



ODYSSEA

Operating a network of integrated observatory
systems in the Mediterranean Sea

Project Deliverable Report

Deliverable Number: 7.4

Deliverable Title: Validated Algorithms

Author(s): Assaf Schuster, Noam Miron, Lorinc Meszaros, Claire LAUDY,
Georgios Sylaios

Work Package Number: 7

Work Package Title: Big Data Management and Tools Development



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727277

ODYSSEA Project Information	
Project full title	ODYSSEA: Operating a network of integrated observatory systems in the Mediterranean Sea
Project acronym	ODYSSEA
Grant agreement number	727277
Project coordinator	Georgios Sylaios, DUTH
Project start date and duration	1 st June 2017, 54 months
Project website	http://odysseaplatform.eu/

Deliverable Information	
Work package number	7
Work package title	Big Data Management and Tools Development
Deliverable number	7.4
Deliverable title	Validated Algorithms
Description	Iterative Progress on Developed Algorithms
Lead beneficiary	Technion
Lead Author(s)	Assaf Schuster
Contributor(s)	Noam Miron (Technion), Lorinc Meszaros (Deltares), Claire LAUDY (Thales), Georgios Sylaios (DUTH)
Revision number	1.0
Revision Date	25-10-2021
Status (Final (F), Draft (D), Revised Draft (RV))	F

Dissemination level (Public (PU), Restricted to other program participants (PP), Restricted to a group specified by the consortium (RE), Confidential for consortium members only (CO))	PU
--	----

Document History			
Revision	Date	Modification	Author
0.1	10-08-2021	Initial draft	Noam Miron
0.2	30-08-2021	Deltares contribution added to the section on 'Additional Iterations of the eutrophication algorithm adaptations'	Lorinc Meszaros
0.3	12-09-2021	Thales Contribution: added the chapter on Semantic and social network harvesting	Claire Laudy
0.4	30-09-2021	DUTH Contribution: added the coastal erosion and the wind resource assessment algorithms	Georgios Sylaios
0.6	10-10-2021	Technion contribution, document structure	Noam Miron
1.0	25-10-2021	Final review	Georgios Sylaios

Approvals				
	Name	Organisation	Date	Signature (initials)
Coordinator	Georgios Sylaios	DUTH	25-10-2021	GS
WP Leaders	Assaf Schuster	Technion	25-20-2021	AS

PROPRIETARY RIGHTS STATEMENT

This document contains information, which is proprietary to the ODYSSEA consortium. Neither this document, or the information contained within may be duplicated, used or communicated except with the prior written permission of the ODYSSEA coordinator.

Table of Contents

1	Executive Summary	9
2	Introduction	10
3	Long Term Wave Power Analysis	11
3.1	The Dataset.....	11
3.2	Data Exploration.....	12
3.2.1	All Years Graphs.....	12
3.2.2	Trend and seasonality	13
3.2.3	Monthly Breakdown.....	15
3.3	Data Analysis	15
3.3.1	Additive and multiplicative	16
3.3.2	Linear regression	16
3.3.2.1	Broken linear regression	17
3.3.3	Anomaly detection.....	18
3.3.4	Difference from the previous day	18
3.3.5	Fourier transform	21
3.3.5.1	Eliminate the annual frequency.....	21
3.3.5.2	Fast Fourier Transform on the filtered data	22
3.4	Conclusion	23
4	Wind Resource Assessment	24
4.1	Introduction	24
4.2	Algorithm Description for Wind Resource Assessment.....	25
4.3	Testing the Algorithm in Thracian Sea.....	26
4.3.1	Wind Scatterometer Data Retrieval and Description	26
4.3.2	Wind Scatterometer Data Quality Control	27
4.3.3	Descriptive Wind Statistics per Sub-Area	29
4.3.4	Spatial Variability in Weibull Fitting Function Parameters	31
4.3.5	Wind Energy Content and Power Density	34
4.3.6	Annual Wind Energy Production	36
5	Costal Erosion Assessment	38
5.1	Introduction	38
5.2	Algorithm Description for Coastal Erosion Assessment	38
5.3	Testing the Algorithm in ODYSSEA Observatories.....	39

5.3.1	The Thracian Sea Observatory	39
5.3.2	The Israeli Observatory	42
5.3.3	The Valencia Observatory	44
5.3.4	The Morocco Observatory	46
5.3.5	The Egyptian Observatory	46
6	Additional Iterations of Eutrophication Algorithm Adaptations.....	49
6.1	Eutrophication algorithm description	49
6.2	Eutrophication product requirements	51
6.2.1	Summary statistics	52
6.2.2	Depths	55
6.2.3	Alerts	56
6.2.4	Classification	57
6.2.5	Confidence	58
6.2.6	Other comments from the interviews	58
7	Semantic and Social Network Harvesting	60
7.1	New needs for Semantic information Analysis from social media.....	60
7.2	Proposed Solutions.....	60
7.2.1	Selection of suitable Information sources.....	61
7.3	Event Extraction and enrichment	61
7.3.1	Ontology definition and event modeling	62
7.3.2	Named Entity Recognition (NER) and Instantiation of the model.....	62
7.3.3	Wikidata based Enrichment	63
7.4	Experiments and results	63
7.4.1	Factual vs. non factual tweet filtering.....	64
7.4.2	Enrichment for Event extraction from texts.....	65
7.5	Concluding remarks for Semantic Information Analysis.....	66
8	References.....	67

Table of Figures

<i>Figure 3.1: The two points of which the data is consisted</i>	<i>12</i>
<i>Figure 3.2: Full time series for P1, P2 and mean of P1, P2.....</i>	<i>12</i>
<i>Figure 3.3: Wave power by year (all years side by side) P1, P2 and mean of P1, P2.</i>	<i>13</i>
<i>Figure 3.4: Year box plot for P1, P2 and mean of P1, P2.....</i>	<i>15</i>
<i>Figure 3.5: Linear regression, full timeseries P1, P2 and mean of P1, P2.</i>	<i>17</i>
<i>Figure 3.6: Broken linear regression over the whole period at different breakpoints and corresponding coefficients</i>	<i>18</i>
<i>Figure 3.7: Difference from prev. day, mean P1, P2.....</i>	<i>19</i>
<i>Figure 3.8: Lag scatter mean P1, P2.....</i>	<i>19</i>
<i>Figure 3.9: Days in the top decile, top percentage and top thousandth mean of P1, P2.....</i>	<i>20</i>
<i>Figure 3.10: FFT mean of P1, P2.....</i>	<i>21</i>
<i>Figure 3.11: FFT, IFFT, year average after eliminating year frequency, mean P1, P2.....</i>	<i>22</i>
<i>Figure 3.12: Linear regression rolling window average, eliminating year frequency, combined.....</i>	<i>22</i>
<i>Figure 4.1: Study area map and CMEMS grid discretization.....</i>	<i>27</i>
<i>Figure 4.2: Density plots histograms of CMEMS wind speed data against wind data from on-site stations in, (a) Chrisoupolis Airport, (b) Lemnos Airport. Dashed line represents the perfect match line, red line the linear regression model fitted on the scattered data and the light red area the 95% confidence interval.</i>	<i>28</i>
<i>Figure 4.3: 6-hourly time series (blue line) and mean-monthly time-series (red line) of wind speed at hub height in Lemnos Plateau (grid point 46).....</i>	<i>30</i>
<i>Figure 4.4: Boxplots for monthly wind speed values at hub height in Lemnos Plateau (point 46).</i>	<i>30</i>
<i>Figure 4.5: Wind frequency roses at hub height over the study area.</i>	<i>31</i>
<i>Figure 4.6: (a) Weibull probability density model, (b) Q-Q plot, (c) Cumulative density function, and (d) P-P plot, as fitted on the NE wind speed data at hub height of point 46, located at Lemnos Plateau.....</i>	<i>33</i>
<i>Figure 4.7: Spatial distribution of the Weibull probability density function parameters, (a) the shape parameter k, and (b) the scale parameter A (in m/s), at the hub level over the study area.....</i>	<i>34</i>
<i>Figure 4.8: Wind turbine power curve.</i>	<i>36</i>
<i>Figure 5.1: Coastal erosion/deposition rates during the 1985 – 2020 period, along the Thracian Sea shoreline.....</i>	<i>40</i>
<i>Figure 5.2: Coastal erosion/deposition rates during the 1985 – 2020 period, along the western Thracian Sea shoreline.....</i>	<i>41</i>
<i>Figure 5.3: Coastal erosion/deposition rates during the 1985 – 2020 period, along the eastern Thracian Sea shoreline.....</i>	<i>41</i>
<i>Figure 5.4: Coastal erosion/deposition rates during the 1985 – 2020 period, along the shoreline of Israel.</i>	<i>42</i>

Figure 5.5: Coastal erosion/deposition rates during the 1985 – 2020 period, along the shoreline of southern Israel..... 43

Figure 5.6: Coastal erosion/deposition rates during the 1985 – 2020 period, along the shoreline of northern Israel..... 43

Figure 5.7: Coastal erosion/deposition rates during the 1985 – 2020 period, along the shoreline of the Valencia Observatory. 44

Figure 5.8: Coastal erosion/deposition rates during the 1985 – 2020 period, along the northern shoreline of the Valencia Observatory. 45

Figure 5.9: Coastal erosion/deposition rates during the 1985 – 2020 period, along the southern shoreline of the Valencia Observatory. 45

Figure 5.10: Coastal erosion/deposition rates during the 1985 – 2020 period, along the shoreline of the Al Hoceima Observatory..... 46

Figure 5.11: Coastal erosion/deposition rates during the 1985 – 2020 period, along the shoreline of the Egyptian Observatory..... 47

Figure 5.12: Coastal erosion/deposition rates during the 1985 – 2020 period, along the western shoreline of the Egyptian Observatory. 47

Figure 5.13: Coastal erosion/deposition rates during the 1985 – 2020 period, along the central shoreline of the Egyptian Observatory..... 48

Figure 6.1: Eutrophication product input variables and derived indices..... 51

Figure 6.2: Presentation of summary statistics in the time series plot. 53

Figure 6.3: Example of geographical subregions identified by the General Fisheries Commission for the Mediterranean (GFCM). Source: https://www.medqsr.org/sites/default/files/inline-files/2017_MedQSR_Online_0.pdf..... 54

Figure 6.6: Example of simple layout, thresholds (Heavy, moderate, light), and notification “Elevated Trophic Index levels starting at xx, lasting xx”. Source: <https://play.google.com/store/apps/details?id=org.yoki.android.drops&hl=en>. 56

Figure 6.7: Trophic Index map legend (overwrite values by classes ‘Oligo-, Meso-, Eu-, Hypertrophic’).... 57

Figure 6.9: Mock-up of confidence (standard deviation) visualization in the time series data (left), and heat map visually showing where confidence reduces due to lack of observation data (Source: <https://leanpub.com/leaflet-tips-and-tricks/read#leanpub-auto-generate-a-heatmap-with-leafletheat>)..... 58

Figure 7.1: Ontology for jellyfish monitoring event extraction. 62

Figure 7.2: Event model graph 62

Figure 7.3: Analysis of polarity and subjectivity of tweets..... 64

Figure 7.4: Evaluation of factual / objective tweet. 64

Figure 7.5: Number of graphs per classes of entities number..... 65

Figure 7.6: Number of graphs per classes of relation number. 65

Figure 7.7: Number of locations linked to the event graphs. 66

Table of Tables

<i>Table 4.1: Descriptive wind statistics [in m/s] at hub height (93 m), per study sub-area.</i>	<i>29</i>
<i>Table 4.2: Weibull probability density function parameters, per directional bin, at hub height for all sub-areas.....</i>	<i>32</i>
<i>Table 4.3: Total annual wind energy content (in kWh/m²) at hub level, per directional bin, for all sub-areas.</i>	<i>35</i>
<i>Table 6.1: List of requested eutrophication algorithm features.....</i>	<i>52</i>
<i>Table 6.2: Further feedback on the eutrophication product from the interviewed experts.....</i>	<i>58</i>

1 Executive Summary

ODYSSEA intends to develop, operate and demonstrate an interoperable and cost-effective platform that fully integrates networks of observing and forecasting systems across the Mediterranean basin, addressing both the open sea and the coastal zone.

In the previous deliverables we have outlined the algorithmic development and summarized the different stages that each algorithm had attained. This deliverable provides a summary of the efforts in the different areas of interest to third parties, based on the user needs identified by ODYSSEA outreach and communication efforts.

As outlined in D7.3 the development process is broken into three parts:

- identification and specification
- development and integration
- validation and maintenance

with each algorithm running indefinitely through these stages. We have completed the first iteration for the first services and the work now consists of adding new algorithms to the process, as well as reiterating the stages for the completed algorithms, and adapt them as new user needs arise.

2 Introduction

In ODYSSEA the WP7 deals with the algorithmic processing of data derived from external data sources, produced by the downscaled ODYSSEA models and collected by the deployed ODYSSEA sensors. The previous reports outlined the input datasets and the beginning of the development and integration efforts. In this report we look toward the output, simultaneously validating that the algorithms, already developed and integrated, reacting as expected in the platform environment and making sure that the developed services answer the needs identified by the ODYSSEA outreach and communication efforts.

We will also examine in depth the data already made available on the platform and try new and more in-depth methods to manipulate the data to provide new kinds of information to users in various areas of interest.

The main chapters in this document are:

- Long term wave power analysis
- Offshore wind resource assessment
- Coastal erosion assessment
- Eutrophication
- Semantic and social network harvesting

Each chapter outlines the efforts made in that area to bring forth new manipulation of the available data in order to respond to a specific user's need, achieve new forms of information, add some value to the data itself and contribute to a deeper understanding of the Mediterranean Sea.

3 Long Term Wave Power Analysis

In the last few decades, a question being asked frequently is “are we in the midst of a catastrophic process going through the Earth's climate?” In this section we have examined whether a certain change in the nature of the sea wave power can be discerned.

Research shows that upper-ocean warming, a consequence of anthropogenic global warming, is changing the global wave climate, making waves stronger [1]. This identifies wave power as a potentially valuable climate change indicator. To that end, we have analyzed the time series of the wave power of the last 25 years, from 1994 to 2018, at two different points in the Mediterranean. The samples were collected with a 3-hourly temporal interval (0:00, 3:00, 6:00, ... , 21:00).

We looked for whether it was possible to detect an increase or decrease in the wave power, or whether over the years, the stability of wave power between consecutive days has changed. We also found that probably the number of annual turbulent systems has changed considerably.

3.1 The Dataset

The data was taken from the Copernicus site, GLOBAL_REANALYSIS_WAV_001_032 dataset. The first point (P1) was measured in the coordinates 34.64583°N; 18.75°E, and the second point (P2) was measured in the coordinates 34°N; 29.6°E. From this dataset we taken the VHM0 parameter (sea_surface_wave_significant_height) and the VTM10 parameter (sea_surface_wave_mean_period_from_variance_spectral_density_inverse_frequency_moment).

The formula we have used for calculating the wave power (in J/m): $P = 0.49 * VHM0^2 * VTM10$

The formula holds in deep waters where the units of 0.49 are $[\frac{kw}{m^3 * sec}]$. The variables VHM0[m] and VTM10[s] denote spectral significant wave height (Hs) and spectral moments (-1,0) wave period ($T_e = T_{-10} = \frac{m_{-1}}{m_0}$, where m_n is the $m^t h$ spectral moment), respectively.

Another interesting point of view is the mean of those two point samples, since each point itself can show different behavior.



FIGURE 3.1: THE TWO POINTS OF WHICH THE DATA IS CONSISTED

The data transfer from the Observatory server to main Marinomica platform has been facilitated by a set of utilities and tools, deployed at each model Observatory, ensuring the operation of this task with the minimum human involvement.

3.2 Data Exploration

To better understand the data, let's look at several general graphs of the data.

3.2.1 All Years Graphs

First, let's look at the full time series:

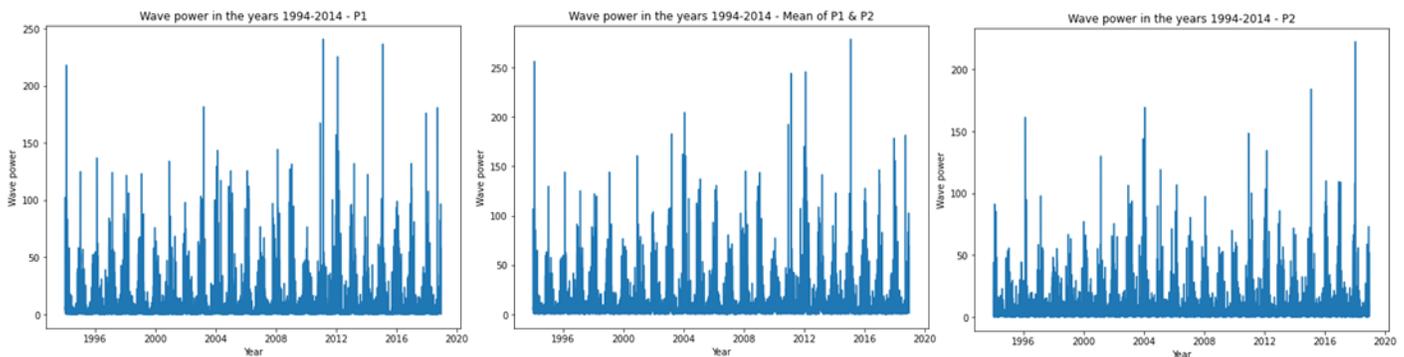


FIGURE 3.2: FULL TIME SERIES FOR P1, P2 AND MEAN OF P1, P2.

Since it is a monthly time series and follows a certain repetitive pattern every year, we can plot each year as a separate line in the same plot. it will let us compare the year-wise patterns side-by-side.

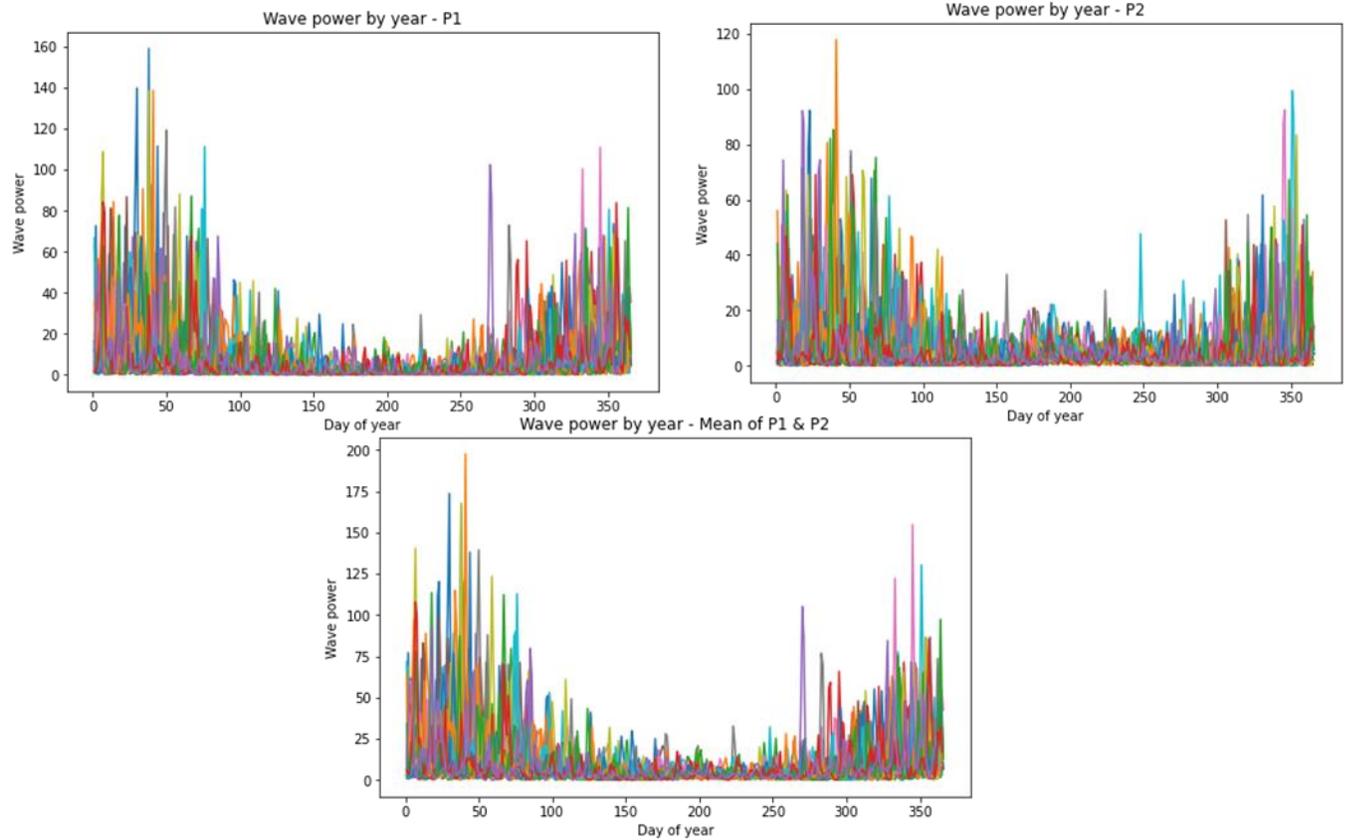


FIGURE 3.3: WAVE POWER BY YEAR (ALL YEARS SIDE BY SIDE) P1, P2 AND MEAN OF P1, P2.

We can see that, as we would expect, the wave power appears to be drastically stronger in the winter months, compared to the summer months.

3.2.2 Trend and seasonality

We will try to visualize the trend and how it varies each year using a nice year-wise boxplot. Likewise, we will do a month-wise boxplot to visualize the monthly distributions and variability. Also, we will plot some statistics by year and by month.

Boxplots are a standardized way of displaying the distribution of data based on a five-number summary (minimum, first quartile (Q1), median, third quartile (Q3) and maximum).

median Percentile)	(Q2/50th	the middle value of the dataset	the line that is in the center of the box
first quartile Percentile)	(Q1/25th	the middle number between the smallest number (not the "minimum") and the median of the dataset	the line at the bottom of the box
third quartile Percentile)	(Q3/75th	the middle value between the median and the highest value (not the "maximum") of the dataset	the line at the top of the box
interquartile range (IQR)		25th to the 75th percentile	all the range of the box
"maximum"		$Q3 + 1.5 * IQR$	the line that is above the box
"minimum"		$Q1 - 1.5 * IQR$	the line that is below the box
outliers		greater than "maximum" or lower than "minimum"	the dots

We will group the data at seasonal intervals and see how the values are distributed within a given year or month and how it compares over time.

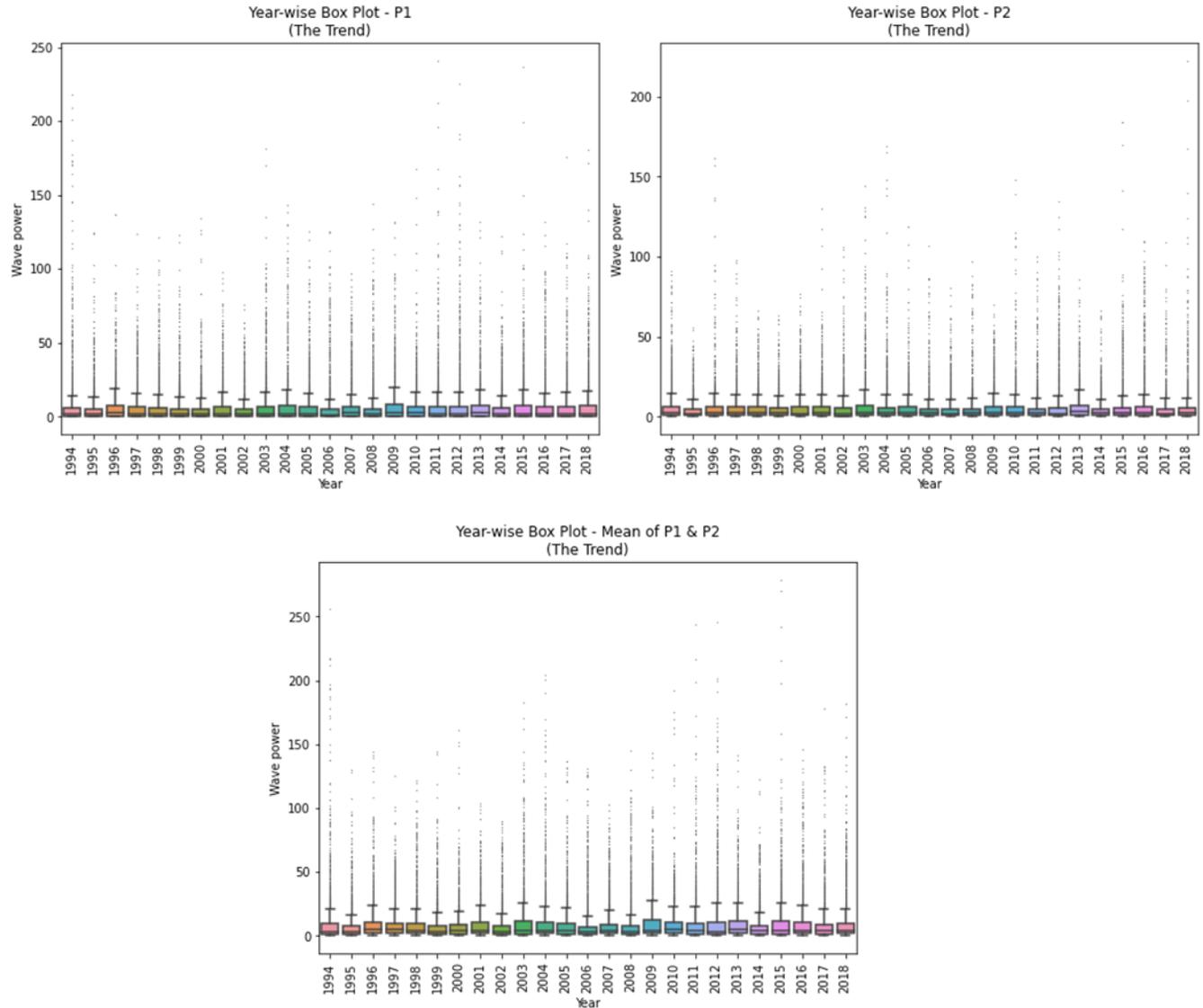


FIGURE 3.4: YEAR BOX PLOT FOR P1, P2 AND MEAN OF P1, P2.

In the course of exploring the preliminary data in a yearly manner, no visible trend was found. Therefore, we will explore further with more versatile tools in the next part.

3.2.3 Monthly Breakdown

As we saw before, those graphs show us the month-wise distribution evident. As we have seen earlier, in the winter months the wave power is significantly higher from the summer months with the peak in January.

3.3 Data Analysis

Any time series may be split into the following components: (Base Level +) Trend + Seasonality + Error



- **Trend:** how things are changing in the long-term
- **Seasonality:** how things change within a given period e.g. a year, month, week, day
- **Error:** activity not explained by the trend or the seasonal value

A trend is observed when there is an increasing or decreasing slope observed in the time series. Whereas seasonality is observed when there is a distinct repeated pattern observed between regular intervals due to seasonal factors. It could be because of the month of the year, the day of the month, weekdays, or even an hour of the day.

As discussed in the previous part, we can clearly observe the annual seasonality of the data. Our main challenge is to eliminate the seasonal effect of the data and see if we can detect an obvious trend and discover what is the relationship between the trend, seasonality, and the error (noise).

3.3.1 Additive and multiplicative

Depending on the nature of the trend and seasonality, a time series can be modeled as an additive or multiplicative, wherein, each observation in the series can be expressed as either a sum or a product of the components:

- Additive time series: $\text{Value} = \text{BaseLevel} + \text{Trend} + \text{Seasonality} + \text{Error}$
- Multiplicative Time Series: $\text{Value} = \text{BaseLevel} * \text{Trend} * \text{Seasonality} * \text{Error}$

In a multiplicative time-series, the components multiply together to compose time series. If you have an "increasing" trend, the amplitude of seasonal activity increases. Everything becomes more exaggerated.

In an additive time-series, the components add together to build the time series. If you have an "increasing" trend, you still see roughly the same size peaks and troughs throughout the time series. This is often seen in indexed time series, where the absolute value is growing but changes stay relative.

We can have a time series that is somewhere in-between the two, but because we are interested in attaining a quick classification, we won't be handling this complication here.

The trend component of a time series is identified using a moving average filter. Since there are 365 days in a year, we need to use a moving average filter of length 365×8 (samples per day). This means that at each point we want to average the 182 steps behind and 182 steps in front of the position we are calculating an average value for.

After estimating the trend by applying a convolution filter to the data, the trend is then removed from the series and the average of this de-trended series for each period is the returned seasonal component.

In both of the models, we can see the obvious seasonality, but the trend does not seem clear.

As the error has a more random distribution in the Multiplicative Decompose than in the Addictive Decompose, we can say that the wave power time series has a multiplicative nature.

3.3.2 Linear regression

We can do a linear regression over all the data, this will give us a broader perspective of the data.

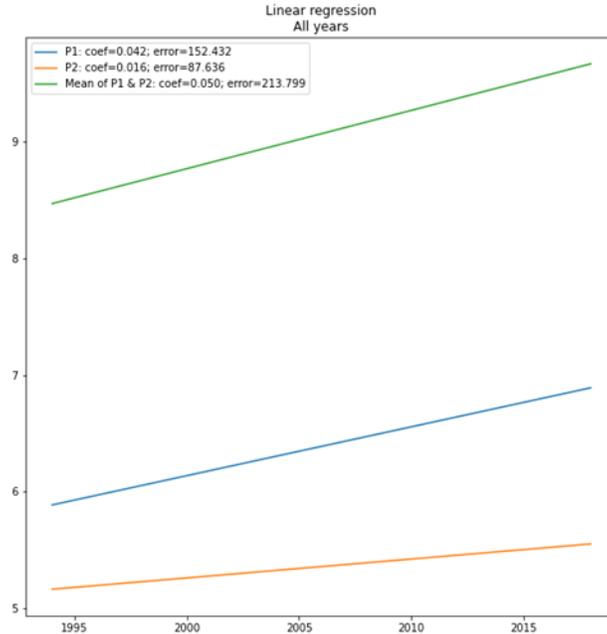


FIGURE 3.5: LINEAR REGRESSION, FULL TIMESERIES P1, P2 AND MEAN OF P1, P2.

We can see that if we try to get the most profitable linear function over all samples, we get an increasing line with a positive coefficient.

The "coef" here represents the coefficient of the linear line of wave power, and the "error" is the linear regression mean square error (MSE).

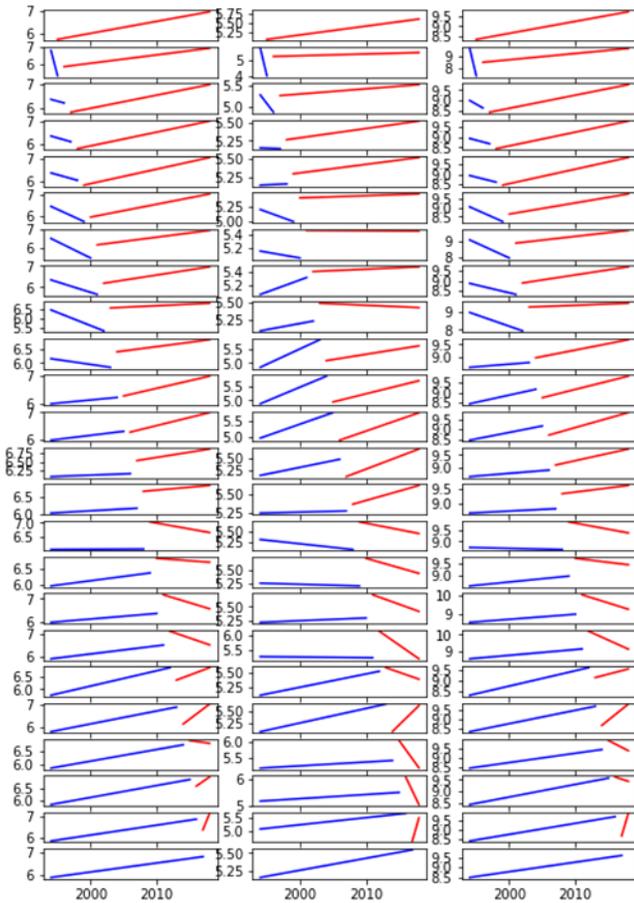
Here we can see that the coefficient is always positive, and for the mean between P1 and P2 we see the greatest coefficient, but also the greatest error.

Positive coefficient means that the linear regression found has an "increasing" trend.

3.3.2.1 Broken linear regression

In order to prevent a finer profit, now we allow breaking the linear function into two different functions. it will give us a better view in case the trend has changed over the years.

Linear regression by splitting to two time series



Linear regression coefficients

1994	0.000	0.052	0.000	0.023	0.000	0.063
1995	-1.435	0.047	-1.743	0.011	-2.307	0.053
1996	-0.077	0.057	-0.204	0.012	-0.179	0.063
1997	-0.085	0.061	-0.004	0.013	-0.087	0.068
1998	-0.083	0.065	0.004	0.012	-0.081	0.071
1999	-0.142	0.060	-0.042	0.004	-0.163	0.062
2000	-0.178	0.047	-0.017	-0.001	-0.187	0.046
2001	-0.098	0.048	0.037	0.005	-0.079	0.050
2002	-0.140	0.017	0.018	-0.004	-0.131	0.014
2003	-0.036	0.032	0.113	0.039	0.021	0.052
2004	0.022	0.054	0.102	0.061	0.073	0.084
2005	0.030	0.061	0.074	0.073	0.067	0.097
2006	0.007	0.032	0.030	0.055	0.022	0.060
2007	0.011	0.017	0.002	0.025	0.012	0.030
2008	0.001	-0.041	-0.016	-0.028	-0.007	-0.055
2009	0.029	-0.016	-0.003	-0.034	0.027	-0.033
2010	0.024	-0.087	0.005	-0.042	0.027	-0.109
2011	0.036	-0.102	-0.002	-0.159	0.035	-0.182
2012	0.062	0.100	0.022	-0.038	0.073	0.081
2013	0.060	0.231	0.031	0.136	0.075	0.299
2014	0.042	-0.033	0.012	-0.283	0.048	-0.175
2015	0.046	0.176	0.016	-0.537	0.054	-0.093
2016	0.045	0.736	0.028	0.898	0.058	1.185
2017	0.039	0.000	0.017	0.000	0.048	0.000
	P1 firsts	P1 lasts	P2 firsts	P2 lasts	Mean of P1 & P2 firsts	Mean of P1 & P2 lasts



FIGURE 3.6: BROKEN LINEAR REGRESSION OVER THE WHOLE PERIOD AT DIFFERENT BREAKPOINTS AND CORRESPONDING COEFFICIENTS

These graphs show us that when we split the data into two different time series, we get higher coefficients that illustrate a significant increase in the wave power over the years.

If we look at P2, when the "split" year is 2005, we can see the coefficients of the two linear functions is more than 0.07.

3.3.3 Anomaly detection

In this part, we will try to identify other trends of the wave power time series from increasing or decreasing.

3.3.4 Difference from the previous day

The first try will be to see if each day is different from the previous one. It will point to the abnormal behavior of the system.

Let's start with a simple graph that shows the difference over all years. it will not be an informative one, but only for a graphical feeling.

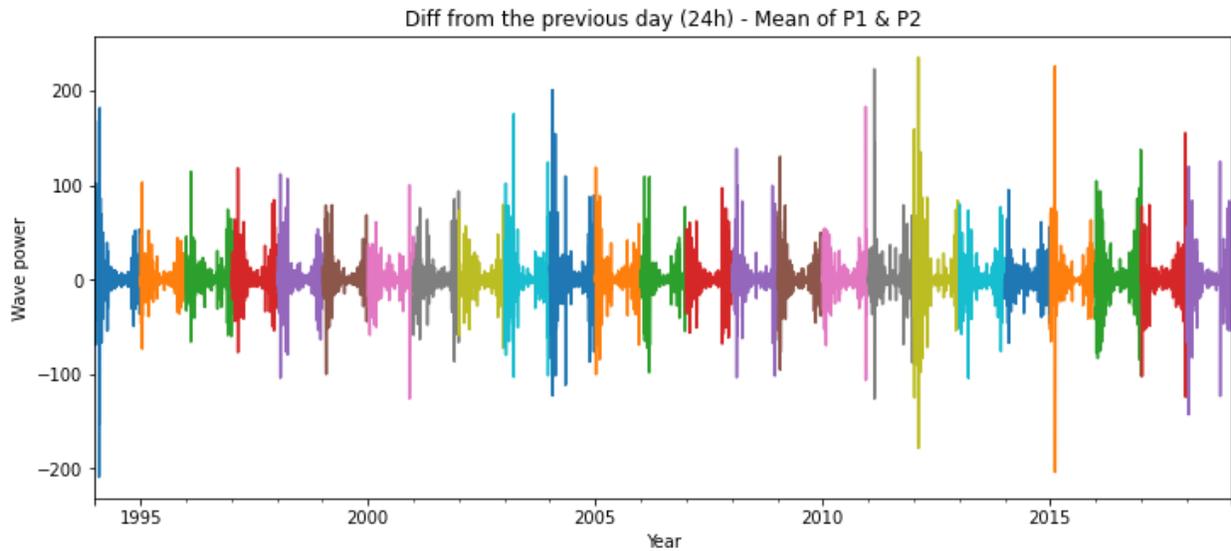


FIGURE 3.7: DIFFERENCE FROM PREV. DAY, MEAN P1, P2.

Another graph we can show is the Lag Scatter graph. This graph will indicate the relationship between each day. The X-axis is the value of the current day sample, and the Y-axis is the value that belongs to the following day.

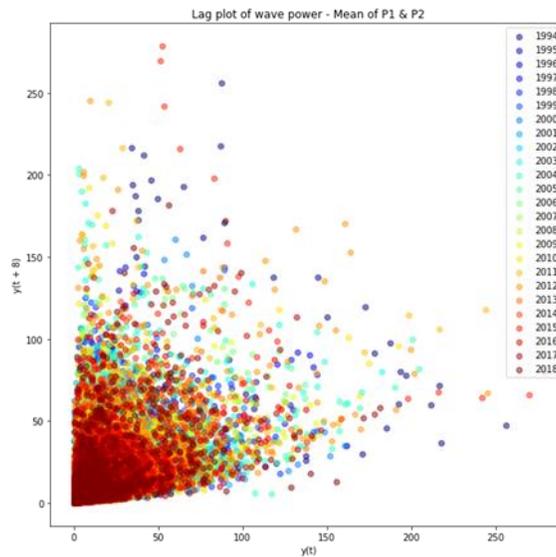
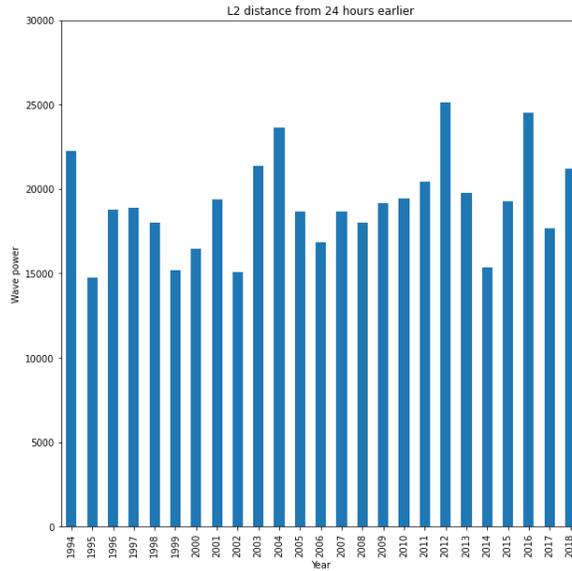


FIGURE 3.8: LAG SCATTER MEAN P1, P2.

To get a more informative graph we will take the average of the L2 distance from every 2 days in every year. This, will firstly will ignore the sign of the difference (positive/negative), and secondly, it will emphasize in the difference in relation to its size.



From this graph, we do not see any significant trend of abnormality over the years.

Another question we can ask is whether a trend can be seen in the number of days that their minimum sample belongs to the top decile of the wave power samples, for each year. Afterwards we can ask about the top percentage and the top thousandth.

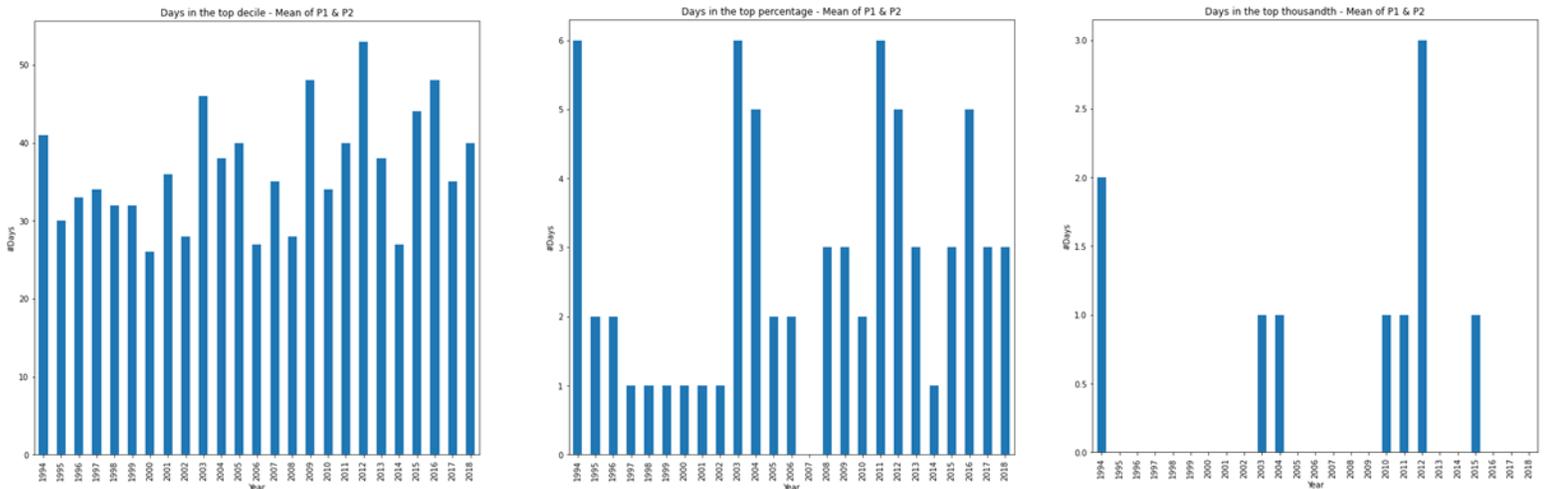


FIGURE 3.9: DAYS IN THE TOP DECILE, TOP PERCENTAGE AND TOP THOUSANDTH MEAN OF P1, P2

All over the datasets, we cannot see any significant trend, no matter if we look at the top decile, top percentage, or top thousandth.

3.3.5 Fourier transform

Another way to analyze the data will be by using the Fast Fourier Transform. The Fast Fourier Transform is a mathematical transform that decomposes functions depending on space or time into functions depending on spatial or temporal frequency, and in our case, it will convert the data point of view from time units to frequency units.

As we saw, the yearly seasonality of the data is very high and prevents us to identify the trend that maybe hide behind the seasonality or another seasonality of the data. We will try to identify the accurate seasonality and eliminate it, and maybe then we will see a significant trend or other seasonal trend.

First, we will simply apply the Fast Fourier Transform on the data and check the frequency graphs. We will use the FFT formula (Fast Fourier Transform).

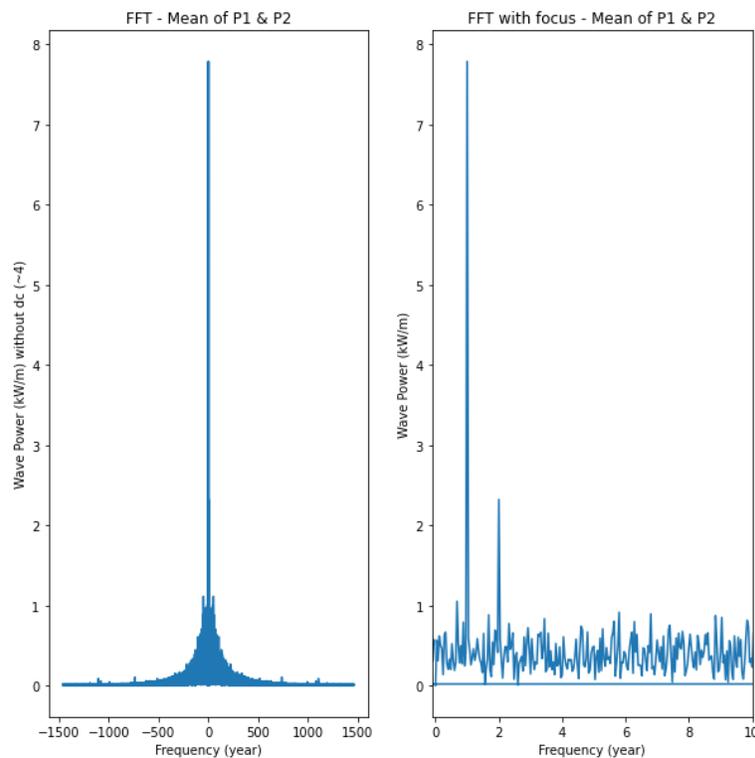


FIGURE 3.10: FFT MEAN OF P1, P2.

As we expected, we see the obvious peak at "1 year", and a smaller peak at the "2 years" (but is only a side effect of the annual frequency).

From here we cannot identify another peak that can represent significant seasonality.

3.3.5.1 Eliminate the annual frequency

Now, we will eliminate the annual frequency by reducing the high score of the result of FFT around the peak and then apply the inverse function (IFFT) to get again the data as a time series. Then, we will apply again the FFT to see if there are more frequencies we missed earlier. Also, we will calculate the sliding

window average, like we did when we started to look at the trend and seasonality of the data. Then we will compare it to the sliding window average of the original data.

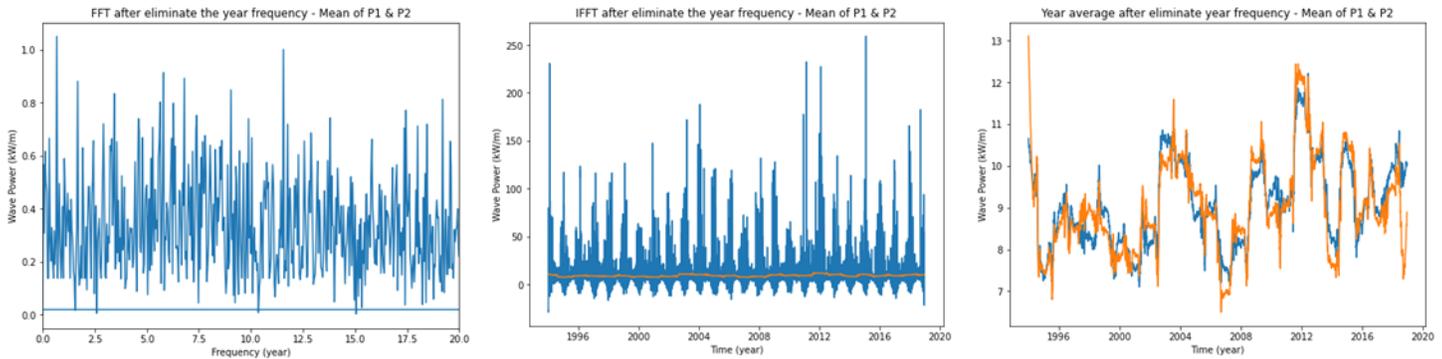


FIGURE 3.11: FFT, IFFT, YEAR AVERAGE AFTER ELIMINATING YEAR FREQUENCY, MEAN P1, P2.

The results for these actions show that the elimination of the annual frequency didn't affect the rolling window average of P2, compared to the effect on P1.

3.3.5.2 Fast Fourier Transform on the filtered data

In the last stage, we will try to apply the FFT on the data after it was filtered with the sliding window, with and without removing the annual frequency produced with the FFT+IFFT.

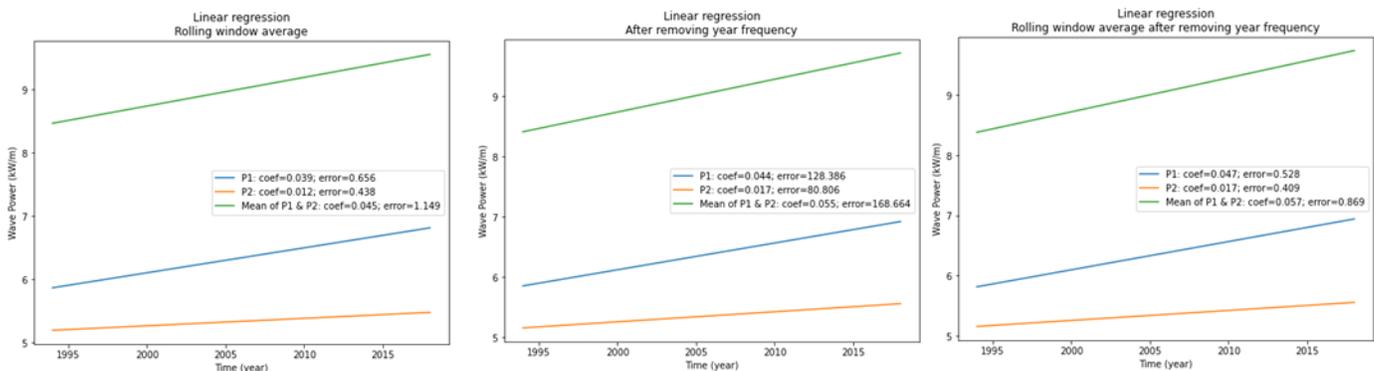


FIGURE 3.12: LINEAR REGRESSION ROLLING WINDOW AVERAGE, ELIMINATING YEAR FREQUENCY, COMBINED.

To sum it up, we saw 3 kinds of linear regression: firstly the one applied directly on the data, secondly that applied on the rolling window average, and thirdly that applied after removing the annual frequency with FFT+IFFT. Finally, there is the combination of removing the annual frequency and then applying a rolling window average.

The coefficients are:

Attempt	First	Two	Third	Fourth
P1	0.042	0.039	0.044	0.047
P2	0.016	0.012	0.017	0.017
Mean of P1&P2	0.050	0.045	0.055	0.057

It is clear that the elimination of the annual frequency exposes more than the slight increase of the wave power over the years, since the two last attempts show greater coefficients compared to the first two, and they are the attempts that include this elimination.

3.4 Conclusion

In this data analysis we tried to explore the presence of trend in the wave power time-series at two locations in the Mediterranean Sea. As explained before, it can indicate climate change and their side effect on sea levels and wave power. We saw the obvious presence of annual seasonality in the wave power and we attempted to eliminate it in order to exploit the trend. We explored the data by using the following techniques: sliding window average, linear regression, FFT + IFFT, and more.

In the first attempts, only the linear regression showed an increasing trend of the wave power over the years. Using the other tools, we didn't find any evidence of trend. Only in the last stage, we saw that if we eliminate the annual frequency from the data, we can expose the slight increasing trend that the linear regression shows better.

4 Wind Resource Assessment

4.1 Introduction

Through the European Green Deal, the European Union (EU) has set the target to reach total decarbonization and achieve energy efficiency for its members by the year 2050 (EU, 2019). To achieve this ambitious goal, the power production sector should follow the Clean Energy Transition pathway, with the renewable energy sources being at the epicenter of such conversion. In this gradually changing energy mix, the offshore wind industry is expected to play a significant role, experiencing a considerable increase in the coming decades (IRENA, 2021; Global Wind Energy Council, 2020). The EU plans to install in all European Sea at least 240 gigawatts (GW) of global offshore wind power capacity by 2050 (European Parliamentary Research Service, 2020). Current developments illustrate the exponential growth in offshore wind installations, i.e., from 1% of annual capacity additions in global wind installations in 2009, the offshore wind grew to over 10% in 2019 (Global Wind Energy Council, 2020).

Technological progress, recent developments in floating technologies and significant cost reductions, in conjunction with the local, low level and controllable environmental impacts, appear the main factors driving the offshore wind energy transformation into a safe and commercially viable form of clean power generation (Aslan, 2020). In any case, the total offshore installations reached 29.1 GW by the end of 2019, representing only 5% of total global wind capacity, while generating barely 0.3 % of global electricity production. In the EU, approximately 10 million households are now being served by offshore wind energy. In the U.S., the first commercial Offshore Wind Farm (OWF) started its operation in December 2016. However, until to date development activity remains impressively high, and sixteen active commercial leases for offshore wind development have been procured (American Clean Power Association, 2021). In Southeast Asia, countries like China, Japan and Taiwan lead the market, with China surpassing the 1 GW in annual offshore wind installation (Global Wind Energy Council, 2020).

The above indicate the enormous potential for offshore wind capacity growth. On this account, a large amount of new OWFs will be designed, installed and become operational, especially in Europe, since the European Commission (EC) forecasts that the total offshore wind installations will range between 240 and 450 GW by 2050 (European Parliamentary Research Service, 2020).

Although all OWFs are concentrated in the North and the Irish Seas, a clear tendency from the private sector to harvest the Mediterranean wind power potential is also seen. A 30 MW wind farm comprised of 10 monopole wind turbines is expected to be installed in the Apulia region, southern Italy, as the first Mediterranean offshore wind project to be implemented. Even though 1 GW of offshore wind power is equivalent to emissions of 3.5 MT CO₂ (Carbon dioxide), several technological, administrative, legislative, environmental, socio-economic, and financial barriers exist to develop OWF projects, summarized by Soukissian et al. (2017). The Geographic Information System (GIS) mapping of offshore marine and maritime uses could assist the selection of proper location and placement of the turbines (Saleous et al., 2016).

The most crucial suitability selection criterion for wind farm siting, the wind resource availability (Díaz et al., 2018), in conjunction with the presence of a wide continental shelf, ensuring relatively shallow depths and the appropriate distance from shore (Latinopoulos and Kechagia, 2016), maybe met over the Thracian Sea, in the Northern Aegean Sea. Several investigators have assessed the wind power potential in the broader area, especially in Çanakkale (Aslan, 2020) and Imvros (Satir et al., 2018), the Samothraki Island (Majidi Nezhad, 2021), and the whole Aegean Sea (Bagiorgas, 2012). Most studies utilize data from meteorological stations (Aslan, 2020; Satir et al., 2018). Bagiorgas et al. (2012) used wind data from

offshore buoys. Soukissian et al. (2017) downscaled the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis data, using a high-resolution meteorological model (15-years period, $0.10^\circ \times 0.10^\circ$) validated by offshore buoy data, while Majidi Nezhad et al. (2021) utilized the ERA-Interim reanalysis dataset (40 years period, mean monthly data).

In ODYSSEA we utilized the gridded 6-hourly wind data collected from ASCAT L2b scatterometer on-board of METOP satellites, combined with the ECMWF ERA-Interim atmospheric reanalysis, as provided by Copernicus Marine Environmental Monitoring Service (CMEMS), in order to assess the offshore wind power potential over the whole Thracian Sea and the Lemnos Plateau. This is an area of significant interest for wind offshore energy development, especially along the NNE-SSW axis following the wind exiting from the Dardanelles Straits (Zafeiratou et al., 2016; Konstantinidis et al., 2014).

Scatterometer data have been widely used in literature for large-scale wind resource assessments, filling the gap in the absence of offshore meteorological stations, while providing continuous, systematic, long-term, relatively-accurate wind data. However, data reliability suffers from low pixel resolution, together with errors related to sensor malfunctioning, wind retrieval algorithm, rain contamination, land contamination, etc. (Arun Kumar et al., 2019). Several global and regional wind resource assessment studies exist using scatterometer data, mostly using QuickSCAT (e.g., Pimenta et al., (2008) for offshore SE Brazil; Mostafaeipour et al., (2010) for Persian Gulf and Gulf of Oman; Karamanis et al. (2011) for Ionian Sea and Fuverik et al. (2011) for the whole Mediterranean Sea). To minimize errors induced by the above factors, recent studies explore offshore wind resources utilizing multiplatform datasets, like QuickSCAT, rapidSCAT, METOP-A and METOP-B, OCEANSAT-2 and others (Young et al., 2020).

4.2 Algorithm Description for Wind Resource Assessment

A series of algorithms and tools were developed in R and Python programming languages by DUTH. These algorithms perform the following tasks:

1. Provide the geographical boundaries and time period and retrieve the scatterometer data from the CMEMS platform.
2. Average the 6-hourly wind data per grid point into daily mean values.
3. Provide the WMO code for each meteorological station operating within the study area and the retrieve the daily-mean meteorological data from the in-situ station.
4. Merge the scatterometer data and the wind speed and direction data for the same time period into a single data frame.
5. Compare the scatterometer data to the in-situ wind speed and direction values; Make scatterplots; Compute the values of statistical parameters (Bias, RMSD, STD, correlation coefficient, slope).
6. Plot the time-series daily and mean-monthly wind speed scatterometer data per grid point.
7. Using the scatterometer data, establish the wind speed profile, i.e., compute the wind speed at the wind turbine hub height based on the 10 m wind speed.
8. Create a metmast object with all wind profile data per grid point.
9. Calculate the frequency and the mean wind speed per wind direction sector, using 16 sectors; store results in an Excel file named “i_processednew” (i is the grid point number); store the frequency table into the Excel worksheet named “frequency”.
10. Calculate the summary statistics per grid point (min, Q1, median, mean, Q3, max, skewness, kurtosis); store all statistical parameters into the Excel worksheet named “summary”.
11. Plot the frequency rose per grid point; store the frequency rose into the working directory.
12. Calculate the monthly-mean values of wind speed per grid point; store all results into the Excel worksheet “monthly”.

13. Plot the monthly frequency roses and store graphs into the working directory.
14. Calculate the modelled Weibull parameters being fitted on the frequency distribution of wind speed data per grid point; store the extracted Weibull parameters per direction into the Excel worksheet named "Weibull".
15. Calculate the total wind energy content (all values in kWh/m²/yr) per direction; store all energy values into the Excel worksheet named "Energy".
16. Calculate the modelled wind profile per wind direction and per grid point; store results into the Excel worksheet named "Profile".
17. Import the wind turbine power curve (e.g., Siemens_SWT-2.3MW-93m, Enercon_E126_7.5MW).
18. Based on the selected wind turbine, calculate the Annual Wind Energy Production (in MWh/yr) per directional bin; store results into the Excel worksheet named: "AnnualEnergyProduction".
19. Plot the AEP wind rose; store the rose into the working directory.
20. Estimate the wind energy capacity factor; store the results into the Excel worksheet named "CapacityFactor".
21. Calculate the total wind energy content per directional bin based on the Weibull data (all values in kWh/m²/yr); store the wind energy content results into the Excel worksheet named "EnergyTotal".
22. Transform daily values into monthly; apply a linear model on monthly time-series.
23. Subset the daily data to process wind speed within a certain directional window (e.g., from 45 to 75 degrees); apply a Weibull model on the frequency distribution and extract the model parameters.
24. Using the Weibull model parameters, estimate the cumulative probability of an event, e.g., the probability to obtain wind higher than 30 m/sec.

4.3 Testing the Algorithm in Thracian Sea

4.3.1 Wind Scatterometer Data Retrieval and Description

The 6-hourly data of wind speed (eastings and northings) measured 10 m above sea level with a spatial resolution of 0.25° × 0.25° were retrieved from the Copernicus Marine Environmental Monitoring Service (CMEMS). The data product used was encoded as WIND_GLO_WIND_L4_REP_OBSERVATIONS_012_006 (<http://marine.copernicus.eu/documents/PUM/CMEMS-WIND-PUM-012-006.pdf>, accessed on 26/4/2021), referring to a set of time-series comprised of level 4 reprocessed hindcasted wind observations, assimilated on a global ocean model. Data were estimated from the global wind fields derived from ASCAT scatterometers on-board METOP-A and METOP-B satellites, combined with the ECMWF ERA-Interim atmospheric reanalysis.

The dataset consists of six meteorological variables: the wind speed, the zonal and meridional wind components, wind stress amplitudes, and the associated components. The present analysis covered the period from January 2011 to December 2019. The resulting fields were estimated on a daily and monthly basis, as equivalent neutral-stability 10-m winds, having spatial resolutions of 0.25° in longitude and latitude over the study area (Figure 4.1).

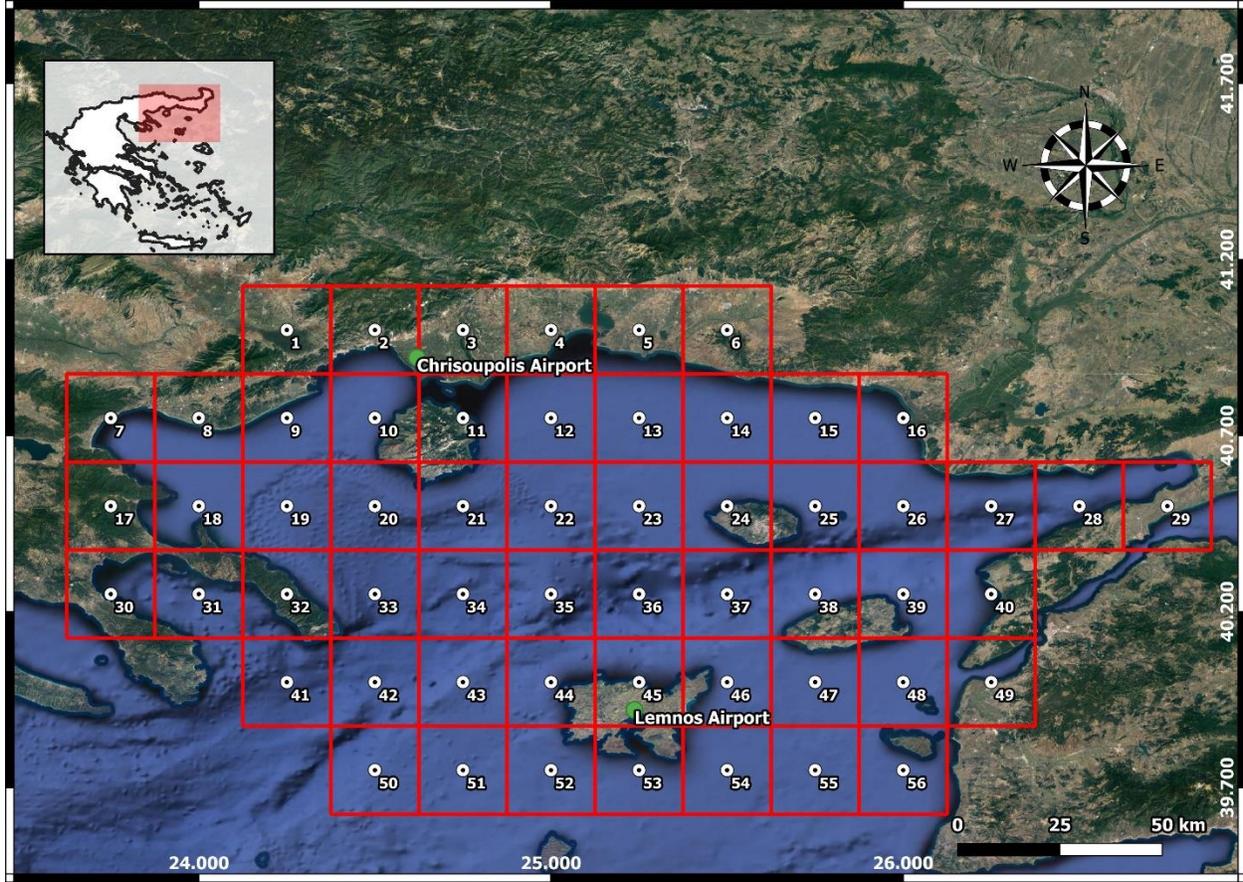


FIGURE 4.1: STUDY AREA MAP AND CMEMS GRID DISCRETIZATION.

4.3.2 Wind Scatterometer Data Quality Control

In total 56, grid points were analyzed, while in-situ daily mean wind data were retrieved for the above defined period from the World Meteorological Organization (WMO) stations located at Lemnos Airport and Chrisoupolis Airport (Hellenic Meteorological Service, Figure 4.1). These data were used to assess the consistency of the CMEMS remotely-sensed wind dataset in the study area.

A set of statistical parameters were used to test the quality of CMEMS scatterometer datasets. These include the difference between temporal means (defined as the bias) and the Root Mean Square Difference (RMSD) between the in-situ (considered as ground-truth) and the satellite data products, the scalar (r) and the regression coefficient slope (b_s). A similar analysis was also performed by Bentamy et al. (2021) between CMEMS and offshore wind data from buoys in California, Canary and Benguela zones. These statistical measures are estimated as:

$$Bias = \overline{X - Y} \quad (1)$$

$$RMSD = \sqrt{\overline{(X - Y)^2}} \quad (2)$$

$$STD = \sqrt{\overline{(X - Y - \overline{X - Y})^2}} \quad (3)$$

$$\rho = \frac{(\overline{X - \bar{X}}) - (\overline{Y - \bar{Y}})}{STD(X) - STD(Y)} \quad (4)$$

$$b_s = \sqrt{\frac{Y^2}{X^2}} \quad (5)$$

where X is the wind speed measured by the meteorological station and Y the CMEMS wind speed.

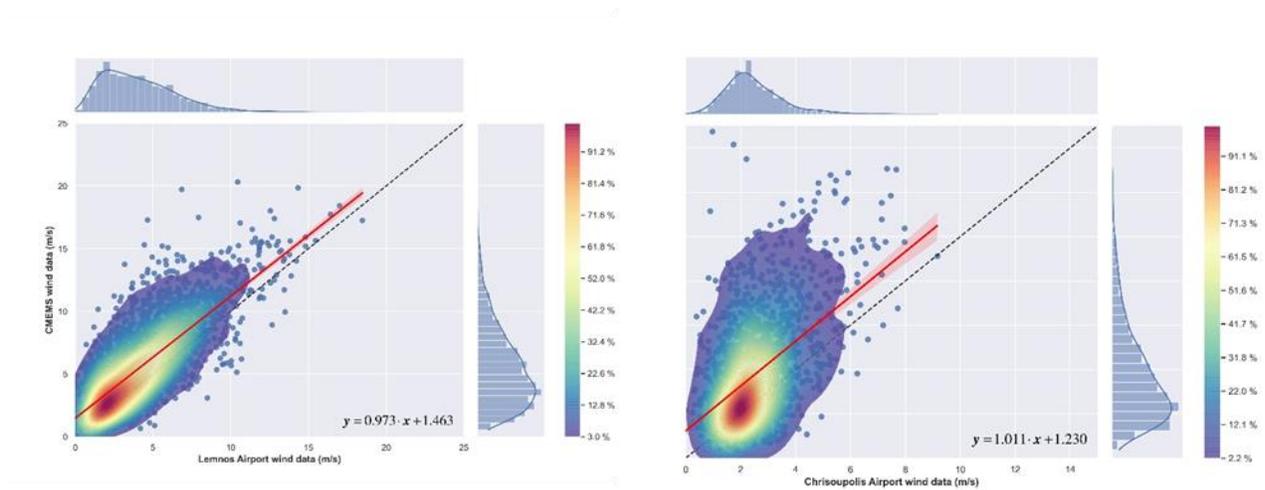


FIGURE 4.2: DENSITY PLOTS HISTOGRAMS OF CMEMS WIND SPEED DATA AGAINST WIND DATA FROM ON-SITE STATIONS IN, (A) CHRISOUPOLIS AIRPORT, (B) LEMNOS AIRPORT. DASHED LINE REPRESENTS THE PERFECT MATCH LINE, RED LINE THE LINEAR REGRESSION MODEL FITTED ON THE SCATTERED DATA AND THE LIGHT RED AREA THE 95% CONFIDENCE INTERVAL.

The intercomparison of the satellite-derived wind data products against “ground-truth” data collected from meteorological stations leads to the assessment of regional accuracy in the satellite wind analysis. Unfortunately, offshore buoy data were not available. Thus, comparisons were made against land-based stations of low altitude and in proximity to the shore, on a daily-mean basis. Figure 4.2(a, b) illustrate the scatter and fitted line plots between the 10-m wind speed retrieved from CMEMS (grid points 45 and 2) and the wind data collected from the Lemnos and Chrisoupolis meteorological stations, respectively.

These figures illustrate the rather good correlation with slight over-estimation of CMEMS wind speed data at the open Thracian Sea area (Lemnos: $n = 3,287$; bias: -1.35 ; RMSD = 2.43 ; STD = 2.02 ; $\rho = 0.76$; $b_s = 1.31$), and a moderate overestimation at the Thracian coastal zone (Chrisoupolis: $n = 1825$; bias: -1.25 ; RMSD = 2.33 ; STD = 1.97 ; $\rho = 0.50$; $b_s = 1.59$), in relation to the in-situ meteorological datasets. In Lemnos, agreement is higher at high wind speeds (15-20 m/s and > 20 m/s, bias: -1.03 ; RMSD = 1.76 ; STD = 1.37 ; $\rho = 0.78$; $b_s = 1.02$). Regression equations for both areas were defined, as:

$$\text{CMEMS scatterometer data} = 1.011 \times \text{Meteorological station data} + 1.230 \quad (6)$$

for Chrisoupolis airport, and

$$\text{CMEMS scatterometer data} = 0.973 \times \text{Meteorological station data} + 1.463 \quad (7)$$

for Lemnos airport.

Errors and biases are attributed to the coarse resolution of data product, exhibiting the tendency of satellite-derived ASCAT data to overestimate offshore winds (Carvalho et al., 2017). Similar findings were also reported by Alvarez et al. (2014), showing that similar satellite data, like QuikSCAT, CCMP and CFSR datasets overestimated the wind, especially at high wind speeds (>4 m/s).

4.3.3 Descriptive Wind Statistics per Sub-Area

To be able to analyze the wind data at hub level (93 m) and provide analytical descriptive statistics, data from grid points were spatially-aggregated according to the main physiographic units of the study area. Table 1 presents the summary values for these sub-areas. Results indicate that along the Thracian Sea continental shelf, a gradient in wind speed values exists, with higher mean, median and quartile values being exhibited towards the Eastern Thracian Sea. Further, the highest offshore wind statistical parameters are shown in Lemnos Plateau and the Dardanelles area; although the maximum wind speed is lower than that in the West Thracian Sea. In all areas data are positively skewed, highly skewed in the west and central Thracian Sea and in Mt Athos (skewness > +5), characterized by increased maximum speeds under extreme events. Leptokurtic curves prevail over the Thracian Sea and Mt Athos area (kurtosis ~ 1.3), and mesokurtic at the Lemnos Plateau and the Dardanelles.

TABLE 4.1: DESCRIPTIVE WIND STATISTICS [IN M/S] AT HUB HEIGHT (93 M), PER STUDY SUB-AREA.

Sub-Area	Q1	Median	Mean	Q3	Max
West Thracian Sea [7,8,9,17,18,19]	2.16	3.65	4.37	5.86	46.88
Central Thracian Sea [10-13,20-23]	2.45	4.14	4.90	6.62	35.19
East Thracian Sea [14,15,16,24-29]	3.17	5.32	6.05	8.25	29.66
Lemnos Plateau [43-46,52-54]	3.60	6.20	6.99	9.57	32.72
Dardanelles' zone of Influence [47,48,49,55,56]	3.82	6.48	7.15	9.81	29.30
Siggitikos Gulf & Mt Athos [30-33,41]	2.86	4.88	5.82	7.89	31.73

An indicative time-series diagram illustrating the 6-hourly wind speed variability in Lemnos Plateau (grid point 46), at the hub height, is shown in Figure 4.3. Winds under extreme stormy conditions exceed the limit of 20 m/s, originating mainly from the Dardanelles, affecting the northern part of the Aegean Sea. Data exhibit seasonality showing higher winter values, with regular incidents exceeding 20 m/s. Mean monthly values indicate that the seasonal component oscillates with an amplitude of 6 m/s and reveals a slightly upward trend (~0.008 m/s), over the years examined.

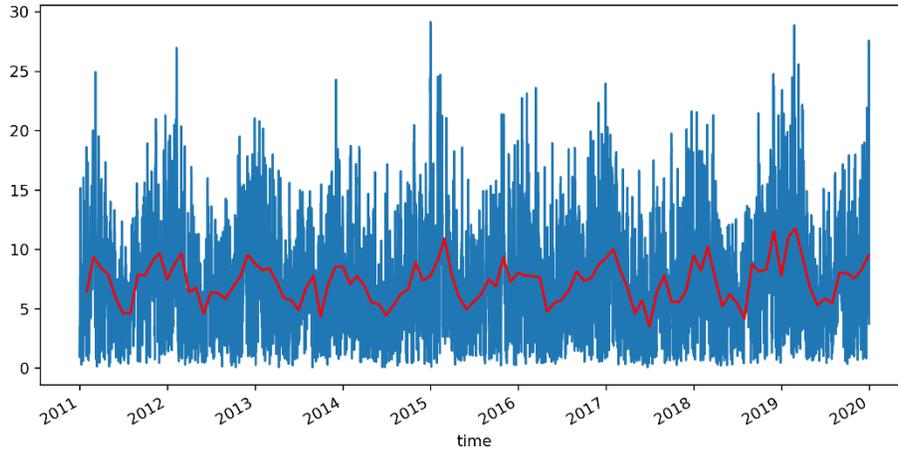


FIGURE 4.3: 6-HOURLY TIME SERIES (BLUE LINE) AND MEAN-MONTHLY TIME-SERIES (RED LINE) OF WIND SPEED AT HUB HEIGHT IN LEMNOS PLATEAU (GRID POINT 46).

The wind speed exhibits intra-annual variability with higher values in winter, especially in February, and significantly lower in spring and summer (April to July). A representative boxplot diagram of monthly-mean wind speed values at the hub level (93 m) at point 46 (Lemnos Plateau) is shown in Figure 4.4.

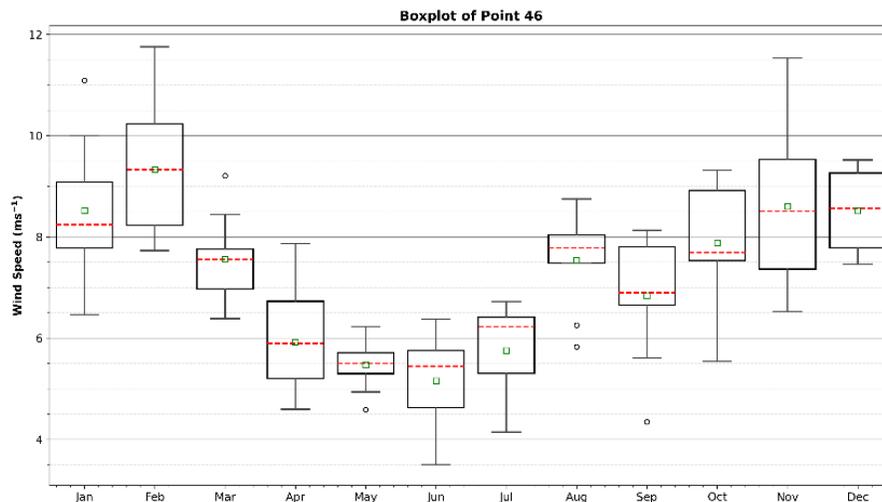


FIGURE 4.4: BOXPLOTS FOR MONTHLY WIND SPEED VALUES AT HUB HEIGHT IN LEMNOS PLATEAU (POINT 46).

The spatial variability of frequency distributions in daily-mean wind speeds, per wind directional sector, are shown in Figure 4.5. It is apparent that NE winds prevail in the study area, followed by ENE at the nearshore parts of the Thracian Sea and Mt Athos, and by NNE winds at the offshore Thracian Sea, Lemnos Plateau and Dardanelles. Wind speeds and frequencies per directional bin are more dispersed in the West and Central Thracian Sea and Mt Athos area, with mean wind speeds of 5.6 m/s, 6.0 m/s and 7.5 m/s (~30%, 36% and 35%) from NE and ENE directions, respectively.

Eastwards and offshore, wind speeds are significantly higher, of higher frequency and appear confined along the NE direction, as in point 46 (Lemnos Plateau), having a mean NE wind speed of 9.5 m/s and 33%

frequency of occurrence. This is attributed to the impact of orographic effects on the cyclonic synoptic circulation of surface wind field over the Black Sea and the funneling effect along the Turkish Straits. In parallel, these offshore points illustrate the influence of moderately strong S winds (~7.5 m/s, 8%).

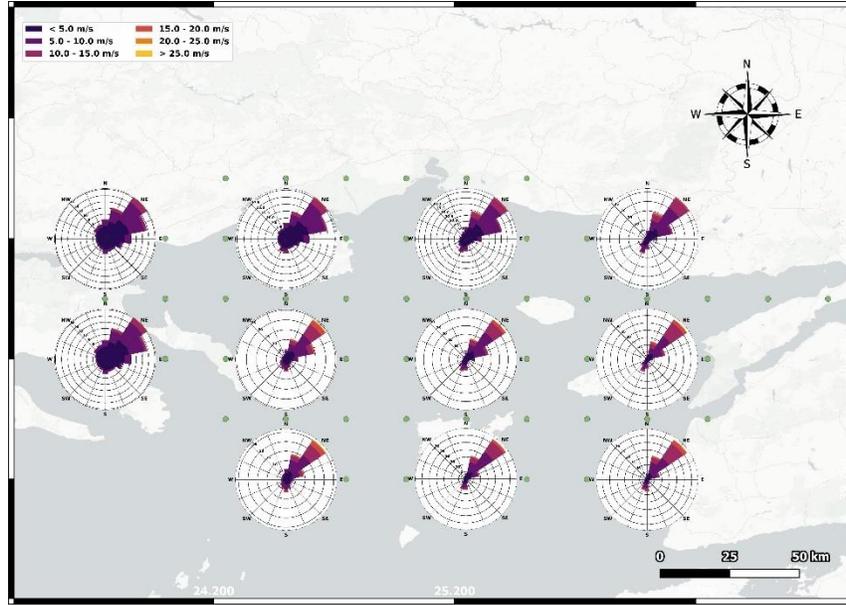


FIGURE 4.5: WIND FREQUENCY ROSES AT HUB HEIGHT OVER THE STUDY AREA.

4.3.4 Spatial Variability in Weibull Fitting Function Parameters

To reach a clear view of the available wind potential of an area, we may not rely only on the description of the instantaneous and mean wind speeds. The statistical parameters k and A of the fitted Weibull probability density function will provide a better and comparable to other areas understanding of wind dynamics.

Several probability density functions are available in the literature to be fitted on the distributions representing the wind speed frequency curve per directional sector for the prediction of randomly distributed wind speed data. The Weibull probability density function depicts an acceptable accuracy level in numerous wind power studies worldwide, expressed mathematically as:

$$f(W_b) = \frac{k}{A} \left(\frac{W_b}{A} \right)^{k-1} e^{-\left(\frac{W_b}{A} \right)^k} \quad (8)$$

Where $f(W)$ is the frequency of occurrence of wind speed W , A is the scale parameter (measure for the wind speed) and k is the shape parameter (description of the shape of the distribution) per directional bin. The Weibull distribution parameters were estimated by:

$$k = \left[\frac{\sum_{b=1}^n W_b^k \ln(W_b) f(W_b)}{\sum_{b=1}^n W_b^k f(W_b)} - \frac{\sum_{b=1}^n \ln(W_b) f(W_b)}{f(W_b \geq 0)} \right]^{-1} \quad (9)$$

$$A = \left(\frac{1}{f(W_b \geq 0)} \sum_{b=1}^n W_b^k f(W_b) \right)^{1/k} \quad (10)$$

where W_b is the mean wind speed per directional bin b , n is the number of bins, $f(W_b)$ the frequency for wind speed ranging within bin b and $f(W_b) \geq 0$ is the probability for wind speed equal to or exceeding zero. To estimate the Weibull distribution parameters k and A , the analysis was performed in R programming language (fitdistrplus package), using the maximum likelihood estimation method per directional bin.

The probability of occurrence, and therefore the fraction of time for each wind speed range, per directional sector, prevailing in the study area, maybe derived through this function. Table 2 presents the annual variation in Weibull parameters per directional bin, for all study area sub-regions. For all bins, the Weibull shape parameter k varies between 1.40 in the West Thracian Sea and 1.73 in the Dardanelles region of influence, with a mean value of 1.61, throughout the gridded data, at hub level ($z = 93$ m). At the nearshore Thracian Sea area, k -mean values range from 1.39 from the N direction to 1.63 from the WSW direction.

In terms of k -distribution over the various directional bins, higher values occur at the NE direction in the East Thracian Sea, Lemnos Plateau, Dardanelles and Mt Athos areas (ranging from 1.88 to 2.45), at the ENE direction in the central Thracian Sea ($k = 2.00$) and the E direction in western Thracian part ($k = 1.88$). In parallel, the Weibull scale parameter (A) exhibits a gradual increase from the western nearshore zone (4.79 m/s) towards the east (6.77 m/s) and then offshore, until Lemnos Plateau (7.81 m/s) and the highly dynamic Dardanelles area (8.02 m/s). The NE direction displays the higher A -values in all sub-areas, except for the East Thracian Sea, where the NNE direction prevails. The highest NE-bin A -value is seen at the Lemnos area (10.42 m/s), followed by the Dardanelles region (10.39 m/s).

TABLE 4.2: WEIBULL PROBABILITY DENSITY FUNCTION PARAMETERS, PER DIRECTIONAL BIN, AT HUB HEIGHT FOR ALL SUB-AREAS.

Sub-Areas/ Directional Bins	West Thracian Sea		Central Thracian Sea		East Thracian Sea		Lemnos Plateau		Dardanelles		Siggitikos/Mt Athos	
	k	A (m/s)	k	A (m/s)	K	A (m/s)	k	A (m/s)	k	A (m/s)	k	A (m/s)
N	1.38	4.71	1.45	4.67	1.35	5.08	1.42	5.03	1.47	5.05	1.46	5.41
NNE	1.50	5.87	1.61	7.08	1.79	8.79	1.71	8.80	1.82	9.16	1.56	7.49
NE	1.75	6.84	1.93	7.50	2.27	8.75	2.23	10.42	2.45	10.39	1.88	9.23
ENE	1.74	5.70	2.00	6.00	2.17	6.57	1.87	7.61	2.02	7.25	1.77	7.58
E	1.88	4.45	1.94	4.47	1.98	4.38	1.65	4.99	1.81	4.96	1.71	5.28
ESE	1.21	3.42	1.65	3.49	1.76	3.54	1.57	4.36	1.70	4.48	1.50	4.29
SE	1.50	3.32	1.28	3.24	1.60	3.81	1.46	4.75	1.55	4.56	1.51	4.23
SSE	1.48	3.62	1.38	3.79	1.51	4.38	1.57	5.89	1.65	5.76	1.46	4.81
S	1.27	4.53	1.35	5.40	1.56	6.70	1.67	8.13	1.75	8.42	1.40	6.13
SSW	1.44	4.02	1.53	5.26	1.68	6.74	1.73	7.43	1.79	7.68	1.51	5.64
SW	1.31	3.35	1.59	3.99	1.89	5.01	1.84	5.69	1.82	5.86	1.61	4.79
WSW	1.61	3.63	1.76	3.74	1.60	4.14	1.60	4.95	1.52	4.57	1.62	4.75
W	1.77	3.24	1.68	3.07	1.66	3.36	1.46	3.84	1.40	3.37	1.69	4.31
WNW	1.51	3.42	1.60	3.20	1.52	3.17	1.43	3.55	1.38	3.41	1.59	4.31
NW	1.45	3.87	1.40	3.33	1.34	3.31	1.31	3.95	1.19	3.41	1.40	4.68
NNW	1.50	4.09	1.29	3.29	1.30	3.50	1.56	3.93	1.59	3.54	1.48	4.81
all	1.40	4.79	1.48	5.42	1.64	6.77	1.63	7.81	1.73	8.02	1.47	6.44

The Weibull probability density function, fitted on the NE wind speed data, at a specific grid point located at Lemnos Plateau, together with the cumulative probability density function and the relevant q-q and p-p plots are shown in Figure 4.6. Based on this analysis and the wind turbine power curve it can be deduced

that the probability of wind speed from the NE direction within the turbine operational (> 5 m/s) window is 79.81%.

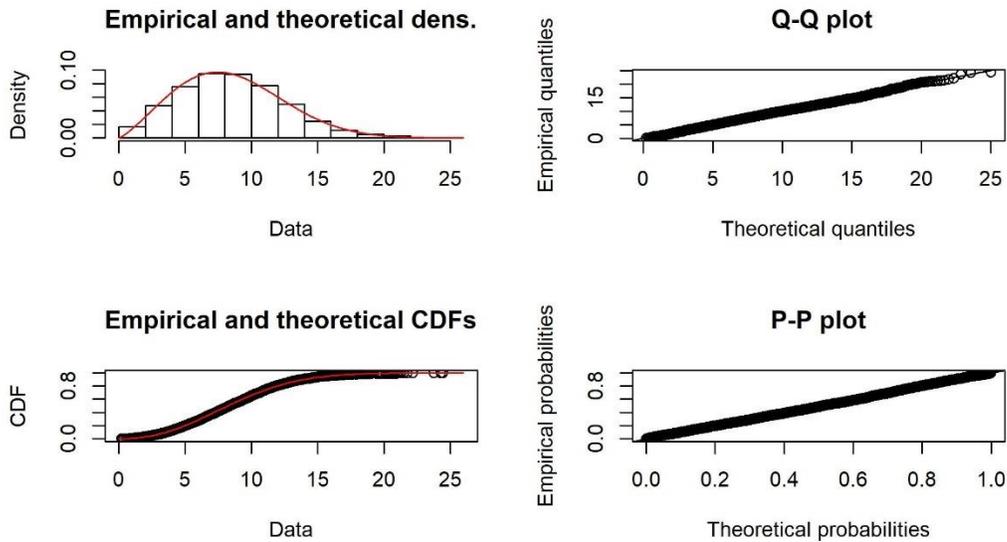
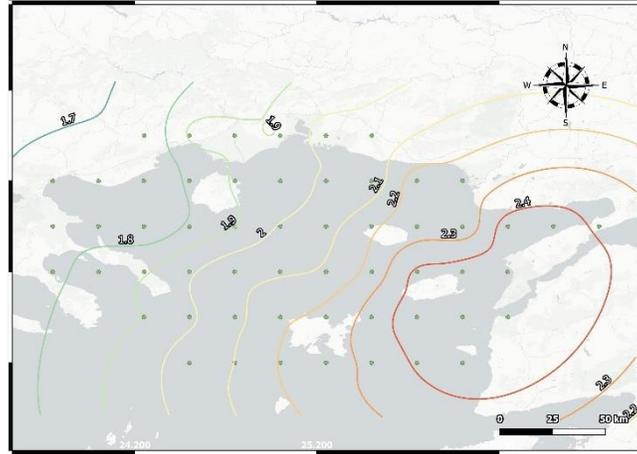
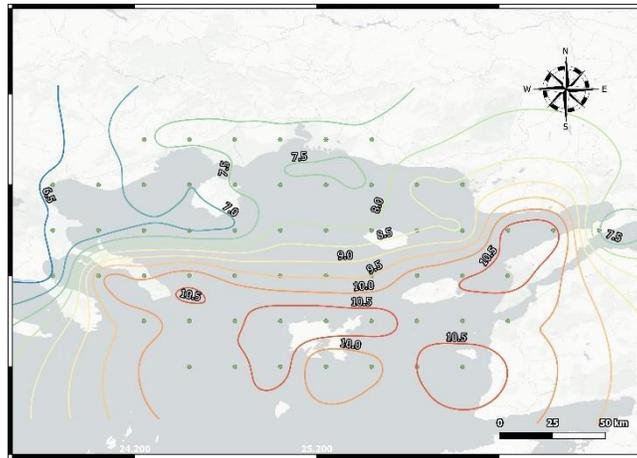


FIGURE 4.6: (A) WEIBULL PROBABILITY DENSITY MODEL, (B) Q-Q PLOT, (C) CUMULATIVE DENSITY FUNCTION, AND (D) P-P PLOT, AS FITTED ON THE NE WIND SPEED DATA AT HUB HEIGHT OF POINT 46, LOCATED AT LEMNOS PLATEAU.

The iso-lines connecting points of equal k and A values, as extracted from the Weibull probability distribution, for the NE wind direction is shown in Figure 4.7. Based on Figure 4.7(a) it is evident that k -values > 2.4 occur in the Dardanelles area, and that k reduces gradually towards the WNW direction, with a stable rate of 0.1 per 20km. On the other hand, the spatial distribution of the scale parameter A seems more complex, with local peaks (> 10.5 m/s) at Bozcaada Island and at Saros Gulf, and a general W-E isolines orientation, indicating a sharp reduction in A towards the nearshore and on-shore Thracian Sea grid points (Figure 4.7(b)).



(a)



(b)

FIGURE 4.7: SPATIAL DISTRIBUTION OF THE WEIBULL PROBABILITY DENSITY FUNCTION PARAMETERS, (A) THE SHAPE PARAMETER K, AND (B) THE SCALE PARAMETER A (IN M/S), AT THE HUB LEVEL OVER THE STUDY AREA.

4.3.5 Wind Energy Content and Power Density

Using the estimated Weibull probability density function, the total wind energy content per directional bin was computed. The total wind energy content (in kWh/m²/yr) can be perceived as the theoretic energy potential of a particular site. Therefore, it is a useful metric for the resource assessment of an area and for comparative purposes among areas, being independent of the characteristics of the wind turbine. The available wind energy content per wind speed increment and wind direction at each gridded point of the Thracian Sea was assessed using the R-package bReeze, by:

$$E(W) = \frac{1}{2} \rho_{air} H \sum_{b=1}^n W_b^3 f(W_b) \quad (11)$$

Where ρ_{air} is the density of air at the sea level under a mean temperature of 15 °C and one atmospheric pressure (= 1.225 kg/m³), n is the total number of directional bins (=16), H is the number of hours of the

desired period (=8,760 per year), W_b is the wind speed per directional bin and $f(W_b)$ is the probability of that bin, estimated by the Weibull distribution described in Eq. [6].

The wind power density is an important factor when assessing the wind potential of a location. It designates the available amount of energy per unit of time and swept area of the blades at the selected location. It is this amount of energy that will be converted to electricity by the wind turbine. The estimation of wind power density per directional bin is achieved by fitting the Weibull distribution to the respective dataset, expressed mathematically as:

$$P(W) = \frac{1}{2} \rho_{air} \sum_{b=1}^n W_b^3 f(W_b) \quad (12)$$

Using the parameters of the Weibull distribution per grid point and integrating spatially, Table 3 presents the wind energy content per directional sector, averaged over the main sub-areas of the study region. The analysis suggests that the highest wind energy content occurs in the Lemnos Plateau area (4,455 kWh/m²/yr), followed by the Dardanelles (4,398 kWh/m²/yr), Sigitikos/Mt Athos (3,091 kWh/m²/yr) and East Thracian Sea (2,964 kWh/m²/yr).

The Table indicates that Lemnos Plateau and the Dardanelles region have a high wind energy content spread over three directional bins (NNE, NE and ENE), representing an annual wind energy content of 3,496 kWh/m² and 3,431 kWh/m², respectively. This energy content is equivalent to the power density of 399 W/m² and 391 W/m², respectively. Approximately 22% of this sectorial energy content is being produced by winds in the 0-5 m/s range, 43% within the 5-10 m/s, 26% in the 10-15 m/s range, 7% in the 15-20 m/s and only 2% by winds higher than 20 m/s. The contribution of the S sector in the total wind energy content of these two areas also seems quite considerable.

TABLE 4.3: TOTAL ANNUAL WIND ENERGY CONTENT (IN KWH/M2) AT HUB LEVEL, PER DIRECTIONAL BIN, FOR ALL SUB-AREAS.

Sub-Areas/ Directional Bins	West Thracian Sea	Central Thracian Sea	East Thracian Sea	Lemnos Plateau	Dardanelles	Sigitikos/Mt Athos
N	82	62	80	73	59	91
NNE	214	389	775	887	973	436
NE	435	664	1,256	2,256	2,206	1,308
ENE	222	238	262	353	252	556
E	60	47	30	42	32	96
ESE	60	15	9	19	14	45
SE	18	30	12	27	21	33
SSE	23	23	25	72	58	62
S	69	112	183	358	395	168
SSW	27	87	227	235	272	89
SW	42	26	52	56	61	40
WSW	14	14	21	29	22	34
W	9	8	6	13	7	21
WNW	16	9	5	10	7	24
NW	29	14	8	17	11	40
NNW	36	37	13	15	9	47
all	1,354	1,774	2,964	4,455	4,398	3,091

4.3.6 Annual Wind Energy Production

The estimation of the annual wind energy production follows:

$$AEP = A_{turb} \frac{\rho}{\rho_{pc}} H \sum_{b=1}^n f(W_b) P(W_b) \quad (13)$$

Where A_{turb} is the average availability of the turbine, ρ_{air} is the density of air ($= 1.225 \text{ kg/m}^3$), ρ_{pc} is the specific air density for power curve definition, $f(W_b)$ is the probability of the wind speed bin W_b , estimated by the Weibull distribution and $P(W_b)$ is the power output for that wind speed bin. Finally, H is the number of operational hours ($= 8,760$ hours).

The Capacity Factor (CF) represents the productive suitability of the wind turbine, i.e., an indicator to assess the field performance of the turbine. It is defined as the ratio between the average output power (Pout) of the wind turbine represented by the AEP, to the theoretical maximum power output on annual basis, and is defined as:

$$CF = \frac{AEP}{P_{th} H} \quad (13)$$

P_{th} is the wind turbine's theoretical power, defined as being proportional to the wind speed cubed for wind speeds lower than the rated wind speed and equal to the turbine rated power for higher wind speeds. In this work, the annual energy production and the capacity factor were assessed based on the Siemens SWT 2.3 MW wind turbine of 93 m height. This turbine was selected as a potential monopile system to be deployed at an offshore wind farm in NE Lemnos. The power curve for this turbine, consisting of wind speed and power pairs, starting at the cut-in wind speed of the turbine and ending with the cut-out wind speed, is shown in Figure 4.8.

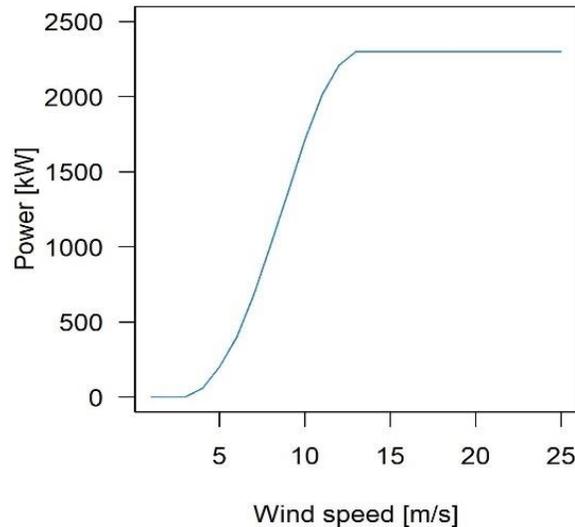


FIGURE 4.8: WIND TURBINE POWER CURVE.

Considering the wind profile produced from $z_1 = 10\text{m}$ (CMEMS data) to hub height ($z_2 = 93\text{m}$), the wind turbine power curve and dimensions, the annual wind energy production (in MWh/yr) was estimated, following equation (13). The highest wind energy may be produced in the Dardanelles region, with a spatially-averaged AEP-value of 7,546 MWh/yr. Approximately 75% of this energy (5,684 MWh/yr) is concentrated along the NNE, NE and ENE directional sectors, with the NE-AEP being the highest of all (48.8%). In parallel, most of the total AEP in Dardanelles is being produced from winds in the range of 10-15 m/s (46% or 3,532 MWh/yr) and 5-10 m/s (37% or 2,859 MWh/yr).

In Lemnos Plateau, the spatially-averaged estimated AEP-value reaches 7,212 MWh/yr, mostly provided by the same directional bins (NNE, NE and ENE) producing in total 5,342 MWh/yr (i.e., 74% of total AEP). As previously, most of the energy is produced by winds in the range of 10-15 m/s and 5-10 m/s, with values of 3,261 MWh/yr and 2,701 MWh/yr, respectively. East Thracian Sea is another area of significant interest, as the spatially-mean AEP approximates 5,620 MWh/yr, 75% of which is produced from the NNE, NE and ENE sectors. Another interesting feature is the rising contribution of the S and SSW directions (5.7% and 7.9%, respectively). The Siggitikos Gulf and the area of Mt Athos exhibit AEP of the order of 5,241 MWh/yr, while the Central Thracian and the West Thracian Sea have values of 3,743 MWh/yr and 2,939 MWh/yr, respectively.

Based on the above AEP-estimates, the capacity factor of turbine performance in the Dardanelles reaches 37.44%, in Lemnos Plateau is 35.80%, in East Thracian Sea is 27.89% and in Mt Athos area 26.02%. The capacity factor in the western and central Thracian Sea was assessed at 14.58% and 18.58%, respectively.

5 Coastal Erosion Assessment

5.1 Introduction

The coastal zone is a very dynamic geomorphologic system where changes occur at diverse temporal and spatial scales, mostly related to erosion action, resulting from natural and/or anthropogenic activities. For coastal zone monitoring, satellite-borne shoreline extraction and detection of change rates at various times are indispensable. These changes in shoreline occur over a wide range of time scales from geological to short-lived extreme events. Changes are basically associated with waves, tides, winds, storms, sea level change and human activities affecting the geomorphic processes of erosion and accretion.

Coastal erosion is directly linked to economic losses due to coastal retreat and loss of land, ecological damage (especially of valuable coastal habitats) and societal problems. Especially, as coastlines and beaches represent valuable buffer zones protecting backshore from marine flooding while shaping the socio-economic environment. Globally, coastal tourism and seashore recreational activities contribute by 5% to Global GDP and by 6-7% to global employment (Hall et al., 2013). Reliable assessments identifying erosion-prone areas, quantifying over-decadal shoreline change rates and prioritizing potential solutions in beaches formulate the basic elements of effective spatial planning, sustainable coastal development and coastal engineering projects (Luijendijk et al., 2018). The robust detection of erosion ‘hot-spots’ and the estimation of shoreline change rates requires long-term information on shoreline position, thus, the satellite imagery, accompanied by automated image processing techniques, seems the only available and low-cost option (Garcva-Rubio et al., 2015).

In ODYSSEA project, DUTH developed a series of algorithms to retrieve and process Landsat satellite images for the period 1985 to 2020, in order to assess the coastline changes along the sandy beaches of the various Observatories. The work focused on the historic shoreline change rates at five ODYSSEA Observatories: a) the Thracian Sea coastline in Northern Greece, b) the Israel coastline, c) the Alexandria – Port Said shoreline in Egypt, d) the Gulf of Valencia, Spain, and e) the Al-Hoceima shoreline, Morocco. Coastal erosion was examined and assessed utilizing satellite-borne shoreline change rates during the latest thirty-five years (1985 - 2020). The methodology employed in this study entailed the shoreline delineation, using a semi-automatic image classification technique (SCP plugin for QGIS). All shorelines were extracted by processing Landsat and Sentinel 2 satellite imageries retrieved from the Earth Explorer and the Sentinel Hub, respectively. In order to evaluate the shoreline evolution, an analysis was performed by the Digital Shoreline Analysis System software (DSAS) provided by the United States Geological Survey (USGS). The DSAS procedure was set using transects every 500 m. The erosion and accretion results produced by DSAS were presented through shape files (.shp) to the Marinomica platform.

5.2 Algorithm Description for Coastal Erosion Assessment

A series of algorithms and tools were developed in Python programming language by DUTH. These algorithms perform the following tasks:

1. The geographical boundaries of the Region of Interest should be provided by the user in the “input” folder of the working directory, as a polygon in .kml or .shp format.
2. The technical specifications of the analysis should be provided by the user in the “Settings.txt” file stored in the working directory. For example, the satellite sensor, the dates of the satellite images, the transect distance, the transect length, the uncertainty, the sand color etc.

3. The algorithm selects and downloads the satellite images based on the specifications defined in the file "Settings.txt". The suitable satellite images are retrieved from the Google Earth Engine.
4. Optionally, the algorithm allows additional pre-processing of the images (e.g., cloud masking, pansharpener/down-sampling).
5. A pop-up window opens to manually create a reference shoreline (helps to identify outliers and false detections). This step applies only once.
6. The algorithm estimates the NDWI index combining the bands green and NIR and applying the equation: $NDWI = \frac{GREEN - NIR}{GREEN + NIR}$ in each pixel of the satellite image. The image is saved in the output folder of the working directory as "{DATE}_NDWI.tif".
7. A machine learning algorithm is being trained manually with a series of sub-samples indicating the reflectances in land and water. This step is applied only in the first classification process.
8. The image is classified with the minimum distance algorithm and saved in the output folder of the working directory as "{DATE}_class.tif".
9. The classified image is vectorized (raster to vector process) to a polygon and then a polyline is created using the vectorizing tool. The polyline is saved in the output folder of the working directory as "{DATE}_shoreline.shp".
10. The same process follows for every historical satellite image and then the historical shorelines are generated. The files are saved in the output folder of the working directory.
11. A new reference line (baseline) is created offshore and parallel to the most recent coastline.
12. Vertical transects to the baseline are generated, intersecting the historical shorelines.
13. The intersect points of the vertical transects are used to estimate the statistical parameters to describe the shoreline change through the years (such as EPR, WLR, SCE etc).
14. The statistical parameters are saved as .csv file and the shorelines as .shp files in the output folder of the working directory.
15. An external algorithm is used for the shoreline validation, using in-situ data or high-resolution satellite images.

5.3 Testing the Algorithm in ODYSSEA Observatories

5.3.1 The Thracian Sea Observatory

The coastal erosion assessment algorithm was run for the continental shoreline of the Thracian Sea. This is a shoreline of 200 km consisting the northern borderline of the Aegean Sea. Overall, 3,404 transects perpendicular to the shoreline were determined distanced to each other by 500 m.

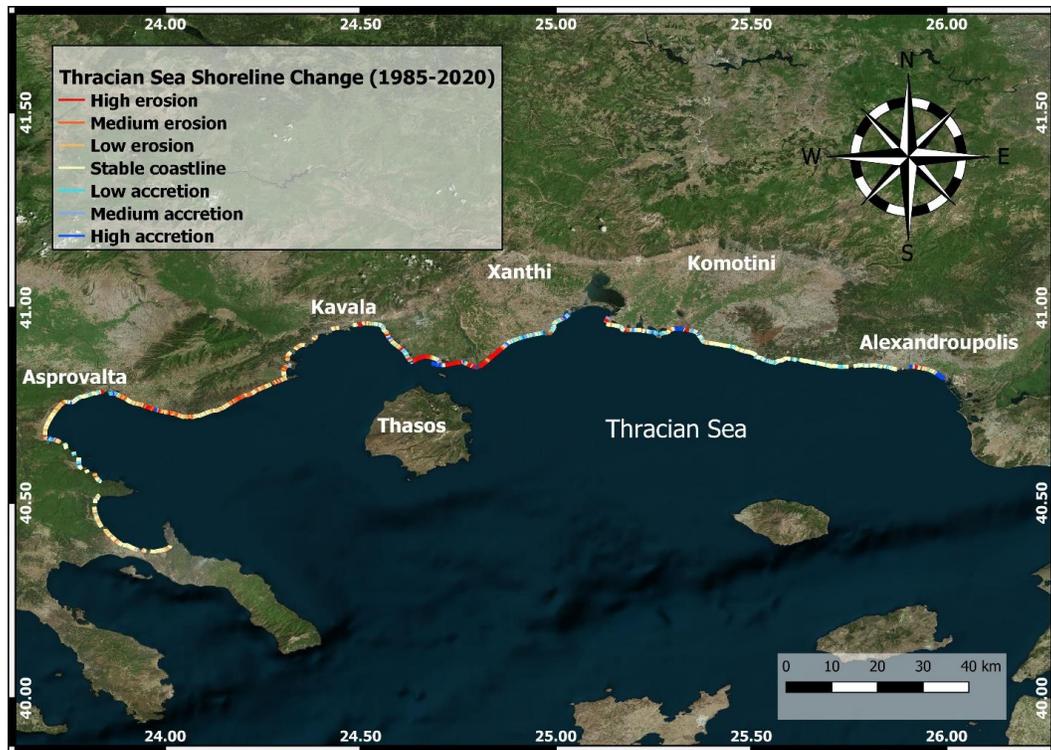


FIGURE 5.1: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE THRACIAN SEA SHORELINE.

The coastline in Strymonikos Gulf seems rather stable with accretion tendency at the north-western segments, due to the sediment transport provided from Strymon River. In Kariani the shoreline shows clear erosive rates of the order of 2,1 m/year at the lee side of a newly-built marina. The upstream side of the marine entraps the longshore transported sediments and the area accretes at rates 1.70 – 2.10 m/year (Figure 5.2). Another coastal erosion “hotspot” exists at the Nestos delta mouth, an area of 15 km in length, where due to the Nestos River damming the highest erosion rates are found of the order of 1,4 – 3,5 m/year. In Fanari, due to the construction of a marina, the shoreline retreats at rates 3,8 – 5,4 m/year, while in Alexandroupolis the expansion of the commercial harbor led to the erosion along the western coast at rates of 0.4 to 0.9 m/year (Figure 5.3).



FIGURE 5.2: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE WESTERN THRACIAN SEA SHORELINE.



FIGURE 5.3: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE EASTERN THRACIAN SEA SHORELINE.

5.3.2 The Israeli Observatory

The shoreline of Israel seems rather stabilized, due to the extensive constructions, mostly breakwaters and groins built along its coastal zone (Figure 5.4). At the southern part, in Ashqelon and Ashdod there exist accretion ‘hotspots’ mostly due to breakwaters and tombolos (Figure 5.5). A series of breakwaters dissipate the incident on the coast wave energy while the groins stabilize the sandy beaches of Tel Aviv. Low to medium accretion prevails in Netanya and Haifa (Figure 5.6).



FIGURE 5.4: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE SHORELINE OF ISRAEL.

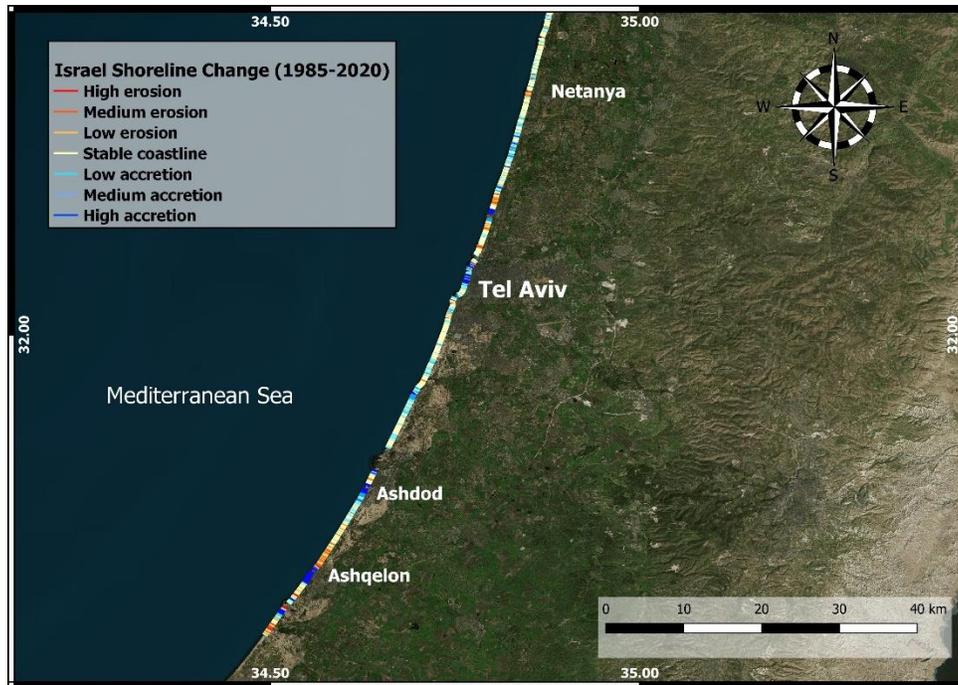


FIGURE 5.5: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE SHORELINE OF SOUTHERN ISRAEL.



FIGURE 5.6: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE SHORELINE OF NORTHERN ISRAEL.

5.3.3 The Valencia Observatory

In Valencia, the coastline is mostly shaped by the sediments transported by Ebro River and the dominant longshore currents and incident waves carrying this material along the coastal zone. At the northern part from Vinaros to Peniscola, the presence of Ebro delta exhibits medium to high sedimentation rates of the order of 0.5 to 0.9 m/year. The coastline is stable along the segment from Alcossebre to Oropesa del Mar and shows high accretion towards the Castello de la Plana. Moving further south, the shoreline shows stabilized trends, with some areas of strong erosion alternating with segments of high sedimentation up to Sagunt harbor. Similar patterns are shown along the coastline towards the Valencia harbor. To the south of the Valencia Harbor, coastal erosion of the order of up to 1.4 m/year prevails as a result of sediment entrapment at the harbor upstream. The remaining coastal zone is characterized by stabilized segments exhibiting that the coastal zone remains in sedimentary balance.

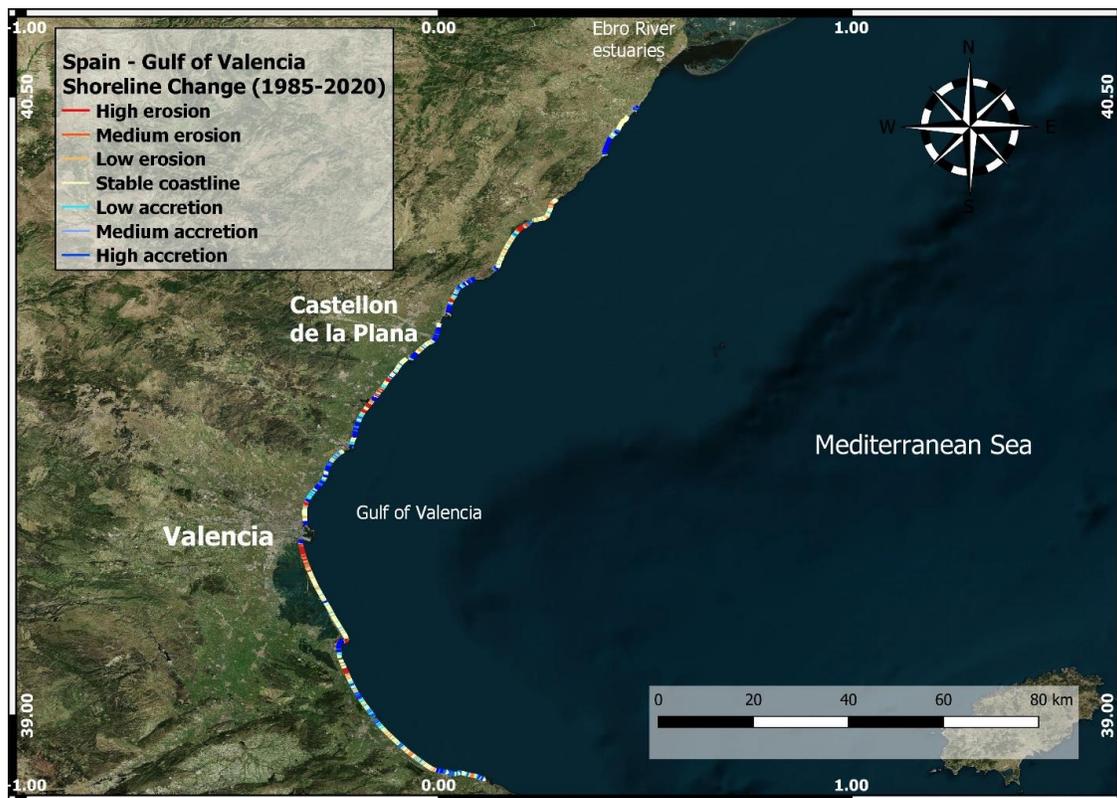


FIGURE 5.7: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE SHORELINE OF THE VALENCIA OBSERVATORY.

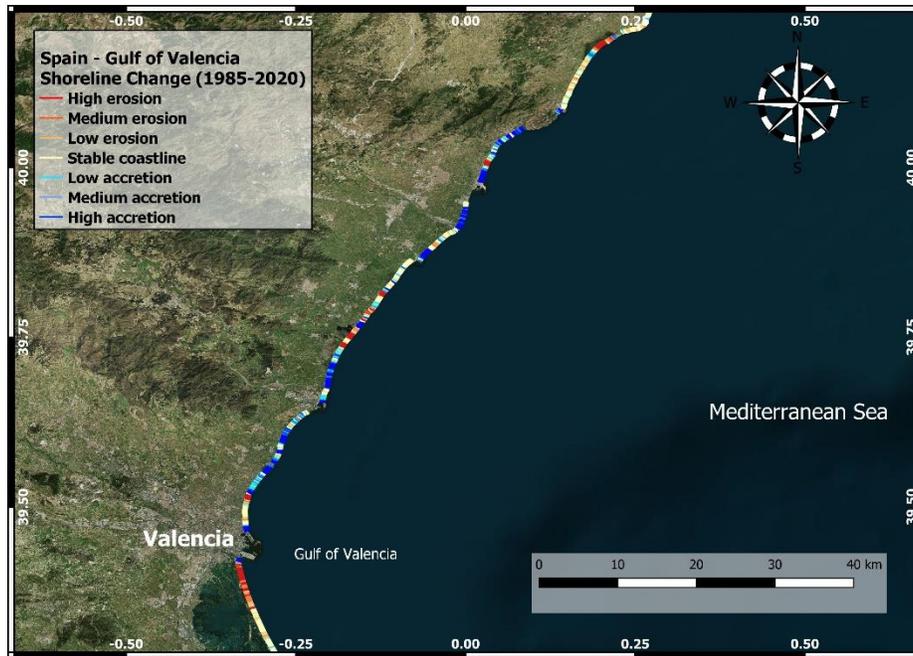


FIGURE 5.8: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE NORTHERN SHORELINE OF THE VALENCIA OBSERVATORY.



FIGURE 5.9: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE SOUTHERN SHORELINE OF THE VALENCIA OBSERVATORY.

5.3.4 The Morocco Observatory

The shoreline along the Al Hoceima Observatory is mostly rocky with high cliffs and therefore coastal erosion is not evident. The western part of the Al Hoceima Bay is characterized by medium erosion to medium accretion. Moving towards the central part of the bay, sandy beaches exist and coastal erosion is prevalent. Therefore, medium to high erosion is seen from Place Shifa to Place Souani and finally to Place R’Hach with coastal erosion rates of 0.3 to 0.6 m/year. Towards the east, deposition is again the dominant feature (rate up to 0.14 m/year) (Figure 5.10).

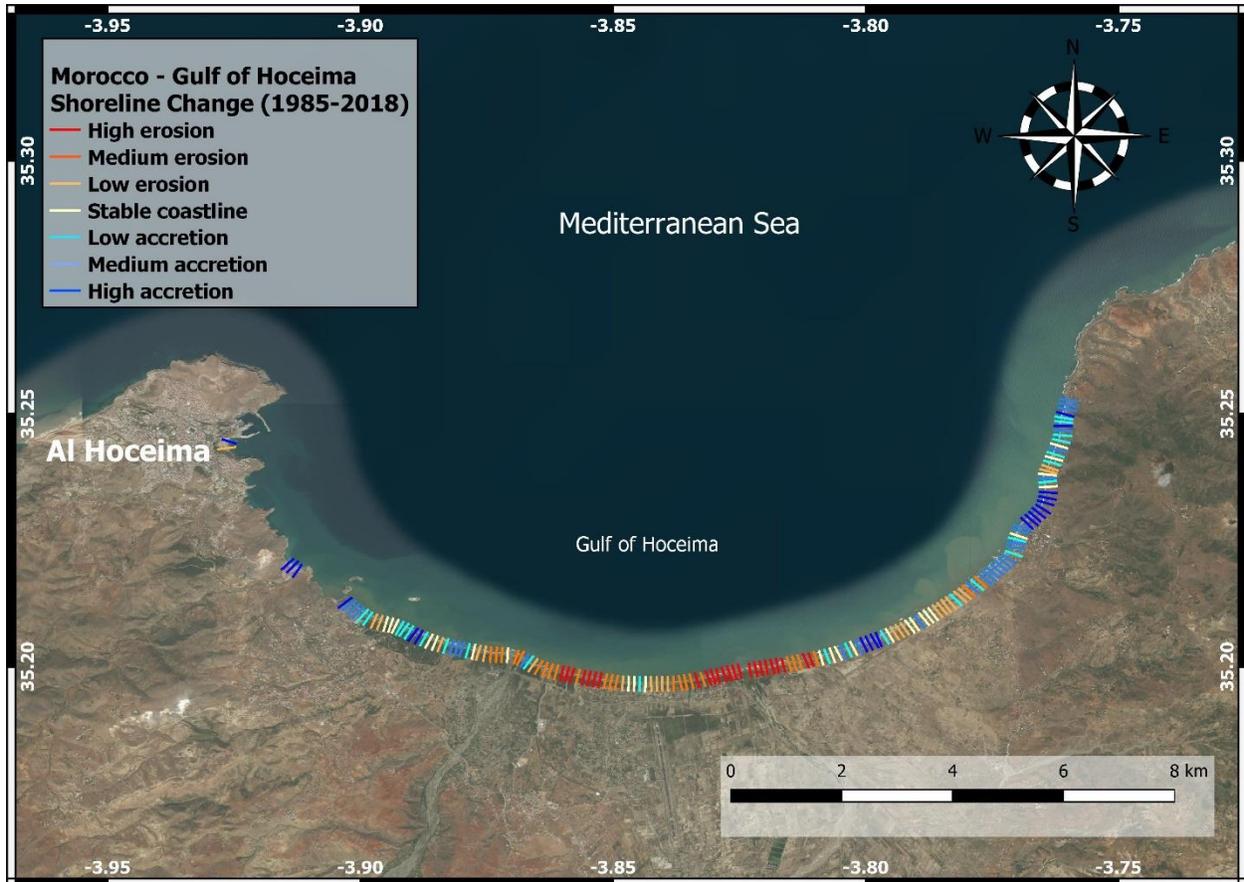


FIGURE 5.10: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE SHORELINE OF THE AL HOCEIMA OBSERVATORY.

5.3.5 The Egyptian Observatory

The coastal zone is stabilized or even accreting at specific segments to the west of Alexandria, in Egypt. The central and eastern parts of Abu Quir Bay are characterized by strong erosion, mostly associated with the damming of the Nile River. High coastal erosion rates followed by accumulation zones are seen in the zone to the west of the city of Baltim. A series of breakwaters, tombolos and groins stabilize the coast between Baltim and Port Said (Figure 5.11). The coastline appears heavily eroded to the east of Port Said.



FIGURE 5.11: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE SHORELINE OF THE EGYPTIAN OBSERVATORY.

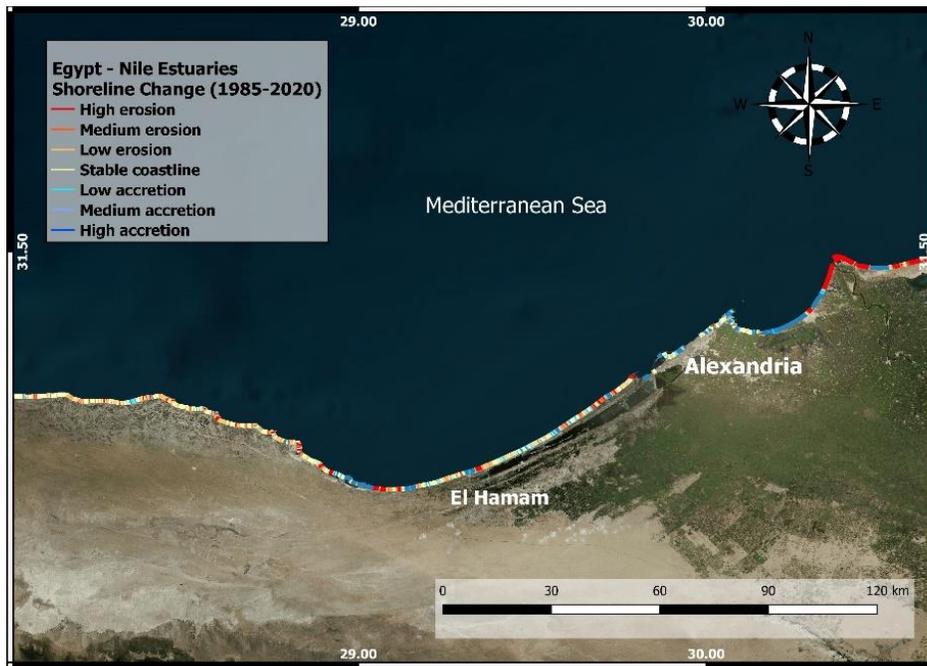


FIGURE 5.12: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE WESTERN SHORELINE OF THE EGYPTIAN OBSERVATORY.

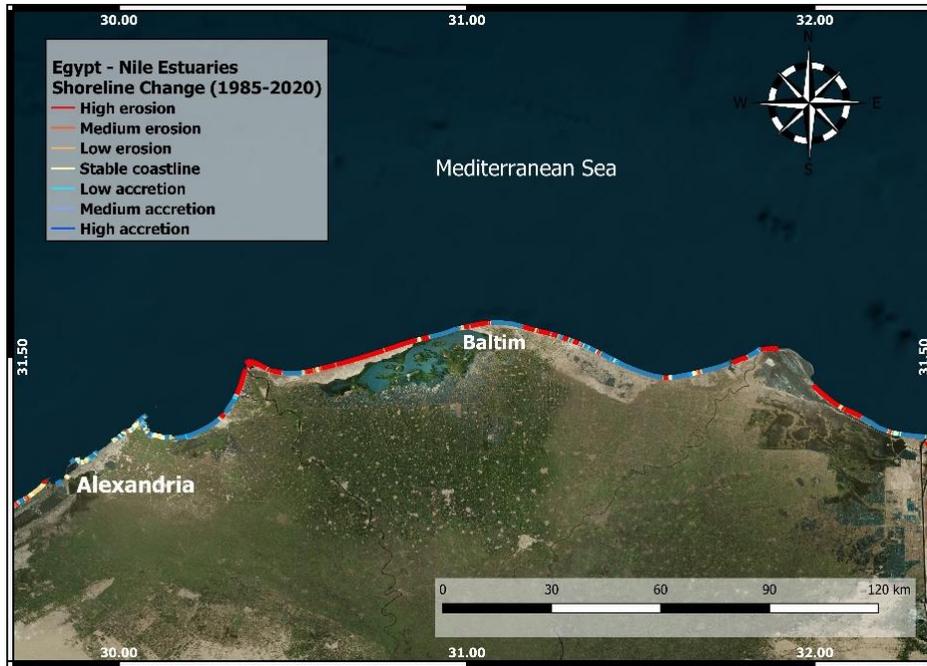


FIGURE 5.13: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE CENTRAL SHORELINE OF THE EGYPTIAN OBSERVATORY.



FIGURE 5.14: COASTAL EROSION/DEPOSITION RATES DURING THE 1985 – 2020 PERIOD, ALONG THE EASTERN SHORELINE OF THE EGYPTIAN OBSERVATORY.

6 Additional Iterations of Eutrophication Algorithm Adaptations

The project delivers data services following the concept of Minimum Viable Products (MVPs). This essentially means that the algorithm development team first focuses on a release which satisfies the basic requirements and therefore avoids spending unnecessary effort to implement further sophisticated features that are not requested by end-users. The MVP has the features to be usable by early customers who can then provide feedback for future product development and fine-tuning. The MVP for the eutrophication product is already released and user validation and expert interviews took place in the past period to gather feedback for additional development iterations. In this section of the deliverable we summarize the outcome of the expert interviews and provide an inventory of next features that could be potentially implemented.

When considering the viability/probability of implementing these requested features for the eutrophication algorithm we should consider certain influencing factors. The output of the algorithm is part of the broader Marinomica Water Quality Service (targeting a wide range of industry, policy and scientific users), however, the eutrophication indicator itself is most useful to policy users. From Marinomica's business sustainability perspective the allocation of future development hours will prioritize those algorithms and features that are producing services for end-users which are most likely to become paying customers.

6.1 Eutrophication algorithm description

Before further elaborating on the eutrophication algorithm features, we provide a brief summary in the below table.

Eutrophication product summary

The eutrophication algorithm calculates four indices:

1. **Eutrophication Index in sea water (Eut index):** computed through the Principal Component Analysis of a combination of five parameters: chlorophyll-a (Chl), nitrate (NO₃), nitrite (NO₂), ammonia (NH₃), and phosphate (PO₄). The first principal component is considered as the eutrophication index;
2. **Trophic Index in sea water (TRIX):** computed as a linear combination of four state variables: chlorophyll (Chl), oxygen saturation (dissolved oxygen - DO), mineral and total nitrogen (dissolved inorganic nitrogen - DIN), and phosphorus (total phosphorus - TP). Initially developed for northern Adriatic Sea;
3. **Unscaled Trophic Index in sea water (UNTRIX):** Unscaled version of TRIX. It is computed by the log of the product of four eutrophication-related parameters: chlorophyll-a (Chl), oxygen saturation (dissolved oxygen - DO), mineral and total nitrogen (dissolved inorganic nitrogen - DIN), and phosphorus (total phosphorus - TP);
4. **Efficiency Coefficient in sea water (Eff coeff):** defined as the logarithm of the ratio between the two aggregated main components of the TRIX index. It is computed by the combination of four parameters: chlorophyll-a (Chl), oxygen saturation (dissolved oxygen - DO), mineral and total nitrogen (dissolved inorganic nitrogen - DIN), and phosphorus (total phosphorus - TP). It can be considered a supplementary index with which to evaluate the nutrient utilization of the system.

These indices are used to inform users on the trophic states of marine ecosystems. The inputs to these indices are chlorophyll a, dissolved oxygen saturation, and nutrients (phosphate and nitrate) to determine the trophic state.

The algorithm code was initially industrialized in collaboration between Thales and GTD. The work consisted in adding configuration support, integrating automatic source data download from CMEMS, adding error handling & console logging, and generating NetCDF compliant output files.

Later, the industrialized algorithm was taken by GTD, and wrapped into a Docker image based on the WPS 2.0 framework developed by CLS. This is the image that has been finally integrated into Marinomica in order to launch the product automatically.

Assumptions

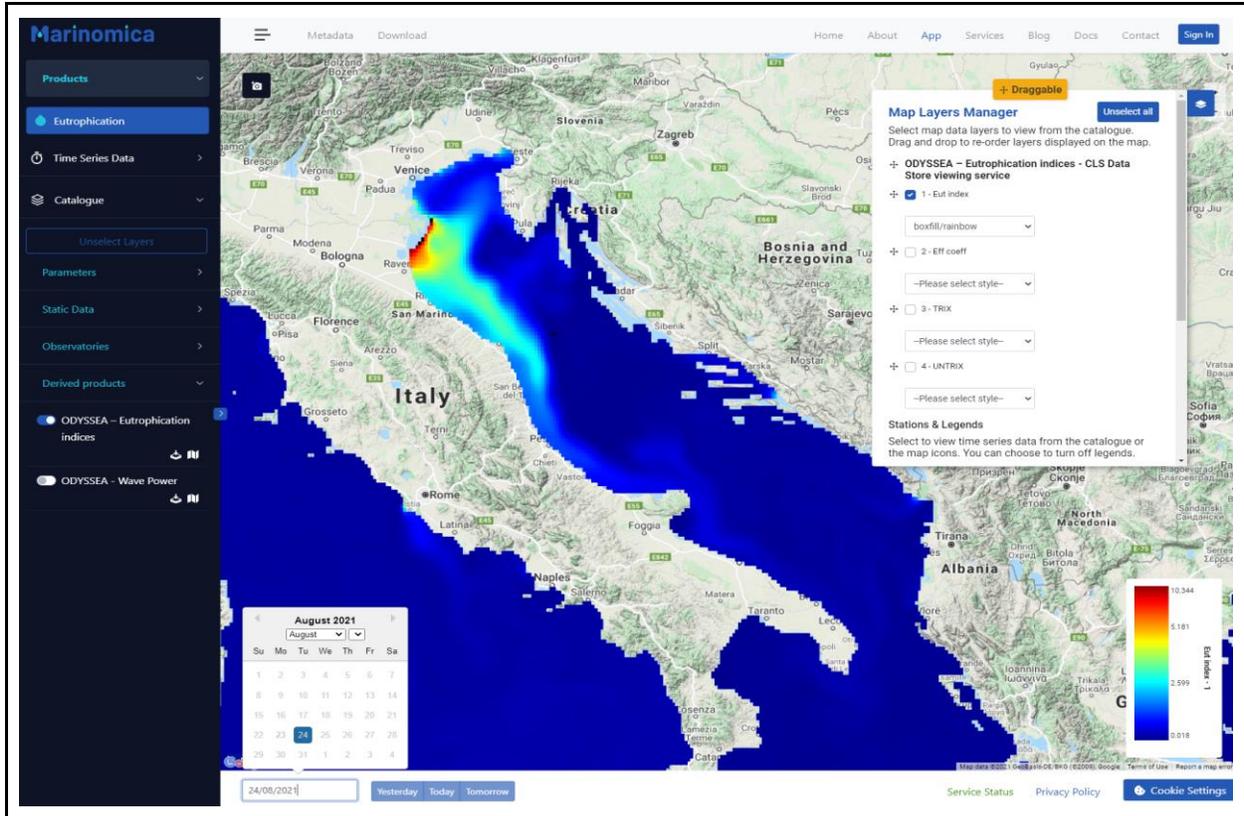
For the development and integration of these indices in the Marinomica platform, several assumptions were made in order to use the operational data available through CMEMS.

Nitrates are the main nitrogen species contribution to the total DIN, due to the lack of ammonia data. In open sea, ammonia and nitrites are of relative low concentration and at the sea surface ammonia rapidly oxidizes to nitrates. When the eutrophication product will be calculated based on the ODYSSEA models, ammonia concentration will be available and the algorithm will be changed slightly to include it especially in areas where wastewater treatment plants outfall. Orthophosphates are the main component of total phosphorus. This is true for the inorganic component, perhaps not true for the organic one, but it is very hard to find data on organic phosphorus (particulate and dissolved) and this is a commonly made assumption.

For the computation of deviation of current DO level from saturation a step-by-step procedure was provided in the deliverable 13.2.

Finally, the coefficients of TRIX calculation are based on data from the Adriatic, general applicability to other areas might not be guaranteed. This means that a rescaling will be needed at a later stage.

Visual representation of the algorithm output



Primary variables

- Nutrients (phosphate and nitrate)
- Dissolved oxygen
- Chlorophyll-a concentration

Secondary (derived) indices

- Eutrophication Index in sea water
- TRophic IndeX in sea water
- UNScaled TRophic IndeX in sea water
- Efficiency Coefficient in sea water

Primary variables



Derived indices

FIGURE 6.1: EUTROPHICATION PRODUCT INPUT VARIABLES AND DERIVED INDICES.

6.2 Eutrophication product requirements

As mentioned above, after the release of the Minimum Viable Product for eutrophication, interviews took place with scientific and policy experts in the field of eutrophication to discuss additional features that

could make the product more tailored to end-user needs. The requested features are listed in the table below and described in this section.

TABLE 6.1: LIST OF REQUESTED EUTROPHICATION ALGORITHM FEATURES.

Feature name	Description
Summary statistics	Complementing standard plots with basic statistics (e.g. mean, standard deviation)
Area sub-setting (sub-regions)	Ability for users to look at the indicators in specific areas. This will require the ability to manage the drawing, saving and sharing of areas for which the product can be run.
Trends	Ability for users to fit various trends based on the values of eutrophication indices
Alerts	Ability to set alerts for when the values of eutrophication indices are above / below set values
Depths	Display data for specific depths. Ability for users to look at eutrophication indices at specific depths. This will require the ability to select depth layers for which the product can be run.
Confidence	Visual presentation of confidence in the time series and the spatial plots
Classification	Working with meaningful value classes (intervals) rather than numeric values

6.2.1 Summary statistics

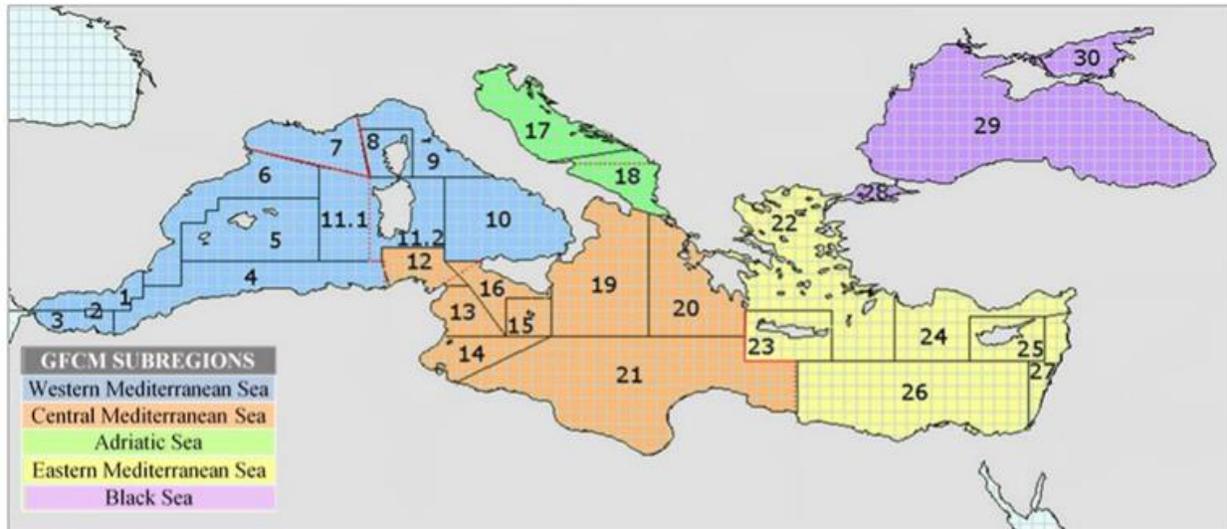
A basic requested feature was to complement traditional time series plots with summary statistics (e.g., mean and standard deviation). This allows users to get a quick overview of the mean behavior of the variable of interest and its variability in the selected time interval. This feature has already been implemented in the latest platform release, as shown in the figure below.



FIGURE 6.2: PRESENTATION OF SUMMARY STATISTICS IN THE TIME SERIES PLOT.

The ability for users to look at the indicators in specific areas was requested. This will require the platform to manage the drawing, saving and sharing of areas for which the product can be run. There should be two ways of defining sub-regions:

- 1) Where an administrator of the platform can define a range of areas that can be made publicly available. For example, the Mediterranean could be divided up into geographic areas and then sub-areas to define a hierarchy of areas that may be used for different purposes (policy, industrial, scientific, etc.). An example of pre-defined geographical sub-regions is shown in the figure below.



--- FAO Statistical Divisions --- GFCM Geographical Subareas (GSAs)

GFCM GSAs

01 - Northern Alboran Sea	07 - Gulf of Lion	13 - Gulf of Hammamet	19 - Western Ionian Sea	25 - Cyprus
02 - Alboran Island	08 - Corsica	14 - Gulf of Gabes	20 - Eastern Ionian Sea	26 - South Levant Sea
03 - Southern Alboran Sea	09 - Ligurian Sea and Northern Tyrrhenian Sea	15 - Malta	21 - Southern Ionian Sea	27 - Eastern Levant Se
04 - Algeria	10 - South and Central Tyrrhenian Sea	16 - Southern Sicily	22 - Aegean Sea	28 - Marmara Sea
05 - Balearic Islands	11.1 - Sardinia (west) 11.2 - Sardinia (east)	17 - Northern Adriatic Sea	23 - Crete	29 - Black Sea
06 - Northern Spain	12 - Northern Tunisia	18 - Southern Adriatic Sea	24 - North Levant Sea	30 - Azov Sea

FIGURE 6.3: EXAMPLE OF GEOGRAPHICAL SUBREGIONS IDENTIFIED BY THE GENERAL FISHERIES COMMISSION FOR THE MEDITERRANEAN (GFCM). SOURCE: [HTTPS://WWW.MEDQSR.ORG/SITES/DEFAULT/FILES/INLINE-FILES/2017 MEDQSR ONLINE 0.PDF](https://www.medqsr.org/sites/default/files/inline-files/2017-medqsr-online-0.pdf).

- 2) A user should be able to create their own areas. This would be particularly useful where a user is interested in a very specific defined area (a mussel farm perhaps). These should be private, to allow fishing areas to be defined, as another example, so that the user can be confident that other users cannot see the information.

Once the areas are defined, these areas can be used as a basis of an alert system. Popup plots should average the values over the area. Function buttons will be added to enable the user to Add, Edit, Delete or share the areas. A user should be able to either draw the points onto a map or enter them as a table of values. Each point can be edited by dragging or by double clicking to edit the location more precisely.

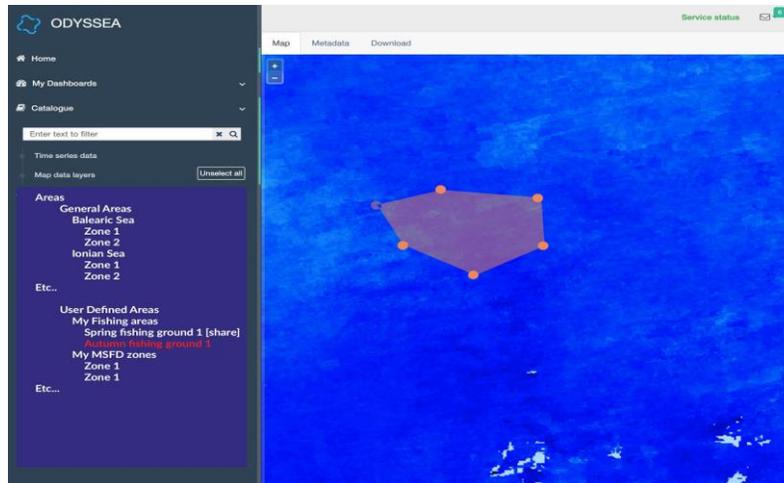


FIGURE 6.4: MOCK-UP OF AREAS NAVIGATION AND A SELECTED AREA.

6.2.2 Depths

The map layer currently loads a single depth. Additional functionality is needed to allow for the user to select the depth at which they are interested in working. It should default to the surface or nearest to the surface if no surface layer exists (note that this may not be appropriate for meteorological parameters). As depth is allowed to be selected, the depth should be shown on the plot (see figure below). A later iteration could allow the user to select the depth from the plot screen or to show plots for all available layers. Note that this plot only relates to a specific pixel size from the map layer and not from an area in general.

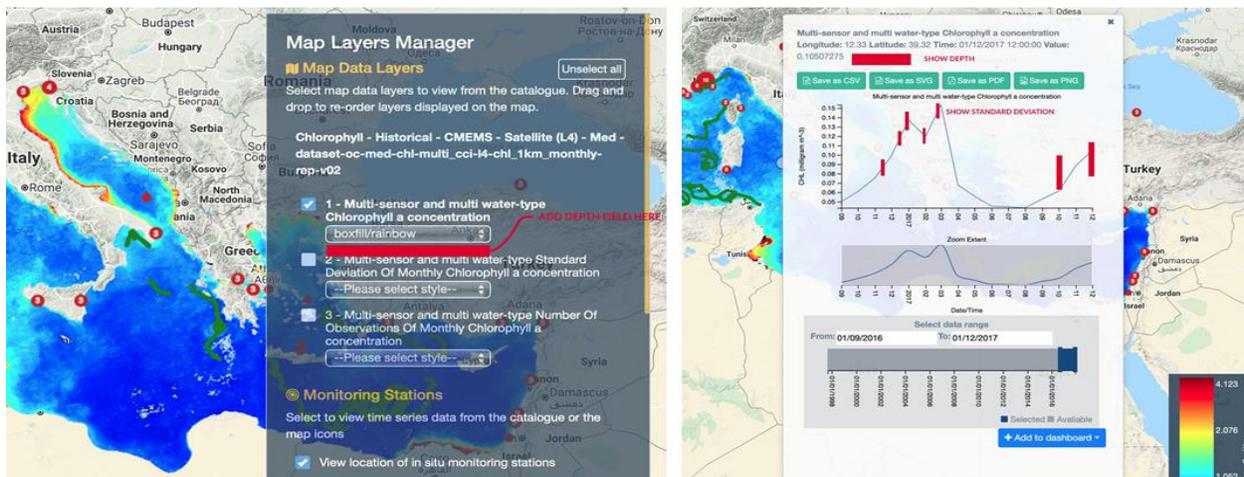


FIGURE 6.5: MAP LAYERS MANAGER SHOWING THE AVAILABLE LAYERS IN THE PRODUCT SELECTED (LEFT). IN THIS CASE AN EXAMPLE OF CHLOROPHYLL. BELOW THE LAYER COLOUR SELECTION, A NEW FIELD IS TO BE ADDED WHICH WILL DISPLAY TO THE USER THE AVAILABLE DEPTHS. DEPTH DISPLAYED ON THE TIME SERIES PLOT (RIGHT).

6.2.3 Alerts

One of the strengths of Marinomica will be the ability for a user to set up alerts for specific events. The eutrophication product is an ideal example of this. Alerts remain an undeveloped part of the application and these notes propose only how they might apply to the eutrophication product but a more general analysis will be needed to design the alerts system for Marinomica.

A user should be able to add / edit / delete multiple alerts and the method of delivery. Delivery method could include email, SMS, push notifications, via the Marinomica mobile app. In addition, not specifically related to eutrophication, administrator alerts could be set to publish automatically to social media or other mechanisms. An example of mobile app visualization of thresholds and alerts (from another domain delivered by a third-party app) is shown below.



FIGURE 6.6: EXAMPLE OF SIMPLE LAYOUT, THRESHOLDS (HEAVY, MODERATE, LIGHT), AND NOTIFICATION “ELEVATED TROPIC INDEX LEVELS STARTING AT XX, LASTING XX”. SOURCE: [HTTPS://PLAY.GOOGLE.COM/STORE/APPS/DETAILS?ID=ORG.YOKI.ANDROID.DROPS&HL=EN](https://play.google.com/store/apps/details?id=ORG.YOKI.ANDROID.DROPS&hl=en).

The alert should be based on an area and inform the user when a condition is in a higher or lower state or within a range. The user should be able to show how often to test for this. For example:

Alert title	Area	Frequency	Indicator	Cond.	Value	When
Too high now	Zone 1	Daily	TRIX	<	4	Now
Too high tomorrow	Zone 1	Daily	TRIX	<	4	Today +1
Hyper	Zone 2	Weekly (Friday)	TRIX	>	6	Today +2

6.2.4 Classification

Since the targeted users of the services do not exclusively include scientists, more informative value classes (intervals) could be displayed instead of harder to interpret numerical values. This information is more important to end-users rather than the actual numeric value. In the figure below we provide two mock-ups on how we envisage the maps and time series plots with values classes. The mock-ups were created for the Trophic index and therefore the values classes are 'Oligo-, Meso-, Eu-, Hypertrophic'. If we consider a more generic water quality indicator, we could use a simpler classification such as 'Poor', 'Average', 'Good', 'Excellent', for instance.

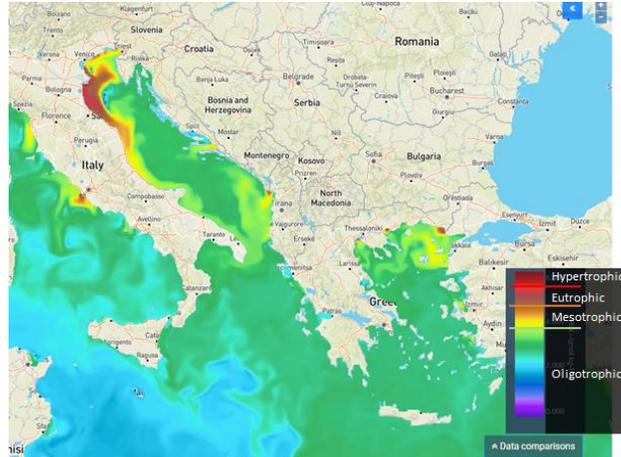


FIGURE 6.7: TROPIC INDEX MAP LEGEND (OVERWRITE VALUES BY CLASSES 'OLIGO-, MESO-, EU-, HYPERTROPHIC')

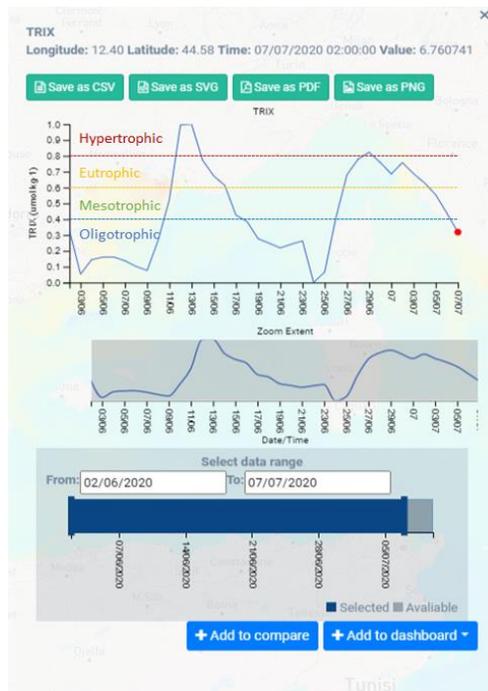


FIGURE 6.8: TROPIC INDEX TIME SERIES (DISPLAY CLASSES 'OLIGO-, MESO-, EU-, HYPERTROPHIC')

6.2.5 Confidence

Another requested feature (most important to scientific users) was to add uncertainty estimates to the products thereby expressing confidence. Visual presentation of confidence can be done in both the time series and the spatial plots. In the following figure we present mock-ups and examples from other applications.

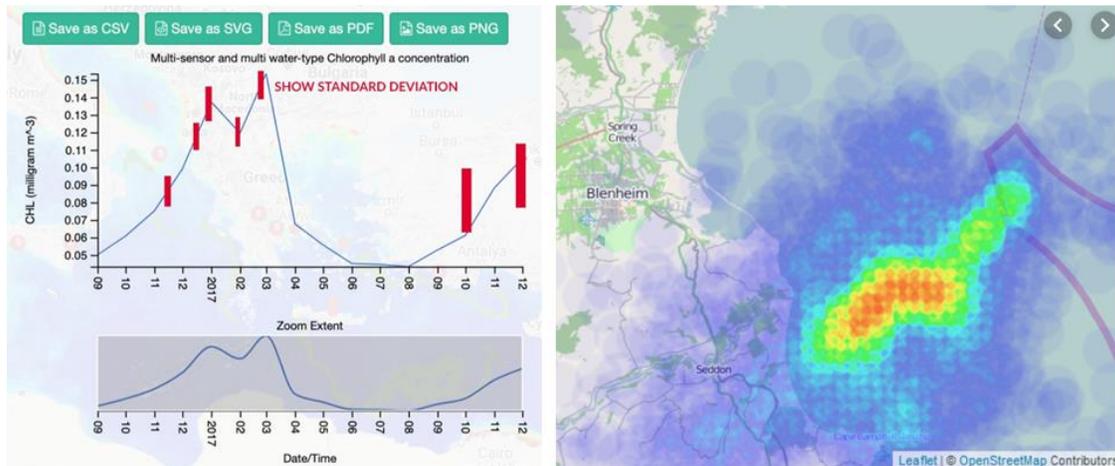


FIGURE 6.9: MOCK-UP OF CONFIDENCE (STANDARD DEVIATION) VISUALIZATION IN THE TIME SERIES DATA (LEFT), AND HEAT MAP VISUALLY SHOWING WHERE CONFIDENCE REDUCES DUE TO LACK OF OBSERVATION DATA (SOURCE: [HTTPS://LEANPUB.COM/LEAFLET-TIPS-AND-TRICKS/READ#LEANPUB-AUTO-GENERATE-A-HEATMAP-WITH-LEAFLETHEAT](https://leanpub.com/leaflet-tips-and-tricks/read#leanpub-auto-generate-a-heatmap-with-leafletheat))

6.2.6 Other comments from the interviews

Apart from the above-described features, here we summarize other important comments of the interviewed experts regarding the Eutrophication product.

TABLE 6.2: FURTHER FEEDBACK ON THE EUTROPHICATION PRODUCT FROM THE INTERVIEWED EXPERTS

Feedback	Description
Review of indicator variables	The chosen eutrophication indices themselves are not used as policy indicators but the underlying variables are used. Thus, the eutrophication product and dashboard should include the evolution in all underlying variables (chlorophyll a, nitrate, phosphate, dissolved oxygen)
Long term data for assessment	For policy assessment purposes, produce long term trends next to the near real time and short-term forecast data, e.g. use the following products: <ul style="list-style-type: none"> ● MEDSEA_REANALYSIS_BIO_006_008: monthly values starting from 1999-01-01 (until 2018-11-01) ● MEDSEA_ANALYSIS_FORECAST_BIO_006_014: daily and monthly values starting from 2018-05-01



Most relevant indicator for operational purposes	For operational forecasting oxygen concentration is the most important variable
Aggregation, additional statistics	For historical assessment provide long-term yearly/seasonal averages (or other statistics like P90 for chlorophyll)
Data source and reliability	It needs to be clearly communicated that the eutrophication products are based on numerical models and these are less reliable along the coast (e.g., if relevant pollution discharges are not present)

7 Semantic and Social Network Harvesting

In deliverables D7.1 and D7.3, we presented the processing chain that we developed in order to obtain information from citizens. The aim is to acquire information from citizen on subjects and in geographically not sufficiently observed types of events. The processing chain relies on the use of an ontology and the conceptual graphs formalism, to provide a pivot representation framework. Thus, information items, coming from different sources can be merged.

D7.3 described a use case on jellyfish invasion monitoring. We presented the domain ontology, as well as InSyTo, a semantic information fusion platform. The adaptation of InSyTo to the use case was described, as well as the obtained results. The use-case developed focused on helping the detection of jellyfish invasions, using unconventional and unusual sources of information for users. The data sources selected were the social network Twitter as well as an online application, jellywatch.com.

Following the completion of the first case study on jellyfish invasion detection, we held demonstrations of the information fusion capabilities developed. An interactive demonstrator has been developed, so as to allow end users to understand the functions offered, by modifying parameters, data sources etc. The work achieved was presented at the FUSION 2020 conference and a film was realized in order to present the results to the project partners and potential Marinomica platform users.

During the various demonstrations, new needs appeared, related to the limitations of the demonstrator. We therefore reiterated the loop of activities by reinitializing a needs collection phase, more targeted this time around the imperfections highlighted in the demonstrator. We detail this phase below.

7.1 New needs for Semantic information Analysis from social media

The demonstrations and workshops set up around the realization of the case study on the detection and management of jellyfish invasion through citizen information feedback revealed a significant lack of information on data sources. of social media types. Indeed, the data acquired on Twitter, in particular, we observed two types of imperfections:

1. Reported events are rarely geo-located and dated. If we can consider using the date of the tweet as the date of the event, this is still risky, as citizens can sometimes intervene about events on dates after the event. Regarding geolocation, the problem is even more significant. First, a very small percentage of tweets are geo-located because phone users do not activate the geolocation option. Second, the interventions and comments of events reported on Twitter can have perpetrators located hundreds of kilometers away.
2. A large amount of information collected is not relevant to monitoring the progress of a situation. This is due to several factors. First, the acquisition of twitter messages is done on the basis of a filter from keywords, or combinations of keywords. As complex as the keyword combination is, however, it is difficult to target exactly one particular event or type of event. In addition, citizens' interventions on an event often contain personal comments and opinions, sometimes controversial or even leading to false information. All of these contributions are irrelevant to situation monitoring, and drown out useful information, making it difficult to access.

7.2 Proposed Solutions

These findings led us to study approaches that would improve the quality of information acquired on social networks. The approach envisaged to do this is to keep track of the source of the information so as to

eventually be able on the one hand to associate a level of trust with it and, on the other hand, to filter the incoming information in order to keep only relevant information for monitoring the situation. From previous work on the traceability of information sources from social networks, we focused, within the framework of ODYSSEA on the second axis: filtering of relevant information. We detail the background and the study conducted below.

7.2.1 Selection of suitable Information sources

The purpose of the new case study is to improve the overall quality of the information used within the case study previously described in the deliverable D7.3. For this reason, we kept the same data sources and added new ones aimed at compensating for their shortcomings.

The data sources for this new phase of the case study can be associated with three media classes. The first two of which were already used in the preliminary study. The first class is made up of all social networks, on which users share with their connections, testimonials including text descriptions, images and videos. These testimonies are dated and sometimes geolocated. In this class, we used the social network Twitter as a source of information. The second class is made up of dedicated applications (mobile or not). In these applications, users are much more guided in entering their testimonial, and testimonials are, in general, geo-localized. This results in much more structured databases, possibly including free text comment fields. As a representative of this class, we used the jellywatch.com website as a source of information. In these first two media classes, the language used for entering comments depends on the user. As a result, the information bases collected are multilingual.

The third class of media considered consists of encyclopedic resources. Among the resources available, we were particularly interested in the use of the Wikidata ontology. Wikidata is a free, collaboratively edited knowledge base hosted by the Wikimedia Foundation. We chose Wikidata, to get away from extracting information from texts, which we should have done if we had, for example, used the encyclopedia Wikipedia. The information in Wikidata is already in a workable format.

Our goal, through this new study, is to use Wikidata to find the precise references of the locations mentioned in the tweets, in order to identify the locations of the events that are reported. The following sections describe how we propose to do that.

7.3 Event Extraction and enrichment

The content of the information conveyed in each tweet must be represented as an information graph, that follows the Conceptual graphs formalism. Our approach relies on the use of bipartite graphs, more specifically the Basic Conceptual Graphs (Sowa 1984, Chein and Mugnier 2008) to represent soft data and knowledge. The conceptual graphs formalism is a model that encompasses a basic ontology (called vocabulary), graph structures and operations on the graphs. The vocabulary defines the different types of concepts and relations that exist in the modelled application domain, while the graphs provide a representation of the observations, which are provided by the information sources.

The event extraction from tweets is made of the three following 3 steps that we detail hereafter. The two first steps were already part of the first proposition described in D7.3. The third one was added to improve the results.

1. Ontology definition and modelling,
2. Named entities recognition
3. Wikidata based enrichment

7.3.1 Ontology definition and event modeling

An ontology that describes the domain of interest (events that have to be monitored for crisis management in our case) is defined. Figure 7.1 depicts the ontology that we defined for the present use-case.

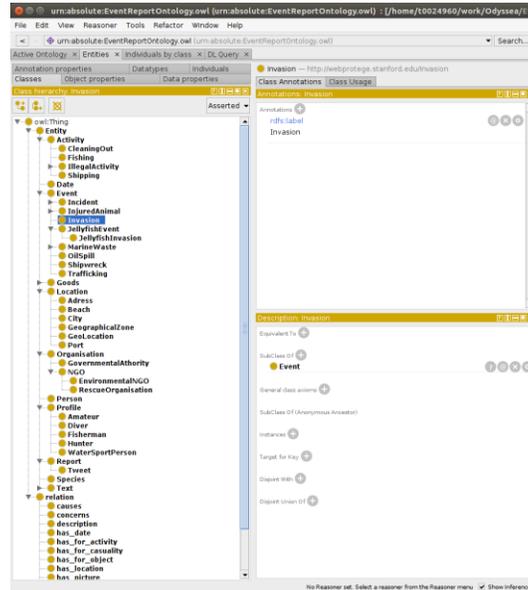


FIGURE 7.1: ONTOLOGY FOR JELLYFISH MONITORING EVENT EXTRACTION.

Furthermore, as we aim at extracting events from the textual description, we use a generic model of graph event description in which we project the extracted entities. This process was detailed in D7.3 so we do not repeat it here.

The model of event, defined as a graph which nodes are labelled with types of the domain ontology is defined and used as an empty graph that will be filled with information acquired in the following sub-processes. The model used for the use case is depicted on Figure 7.2.

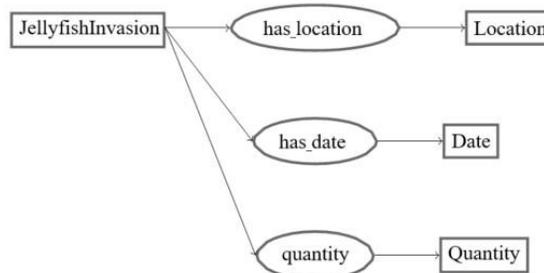
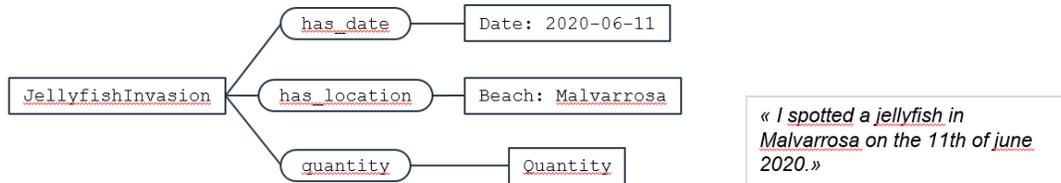


FIGURE 7.2: EVENT MODEL GRAPH

7.3.2 Named Entity Recognition (NER) and Instantiation of the model

To analyze the natural language parts of the tweets (main texts of the tweets), we rely on the Named Entity recognition (NER) function of Spacy. Each entity extracted is associated to its corresponding type in the domain ontology and added to the event graph.

The extracted Named entities are then projected in the event model. This results in small graphs, that have the form of the event model graph, but which contain some instantiated nodes and some instantiated ones. An example of such a partially instantiated graph is given hereafter.



In the first proposition (D7.3), we used the tweets meta-data in order to complete the description of events extracted from each tweet. However, most of the tweets are not geolocalised, and in many cases, the tweets are re-tweeted later on and by people not geolocalised at the same place. Therefore, it would improve the quality of the event extracted if one could correlate the geolocation of the tweet, when available, with the location related in the text, or even more often, use the location related in the text because no geolocation is available in the tweets metadata. For this reason, we study here the possible enrichment using Wikidata.

7.3.3 Wikidata based Enrichment

The event graph is then further enriched with entities and relations queried from Wikidata. Mixing the use of *EntityIndexes* and *PropertyIndexes* from the project Falcon2.0 and requests on Wikidata between pairs of extracted entities using the *pywikibot* python library, we are able to enrich the event graph with both new information and relations between several information items inside the event graph. This enables us improving the level of semantics embedded in the information that represent the tweets.

After adding concepts and entities, the information contained in each Twitter post is enriched with related items extracted from Wikidata. Wikidata is a free knowledge base built to capture and represent general knowledge about the world that can be read and edited by humans and machines alike. It provides data in all languages of the Wikimedia projects, and allows for central access to the data stored. Entities and concepts described in the Wikidata knowledge base are called items and can have labels, descriptions and aliases in several languages. We aim at providing meaningful and enriched information. Therefore, we use the enriched the framework developed earlier with this new capacity.

7.4 Experiments and results

In this section, we first present the data sets that we used and our experimental settings. We then emphasize on the clustering and topic extraction tasks. The impact of the enrichment and filtering methods was investigated by considering distinct content analysis tasks: the impact of adding *WordNet* concept and *DBPedia* entities was analyzed by considering to data clustering and topic extraction before and after performing the enrichment. Filtering factual tweets was analyzed independently, as the task requires in-depth analysis of results by experts. Preliminary results are presented here-under. It was also used as a preliminary step to event extraction. However, we have not yet studied the impact of the use of filtering prior to the events extraction in details. Finally, we evaluate the impact of using Wikidata resources to enrich the graph events extracted from the tweets

7.4.1 Factual vs. non factual tweet filtering

The question raised in this processing step is whether subjectivity in a good indicator for factual/ non factual contents, since there are several approaches developed for subjectivity analysis without considering the objectivity aspects.

In the context of this work, tweets of the first data set (autonomous car) were analysed with TextBlob library, which uses SenticNet to implement a lexicon-based approach for sentiment analysis. The algorithm returns polarity and subjectivity degrees, whose correlation is shown in Figure 7.3.

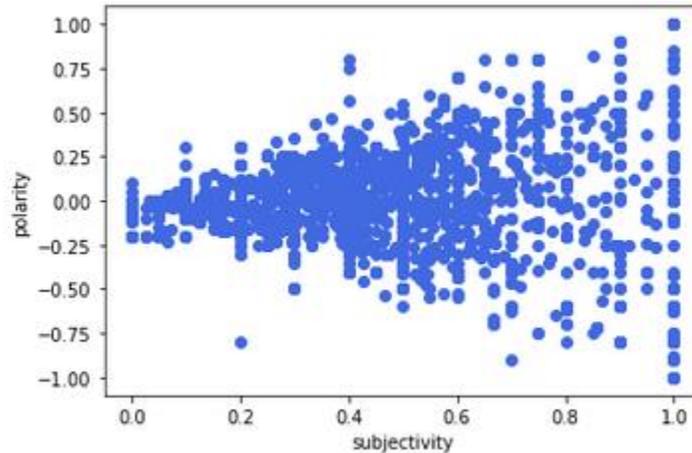


FIGURE 7.3: ANALYSIS OF POLARITY AND SUBJECTIVITY OF TWEETS

To address this issue, the results provided by TextBlob were analyzed. First, two degrees of objectivity (i.e., high and low or objective and non-objective) were defined. Then, the results were analyzed manually by the authors. As tweets can provide both information and opinions or other subjective features, the analysis is carried out at sentence level, in an effort to accurately separate the factual and non-factual parts.

In order to evaluate the quality of the automatic annotation, several examples for each category (high factual & low factual) were selected. Several sentences were also selected for ambiguous cases (i.e., not clearly factual). Those sentences were then analyzed by three annotators who indicated the factual and non-factual features. Figure 7.4 shows the results of the manual evaluation.

Annotator	% factual tweets	% objective tweets
Annotator 1	0.835	0.72
Annotator 2	0.765	0.715
Annotator 3	0.42	0.425
Mean value	0.67	0.62

FIGURE 7.4: EVALUATION OF FACTUAL / OBJECTIVE TWEET.

This evaluation is a first step. We can observe that there is no common consensus over the human annotators, and thus the assessment of the quality of the automatic results depend a lot on the referenced human annotator. Further evaluation is envisioned, for which a first step of sound evaluation framework and methodology will be defined.

7.4.2 Enrichment for Event extraction from texts

In order to evaluate the impact of using Wikidata as a resource for event extraction from tweets, we extracted the graph events from the texts of the tweets from the disaster data set, with and without requesting Wikidata for additional information. We selected the size of the resulting graphs as an indicator of the effect of using Wikidata. The size of one graph is processed as the number of its entities and relation nodes.

Figure 7.5 and Figure 7.6 depict the evolution of these numbers without and with the use of wikidata. On Figure 7.5, the X axis represents the number of entity nodes (respectively relation nodes on Figure 7.2) while the Y axis represents the number of event graphs containing this number of entities (respectively relations). As expected, the size of the graphs, thus the quantity of information conveyed increases with the use of Wikidata as additional source of information.

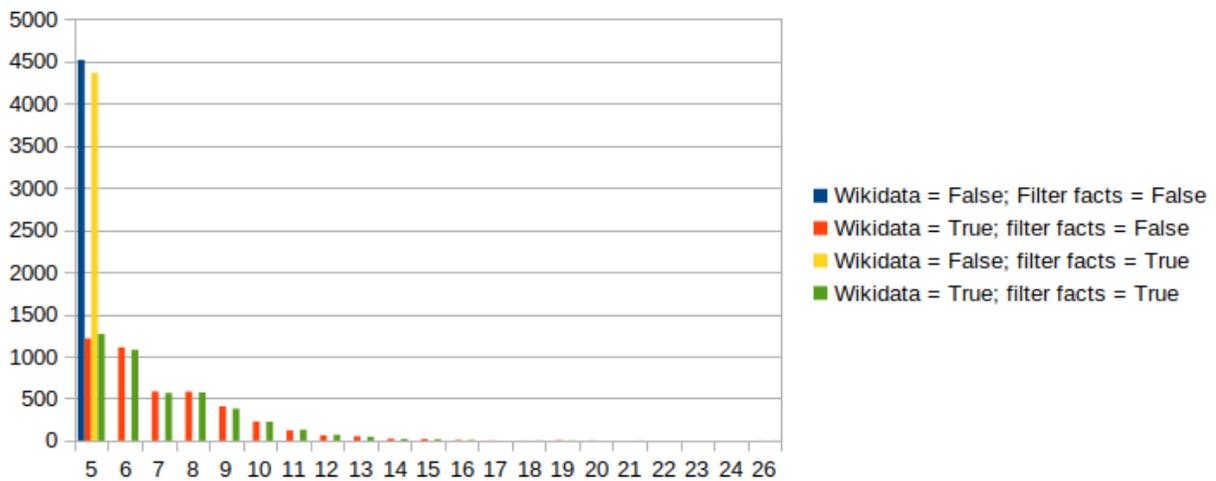


FIGURE 7.5: NUMBER OF GRAPHS PER CLASSES OF ENTITIES NUMBER.

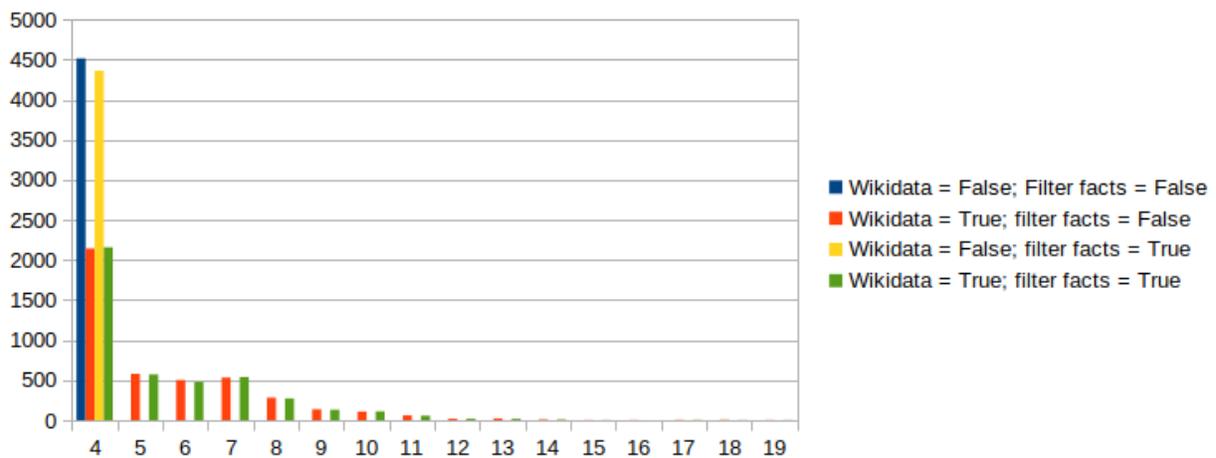


FIGURE 7.6: NUMBER OF GRAPHS PER CLASSES OF RELATION NUMBER.

Another of our concern was to obtain location information from the texts of the tweets in order to overcome the limitation due to the small amount of geo-localized tweets. Therefore, we also compared the locations found with and without coupling the named entity recognition (NER) step to Wikidata.

Figure 7.7 depicts the results of this evaluation. Each bar represents the number of graphs in each class. The classes being identified by the number of “has_location” relation found in the graphs. The coupling first enables us to link locations found in the texts to the event graphs and between themselves inside the graph itself. This last linkage occurs when the text references linked locations such as city and corresponding country for instance. Secondly, coupling NER and Wikidata enables us to find additional locations, linked to the ones referenced in the text (country of a cited city, city of a cited airport etc.)

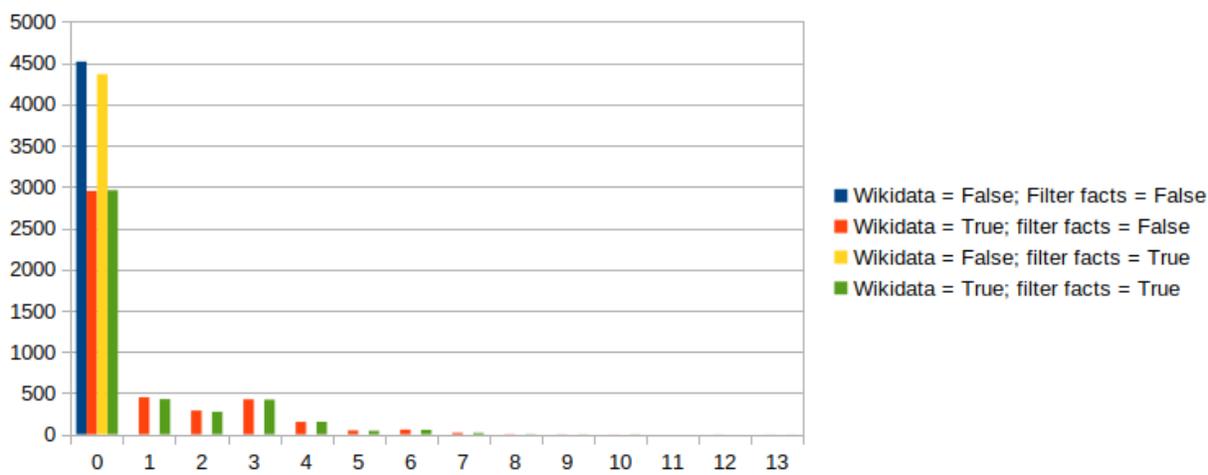


FIGURE 7.7: NUMBER OF LOCATIONS LINKED TO THE EVENT GRAPHS.

Further in-deep manual evaluation of the quality of the enriched graphs with regards to initial ones should also be carried in the near future. We will focus on the quality of the information added to the event graphs to complete the initial quantitative evaluation. To do so, we will need to rely on the evaluation of the quality by experts.

7.5 Concluding remarks for Semantic Information Analysis

Although the volume of tweets is higher in case of crisis, disasters or emergencies, extracting meaningful information from feeds is a challenging task as user post come with several limitations: contextual information is missing, the semantic of the content is poor and factual items are woven with personal beliefs, appreciations and opinions. To overcome those limitations, the approach presented in this paper enriches the Twitter feeds by adding concepts, contextual information and relations. Three external sources support the enrichment process: WordNet, DBPedia and Wikidata. Nonfactual tweets are also filtered in order to enhance the semantics of tweets. The framework is developed to support crisis management and to improve the quality and relevance of items analyzed. The first direction for future work will address the analysis of results. More specifically, we will investigate the impact of enrichment on the quality of events extracted.

8 References

- Borja G. Reguero, Iñigo J. Losada, Fernando J. Méndez. A recent increase in global wave power as a consequence of oceanic warming. *Nature Communications*, 2019; 10 (1) DOI: 10.1038/s41467-018-08066-0
- Claire Laudy, Lőrinc Mészáros, Sonja Wanke, Mercedes de Juan. Mixing social media analysis and physical models to monitor invasive species. FUSION2020 Virtual conference, Pretoria, South Africa. 6-9 July 2020
- Garcva-Rubio G., Huntley D., Russell P., 2015. Evaluating shoreline identification using optical satellite images. *Marine Geology* 359, 96–105.
- GIOVANARDI, F., & Vollenweider, R. A. (2004). Trophic conditions of marine coastal waters: experience in applying the Trophic Index TRIX to two areas of the Adriatic and Tyrrhenian seas. *Journal of Limnology*, 63(2), 199-218
- Hall C.M., Scott D., Gössling S., 2013. The Primacy of Climate Change for Sustainable International Tourism. *Sustainable Development* 21, 112–121.
- Luijendijk A., Hagenaars G., Ranasinghe R., Baart F., Donchyts G., Aarninkhof S., 2018. The State of the world's beaches. *Scientific Reports* 8, 6641.
- Maurizio Pettine, Barbara Casentini, Stefano Fazi, Franco Giovanardi, Romano Pagnotta (2007), A revisitation of TRIX for trophic status assessment in the light of the European Water Framework Directive: Application to Italian coastal waters, *Marine Pollution Bulletin*, Volume 54, Issue 9
- Primpas, I., Tsirtsis, G., Karydis, M., Kokkoris, G.D., D., 2010. Principal component analysis: development of a multivariate index for assessing eutrophication according to the European water framework directive. *Ecol. Indic.* 10, 178–183.
- Vollenweider, R. A., Giovanardi, F., Montanari, G., & Rinaldi, A. (1998). Characterization of the trophic conditions of marine coastal waters with special reference to the NW Adriatic Sea: proposal for a trophic scale, turbidity and generalized water quality index. *Environmetrics: The official journal of the International Environmetrics Society*, 9(3), 329-357.