2010; Spilmont et al., 2009). Furthermore, they can detrimentally cause economic losses in sectors such as fish farms, shellfish aquaculture, tourism, and recreational activities, as well as public health (Derot et al., 2020; Hallegraeff et al., 2021).

1 Introduction

For millennia, the marine environment has been subjected to various sources of pollution. Major inputs of nitrate, phosphate, sulphate, metals, and others have been causing detrimental effects on the marine environment, including harmful algal blooms (HAB), and eutrophication (Le Moal et al., 2019). Since phytoplanktons are at the base of food webs, their blooms can affect the entire trophic levels, and can cause serious changes in the marine biodiversity and water quality (e.g. oxygen deficiency). HABs can produce toxins that degrade water quality and may cause health problems in humans and marine animals, in addition to their ability to form high biomass, which leads to foam accumulation with direct and indirect impacts. (Ross Brown et al., 2 2022). Furthermore, they can detrimentally cause economic losses in sectors such as fish farms, shellfish aquaculture, tourism and recreational activities, as well as public health (Derot et al., 2020).

Understanding the processes underlying HABs and eutrophicationthese problems necessitates continuous monitoring of the marine environmentenvironments in order to prevent the associated deterioration effects and help managers and stakeholders achieve optimizedoptimised environmental assessment and management strategies. Traditionally, monitoring aquatic and marine ecosystems was done using low frequency in -situ measurements (weekly to monthly sampling frequency). ItThis was performeddone by collecting water samples through Niskin bottles, and then performing several laboratory analysesanalysis to determine various physical, chemical,physiochemical and biological parameters, including salinity, temperature, conductivity, organic and inorganic matter, as well as phytoplankton biomass, abundance and diversity.analysis. Despite the fact that these datatests helped scientists to have an overview of the processes taking place in the marine environment, they arefailed to enhance their knowledge and understandings of insufficient temporal resolution to advance understanding ofmarine ecosystems, particularly phytoplankton dynamics and eutrophication. because of their too low sampling resolution.

In order to implement proper management strategies that prevent further deterioration of the marine ecosystem, it is crucial to enhance our understanding of algal blooms, eutrophication, recurrent, rare, and extreme events, as well as phytoplankton dynamics. Thus, it is necessary to collect continuous measurements not only on a monthly or weekly basis but also on an hourly or even sub hourly scale. Satellites and models can provide data of high spatio-temporal resolution (Chai et al., 2020), but such data must be validated with in situ data (Lefebvre and Schmitt, 2016). This motivated scientists to study the marine environment using in situ high frequency (high temporal resolution) monitoring systems such as buoys, ferry boxes, etc. (Dickey and Bidigare, 2005).

In order to be able to set proper management to prevent further deterioration of marine ecosystems, continuous measurements are needed to derive the most relevant information, not just on a monthly or weekly scale, but rather on a daily or hourly scale. In other words, high frequency measurements are needed in order to enhance our understanding of harmful algal blooms, their dynamics, as well as processes such as eutrophication. Although satellite and earth modelling data provide high frequency data, they alone, remain incapable of providing all the needed information required to set better management practices. Indeed, in-situ data is essential to calibrate and validate algorithms used by these two complementary data sources. This urged scientists and stakeholders to study the marine environment using high frequency in-situ monitoring systems, such as ferry boxes, buoys etc.

Over the past decades, the advancement of sensor technology and data science has shed light on the importance of time series in marine research. This urged the construction of autonomous systems capable of supporting long-term time series for key physical, chemical, and biological parameters. TheIn other words, the implementation of such automated systems enabled the measurement of essential ocean variables (EOV) and essentialimportant biodiversity variables (EBV) at high frequency, which aided in reorienting marine research from low frequency measurements to high frequency measurements (Blain et al., 2004). In the Eastern English Channel (EEC), HABs are mainly caused by the Prymnesiophyceae, *Phaeocystis globosa*, which is often associated with *Pseudo-nitzchia* (Karasiewicz and Lefebvre, 2022). When the temperature of the water rises in the spring and summer and nutrient concentration is optimal, *P. globosa* forms a large biomass.(Blain et al., 2004). In the Eastern English Channel (EEC), HABs are mainly caused by the Prymnesiophyceae Phaeocystis globosa, which is often associated with Pseudo-nitzchia causing severe paralytic shellfish poisoning (Karasiewicz & Lefebvre, 2022). When the temperature of the water rises in the spring and summer, P. globosa forms large biomass. In fact, *P.globosa* was identified as a potentially harmful species for several reasons. First, it releases dimethyl sulfide gas (DMS), which can irritate people's eyes, skin, and respiratory system (Riegman and Van Boekel, 1996). Second, mucopolysaccharides are abundant in its colonies (Zhu et al., 2021).systems (Riegman & Van Boekel, 1996). Second, mucco-polysaccharides are abundant in its colonies. These polysaccharides are broken up by external factors like turbulence as well as internal factors like lysis and aging, which cause the accumulation of a thick, odorous foam on the coast. Besides, needle-shaped *Pseudo-nitzschia* complex needles can stick into *P. globosa* colonies and form structures that irritate filter feeders during *P. globosa* blooms (Sazhin et al., 2007). These structures can also injure fish, making them more susceptible to bacterial and viral infections (Lefebvre and Devreker, 2023). Moreover, the neurotoxin domoic acid (DA) produced by *Pseudo-nitzschia* is responsible for the neurological disorder known as amnesic shellfish poisoning (ASP) in humans (Bates et al., 2018; Petroff et al., 2021). Additionally, marine mammals and seabirds may get poisoned if they consume DA-contaminated planktivorous prey (Delegrange et al., 2018).globosa blooms. These structures' lesions may promote viral and bacterial infections in fish, thereby affecting higher trophic levels, and reducing biodiversity (Alain & David, 2022).

The French monitoring of phytoplankton population and associated environmental factors in the Eastern English Channel

(ECC) started in 1979 with RNO (Réseau National d'Observation) or RNC (Réseau Nationale de Contrôle). Then, in 1984, a national network called REPHY (le REseau de surveillance du PHYtoplankton et des phycotoxines) was established by Ifremer to estimate the abundance and taxonomic composition of phytoplankton, describe their spatio-temporal dynamics, detect toxin-producing species, monitor, and alert for harmful blooms (https://doi.org/10.17882/47248). After that, in 1992, the Artois-Picardy Water Agency and Ifremer decided to establish SRN (Suivi Régional des Nutriments) to accurately monitor nutrient concentration (Lefebvre and Devreker, 2023). Although these monitoring networks enhanced our knowledge of phytoplankton dynamics, they remain inadequate to thoroughly understand recurrent, rare and extreme events occurring in the marine environment.

InMoreover, the neurotoxin domoic acid (DA) produced by Pseudo-nitzschia is responsible for the neurological disorder known 3 as amnesic shellfish poisoning (ASP) in humans. Additionally, marine mammals and seabirds can get poisoned if they consume DA-contaminated planktivorous prey (Delegrange et al., 2018).

The french monitoring of phytoplankton populations and associated environmental factors in the English Channel started in 1979 with RNO (Réseau National d'Observation) or RNC (Réseau Nationale de Contrôle). Then, in 1984, a national network called REPHY(le Réseau de Surveillance du Phytoplankton et des phycotoxines) was established by Ifremer, to estimate the abundance and taxonomic composition of phytoplankton, describe their spatio-temporal dynamics, detect toxin-producing species, and monitor and alert for harmful blooms (https://doi.org/10.17882/47248). After that, in 1992, the Artois-Picardy Water Agency and Ifremer decided to establish SRN (Suivi Régional des Nutriments) in response to the need of precise monitoring of nutrient concentration over a longer period of time, and to harmful algal blooms. Despite the fact these studies helped a lot in avoiding the detrimental effects of HABs, they alone remained insufficient to fully understand the dynamics of phytoplanktons and algal blooms (Dickey, 2003).

It was until 2004, thewhen MAREL (Mesures Automatisées en Réseau pour l'Environnement Littoral) Carnot monitoring station washas been installed in the French part of the ECC.English Channel. The MAREL (Mesures Automatisées en Réseau pour l'Environnement Littoral) Carnot station, developed and implemented by Ifremer (French Research Institute for Sea Exploitation), is a moored buoy protected by a tube and equipped with physical, chemical,physicochemical and biological measuring devices and sensors that operate continuously and autonomously. This multi-sensor station is located in the Boulogne-sur-Mer harbor (Eastern English Channel),) which is influenced by both marine and fresh waterwaters. It is equipped with high-performance systems for seawater analysis and data transmission in near real time. It measures the following parameters with a high frequency resolution (20 minutes): estimated sea level, gust wind speed, wind direction relative to true north, horizontal wind speed, photosynthetic active radiationrelative humidity, light irradiance surface PAR, sea water temperature, practical salinity, pH, dissolved oxygen, oxygen saturation, fluorescence, and turbidity. For nutrients, including nitrate, phosphate, and silicate, the sampling frequency is set to 12 hours.

2. Objectives

The purpose of this article is to describe~~introduce~~ the MAREL Carnot dataset and provide an overview of the~~its~~ variability of its physical, chemical, and biological parameters. For future users of the related dataset, we~~. We~~ will offer~~provide~~ a thorough~~detailed~~ description of the MAREL Carnot station, including its deployment and measurements.~~, for any future users of the associated dataset.~~ Based on previous research papers~~publications~~, we aim to demonstrate that the MAREL Carnot dataset can be used to evaluate the~~is useful for evaluating~~ environmental or ecological status and conduct research in~~statuses,~~ marine phytoplankton ecology and~~, physical~~ oceanography~~, turbulence, as well as public policy~~.

3. Materials and Methods

3.1 Location and Study Area

The~~In 2004, the~~ MAREL station was installed on the Carnot sea wall in 2004, hence the name MAREL Carnot. It is located at 50.7405N and 1.5677E~~situated~~ on the French ~~4~~ side of the Eastern English Channel, near the ~~at 50.7405N, and 1.5677W. In other words, this automated channel is situated at the~~ exit of the port of Boulogne-sur-Mer~~ harbour~~, which is France's greatest~~France's first~~ fishing port in terms of annual tonnage. Figure **1**~~. Figure 1~~ below depicts the location of MAREL Carnot station on the map.

There is no seasonal pycnocline in the Eastern English Channel (ECC), and stratification is limited and sporadic depending on freshwater discharge levels. Water can be extremely turbid due to the continental shelf nature of its seabed, which can reach a maximum depth of 180 m depending on tidal regimes. The ECC has a macro-tidal regime in the Dover Strait that varies from 3 m to 9 m during neap and spring tides, respectively (Jouanneau et al., 2013). This regime produces significant residual tidal currents from the English Channel to the North Sea, as well as high tidal currents that are nearly parallel to the shore. Fluvial supplies distributed throughout the French coast from the Bay of Seine to Cape Gris-Nez form a nearshore coastal water mass that is protected from the open ocean by a frontal area (Brylinski et al., 1996). This coastal water mass is tide-dependent and can extend from 3 to 5 miles offshore (Brylinski et al., 1991). The frontal area plays a significant role in structuring biological and non-biological exchange between coastal and offshore water masses. It is more sloped from the vertical during neap tides, resulting in a greater surface of exchange between the two water masses (Brylinski et al., 1991). Thus, particle and nutrient movement between inshore and offshore water masses is greater during neap tides than during spring tides (Lefebvre and Devreker, 2023).

~~In general, there is no seasonal pycnocline in the Eastern English Channel (ECC), and stratification is limited and sporadic depending on freshwater discharge levels. Water can be extremely turbid due to the continental shelf nature of its seabed, which can reach a maximal depth of 180 m depending on tidal regimes. Most importantly, ECC has macro-tidal regime in the Dover Strait which varies from 3 m to 9 m during neap tide and spring tide, respectively. This regime produces significant residual tidal currents from the English Channel to the North Sea, as well as high tidal currents that are nearly parallel to the shore. Fluvial supplies distributed throughout the French coast from the Bay of Seine to Cape Gris-Nez form a coastal water mass that floats near the shore and is protected from the open ocean by a frontal area (Brylinski et al., 1996) . This frontal area~~

~~plays a significant role in structuring biological and non-biological exchange between coastal and offshore water masses. However, particle and nutrient transport, as well as exchanges between inshore and offshore water masses, are tide-dependent, with neap tides being stronger than spring tides.~~

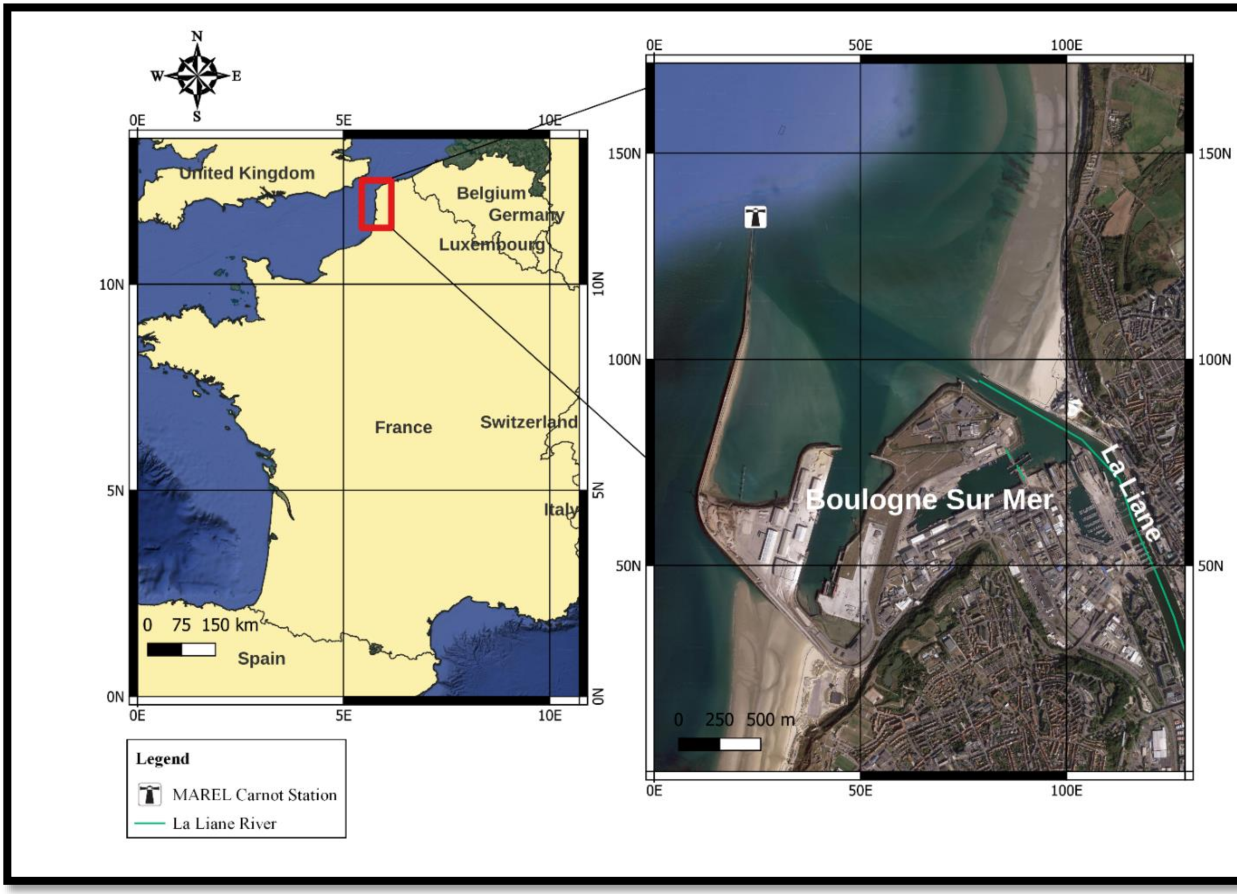


Figure 1 **Location~~The location~~ of MAREL Carnot station in the Eastern English Channel (EEC) (Map data © 2022 Google Satellite).**

3.2 Description of the MAREL Carnot Station

MAREL is a French acronym for Mesures Automatisées en Réseau pour l'Environnement Littoral (automated sampling network for coastal waters). It belongs to a network of fixed platforms extending across the entire French coast called COAST-HF (https://coast-hf.fr), which~~). In fact, COAST HF~~ is a component of the~~ IR~~ ILICO research infrastructure at the French national level (https://www.ir-ilico.fr/). MAREL Carnot station consists of a tube weighing 12 tons~~,~~ and measuring 15 meters in length. Because MAREL Carnot is located in a macrotidal~~megatidal~~ zone, it is encased in a tube to be protected~~protect it~~ from strong currents, frequent storms, and boat collisions near the port. Indeed~~In other words~~, buoys are not designed to withstand~~for~~ such challenging environments, so~~thus~~ an infrastructure to maintain~~retain~~ the buoy in a specific location and provide necessary protection was required. However, such an infrastructure would be ~~very ~~huge and expensive, so the tube was the best solution. Figure 2Figure 2 shows the MAREL Carnot measuring station, consistingwhich consists of the MAREL tube and the lighthouse platform which is used for meterological sensors, along with the light house.

Its sensors are placed on a float inside the tube in order to follow tidal movements. A pulley system is placed in a chamber inside the harbor structure to manage the cables during high and low tides and to easily lift the station for maintenance when needed. Until 2014, it was made up of a measurement cell containing several sensors. The seawater was pumped upwardupwards to be analyzedanalysed. During periods when there wereare no measurement cycles, the system was chlorinated via electrolysis to prevent biofouling. WaterIn other words, water was extractedtaken from the subsurfacesub-surface at an approximate a depth of 1.5 meters, and then sent to a measurement the passage chamber to be redistributed to different sensors.

The first version of the measuring system was constructed using electronic, computer, and mechanical equipment that date back to the 1990s. Some of these elements deteriorated over time, particularly those submerged in seawater, and had to be replaced with new equipment. In 2014, the prior measuring equipment was replaced with a new automated measuring probe. The objective was to conduct direct in-situ In order to make measurements using andirectly in-situ using a multi-parameter probe. Thus,, the system was updated in 2014, and water circulation in the chamber was no longer performedremoved to avoid air intake, which would compromise measurements and data quality. The replacement of The pulley system is placed in a chamber inside the harbor structure, allowing for the management of the cables during the tide's downward and upward movements, as well as the old measuring system with a new one consumed time due to financial and technological challenges, and hence mostraising of the data for 2014 is missing. Table 1 shows the characteristics of the sensors installed on MAREL Carnot from 2004 to 2022. Sensor calibration was performed on a regular basis, usually every three monthsstation for maintenance.

The MAREL Carnot automated station is built with 1990s electronic, computer, and mechanical components. The general aging process, which primarily affects marine-exposed systems, necessitates the replacement of a number of elements that are no longer functional and whose maintenance is impossible due to a lack of spare parts. This explains why the measurement system was replaced in 2014 with a new automated measuring probe. Hence, several data for the year 2014 are missing. The core of the system is now composed of the following elements:

-a PLC type MAREL ESTRAN

-a small in situ circulation pump (pumping of the water on the probe)

- a chlorinator for the production of chlorine by electrolysis

-a multiparameter probe type MP6 nke

-a Systea nutrient analyzer (nitrate, phosphate, silicate)

-Seabird PAR Satlantic to measure the Photosynthetic Active Radiation,

-A PONCPC-EH-10 probe for pH measurement.

Measurements are taken at 3 different levels, numbered -1, 0, 1. Level -1 denotes atmospheric measurements (+ 28 m). Level

~~0 represents water surface measurements, while level 1 represents primary levels of immersion (-1.5 m).~~

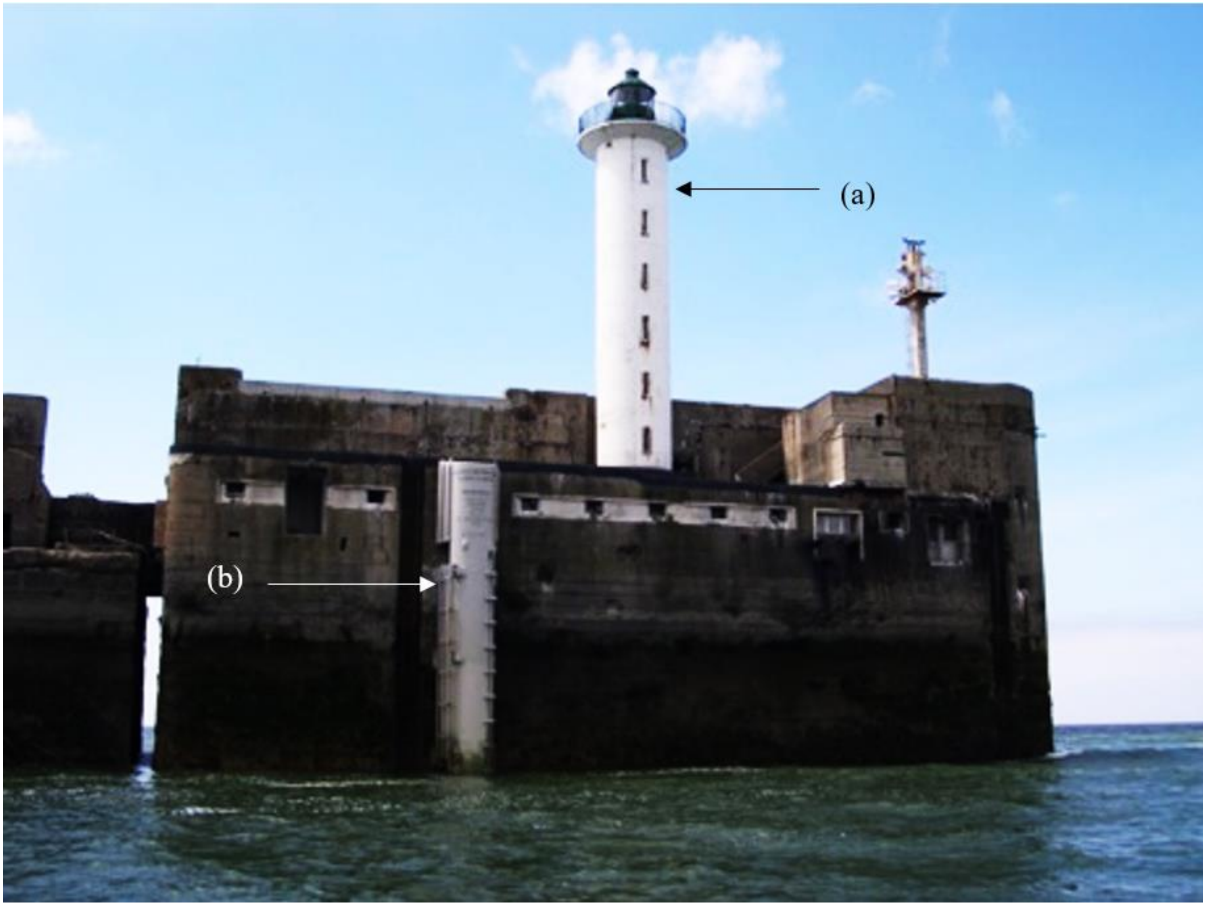


Table 1 The characteristics of the senors installed on MAREL Carnot station from 2004 to 2022

Parameter	Level of measurement	Sensor	Accuracy	Duration
Sea Water Temperature	1.5 m below water (Level 1)	Pt100	± 0.1°C	2004 - 2014
		NKE MP6	± 0.05°C	2014 - 2022
Practical Salinity	1.5 m below water (Level 1)	NKE MP6	± 0.1 PSU	2014 – 2022[(a)]
Electrical Conductivity	1.5 m below water (Level 1)	InduMax H CLS 52	± 0.3mS/cm [(b)]	2004 - 2014
		NKE MP6	± 0.05 mS/cm	2014 - 2022
Turbidity	1.5 m below water (Level 1)	TurbiMax W CUS 31	± 10%	2004 - 2014
		NKE MP6	± 5%	2014 - 2022
Dissolved Oxygen	1.5 m below water	OxyMax W COS 31	± 0.2 mg/L	2004 - 2014
		NKE MP6	± 5%	2014 - 2022
Fluorescence	1.5 m below water (Level 1)	SeaPoint Chlorophyll Fluorometer	± 10%	2004 - 2014
		NKE MP6	± 5%	2014 - 2022
PAR	28 m above water (Level -1)	LI-COR Sensor	± 5%	2004 - 2010
		Seabird PAR Satlantic	± 5%	2010 - 2022
pH	1.5 m below water (Level 1)	Orbisint CPS11	± 0.2	2004 - 2014
		NKE MP6	-[(c)]	2014 - 2022
Nutrients (Nitrate + Nitrite, Phosphate, Silicate)	1.5 m below water (Level 1)	SYSTEA NPA: Nutrient Probe Analyzer	± 5%	2004 - 2010
Wind Speed	28 m above water (Level -1)	ROWIND CV3F wind vane anemometer	± 15% RMS for Wind Speed <3.6 m/s ±6 % RMS for Wind Speed >3.6 m/s	2004 - 2015
		AirMAR 200WX	±5%	2021 - 2022
Wind Direction	28 m above water (Level -1)	ROWIND CV3F wind vane anemometer	± 2°	2004 - 2015
		AirMAR 200WX	±3°	2021 - 2022
Air Temperature	28 m above water (Level -1)	ROWIND CV3F wind vane anemometer	±1.5°C	2004 - 2015
		AirMAR 200WX	±5%	2021 - 2022
Estimated Sea Level	Sea surface (Level 0)	Hydro Ranger PLUS Siemens	-	2005-2014

[(a)]Salinity was derived from Conductivity before 2014

[(b)]The conductivity data before 2014 were deleted by Coriolis

[(c)] pH sensor failure after 2015

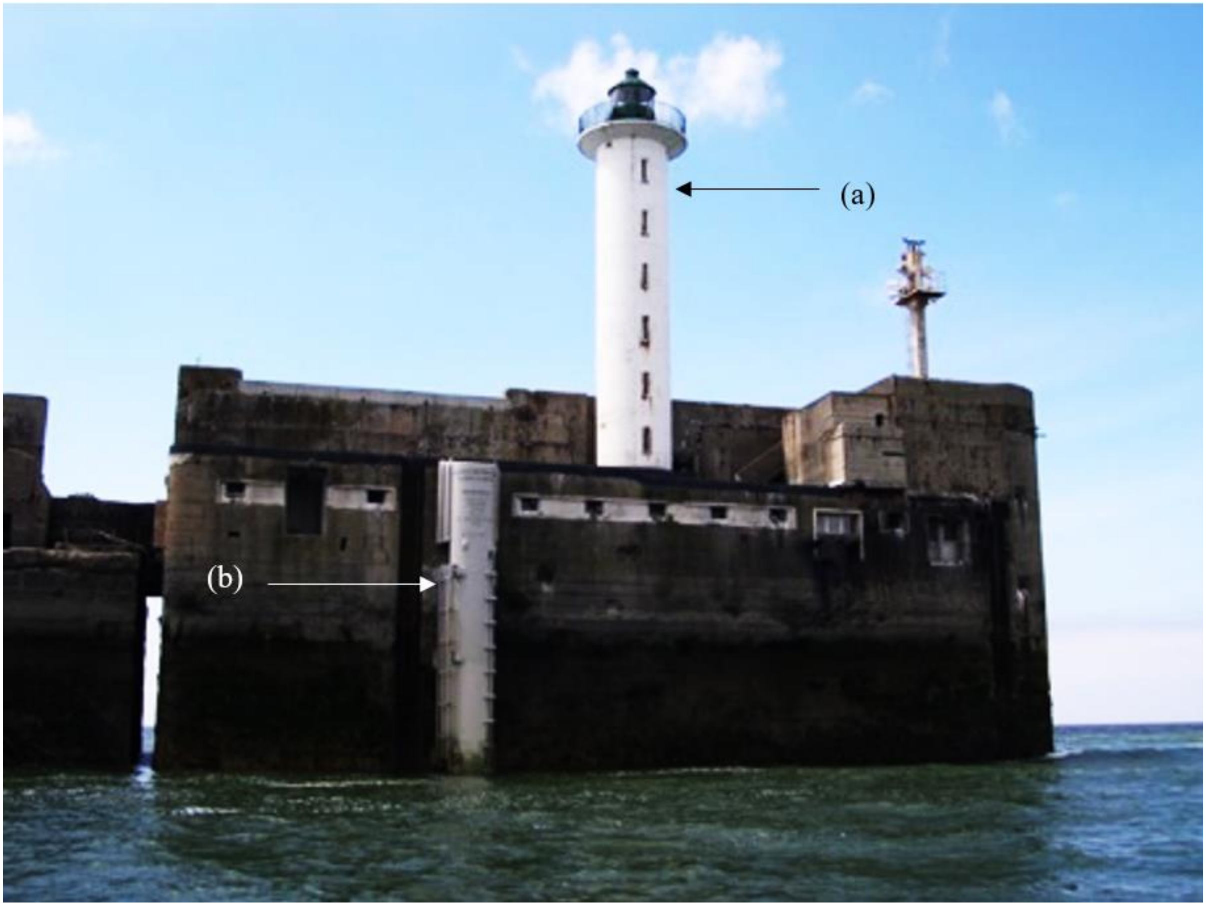


Figure 2 MAREL Carnot station consisting of the lighthouselight house (a)), and the MAREL Carnot tube (b) (photo © -Ifremer).

3.3 Measured and Calculated Parameters

The MAREL Carnot multi-sensormultisensor station measures physical, chemicalphysiochemical and biological parameters in a continuous and autonomous mode. With a sampling frequency of 20 minutes, it is capable of providing high resolution data for conductivity (S.m^{-1}), water and air temperature (°C), pH, fluorescence (FFU), turbidity (NTU), dissolved oxygen concentration (mg.L^{-1}), **P**hotosynthetically **A**ctive **R**adiation or P.A.R (µEµmol of photons.s^{-1}.m^{-2}), wind direction (degree), gust wind direction (degree), wind speed (m.s^{-1}) and gust wind speed (m.s^{-1}), as well as sea level (m). On the other hand, nutrient concentration like nitrate, phosphate, and silicate wereare only measured once every 12 hours in orderdue to limit the limited amount of chemical reagents required for the in situ analysis. Apart from salinity, which was. As a result, taking measurements twice a day allows chemicals to last longer (3 months). In addition to these measurements, certain parameters are calculated from conductivity prior to the installation of the NKE MP6 sensor in 2014, the only estimated parameter is sea level.such as oxygen saturation (%). Table 1Table 1 shows the different physiochemical parameters measured by MAREL

Carnot station, along with their sensor and expert ranges.

Table 1 Sensor and expert ranges of the various parameters measured by MAREL Carnot station

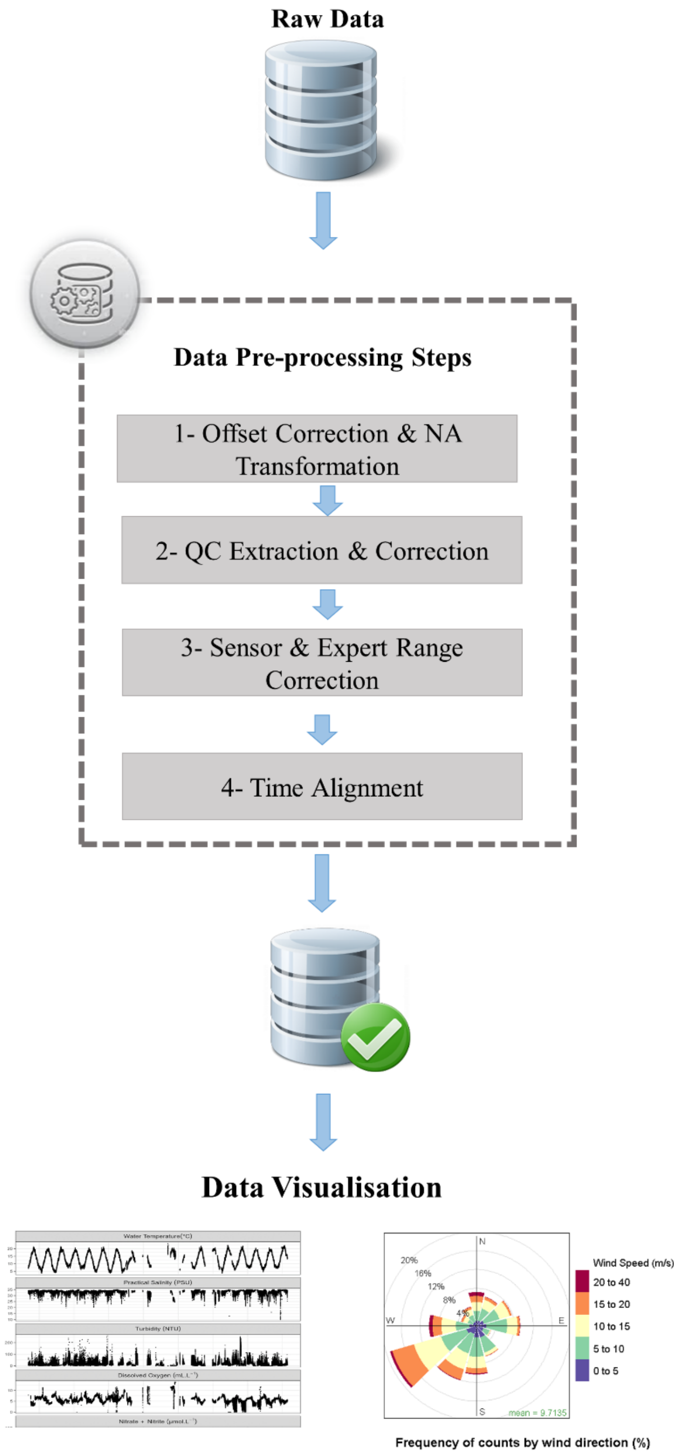
Parameter	Unit	Sensor Range	Expert Range
Fluorescence	FFU	0 – 500	0.03 – 120
pH	-	1 – 14	6.5 – 9.5
Practical Salinity	PSU	2 – 42	5 – 35
Electrical Conductivity	S/m	0 – 7	3 – 6
Sea Water Temperature	°C	-5 – 35	0 – 30
Air Temperature	°C	-20 – 40	-20 – 45
P.A.R (Photosynthetic Active Radiation)	µmol.s^{-1}.m^{-2}	0 – 5000	0 – 2500
Turbidity	NTU	0 – 500	0 – 270
Nitrate +Nitrite Concentration	µmol/L	0 – 100	0 – 100
Phosphate Concentration	µmol/L	0 – 100	0 – 10
Silicate Concentration	µmol/L	0 – 100	0 – 50
Dissolved Oxygen	mL/L	-*	-*
Dissolved Oxygen	mg/L	-*	0 – 20
Oxygen Saturation	%	0 – 150	0 – 150
Horizontal Wind Speed	m/s	0 – 40	0 – 40
Wind Direction Relative True North	degree	0 – 360	0 – 360
Observed Sea Level	m	– **	0 – 20

*: Due to complicated calculations involving several formulas and rules, the sensor range of dissolved oxygen is unavailable in mL/L. Consequently, it is also absent in mg/L

**: There is no specific sensor range for Observed Sea Level

3.4 Pre-processing of MAREL Carnot dataset

Data acquired by MAREL Carnot station are transmitted in near real-time toFigure 3 shows the Coriolis data center. Coastal CORIOLIS, or simply CORIOLIS, is a data portal for all in situ data platforms in Coastal French waters, including MAREL Carnot (https://data.coriolis-cotier.org). After downloadingdifferent steps performed for the dataset before visualization. Briefly, the variables represented in Table 1 are selected and severalsteps of pre-processing steps were performed including offset correctionare Data Correction and NA transformation, quality codeQuality Control (QC) extraction and correction, sensor and expert range correction, andas well as time alignment (Figure 3). The subsections below provide details for. Below is a detailed explanation of each step.


Raw Data
Data Pre-processing Steps
1- Offset Correction & NA Transformation
2- QC Extraction & Correction
3- Sensor & Expert Range Correction
4- Time Alignment
Data Visualisation
Water Temperature(°C)
Practical Salinity (PSU)
Turbidity (NTU)
N
S
W
E
Wind Speed (m/s)
20 to 40
15 to 20
10 to 15
5 to 10
0 to 5
Frequency of counts by wind direction (%)

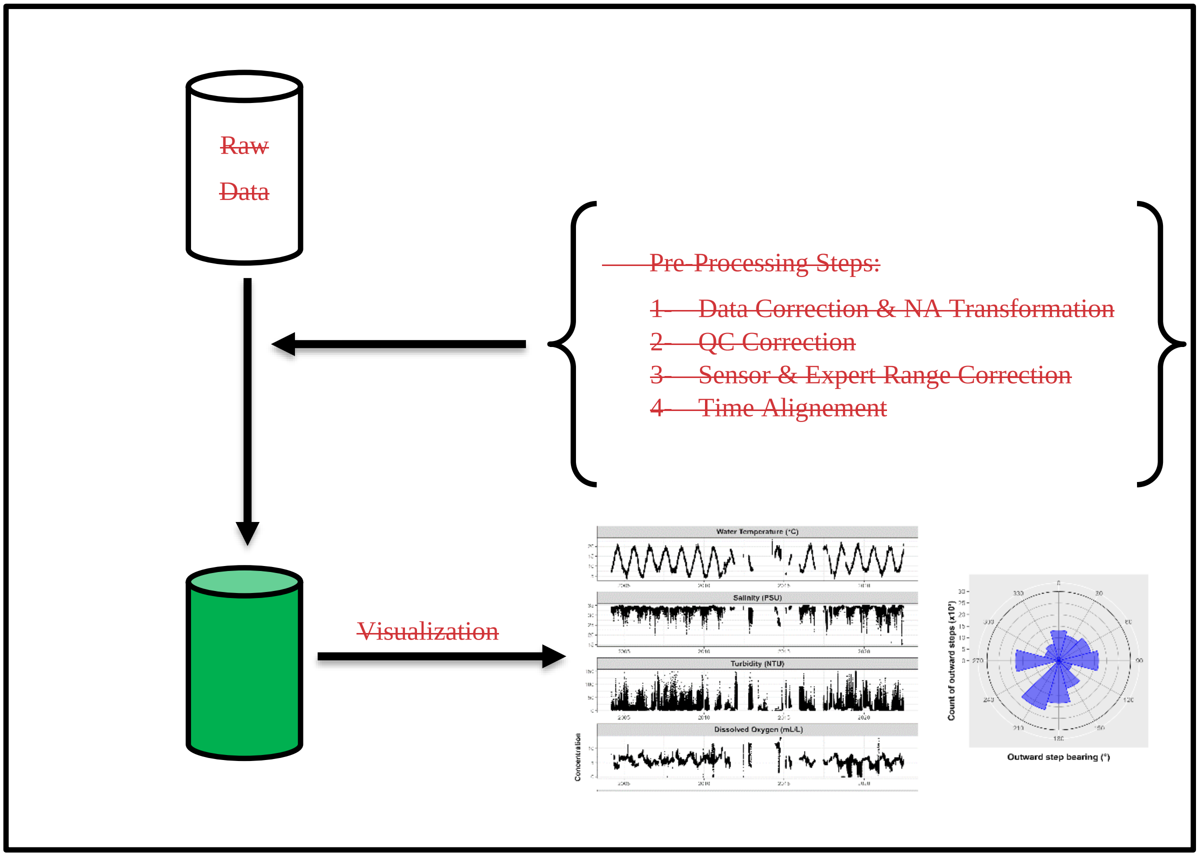


Figure 3 A simplified overview~~Schematic representation~~ of the pre-processing steps and data visualization~~ of the dataset~~

3.4.1 Offset~~Data~~ Correction and NA Transformation

We~~First, we checked the data to see if it contains major errors. After the year 2020, we~~ corrected the~~an~~ offset present in the photosynthetically active~~PAR (Photosynthetically Active~~ Radiation (PAR) and salinity variables.~~) variable.~~ Then, we noticed that~~discovered~~ some nutrient values were present on Level 2. Since no measurements are carried out ~~data for nutrients present~~ at Level 2 in~~. Since~~ MAREL Carnot, ~~doesn't contain Level 2 data, we knew that~~ these measurements were deleted~~nutrients belong to Level 1, but were wrongly introduced into Level 2. Hence, we made sure values are similar at Level 1 and Level 2, and remove all values from Level 2~~.

In addition, missing values in datasets are typically represented as NA, which stands for *Not Available*. However, in some cases, NA values are replaced with other numbers such as 77.77, 7777, 999, 999.999, 9999.99~~... Etc~~. A dataset may also include values like Inf, which stands for infinity, and Nan, which stands for infinity and Not A Number, respectively. Because these types of observations can affect or even obstruct further processing steps, we convert them into something feasible, which is NA.

3.4.2 Quality Code Extraction and Correction

Coastal CORIOLIS, or simply CORIOLIS, is a data portal for all in situ data platforms in Coastal French waters, including MAREL Carnot (https://data.coriolis-cotier.org). CORIOLIS quality control proceduresprocedure provide the users with the quality of each measurement as a Quality Code (QC) (). This code is normally given following the completion of quality control procedures, which are a part of the CORIOLIS harmonized method. During this process, the data is automatically verified using fundamental statistics (minimum, maximum, median, and standard deviation), and is subsequently validated or modified by an expert using more sophisticated methods and based on his environmental expertise. Table 2). Quality codes are assigned according to the Argo Quality Control flag scale (Wong et al., 2022), and are part of the CORIOLIS harmonized procedure applied to all its in situ data platforms.

In the raw dataset, the quality codes are present in one single column and requires deserialization. To extract the quality code of each observation, we deserialized the QC data and returned it into a matrix. According to Argo quality control manual, measurements given a QC 4 are not to be used. A flag '4' is assigned when a relevant real-time QC test has failed, or for bad measurements that are known to be not adjustable, e.g. due to sensor failure (Wong et al., 2022). Thus, all data with QC =4 (Bad data) were deleted, and replaced with *NA* values.

At the end of this step, we converted the dissolved oxygen measurements from mL/L into mg/L according to Aminot & Kérouel, (2004) using the following formula

$$DO\ (mg/L) = 1.429\ \times DO\ (mL/L)$$

Even after QC correction, the data may still contain errors. For instance, a pH measurement of 1 might not have a quality code of 4, and will therefore appear correct despite being false. For this reason, we performed sensor and expert range correction to remove values that are unusual in marine coastal waters. shows the significance of the quality code utilized with MAREL Carnot dataset. Thus, all data with QC =4 (Bad data) were deleted, and replaced with *NA* values.

Table 2 Significance of the quality code (QC)

Quality Code	Significance
0	No Quality Code was performed
1	Good data
2	Probably Good data
3	Probably bad data that are potentially correctable
4	Bad data
5	Value Changed

6	Not Used
7	Not Used
8	Interpolated Value
9	Missing Value

~~It is worth noting that the QC procedure is not always accurate, experts cannot verify all measurements, and a false value may be found under a different code, like QC=2 or QC=3. For instance, an oxygen reading of 0 is wrong but may not be consistent with QC=4. As a result, we sought an additional method for automatically eliminating a sizable portion of the potential false data in addition to QC. Hence, correction is usually performed using both "sensor" and "expert" ranges.~~

3.4.3 Sensor and Expert Range Correction

The sensor range is a range~~an interval~~ of correct values from the highest possible measurement to the lowest possible measurement set~~defined~~ by the manufacturer. The~~, while the~~ expert range is a range~~an interval~~ of correct values set~~defined~~ by a field expert. The~~Indeed, the~~ sensor range was~~ranges were~~ obtained from the information provided by the sensor suppliers and MAREL (Ifremer),~~,~~ whereas the expert range was~~ranges were~~ derived from expert ~~judgment based on specific~~ knowledge acquired in the studied area through previous research activities. For all~~ of the~~ parameters, only the values that fall within the sensor and expert ranges are kept. Values that fall outside of the ranges are replaced with NA~~missing data~~ (*Not Available*~~ or NA~~).

The ~~sensor and~~ expert range and the sensor range~~ranges for MAREL Carnot~~ are represented in **Table 3**.~~Table 1.~~ Indeed, the expert range is more precise than the sensor range. For instance, the sensor may give us a salinity value of 38, but our specialists know that salinity can only reach 35 in the Boulogne sur mer~~Eastern English Channel~~, so the sensor's result is qualified as false and must be remved~~corrected. Indeed, it is worth mentioning that scientists willing to use this dataset for any research objectives shall perform this additional pre-processing step in order to achieve higher levels of accuracy and precision~~.

Table 3 Sensor and expert ranges of parameters measured by MAREL Carnot

Parameter	**Given Name**	**Unit**	**Sensor Range**		**Expert Range**
			Before 2014	**After 2014**	
Fluorescence	Fluorescence_FFU	FFU	-	-	0 - 120
pH	pH	-	0.001 - 14	0 - 14	6.5 - 9.5
Practical Salinity	Salinity_PSU	PSU	-	2 - 42	5 - 35
Electrical Conductivity	Conductivity_S_m	S/m	-	0 – 7	3 - 6
Sea Water Temperature	Water_Temp_degreeC	°C	-5 - 30	-5 - 35	0 - 30

Air Temperature	TempAir_degreeC	°C	-10 - 50	-40 - 80	-10 - 45
P.A.R (Photosynthetic Active Radiation)	PAR_micoE_m2_s1	µE.m^{-2}.s^{-1}	-	-	0 - 2500
Turbidity	Turbidity_NTU	NTU	0 - 4000	0 – 2000	0 - 270
Nitrate + Nitrite Concentration	Nitrate_Nitrite_micromol_l	µmol/L	0 - 100	-	0 - 100
Phosphate Concentration	Phosphate_micromol_l	µmol/L	0 - 100	-	0 - 10
Silicate Concentration	Silicates_micromol_l	µmol/L	0 - 100	-	0 - 50
Dissolved Oxygen	OxyDissolved_mg_l	mg/L	0 - 20	0 - 16	0 – 20*
Oxygen Saturation	OxygenSaturation_percent	%	-	0 - 120	0 - 120
Horizontal Wind Speed	WindSPD_m_s	m/s	0 – 50.93	0 - 40	-
Wind Direction Relative True North	WindDIR_degree	degree	0-359.9	0 – 359.9	0 - 359.9
Gust Wind Speed	Gust_WindSPD_m_s	m/s	0 – 50.93	0 - 40	-
Gust Wind Direction	Gust_WindDIR_degree	degree	-	0 – 359.9	-
Relative Humidity	RelativeHumidity_Percent	%	0 - 100	0 - 100	
Atmospheric Pressure	Atmospheric_Pressure_hPa	hPa	-	300 - 1100	
Estimated Sea Level	SeaLevel_m	m	-	-	0 - 20

* value determined for the entire period from 2004 to 2022

3.4.4 Time Alignment

TheBefore statistical methods can be applied to the dataset, it must have an identical time interval between each measurement. However, the measurements of the various sensors are not taken at the same time, resulting in a time lag that can range from few seconds to several minutes. In addition, the series may contain duplicates in some cases. Before statistical methods can be appliedThus, to eliminate potential replicates and synchronize the dataset, it must have an identicalwe perform a temporal alignment using the average time interval between each measurement.of the measurements.

In order to synchronize the dataset and eliminate potential replicates, we performed a time The alignment protocol calculates the average time step. After extracting the day, month, years and hours initially present in the raw dataset, we extracted the minute's column and set minutes 0 through 19 to 10, 20 through 39 to 30, and 40 through 59 to 50. From this, we generated a time sequence of 20 minutes interval and merged it with the original data. After that, we aggregated the data at the obtained regularThis time step (20 minutes). If multiple measurements of the same creates a regular/no-replicates time variable exist within the same time step. Based on the parameter to be regularized and the goals to be attained, the maximum, /minimum, /or average can be returned. In order to focus on the most critical environmental conditions posing risk of eutrophicationof all subsets of our dataset matching to each regular interval of our ideal time variable is then returned.

~~In this paper~~, the maximum value was~~is~~ chosen for all parameters~~variables~~ except oxygen~~,~~ where the minimum value was chosen. The QC value~~is used. This is because during phytoplankton blooms, the amount~~ of each observation was then retained, and a quality code of 9 was assigned to all NA~~oxygen in the water drops. Hence, it is more interesting to use the minimum~~ values, including those removed by previously mentioned pre-processing steps ~~to highlight this feature of HABs~~.

4 Results and Discussion

~~To summarize, several pre-processing procedures were performed on the displayed dataset, including data correction and NA transformation, quality code correction, sensor range correction, as well as temporal alignment.~~ Table 4 represents the descriptive statistics for the main ~~physiochemical~~ parameters measured by MAREL Carnot from 2004 until 2022. The results show a high percentage of missing data, denoted as NA, or *Not Available*. Missing~~In fact, missing~~ data is a major problem in time series. It is primarily due to sensor failure, communication problems, or sensor maintenance disability.

Table 4 Statistical summary (minimum, first quartile, median, mean, third quartile, maximum, and percentage of NA) of the parameters measured by MAREL Carnot station

Parameters (Units)	Min	Q1	Median	Mean	Q3	Max	percentage of NA
Air Temperature (°C)	-6.18	7.82	11.9~~90~~	11.710~~71~~	16.11	35~~.00~~	56.873~~44.58~~
Gust Wind Direction (°)	7.8	124.55	226.3	205.176	271.5	359.9	94.622
Gust Wind Speed (m/s)~~Water Temperature (°C)~~	0~~3.60~~	7.27~~8.73~~	11.24~~12.80~~	12.282~~94~~	16.24~~17.30~~	~~23.~~50.9	61.532~~23.73~~
Photosynthetic Active Radiation (µE.m^{-2}.s^{-1})~~Salinity (PSU)~~	0~~10.04~~	0~~33.07~~	34.7~~33.64~~	280.850~~33.32~~	361.8~~34.09~~	2497.34~~35.00~~	60.110~~28.45~~
Relative Humidity (%)~~Turbidity (NTU)~~	25.21~~0.00~~	74.07~~4.63~~	82.05~~8.91~~	81.039~~14.62~~	88.99~~17.30~~	100~~259.70~~	63.447~~25.25~~
~~P.A.R (µmol.s^{-1}.m^{-2})~~	~~0.00~~	~~0.00~~	~~25.00~~	~~282.26~~	~~373.20~~	~~2489.03~~	~~47.81~~
Wind Direction (°)~~(Degree)~~	0~~.00~~	92~~.00~~	198~~.00~~	177.989~~99~~	246~~.00~~	359.9~~90~~	61.414~~48.14~~
Horizontal Wind Speed (m/~~.~~s^{-1})	0~~.00~~	5.78~~77~~	9.08	9.737~~71~~	13.14~~12~~	50.82~~39.96~~	61.511~~48.23~~
Estimated Sea Level (m)	5.43	8.08	10.2	10.280	12.32	15.84	73.091
Water Temperature (°C)	3.6	8.73	12.8	12.935	17.3	23.5	30.276
Electrical Conductivity (S/m)	3	3.601	3.944	3.991	4.411	4.959	69.017
Dissolved Oxygen (mg/L)	0	6.844	7.990	8.051	9.210	17.01*	37.049
Fluorescence (FFU)	0~~.03~~	0.51~~52~~	1.07~~08~~	3.004~~04~~	2.46~~48~~	116.59	25.933~~20.93~~

Nitrate + Nitrite (µmol/.L^{-1})	0.0201	5.2381	10.90513.22	15.21317.44	21.94724.06	98.8999.54	99.36665.50
Oxygen Saturation (%)	0	83.69	91.573	88.582	97.11	120	70.150
Phosphate (µmol/.L^{-1})	0.00	0.4448	0.6571	0.71995	0.8696	10.00	99.42666.00
Silicate (µmol.L^{-1})	0.00	2.10	4.44	5.55	7.79	49.04	64.73
Dissolved Oxygen (mL.L^{-1})	0.00	4.80	5.60	5.67	6.46	13.92	28.76
Oxygen Saturation (%)	0.00	83.78	91.68	88.89	97.30	198.72	54.78
pH	6.550	7.92	8.110	8.13714	8.38	9.33	67.18652.66
Practical Salinity (PSU)	8.7748	32.98	33.56	33.225	34.01	35	31.525
Silicate (µmol/L)Conductivity (S.m^{-1})	03.00	1.96253.60	4.1353.94	5.1453.99	7.31754.41	39.254.96	99.32654.10
Turbidity (NTU)Sea Level (m)	05.43	4.6348.08	8.91010.20	14.62410.28	17.312.32	15.84259.7	32.21257.29
Atmospheric Pressure (hPa)	980	1011	1018	1016.964	1024	1044	94.622

*Value obtained before 2014

Figure 4 and Figure 5Figure 5 show the time series of parameters signals collected from MAREL Carnot station from 2004 until 2022. We noticed that some signals have seasonalvisible cycles, such as water and air temperature as well as photosynthetic active radiation (PAR). In addition, the signals contain episodic or continuous missing values over several time periods. For instance, a large number of missing values can be found around the year 2014 almost in most time seriesall signals. This is due to station and sensor alterations that occurred during that time, particularly, the replacement of several sensors with a multi-parameter probe (Lefebvre and Schmitt, 2016). Likewise, the signals of air temperature, PAR, wind speed as well as sea level have been lost for several years while waiting for new funding resources to ensure the renewal of sensors and associated electronic systems. Conductivity data prior to 2015 were deleted by Coriolis data center, probably under the presumption that salinity is more relevant to the scientific community. This highlights the added value of our research, which is to ensure that all observations collected by MAREL Carnot remain permanently available and accessible to everyone. (Lefebvre and Schmitt, 2016). Similarly, signals of air temperature, PAR, wind speed as well as sea level have been lost for several years. However, conductivity signals have only been available since 2015.

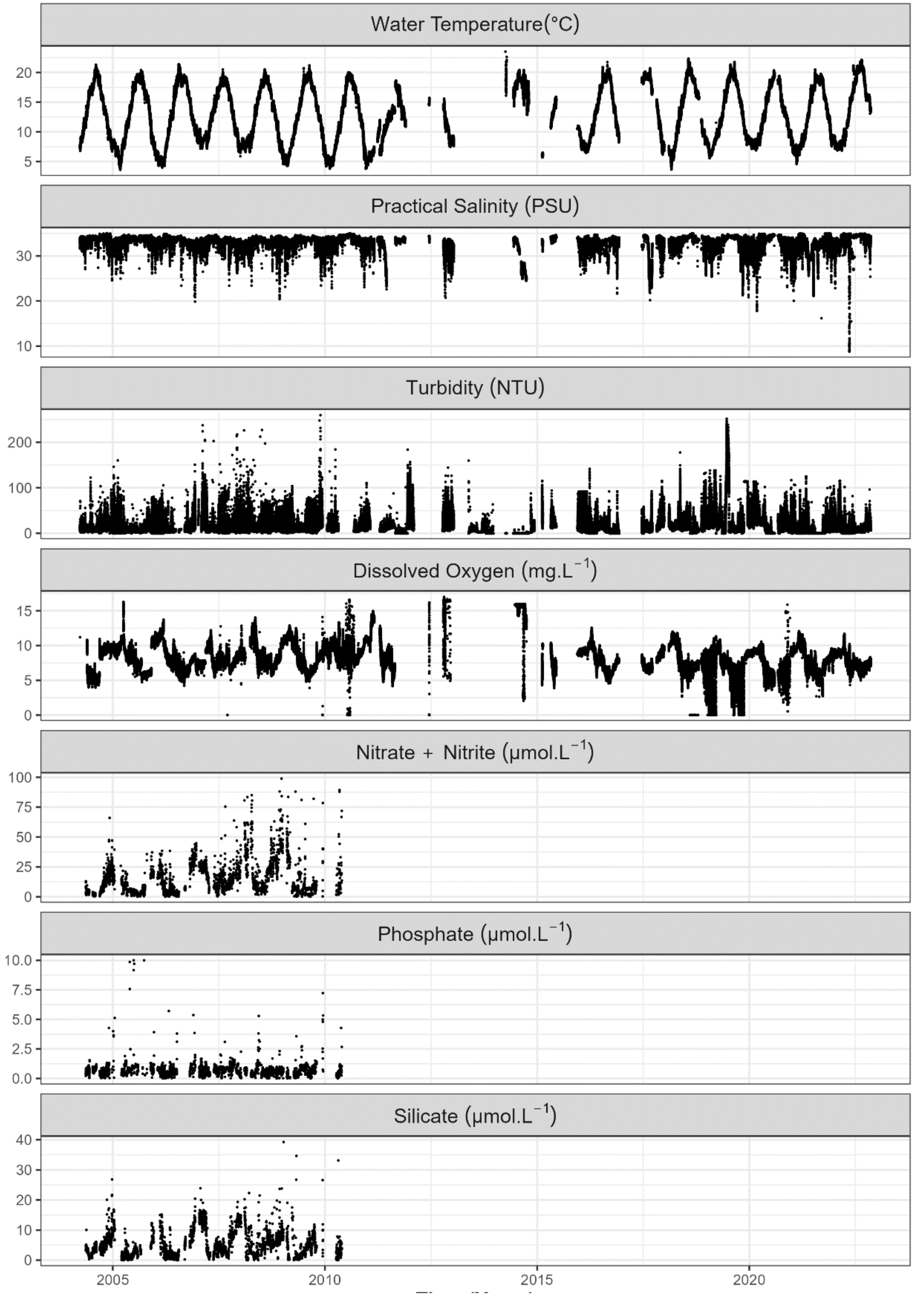
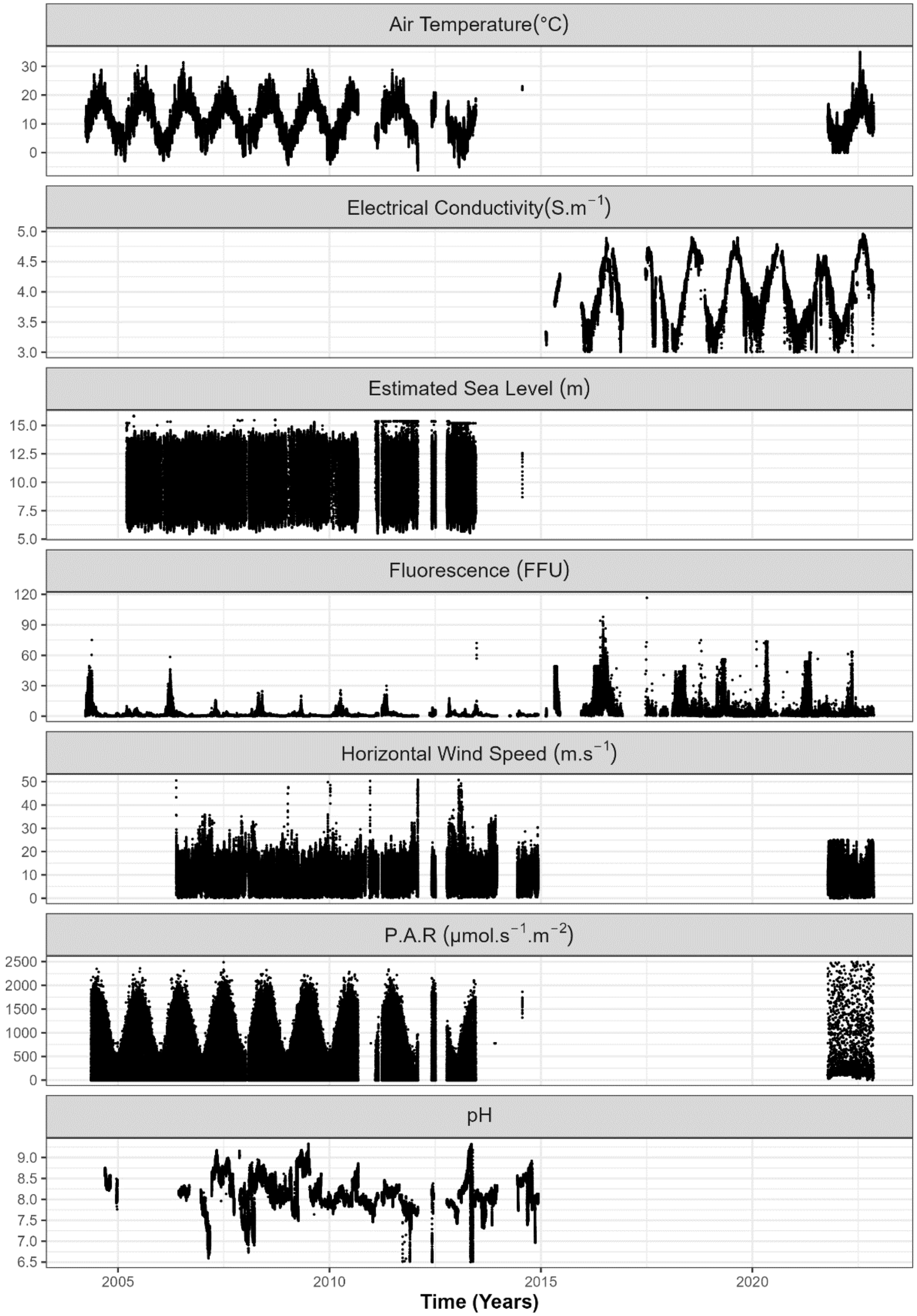
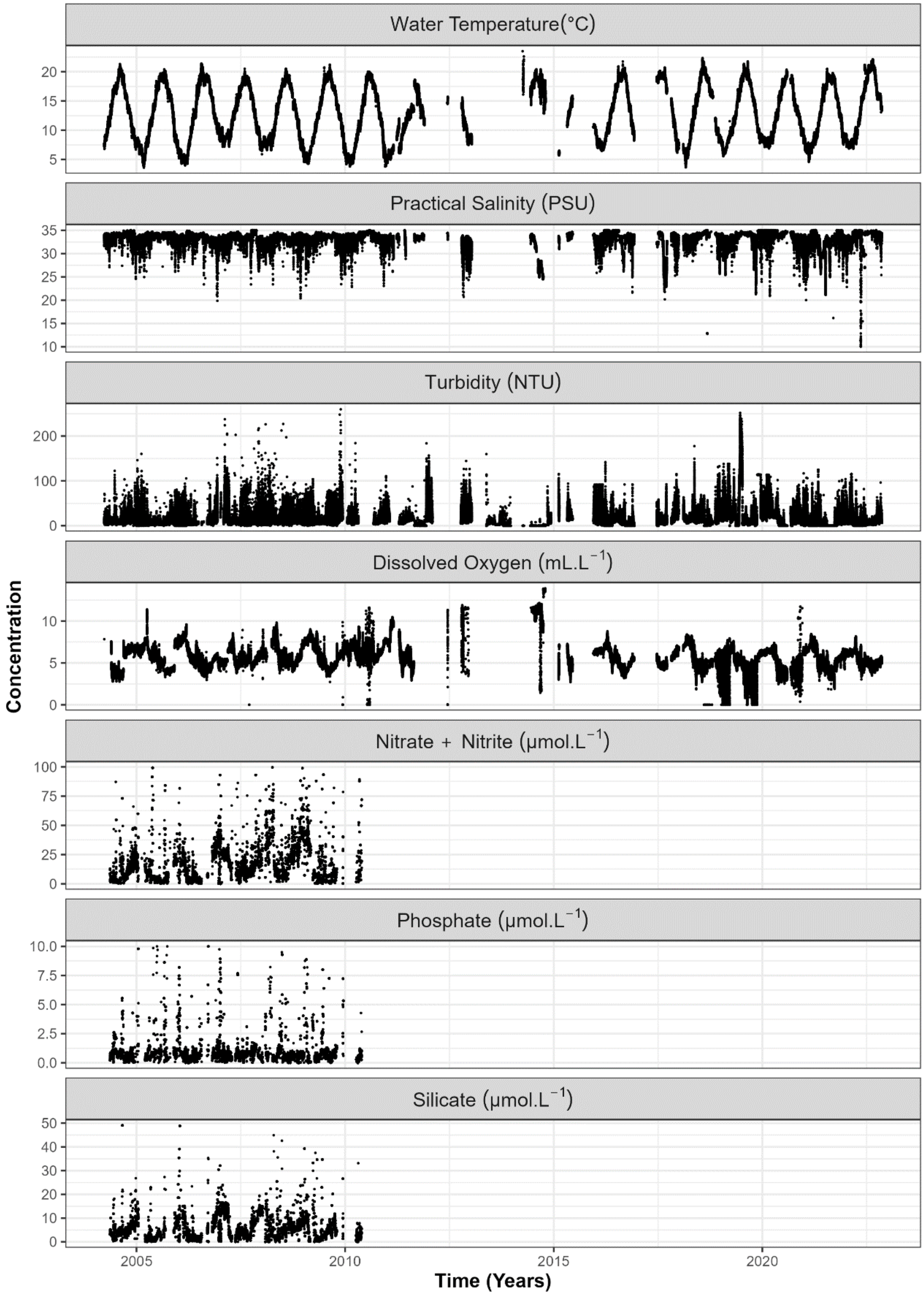


Figure 4 Time series of parameters collected from MAREL Carnot station during the period 2004-2022



Air Temperature(°C)
30
20
10
0
Electrical Conductivity(S.m⁻¹)
5.0
4.5
4.0
3.5
3.0
Estimated Sea Level (m)
15.0
12.5
10.0
7.5
5.0
Fluorescence (FFU)
120
90
60
30
0
Horizontal Wind Speed (m.s⁻¹)
50
40
30
20
10
0
P.A.R (µmol.s⁻¹.m⁻²)
2500
2000
1500
1000
500
0
pH
9.0
8.5
8.0
7.5
7.0
6.5
2005
2010
2015
2020
Time (Years)

350



~~Figure 4 Signals collected from the MAREL Carnot station during the period 2004-2022~~

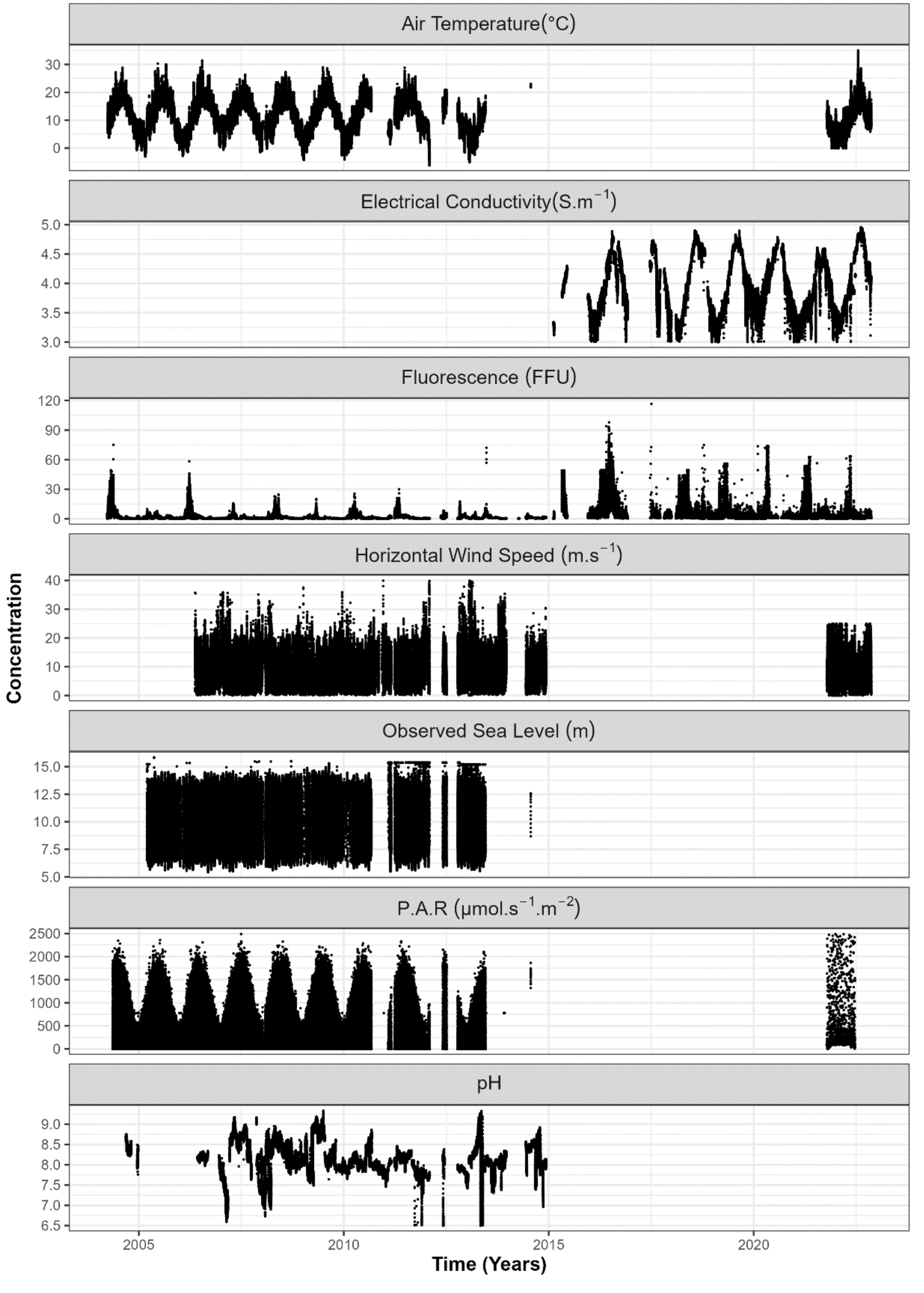
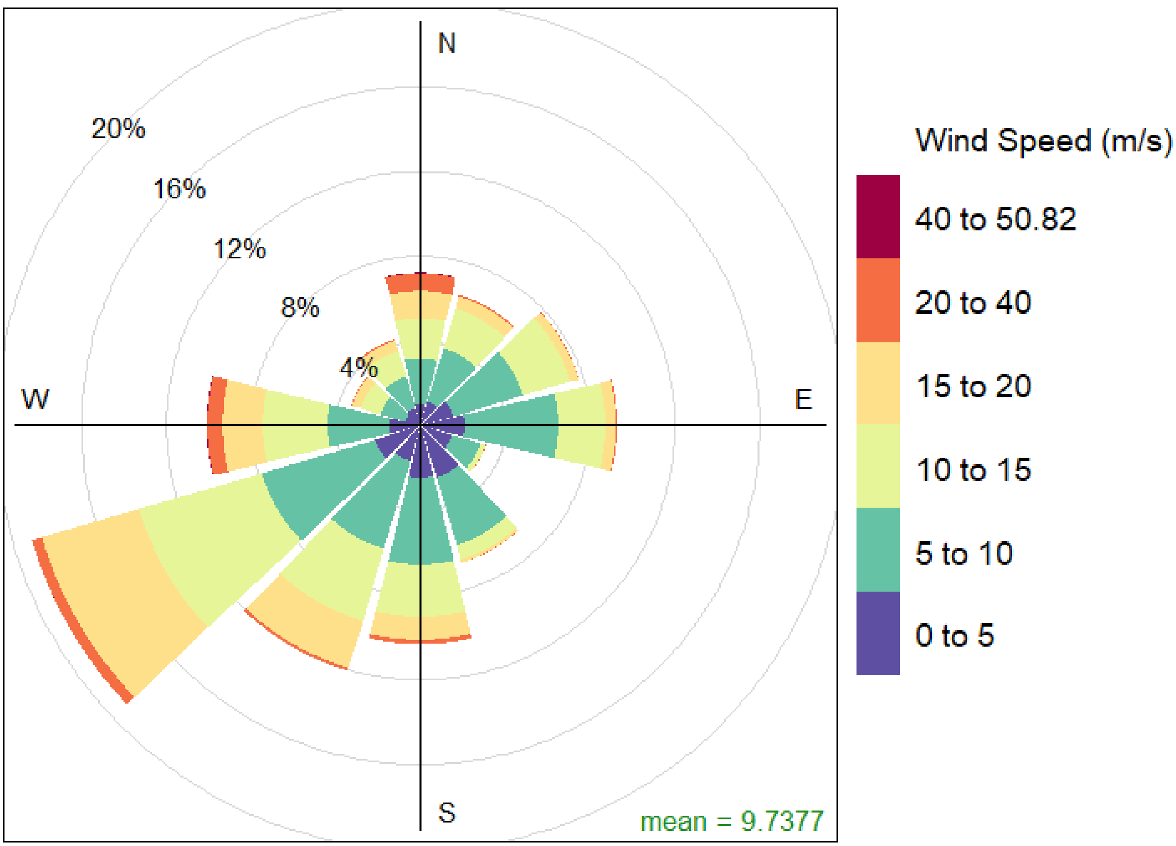


Figure 5 Time series of parameters5 Signals collected from the MAREL Carnot station during the period 2004-2022

Nutrient signals such as phosphate, nitrate, and silicate are only available until 2010. This is caused by a previous sensor failure and the inability to replace it.

Indeed, high frequency fluorescence data from MAREL Carnot station can contribute to the calibration of satellite observations. In addition, MAREL Carnot perform measurements of nutrients that are not provided by current spacecraft techniques. Hence, it is much more effective than satellites at monitoring water quality (Lefebvre and Schmitt, 2016). Nonetheless, nutrient signals such as phosphate, nitrate, and silicate are only available until 2010. This is caused by a previous sensor failure and the inability to replace it.

Figure 6 shows a wind rose showingfor the frequency (%) and, wind speed (m/s) for different wind directions measuredand direction relative true north collected by MAREL Carnot from 2004 until 2022. station, after removing all NA values.



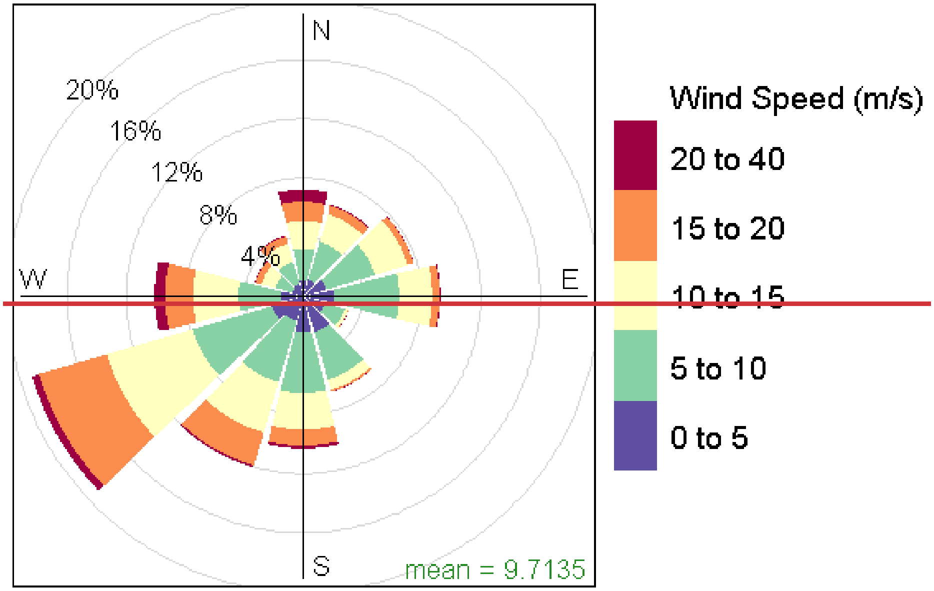


Figure 6 wind rose showing~~Wind Rose representing~~ the frequency (%) ~~wind speed~~ and wind speed (m/s) for different wind directions~~direction~~ measured by MAREL Carnot station from 2004 until 2022

Scientists~~Indeed, scientists~~ from several disciplinary backgrounds have utilized MAREL Carnot data to accomplish a wide range of research objectives. In the following paragraphs, we will go over some of the most significant findings from several research efforts. The scientific community that is interested in the MAREL Carnot dataset may find this evaluation useful in determining which topics may or may not require further study based on the results of this evaluation. This~~In general, this~~ dataset allows researchers to investigate the dynamics of phytoplankton as well as detect blooms caused by human activities and/or climate change.

For instance, Rousseeuw et al., (2015)~~Rousseeuw et al., (2015)~~ developed an unsupervised Hidden Markov Model (uHMM) for monitoring the marine environment, specifically for detecting algal blooms and understanding phytoplankton dynamics. In their~~his~~ unsupervised Hidden Markov Model, uHMM parameters were estimated using spectral clustering rather than the commonly used iterative Expectation Maximization. The results obtained using MAREL Carnot dataset showed that the proposed system is efficient to detect the main productive and non-productive periods, as used for the purposes of the EU Water Framework Directive to assess good environmental status, and refine knowledge about phytoplankton bloom dynamics in a temperate ecosystem, temporarily dominated by a harmful algae, *Phaeocystis globosa*. Thus, the suggested uHMM system successfully characterizes phytoplankton dynamics from new incoming data (in near real-time), and will enable researchers to gain a better understanding of the main controlling or forcing parameters (e.g., nutrient pressure, light availability, turbidity), the environmental status (e.g., phytoplankton biomass), and the direct and/or indirect effects of algal blooms (e.g., oxygen concentration) (Rousseeuw et al., 2015).

Following~~such blooms (e.g., oxygen concentration). Most importantly, the ability of uHMM to establish environmental states~~

represents a clear potential to better understand what a good environmental condition is, as defined and applied for the needs of the WFD, MSFD, or other regional sea conventions such as OSPAR. Even though uHMM was only applied to the MAREL Carnot dataset, it could contribute to the processing of huge multivariate time series generated by high resolution platforms, which are increasingly used for the integrated observation of pelagic ecosystems and biogeochemical cycles in oceans Rousseeuw et al., (2015) unsupervised approach, Grassi et al., (2019)(Rousseeuw et al., 2015). suggested a Multilevel Spectral Clustering (M-SC) to split multivariate time series from general patterns to extreme events without *a priori* knowledge. The results obtained from MAREL Carnot dataset have shown that we can extract knowledge on dynamics of events or environmental states. In addition, it was shown that M-SC allows unsupervised labelling of time series, which is a basic part of machine learning needed to build an event prediction system and improve sampling strategies to become in near real time (Grassi et al., 2019). As a result, scientists should be able to create a HAB early warning expert system to warn shellfish farmers, and prevent both public health risks and commercial losses in the shellfish farming business.

On the other hand, Grassi et al., (2019) suggested a Multilevel Spectral Clustering (M-SC) to split multivariate time series from general patterns to extreme events without *a priori* knowledge. The results obtained from MAREL Carnot dataset have shown that we can extract knowledge on dynamics of events or environmental states. The application of M-SC and uHMM on MAREL Carnot dataset can reveal rare, recurrent and extreme events, which may aid in improving coastal assessment and defining what constitutes a desirable environmental state. This can indirectly help improve management strategies established by the Water Framework Directive (WFD), Marine Strategy Framework Directive (MSFD) and Oslo and Paris Convention (OSPAR).

MAREL Carnot dataset can be also be beneficial to data scientists and machine learning specialists. This dataset contains some missing data due to sensor failure, and harsh weather conditions that prevent immediate sensor maintenance. It was used to evaluate the performance of a proposed Dynamic Time Warping method to fill in successive missing values of univariate time series (Phan et al., 2017b), and low uncorrelated multivariate time series (Phan et al., 2017a). It was also utilized in the application of a fuzzy logic-based similarity measure to impute large gaps of uncorrelated multivariate time series (Phan et al., 2018). These data imputation approaches are published on the Comprehensive R Archive Network (CRAN) and accessible through DTWBI and DTWUMI packages, respectively. The MAREL Carnot high frequency dataset can be used to validate satellite-derived products such as fluorescence. In addition, MAREL Carnot provide measurements for parameters that cannot be measured from space such as nutrient concentration (Lefebvre and Schmitt, 2016).

This dataset can also be utilized to assess the performance of time series analysis methods on marine datasets. For instance, Kbaier Ben Ismail et al., (2016) used four parameters measured by MAREL Carnot to compare the classical techniques of time series analysis to recent ones. Also, Huang and Schmitt, (2014) performed empirical mode decomposition (EMD) to study time dependent intrinsic correlation of temperature and dissolved oxygen time series measured by MAREL Carnot.

Derot et al., (2020) investigated the impact of different sampling frequency on forecasting harmful algal blooms. They applied Random Forest (RF) and sliding window strategy on 12 parameters derived from MAREL Carnot dataset. This research demonstrated that the sampling frequency has a direct impact on the forecast performance of a Random Forest (RF) model as high-frequency datasets might provide useful information to the RF. This type of model sets the groundwork for the creation of a numerical decision-making tool that could help mitigate the impact of algal blooms, and can recreate interactions that closely resemble the real biological processes (Derot et al., 2020).

Moreover, MAREL Carnot dataset might be useful for studying turbulence. Derot et al., (2015) studied the phytoplankton biomass during bloom events by applying Empirical mode decomposition (EMD) on fluorescence dataset from MAREL CARNOT. Results revealed that bloom events include considerable internal variations. Blooms are not smooth and "mountain-like", but exhibit high frequency oscillations due possibly to turbulent advection and complex population dynamics (Derot et al., 2015). Besides, Zongo & Schmitt, (2011) demonstrated that pH fluctuations in marine waters are strongly influenced by turbulent hydrodynamical transport, and may be considered as a turbulent active scalar.

Moreover, the sensors placed on the lighthouse provide valuable data for meteorological research and may improve local weather forecasts by measuring variables including wind speed, wind direction and air temperature. Also, the MAREL Carnot high frequency dataset can be used to validate satellite-derived products such as fluorescence. It also provides measurements for parameters that cannot be measured from space such as nutrient concentration (Lefebvre and Schmitt, 2016). Our dataset may assist fisheries research. For instance, Toomey et al., (2023) incorporated MAREL Carnot water temperature time series in the supplementary material of her study on the impact of temperature on Downs herring.

Overall, the MAREL Carnot station provides automatic, continuous, and long-term observation of various physical, chemical and biological parameters that enhance our knowledge about the environmental state of the coastal environment and bloom events. Hence, MAREL Carnot dataset aligns with objectives of SRN (Suivi Régional des Nutriments in French, Regional Nutrients Monitoring Program), especially by assessing the influence of continental inputs on the marine environment, and their implication on possible eutrophication which can assist in estimating the effectiveness of development and management policies in the marine coastal zone (Lefebvre and Devreker, 2023). To clarify, MAREL Carnot is the first coastal sampling station for the SRN transect. Thus, it assists in understanding phytoplankton dynamics by determining recurrent, extreme and rare events in this highly impacted and vulnerable coastal area.

Furthermore, MAREL Carnot dataset can be complementary~~In addition, it was shown that M-SC allows unsupervised labelling of time series, which is a basic part of machine learning and is needed to build an event prediction system and come up with sampling strategies that work close to real time (Grassi et al., 2019). As a result, scientists will be able to create a HAB early warning expert system to warn shellfish farmers, and prevent both public health risks and commercial losses in the shellfish farming business.~~

~~Nevertheless, datasets, notably MAREL Carnot, are typically incomplete and contain a significant amount of missing data due to sensor failures, communication/transmission difficulties, or poor weather conditions for manual measurements or maintenance. Phan et al., (2018) proposed Dynamic Time Warping method to fill in successive missing values of univariate time series as well as low uncorrelated multivariate time series (Phan et al., 2017).~~

~~Furthermore, to compare alternative approaches to studying and understanding HABs, all researchers must have access to the same datasets. For instance, researchers may now assess the effectiveness of new and old machine learning algorithms in~~

understanding the dynamics and forecasting harmful algal blooms. Thus, a comparative study of clustering approaches applied to spatial or temporal pattern discovery gave promising results in the segmentation of both UCI databases and marine time series compared to other approaches (Grassi et al., 2020). Therefore, we may conclude that the MAREL Carnot dataset is beneficial not just for marine ecologists, but also for machine learning specialists and data scientists. It is worth mentioning that all the above algorithms are available and published on the Comprehensive R Archive Network (CRAN).

Another study performed by Derot et al., (2020) analysed how forecasts of phytoplankton blooms are impacted by different sampling frequencies. They applied Random Forest (RF) and sliding window strategy on 12 parameters derived from MAREL Carnot dataset. This research demonstrated that the sampling frequency has a direct impact on the forecast performance of a Random Forest (RF) model as high-frequency datasets might provide useful information to the RF. Furthermore, this type of model sets the groundwork for the creation of a numerical decision-making tool that could help mitigate the impact of algal blooms, and can recreate interactions that closely resemble the real biological processes (Derot et al., 2020).

Moreover, MAREL Carnot dataset is useful for studying turbulence. In fact, many fields in the marine environment fluctuate over a wide variety of geographical and temporal scales. To study their dynamics and estimate their variations at all scales, high frequency measurements are needed (Huang and Schmitt, 2014). Hence, Derot et al., (2015) investigated phytoplankton biomass during bloom by applying Empirical mode decomposition (EMD) on fluorescence dataset from MAREL CARNOT. Results revealed that bloom events include considerable internal variations. In other words, blooms are not smooth and "mountain-like", but exhibit high frequency oscillations due to turbulent advection and complex population dynamics (Derot et al., 2015). Similarly, (Huang and Schmitt, 2014) analysed time dependent intrinsic correlation analysis of temperature and dissolved oxygen time series using empirical mode decomposition. The anti-correlation between temperature and oxygen showed that higher temperatures may favor larger phytoplankton growth rate, and hence, with a time delay, a lower percentage of oxygen (Huang and Schmitt, 2014). In addition, Zongo & Schmitt, (2011) demonstrated that pH fluctuations in marine waters are strongly influenced by turbulent hydrodynamical transport, and may be considered as a turbulent active scalar (Zongo and Schmitt, 2011).

Overall, the MAREL Carnot station provides automatic, continuous, and long-term observation of various physiochemical and biological parameters that allow for monitoring the general quality of marine environment, detecting harmful algal blooms (HAB) and understanding phytoplankton dynamics. Hence, MAREL Carnot dataset aligns with objectives of SRN (Suivi Régional des Nutriments in French, Regional Nutrients Monitoring Program), especially by assessing the influence of continental inputs on the marine environment, and their implication on possible eutrophication, and can assist in estimating the effectiveness of development and management policies in the marine coastal zone (Alain and David, 2022). In other words, MAREL Carnot is the first sampling station for the SRN transect. Thus, it assists in understanding phytoplankton dynamics by determining recurrent, extreme and rare events.

Furthermore, MAREL Carnot dataset can contribute to both REPHY (Observation and Surveillance Network for Phytoplankton and Hydrology in coastal waters) (https://doi.org/10.17882/47248), and REPHYTOX (Monitoring Network for Phycotoxins in marine organisms) (https://doi.org/10.17882/47251). TheActually, the goal of REPHY is to measure the biomass, abundance, and composition~~, and hydrological parameters~~ of marine phytoplankton as well as hydrological parameters in coastal and lagoon waters. REPHYTOX is designed to find and track three types of toxins that can build up in bivalve mollusks and cause DSP (Diarrheic Shellfish Poisoning), PSP (Paralytic Shellfish Poisoning), and ASP (Amnesic Shellfish Poisoning) (Belin et al., 2021)~~(Belin et al., 2021)~~. Monitoring carried out by MAREL Carnot in parallel with REPHY and REPHYTOX permits continuous adaptation to the objectives, developing analysis strategies with extensive and complex data, thereby ensuring sustainability, which were challenges faced by REPHY and REPHYTOX before.

~~Thus, the contribution of MAREL Carnot to improve assessment based on low frequency renders it important to achieve the objectives of WFD (Water Framework Directive) and MSFD (Marine Strategy Framework Directive). Besides, the lighthouse's sensors provide valuable data for meterological research and may improve local weather forecasts by measuring variables including wind speed, wind direction and air temperature. Additionally, depending on the goals, some of the parameters determined by our station can actually be useful for fisheries research when making the link between the different trophic levels.~~

While MAREL Carnot has made substantial progress toward automating marine ecosystem monitoring, there are still some significant challenges to overcome. Indeed~~In fact~~, it can be interrupted by rough sea conditions, such as strong tidal currents and storms. In addition, biofouling presents a major problem for sensors in the coastal environment, which explains why only a few moored autonomous systems have been deployed in the coastal environment (Blain et al. 2004). Due to sensor failure, phosphate, nitrate, and silicate measurements are not available after 2010. To better explain the large data gap, we should emphasize that we were in an interim phase, facing difficulties in maintaining a system developed and built in the early 2000s, with electronic parts that were no longer available and waiting for the improvement of the smart multisensor marine observation platform, costof2, which was driving all of the sensors and dataflow. ~~despite their ease of maintenance (Blain et al. 2004). Also, new EOV (Essential Ocean Variables) and EBV (Essential Biodiversity Variables) might be added with time. This adds a further obstacle, as it may be necessary to install brand-new sensors for the updated parameters. Above all, the major challenge will continue to be the issue of missing values, especially when it comes to data that has been missing for a long time, as in the case of nutrients, where nitrate, phosphate, and silicate observations have been missing since 2010 due to sensor failure.~~

As our knowledge and understanding of coastal ecosystems is growing with time, the EOV (Essential Ocean Variables) and EBV (Essential Biodiversity Variables) might be updated in the future. This may necessitate the installation of new sensors on MAREL Carnot station to measure these new variables or parameters.

In future work, we plan to use a multi-scale, multi-source, multi-criteria, and multi-parameter approach to characterize and predict harmful algal blooms in the Eastern English Channel caused by *Phaeocystis globosa* and *Pseudo-nitzschia spp.*. We will do this by combining high frequency datasets from MAREL Carnot, satellite, and modeling data with low frequency datasets from other sources. This integrated observing system will be used to identify environmental states present in the region, and develop an early warning system that can anticipate harmful algal blooms in particular, as well as changes in water quality and environmental state in general.

4.1 Data Availability

The raw data are present on the official Coriolis website. "These data were collected and made freely available by the Coriolis project and programmes that contribute to it (http://www.coriolis.eu.org)." The dataset after quality control procedures are present on the SEANOE website (DOI: 10.17882/39754) (Lefebvre A, 2023) in file "2004-2022 Coriolis processed data" . Our~~(Lefebvre, 2015). In fact, our~~ data are made available according to the FAIR approach (Findable, Accessible, Interoperable, Reusable).

5 Conclusion

In conclusion, ~~this~~ high frequency data from MAREL Carnot instrumented station are~~dataset is~~ useful in many scientific fields, such as phytoplankton ecology, data science, and oceanography~~turbulence~~. It can be used to describe the environmental state and forecast~~predict harmful~~ algal blooms in the Eastern English Channel, which is important to warn shellfish farmers and prevent economic losses and health problems. It can also be used with satellite, modeling, and low-frequency in situ data to enhance our understanding of the marine ecosystem. ~~achieve better knowledge and understanding of the marine ecosystem. Most importantly, this data set has been shown to be useful for fulfilling the goals set by the Water Framework Directive (WFD), the Marine Strategy Framework Directive (MSFD), and the Oslo and Paris Conventions (OSPAR). This would help researchers in the future get better results, which would lead to more scientific progress.~~

Competing Interests

The authors declare that they have no conflict of interest

Author Contribution

Raed Halawi Ghosn wrote the paper. Alain Lefebvre led the conceptualization, the writing of the paper and he also led the funding acquisition, the scientific coordination of MAREL Carnot related activities since 2002.We highly appreciate the effort of ~~each of~~ Émilie Poisson-Caillault, Guillaume Charria, Armel Bonnat, and Michel Repecaud for their contribution in data pre-processing. We would also like to sincerely thank ~~each of~~ Jean-Vallery FACQ, Loïc QUÉMÉNER, Vincent DUQUESNE, and Camille BLONDEL for all their efforts~~effort~~ in providing technical information, and maintaining the MAREL Carnot station~~in operational conditions~~.

Acknowledgement

This PhD project is funded by the Office Français de la Biodiversité (OFB) and by Ifremer (grant agreement OFB.21.0578).

This work has been financially supported (i) by the European Union (ERDF), the French State, the French Region Hauts-de-France and Ifremer, in the framework of the project CPER MARCO 2015-2021, (ii) by the European Union's Horizon 2020 research and innovation programme under grant agreement No 654410 in the framework of the project JERICO S3, and (iii) by the Artois Picardy Water Agency.

Bibliography

Aminot, A. and Kérouel, R.: Hydrologie des écosystèmes marins: paramètres et analyses, Ifremer Edition, France, 110–118 pp., 2004.

Bates, S. S., Hubbard, K. A., Lundholm, N., Montresor, M., and Leaw, C. P.: Pseudo-nitzschia, Nitzschia, and domoic acid: New research since 2011, Harmful Algae, 79, 3–43, https://doi.org/10.1016/j.hal.2018.06.001, 2018.

~~Alain, L. and David, D.: How to learn more about hydrological conditions and phytoplankton dynamics and diversity in the eastern English Channel and the southern bight of the North Sea ? the SRN data set (1992-2021)., 1–21, https://doi.org/https://doi.org/10.5194/essd-2022-146, 2022.~~

Belin, C., Soudant, D., and Amzil, Z.: Three decades of data on phytoplankton and phycotoxins on the French coast: Lessons from REPHY and REPHYTOX, Harmful Algae, 102, ~~101733,~~ https://doi.org/10.1016/J.HAL~~j.hal~~.2019.101733, 2021.

Blain, S., Guillou, J., Tréguer, P., Woerther, P., Delauney, L., Follenfant, E., Gontier, O., Hamon, M., Leildé, B., Masson, A., Tartu, C., and Vuillemin, R.: High frequency monitoring of the coastal marine environment using the MAREL buoy, J. Environ. Monit., 6, 569–575, https://doi.org/10.1039/b314073c, 2004.

Blauw, A. N., Los, F. J., Huisman, J., and Peperzak, L.: Nuisance foam events and Phaeocystis globosa blooms in Dutch coastal waters analyzed with fuzzy logic, J. Mar. Syst., 83, 115–126, https://doi.org/10.1016/j.jmarsys.2010.05.003, 2010.

Brylinski, J.-M., Lagadeuc, Y., Gentilhomme, V., Dupont, J.-P., Lafite, R., Dupeuble, P.-A., Huault, M.-F., Auger, Y., Puskarlc, E., Wartel, M., and Cabioch, L.: Le "fleuve côtier" : un phénomène hydrologique important en Manche orientale. Exemple du Pas de Calais, Oceanol. Acta, 11, 197–203, 1991.

Brylinski, J. M., Brunet, C., Bentley, D., Thoumelin, G., and Hilde, D.: Hydrography and phytoplankton biomass in the Eastern English Channel in spring 1992, Estuar. Coast. Shelf Sci., 43, 507–519, https://doi.org/10.1006/ecss.1996.0084, 1996.

Chai, F., Johnson, K. S., Claustre, H., Xing, X., Wang, Y., Boss, E., Riser, S., Fennel, K., Schofield, O., and Sutton, A.: Monitoring ocean biogeochemistry with autonomous platforms, Nat. Rev. Earth Environ., 1, 315–326, https://doi.org/10.1038/s43017-020-0053-y, 2020.

Delegrange, A., Lefebvre, A., Gohin, F., Courcot, L., and Vincent, D.: Pseudo-nitzschia sp. diversity and seasonality in the southern North Sea, domoic acid levels and associated phytoplankton communities, Estuar. Coast. Shelf Sci., 214, 194–206, https://doi.org/10.1016/j.ecss.2018.09.030, 2018.

Derot, J., Schmitt, F. G., Gentilhomme, V., and Zongo, S. B.: Long-term high frequency phytoplankton dynamics, recorded from a coastal water autonomous measurement system in the eastern English Channel, Cont. Shelf Res., 109, 210–221, https://doi.org/10.1016/J.CSR.2015.09.015, 2015.

Derot, J., Yajima, H., and Schmitt, F. G.: Benefits of machine learning and sampling frequency on phytoplankton bloom forecasts in coastal areas, Ecol. Inform., 60, 101174, https://doi.org/10.1016/j.ecoinf.2020.101174, 2020.

Dickey, T. D. and Bidigare, R. R.: Interdisciplinary oceanographic~~.: Emerging ocean~~ observations: The wave of the future, Sci.~~ for interdisciplinary data assimilation systems, J.~~ Mar., 69, 23–42~~. Syst., 40-41, 5-48~~, https://doi.org/10.3989/scimar.2005.69s123, 2005~~1016/S0924-7963(03)00011-3, 2003~~.

Grassi, K., Caillault, E. P., and Lefebvre, A.: Multilevel spectral clustering for extreme event characterization, Ocean. 2019 - Marseille, Ocean. Marseille 2019, 2019-June, https://doi.org/10.1109/OCEANSE.2019.8867261, 2019.

Hallegraeff, G. M., Anderson, D. M., Belin, C., Bottein, M. Y. D., Bresnan, E., Chinain, M., Enevoldsen, H., Iwataki, M., Karlson, B., McKenzie, C. H., Sunesen, I., Pitcher, G. C., Provoost, P., Richardson, A., Schweibold, L., Tester, P. A., Trainer, V. L., Yñiguez, A. T., and Zingone, A.: Perceived global increase in algal blooms is attributable to intensified monitoring and emerging bloom impacts, Commun. Earth Environ., 2, https://doi.org/10.1038/s43247-021-00178-8, 2021.

~~Grassi, K., Poisson-Caillault, É., Bigand, A., and Lefebvre, A.: Comparative study of clustering approaches applied to spatial or temporal pattern discovery, J. Mar. Sci. Eng., 8, https://doi.org/10.3390/JMSE8090713, 2020.~~

Huang, Y. and Schmitt, F. G.: Time dependent intrinsic correlation analysis of temperature and dissolved oxygen time series using empirical mode decomposition, J. Mar. Syst., 130, 90–100, https://doi.org/10.1016/J.JMARSYS.2013.06.007, 2014.

Jouanneau, N., Sentchev, A., and Dumas, F.: Numerical modelling of circulation and dispersion processes in Boulogne-sur-Mer harbour (Eastern English Channel): Sensitivity to physical forcing and harbour design, Ocean Dyn., 63, 1321–1340, https://doi.org/10.1007/s10236-013-0659-4, 2013.

Karasiewicz, S. and Lefebvre, A.: Environmental Impact on Harmful Species Pseudo-nitzschia spp. and Phaeocystis globosa Phenology and Niche, J. Mar. Sci. Eng., 10, https://doi.org/10.3390/jmse10020174, 2022.

Kazmi, S. S. U. H., Yapa, N., Karunarathna, S. C., and Suwannarach, N.: Perceived Intensification in Harmful Algal Blooms Is a Wave of Cumulative Threat to the Aquatic Ecosystems, Biology (Basel)., 11, 1–32, https://doi.org/10.3390/biology11060852, 2022.

Kbaier Ben Ismail, D., Lazure, P., and Puillat, I.: Statistical properties and time-frequency analysis of temperature, salinity and turbidity measured by the MAREL Carnot station in the coastal waters of Boulogne-sur-Mer (France), J. Mar. Syst., 162, 137–153, https://doi.org/10.1016/j.jmarsys.2016.03.010, 2016.

Lefebvre, A. High Frequency measurement of the coastal environment in the eastern English Channel. Data from MAREL CARNOT - COAST-HF (Coastal ocean observing system - High frequency) monitoring programme within the Research Infrastructure ILICO https:doi;org/10.17882/39754, 2023

Lefebvre, A. and Devreker, D.: How to learn more about hydrological conditions and phytoplankton dynamics and diversity in the eastern English Channel and the Southern Bight of the North Sea: the Suivi Régional des Nutriments data set (1992-2021), Earth Syst. Sci. Data, 15, 1077–1092, https://doi.org/10.5194/essd-15-1077-2023, 2023.

Lefebvre, A. and Schmitt, F. G.: In Mesures à haute résolution dans l'environnement marin côtier, CNRS ~~Editi., edited by:~~

~~CNRS Editions, CNRS~~ Editions, 162 pp., 2016.

Le Moal, M., Gascuel-Odoux, C., Ménesguen, A., Souchon, Y., Étrillard, C., Levain, A., Moatar, F., Pannard, A., Souchu, P., Lefebvre, A., and Pinay, G.: Eutrophication: A new wine in an old bottle?, Sci. Total Environ., 651, 1–11, https://doi.org/10.1016/j.scitotenv.2018.09.139, 2019.

Petroff, R., Hendrix, A., Shum, S., Grant, K. S., ~~Phan, T. T. H., Caillault, E. P.,~~ Lefebvre, K. A., and Burbacher, T. M.: Public health risks associated with chronic, low-level domoic acid exposure: A review of the evidence, Pharmacol. Ther., 227, 107865, https://doi.org/10.1016/j.pharmthera.2021.107865, 2021.

Phan, T., Caillault, É. P., Bigand, A., Lefebvre, A., and Bl, L. E. R.:~~.: Which~~ DTW-APPROACH FOR UNCORRELATED MULTIVARIATE TIME SERIES IMPUTATION Univ . Littoral Côte d ' Opale , EA 4491-LISIC , F-62228 Calais , France Vietnam National University of Agriculture , Department of Computer Science Hanoi , Vietnam, 2017 Ieee Int. Work. Mach. Learn. Signal Process., 2017a.

Phan, T. T. H., Poisson Caillault, É., Lefebvre, A., and Bigand, A.: Dynamic time warping-based imputation for~~ method applied to marine~~ univariate time series data, Pattern Recognit. Lett., 139, 139–147~~imputation, Ocean. 2017 - Aberdeen, 2017-Octob, 1–7~~, https://doi.org/10.1016/j.patrec~~1109/OCEANSE~~.2017.08.019, 2017b~~8084598, 2017~~.

Phan, T. T. H., Bigand, A., and Caillault, É. P.: A New Fuzzy Logic-Based Similarity Measure Applied to Large Gap Imputation for Uncorrelated Multivariate Time Series, Appl. Comput. Intell. Soft Comput., 2018, https://doi.org/10.1155/2018/9095683, 2018.

Riegman, R. and Van Boekel, W.: The ecophysiology of Phaeocystis globosa: A review, J. Sea Res., 35, 235–242, https://doi.org/10.1016/S1385-1101(96)90750-9, 1996.

Ross Brown, A., Lilley, M. K. S., Shutler, J., Widdicombe, C., Rooks, P., McEvoy, A., Torres, R., Artioli, Y., Rawle, G., Homyard, J., Tyler, C. R., and Lowe, C.: Harmful Algal Blooms and their impacts on shellfish mariculture follow regionally distinct patterns of water circulation in the western English Channel during the 2018 heatwave, Harmful Algae, 111, 102166, https://doi.org/10.1016/j.hal.2021.102166, 2022.

Rousseeuw, K., Poisson Caillault, E., Lefebvre, A., and Hamad, D.: Hybrid hidden markov model for marine environment monitoring, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 8, 204–213, https://doi.org/10.1109/JSTARS.2014.2341219, 2015.

Sazhin, A. F., Artigas, L. F., Nejstgaard, J. C., and Frischer, M. E.: The colonization of two Phaeocystis species (Prymnesiophyceae) by pennate diatoms and other protists: A significant contribution to colony biomass, Biogeochemistry, 83, 137–145, https://doi.org/10.1007/s10533-007-9086-2, 2007.

Spilmont, N., Denis, L., Artigas, L. F., Caloin, F., Courcot, L., Créach, A., Desroy, N., Gevaert, F., Hacquebart, P., Hubas, C., Janquin, M. A., Lemoine, Y., Luczak, C., Migné, A., Rauch, M., and Davoult, D.: Impact of the Phaeocystis globosa spring bloom on the intertidal benthic compartment in the eastern English Channel: A synthesis, Mar. Pollut. Bull., 58, 55–63, https://doi.org/10.1016/j.marpolbul.2008.09.007, 2009.

Toomey, L., Giraldo, C., Loots, C., Mahé, K., Marchal, P., and MacKenzie, K.: Impact of temperature on Downs herring

(Clupea harengus) embryonic stages: First insights from an experimental approach, PLoS One, 18, e0284125, https://doi.org/10.1371/journal.pone.0284125, 2023.

Wong, A., Keeley, R., Carval, T., and Argo Data Management Team: Argo Quality Control Manual For CTD and Trajectory Data, 1–58, 2022.

Young, N., Sharpe, R. A., Barciela, R., Nichols, G., Davidson, K., Berdalet, E., and Fleming, L. E.: Marine harmful algal blooms and human health: A systematic scoping review, Harmful Algae, 98, 101901, https://doi.org/10.1016/j.hal.2020.101901, 2020.

Zhu, Z., Meng, R., Smith, W. O., Doan-Nhu, H., Nguyen-Ngoc, L., and Jiang, X.: Bacterial Composition Associated With Giant Colonies of the Harmful Algal Species Phaeocystis globosa, Front. Microbiol., 12, https://doi.org/10.3389/fmicb.2021.737484, 2021.

Zongo, S. B. and Schmitt, F. G.: Scaling properties of pH fluctuations in coastal waters of the English Channel: pH as a turbulent active scalar, Nonlinear Process. Geophys., 18, 829–839, https://doi.org/10.5194/NPG-18-829-2011, 2011.