Norwegian
Meteorological
Institute

METreport

No. 3/2026
ISSN 2387-4201

METOCEAN report for Tanzania demo point

(6.5◦S; 41◦E)

Lars R. Hole and Clio Michel

Foto: Danm12/Shutterstock

Norwegian Meteorological Institute

Title **METOCEAN report for Tanzania demo point**	**Date** 2 March 2026
Section FOU-OM	**Report no.** No. 3/2026
Author(s) and Affiliation(s) Lars R. Hole (lrh@met.no), FOU-OM Clio Michel, FOU-OM	**DOI** doi.org/10.60839/ydvr-qq90
Classification ● Free ○ Restricted	**License**
Client(s) and/or project number NORAD	**Client's reference**

Abstract
Metocean-stats is a Python tool for comprehensive statistics and visualization of metocean data (wind, waves, ocean currents, tide levels, air and water temperature, sea ice, and more). The input data is conveniently provided within the tool as a Pandas DataFrame (time series of metocean variables) from a single position.
The tool is designed with cross-platform compatibility in mind. It can smoothly operate on various operating systems, including Linux, Windows, and MacOS. It is also compatible with WEkEO Jupyter Lab, allowing seamless integration and use.
Here, we have applied metocean-stats in order to develop wind and wave statistics for a location outside the coast of Tanzania. The Jupyter notebooks applied are available on request. See also metocean-stats.readthedocs.io. The data source is the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis ERA5.

Keywords
Wind, wave, climate, statistics, offshore

Disciplinary signature

Responsible signature

Meteorologisk institutt
Meteorological Institute
Org.no 971274042
post@met.no

Oslo
P.O. Box 43 Blindern
0313 Oslo, Norway
T. +47 22 96 30 00

Bergen
Allégaten 70
5007 Bergen, Norway
T. +47 55 23 66 00

Tromsø
P.O. Box 6314, Langnes
9293 Tromsø, Norway
T. +47 77 62 13 00

www.met.no

METOCEAN REPORT FOR TANZANIA DEMO POINT

(6.5°S; 41°E)

Prepared by:
Lars R. Hole
Clio Michel

Reviewed by:
Konstantinos Christakos

Date: March 3, 2026, Rev. 0

This page intentionally is left blank.

Contents

1 Scope of the report

This metocean report is prepared to support general operations and project activities related to Tanzania demo point. The report has a general layout and is applicable to related activities where long-term metocean conditions are in question in the same area.

Metocean-stats is a open-source Python tool for comprehensive statistics and visualization of metocean data (wind, waves, ocean currents, tide levels, air and water temperature, sea ice, and more). The input data is conveniently provided within the tool as a Pandas DataFrame (time series of metocean variables) from a single position.

The tool is designed with cross-platform compatibility in mind. It can smoothly operate on various operating systems, including Linux, Windows, and MacOS. It is also compatible with WEkEO Jupyter Lab, allowing seamless integration and use.

Here, we have applied metocean-stats in order to develop wind and wave statistics for a location outside the coast of Tanzania. The Jupyter notebooks applied are available on request. See also metocean-stats.readthedocs.io.

2 The location of interest

Tanzania demo point with coordinates (6.5°S; 41°E) is located east of Dar es Salaam and Zanzibar (Fig. 2.1).

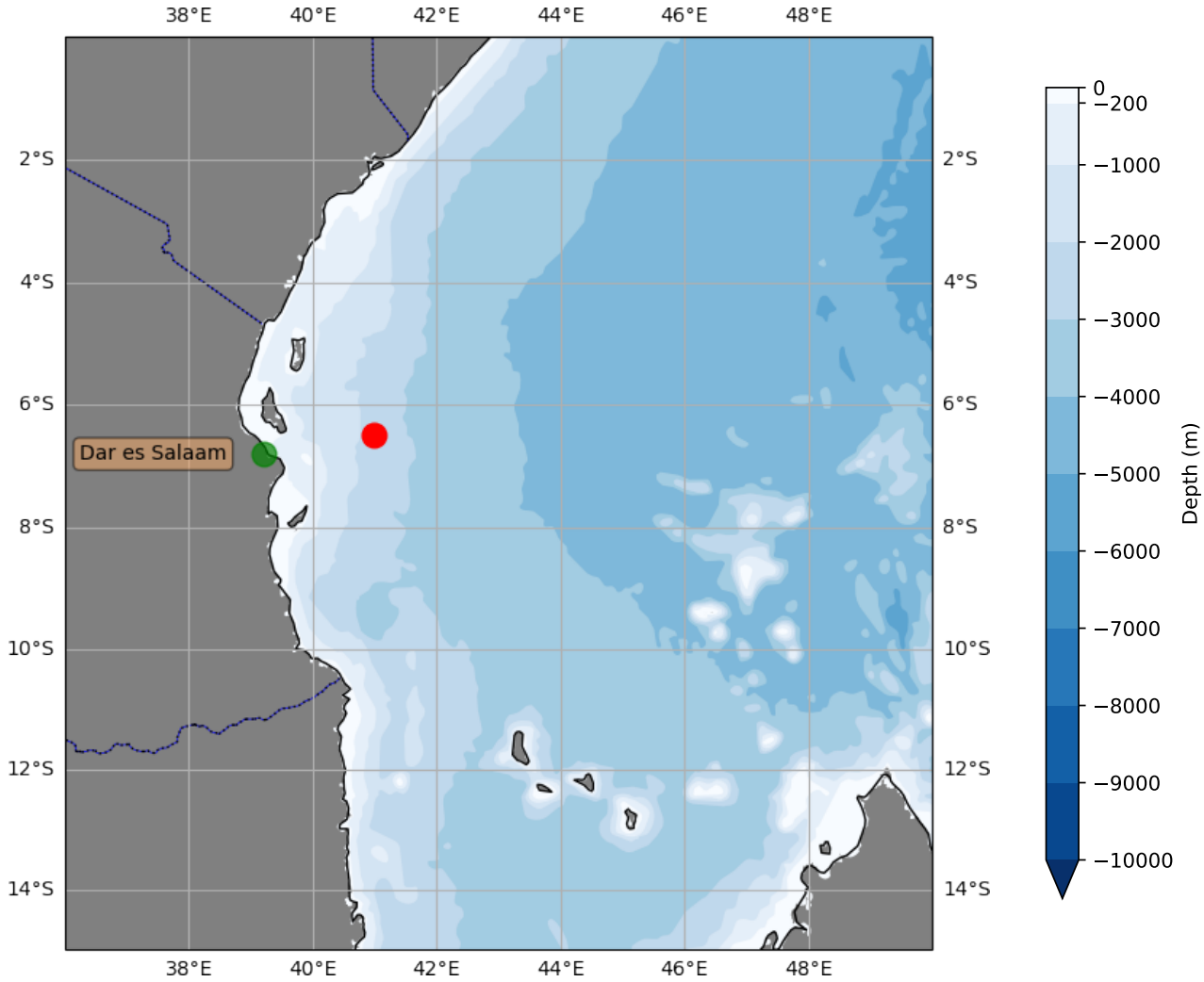

Figure 2.1: The location of Tanzania demo point (6.5°S; 41°E) shown in red. The colored surface represents the bathymetry (in m).

3 Air temperature

3.1 Description of the data

We here use the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis ERA5 (Hersbach et al., 2020). ERA5 here provides hourly data at 31 km resolution for the 55-year period 01.01.1970 - 31.12.2024. The closest point to Tanzania demo point (6.5°S; 41°E) in the hindcast grid is at 6.5°S; 41°E. The variable used is the air temperature at 2 m above the ground/sea level.

3.2 General statistics

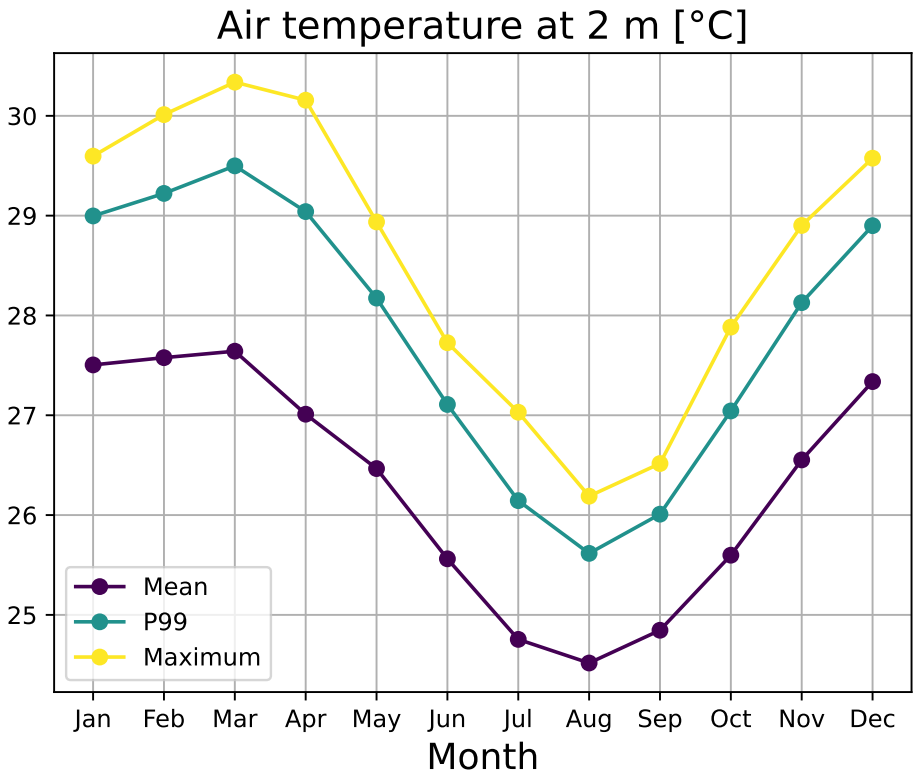

Figure 3.1: Monthly air temperature statistics at Tanzania demo point.

3.3 Long-term trends

Figure 3.2 shows the increasing yearly mean temperature with time.

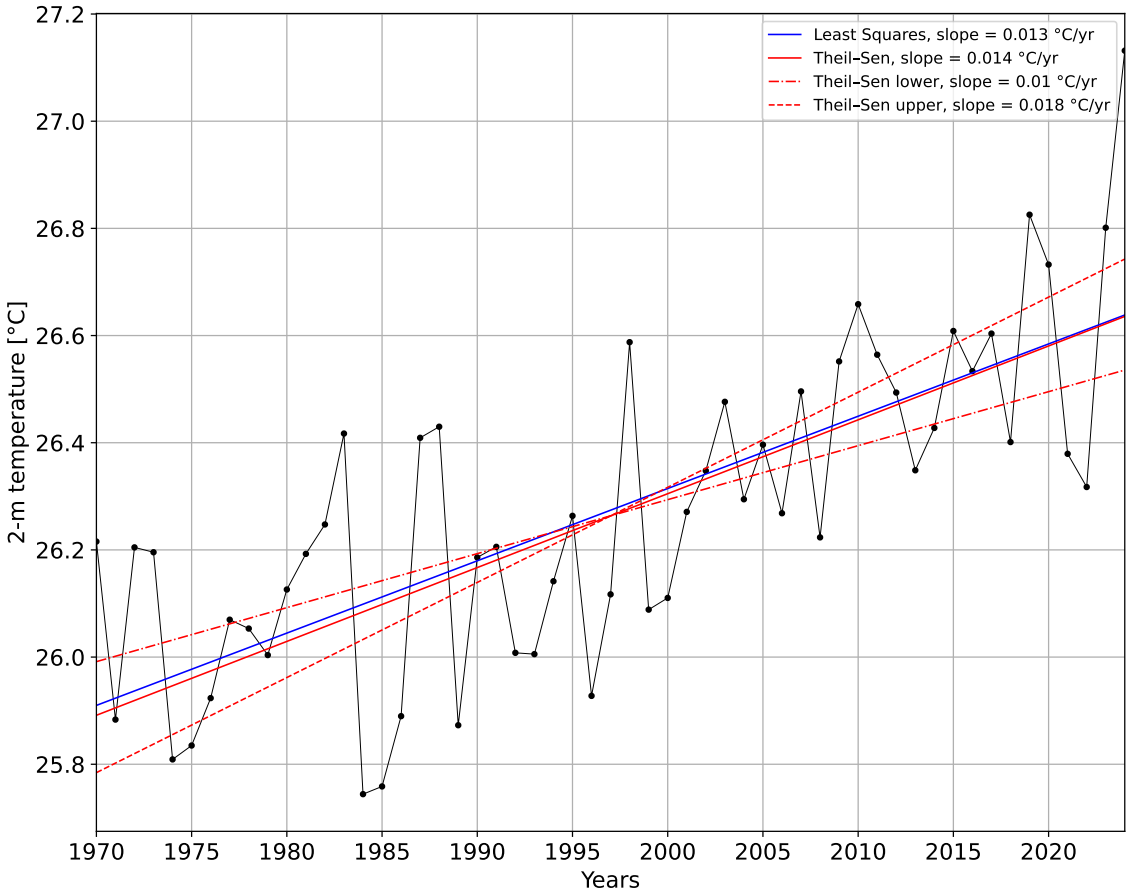

Figure 3.2: Yearly mean of the 2-m air temperature (black line) and its linear trend (blue and red lines).

4 Wind

4.1 Description of the data

We here use the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis ERA5 (Hersbach et al., 2020). ERA5 here provides hourly data at 31 km resolution for the 55-year period 01.01.1970 - 31.12.2024. The closest point to Tanzania demo point (6.5°S; 41°E) in the hindcast grid is at 6.5°S; 41°E. The variable used is the wind speed and direction at 10 m above the ground/sea level.

4.2 General statistics

Figures 4.1 and 4.2 present monthly and directional distribution of mean, P99 and maximum of wind speed. Monthly statistics are shown in Table 4.1. The monthly and annual non-exceedance levels of wind speed at 10 meters are presented in Table 4.2.

Figure 4.3 presents the frequency table of wind speed (sorted by 2 m/s) vs. H_s (sorted by 1 meter).

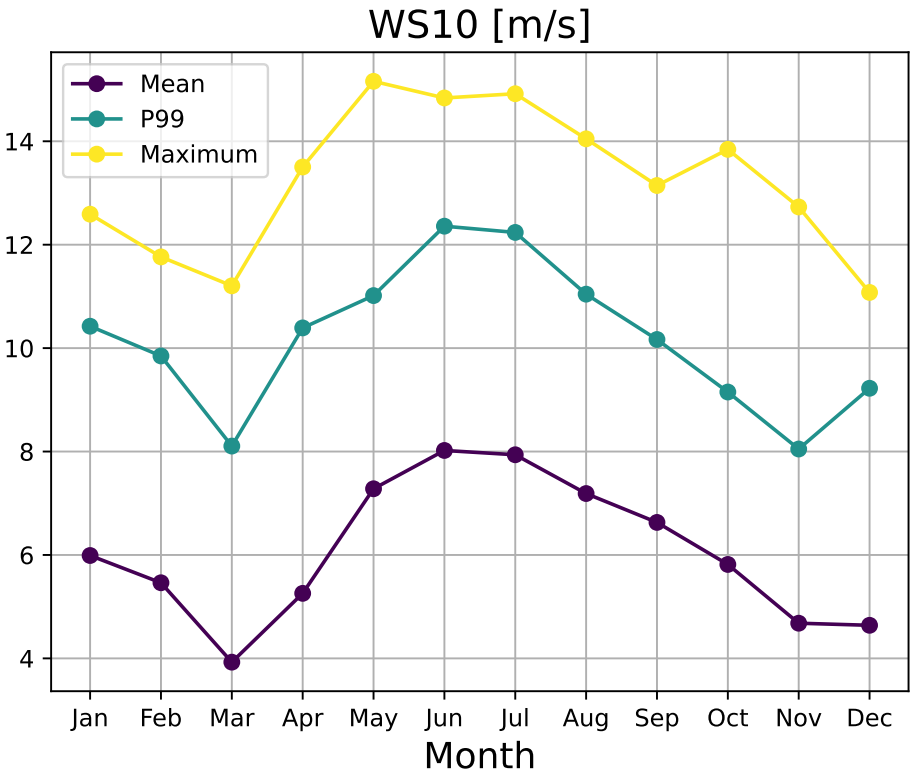

Figure 4.1: Monthly mean, 99th percentile (P99), and maximum of the 10-m wind speed (in m/s).

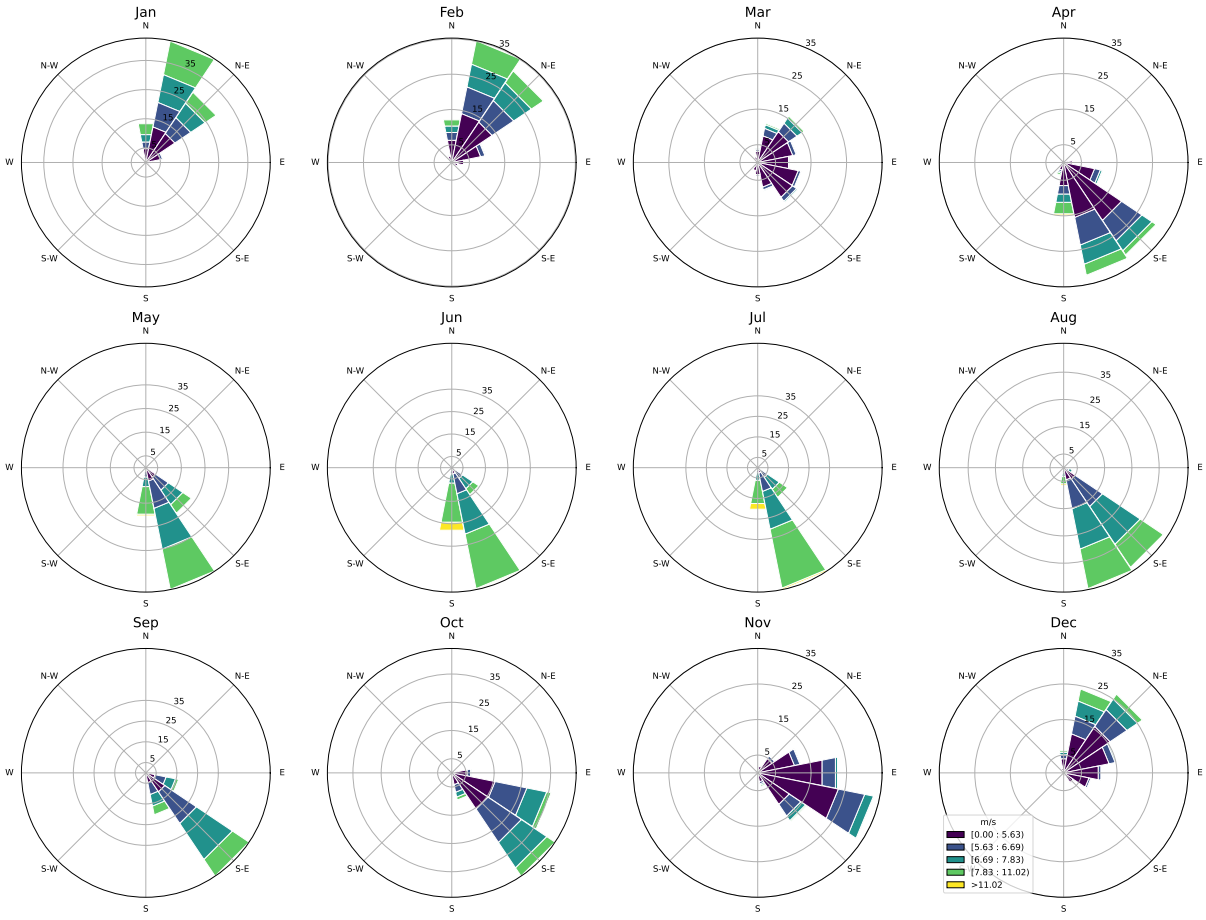

Figure 4.2: Wind roses of the 10-m wind speed (in m/s, coming from) per 30° sectors for each month. The colors show the wind speed intervals and the gray circles the frequency of occurrence as a percentage.

Month	Minimum	Mean	Maximum
Jan	7.9	10.4	12.6
Feb	7.3	9.8	11.8
Mar	6.3	8.3	11.2
Apr	6.7	9.9	13.5
May	8.6	11.2	15.2
Jun	9.5	12.1	14.8
Jul	9.5	12.1	14.9
Aug	9.2	11.0	14.0
Sep	8.3	10.2	13.1
Oct	8.0	9.4	13.8
Nov	7.0	8.3	12.7
Dec	6.7	9.1	11.1
Annual Max.	11.1	13.0	15.2

Table 4.1: Monthly and annual minimum, maximum and mean values of the annual maximum of the 10-m wind speed (WS10 in m/s).

4.3 Extreme value analysis

Figure 4.4 presents the monthly return values of the 10-m wind speed obtained using the 3-parameter Weibull distribution model. The extreme event duration is 1 hour.

ws10-level	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Year
<1.0	0.82	0.95	3.3	0.87	0.01	0.0	0.0	0.0	0.0	0.05	0.66	1.8	0.71
<2.0	3.63	4.61	13.89	4.1	0.1	0.0	0.02	0.01	0.05	0.64	3.96	8.36	3.28
<3.0	9.24	12.21	31.48	12.13	0.51	0.04	0.13	0.08	0.4	3.09	12.85	20.94	8.59
<4.0	18.43	24.03	53.24	25.84	2.09	0.41	0.39	0.6	2.25	10.14	30.81	38.78	17.25
<5.0	31.95	40.74	73.37	45.54	6.9	2.18	2.1	4.28	10.18	27.1	57.55	58.33	30.02
<6.0	48.3	60.22	88.02	66.36	19.61	9.45	8.99	17.89	30.94	53.96	82.11	75.77	46.8
<7.0	65.93	76.83	95.66	82.67	43.04	27.86	27.79	46.0	61.95	80.38	95.48	88.19	65.98
<8.0	81.84	88.81	98.86	92.02	68.77	53.41	55.08	74.99	86.87	94.93	98.93	95.29	82.48
<9.0	92.91	96.13	99.72	96.8	86.64	73.77	77.39	91.58	96.31	98.8	99.76	98.63	92.37
<10.0	97.87	99.22	99.95	98.54	95.54	87.53	89.97	97.27	98.79	99.53	99.92	99.78	96.99
<11.0	99.66	99.9	99.99	99.45	98.98	95.3	96.12	98.96	99.62	99.8	99.95	100.0	98.98
<12.0	99.99	100.0	100.0	99.92	99.8	98.42	98.69	99.57	99.91	99.88	99.97	100.0	99.68
<13.0	100.0	100.0	100.0	99.97	99.97	99.59	99.59	99.92	99.99	99.98	100.0	100.0	99.92
<14.0	100.0	100.0	100.0	100.0	99.98	99.93	99.93	100.0	100.0	100.0	100.0	100.0	99.99
<15.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
<16.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Minimum	0.03	0.07	0.05	0.02	0.39	2.24	1.04	0.96	1.3	0.48	0.11	0.01	0.01
Mean	5.99	5.46	3.93	5.26	7.28	8.02	7.94	7.19	6.63	5.82	4.68	4.64	6.07
P50	6.09	5.47	3.85	5.21	7.26	7.86	7.81	7.12	6.63	5.86	4.73	4.56	6.18
P75	7.54	6.88	5.09	6.49	8.29	9.07	8.87	8.0	7.46	6.76	5.66	5.95	7.51
P95	9.33	8.8	6.87	8.52	9.91	10.94	10.76	9.45	8.74	8.01	6.94	7.95	9.47
P99	10.42	9.85	8.11	10.39	11.02	12.36	12.24	11.04	10.17	9.15	8.05	9.22	11.02
Maximum	12.59	11.76	11.2	13.5	15.16	14.84	14.92	14.05	13.14	13.84	12.73	11.08	15.16

Table 4.2: Monthly and annual non-exceedance levels of the 10-m wind speed (WS10 in m/s) for various WS10 thresholds along with the minimum, mean, maximum and 50th, 75th, 95th, and 99th percentiles of WS10.

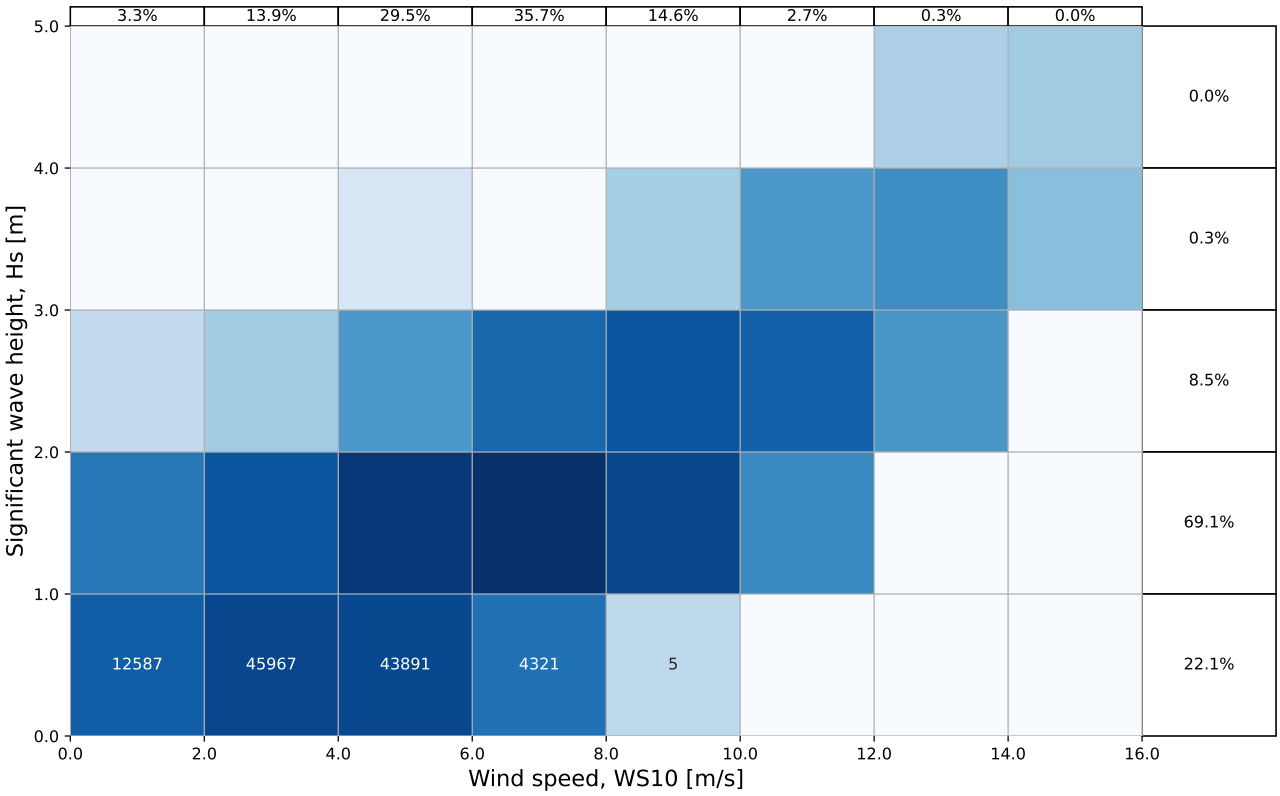

Figure 4.3: Frequency table of the 10-m wind speed (WS10 in m/s, horizontal axis) vs. significant wave height (Hs in m, vertical axis).

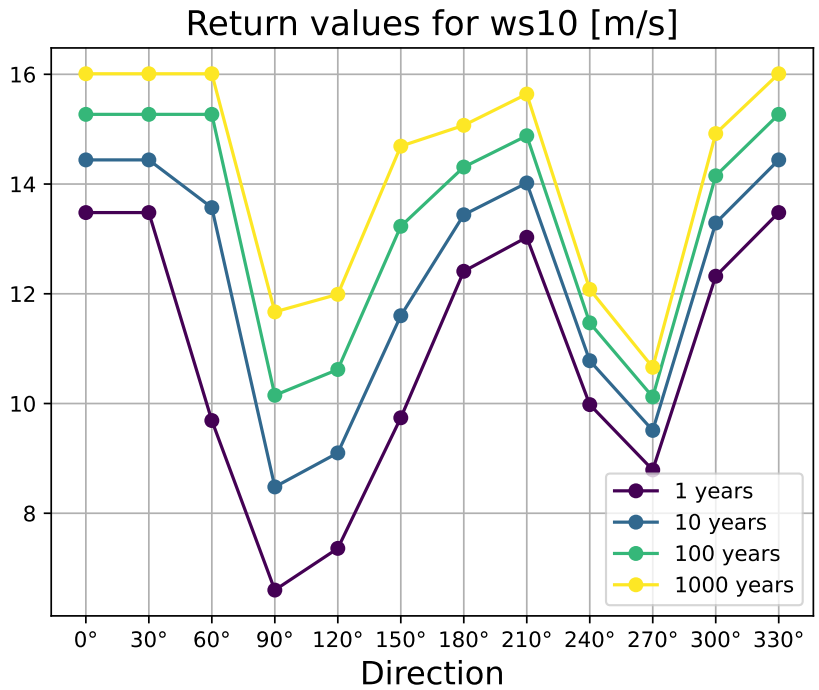

Figure 4.4: 1, 10, 100, and 1000-year return values of the 10-m wind speed (in m/s) for each wind direction sector (in degrees).

5 Waves

5.1 Description of the data

We here use the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis ERA5 (Hersbach et al., 2020). ERA5 here provides hourly data at 31 km resolution for the 55-year period 01.01.1970 - 31.12.2024. The closest point to Tanzania demo point (6.5°S; 41°E) in the hindcast grid is at 6.5°S; 41°E. The variables used are the significant height of combined wind waves and swell, and the mean wave period and direction. Note that the wave directions are defined as the directions the waves are coming from.

5.2 General statistics

Figure 5.1 presents the frequency table of corresponding H_s and T_p in percentage. The marginal distributions of the respective parameters are found at the far right and the top of the table.

Figure 5.2 presents the monthly directional distribution of H_s corresponding to the mean wave direction M_{Dir}, sorted into sectors of 30°.

Figure 5.3 presents the monthly values of mean, the 99th percentile (P99) and maximum of the significant wave height.

Figure 5.4 shows the joint significant wave height and 10-m wind speed distribution.

Figure 5.5 shows the overall probability of non-exceedance for Hs, fitted by the Weibull curve.

Table 5.1 shows monthly and annual non-exceedance levels of significant wave height, sorted by 1-m interval wave.

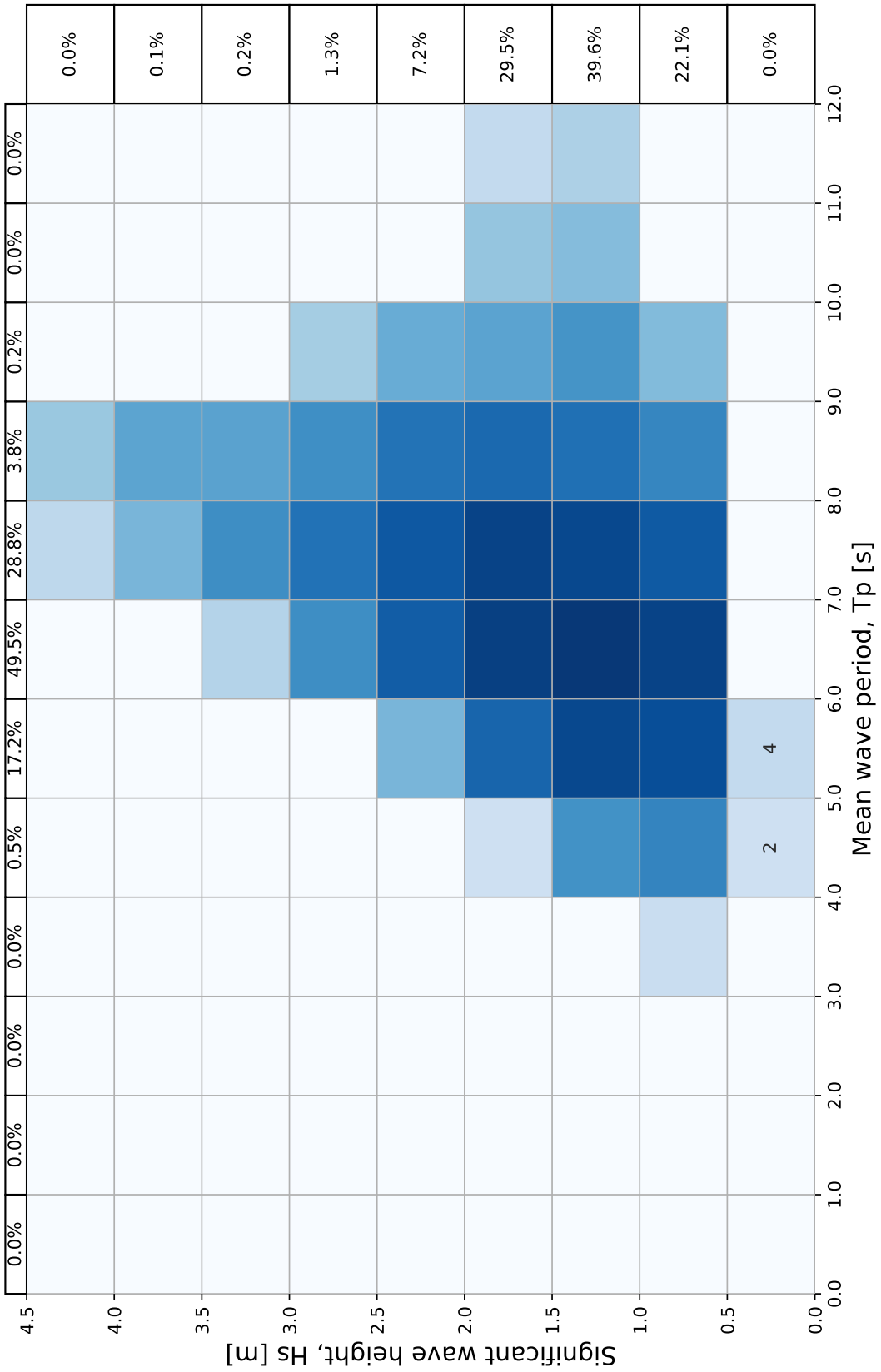

Figure 5.1: Frequency table of the significant wave height (H_s in m, rows) vs. the mean wave period (T_p in s, columns).

swh-level	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Year
<0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0.0	0.0
<1.0	24.01	39.7	69.44	35.12	4.33	0.94	0.01	0.09	0.54	7.51	41.63	43.64	22.25
<1.5	76.67	84.87	98.49	84.36	43.1	20.74	15.25	20.55	38.43	71.57	94.94	93.22	61.85
<2.0	96.05	98.37	100.0	96.6	87.4	72.45	71.46	84.36	92.13	97.4	99.6	99.36	91.26
<2.5	99.81	99.86	100.0	99.33	98.34	93.57	93.5	98.17	98.96	99.61	99.93	100.0	98.42
<3.0	100.0	100.0	100.0	99.95	99.84	98.78	98.49	99.51	99.81	99.85	99.95	100.0	99.68
<3.5	100.0	100.0	100.0	99.98	99.98	99.75	99.6	99.88	100.0	99.96	99.98	100.0	99.93
<4.0	100.0	100.0	100.0	100.0	99.99	99.97	99.97	100.0	100.0	100.0	100.0	100.0	99.99
<4.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Minimum	0.64	0.52	0.53	0.54	0.51	0.8	0.93	0.96	0.84	0.5	0.49	0.53	0.49
Mean	1.27	1.15	0.93	1.18	1.59	1.82	1.86	1.73	1.6	1.37	1.08	1.08	1.39
P50	1.21	1.08	0.89	1.11	1.56	1.76	1.8	1.7	1.58	1.34	1.04	1.04	1.35
P75	1.48	1.33	1.04	1.35	1.8	2.04	2.05	1.89	1.76	1.53	1.2	1.21	1.68
P95	1.94	1.79	1.31	1.85	2.22	2.58	2.58	2.24	2.09	1.87	1.5	1.57	2.16
P99	2.29	2.09	1.55	2.41	2.61	3.06	3.15	2.74	2.52	2.24	1.77	1.92	2.64
Maximum	2.89	2.94	1.99	3.59	4.27	4.14	4.12	3.87	3.41	3.97	3.62	2.36	4.27

Table 5.1: Monthly and annual non-exceedance levels (as percentage) of the significant wave height (H_s in m) for various H_s thresholds along with the minimum, mean, maximum and 50th, 75th, 95th, and 99th percentiles of H_s.

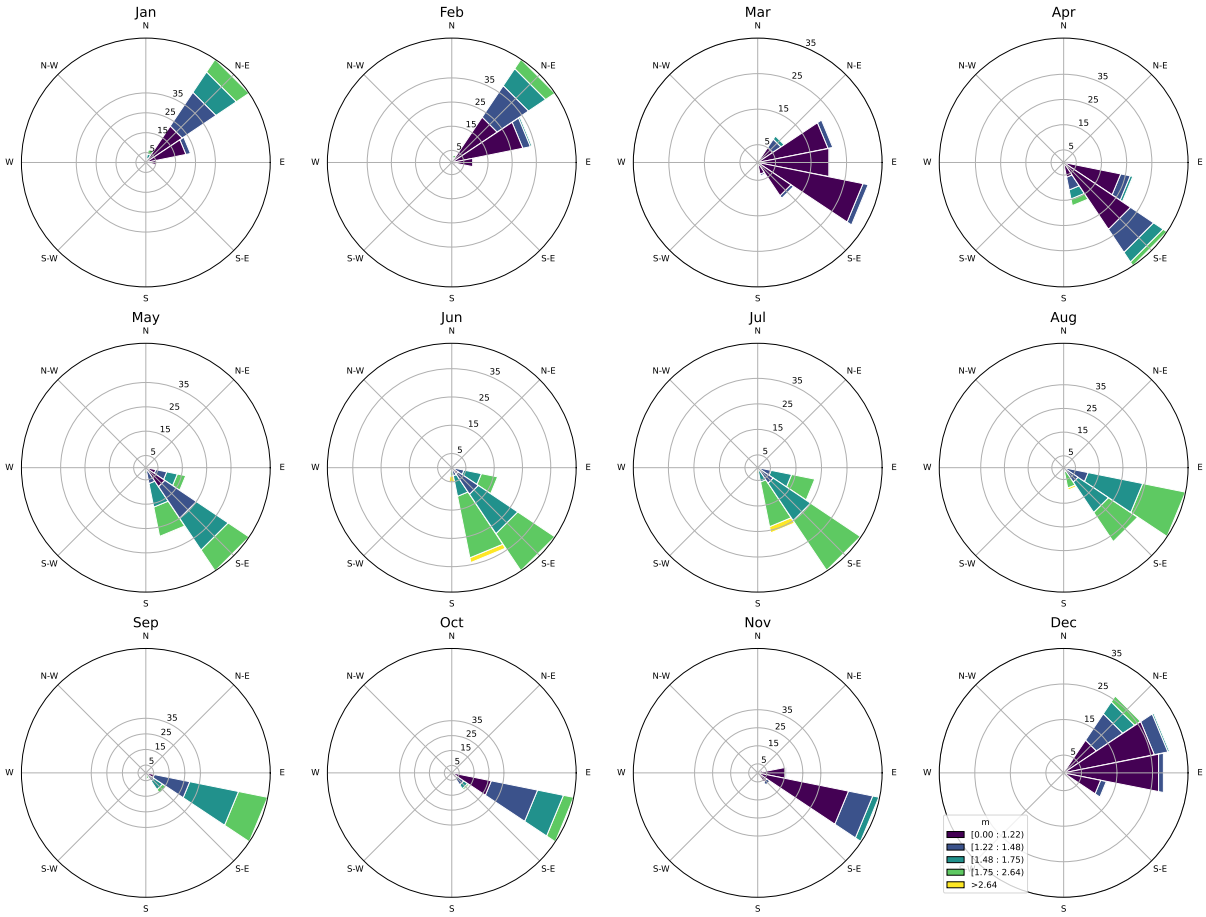

Figure 5.2: Directional distribution of the significant wave height (H_s in m) corresponding to the mean wave direction (coming from, in degrees). The colors show the H_s intervals and the gray circles the frequency as a percentage of the time.

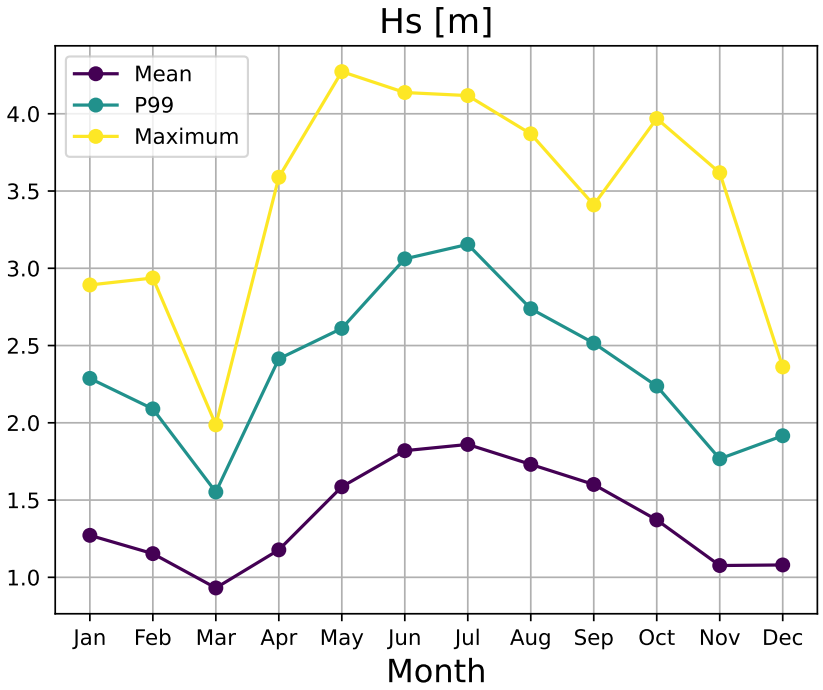

Figure 5.3: Monthly distribution of the mean, 99th percentile (P99), and maximum of the significant wave height (in m).

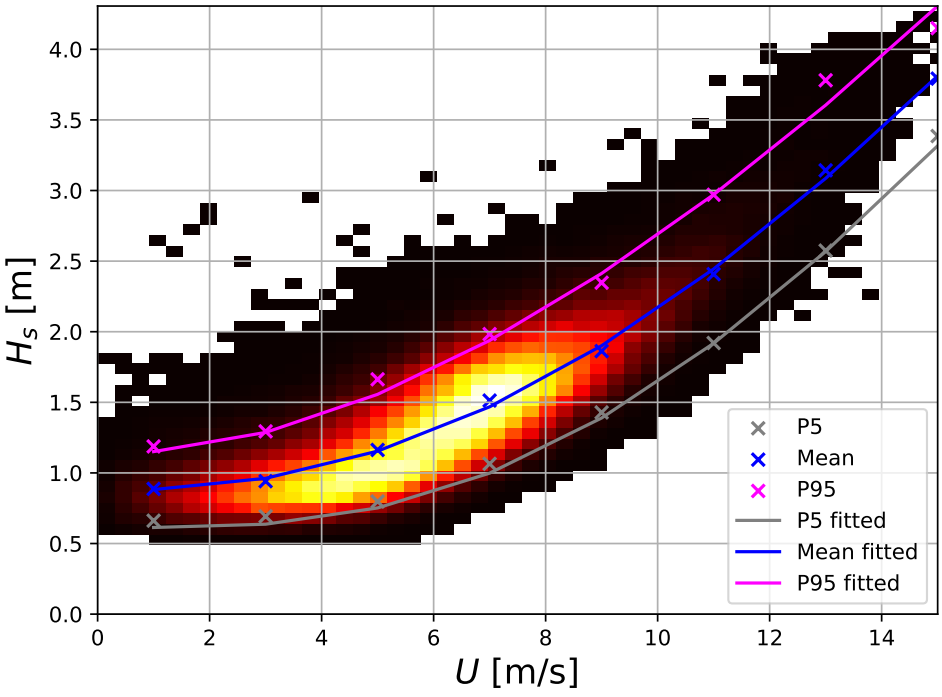

Figure 5.4: Joint distribution of the significant wave height and wind speed at 10 m.

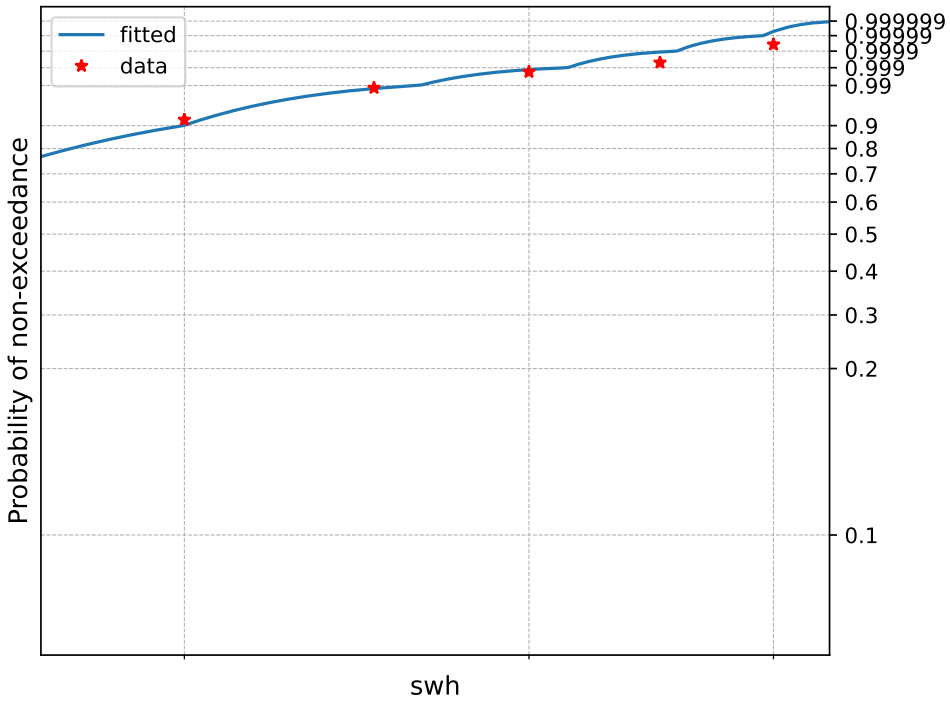

Figure 5.5: Probability of non-exceedance of the significant wave height (swh, x-axis, in m).

5.3 Extreme value analysis

Figure 5.6 shows the monthly return values of the significant wave height for the 1, 10, and 100-year return periods. The values of the monthly and annual return periods of the significant wave height are given in Table 5.2.

Figure 5.7 presents the joint Hs-Tp distribution and the 1, 10, 100, and 1000-year return values in contours. The data are fitted using the LoNoWe model as described in Section 6.4.1.

In general, when considering monthly return value estimates, it is recommended that the reader also consider estimates given for adjacent months. Even though the ERA5 dataset spans 55 years, a single month may be underrepresented in terms of severe storms, simply by chance. Therefore, we advice the reader to consider adjacent months to be conservative. For the directional estimates, there is added an extra conservatism according to NORSOKSTANDARD-N-003 (2017), as explained in Section 6.3.3. Therefore, there is no need to consider adjacent sectors.

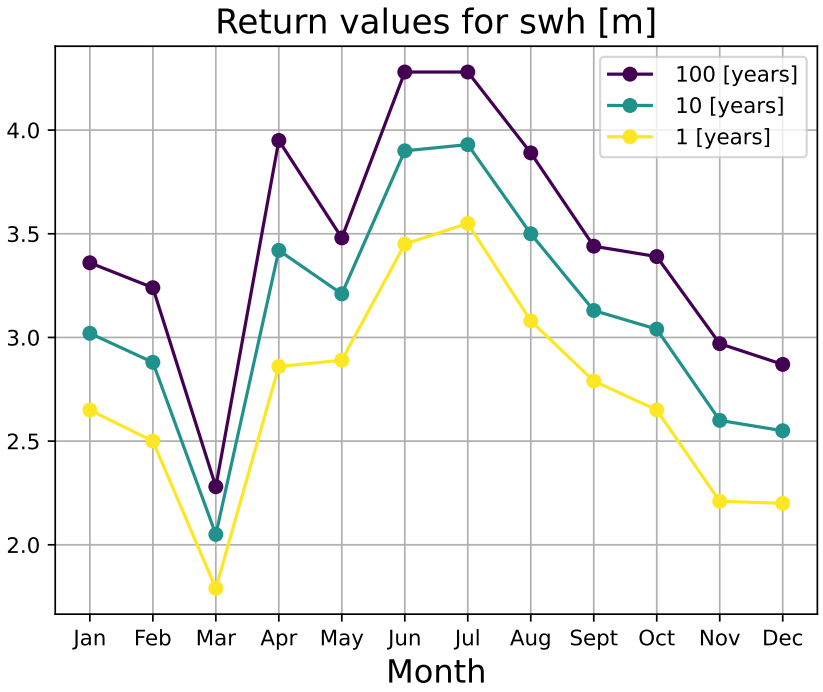

Figure 5.6: Monthly return values of the significant wave height (in m).

Month	Annual prob.	Shape	Scale	Location	Return period: 1 [years]	Return period: 10 [years]	Return period: 100 [years]
-	%	-	m	m	m	m	m
Jan	8.33	1.719	0.653	0.69	2.65	3.02	3.36
Feb	8.33	1.581	0.564	0.65	2.5	2.88	3.24
Mar	8.33	1.497	0.327	0.64	1.79	2.05	2.28
Apr	8.33	1.281	0.489	0.72	2.86	3.42	3.95
May	8.33	2.102	0.832	0.85	2.89	3.21	3.48
Jun	8.33	1.712	0.772	1.13	3.45	3.9	4.28
Jul	8.33	1.368	0.578	1.33	3.55	3.93	4.28
Aug	8.33	1.411	0.465	1.31	3.08	3.5	3.89
Sept	8.33	1.598	0.505	1.15	2.79	3.13	3.44
Oct	8.33	1.469	0.47	0.95	2.65	3.04	3.39
Nov	8.33	1.225	0.304	0.79	2.21	2.6	2.97
Dec	8.33	1.417	0.388	0.73	2.2	2.55	2.87
Year	100.0	1.835	0.884	0.6	3.55	3.93	4.28

Table 5.2: Monthly and annual return values of the significant wave height (in m).

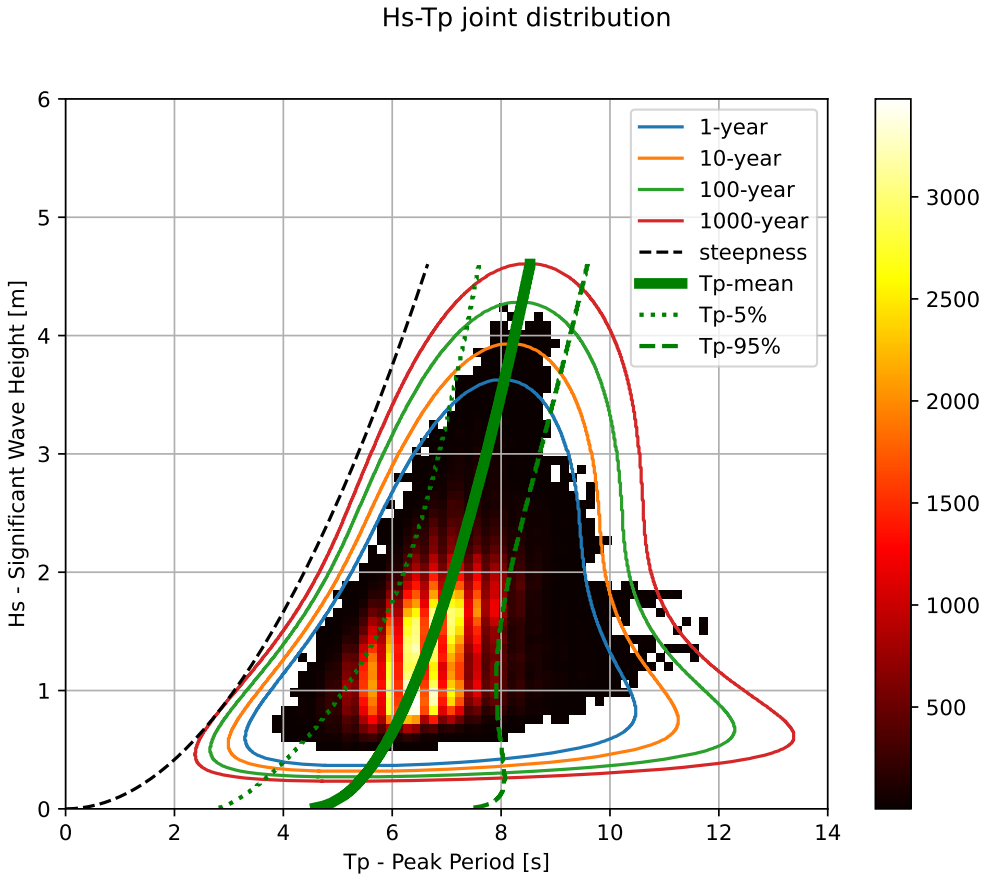

Figure 5.7: Joint distribution of the significant wave height (H_s in m) and mean wave period (T_p in s).

6 Methods

6.1 Performance metrics

Bias (BIAS)

BIAS represents the systematic difference between the time series $X(x_i, i = 1, ..., n)$ and $Y(x_i, i = 1, ..., n)$ where X can be observations and Y modelled data. The bias is the mean of the errors or the difference in the means of X and Y and its unit is the same as X and Y units. The closer to 0, the better, meaning that X and Y are very similar.

$$\text{BIAS} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) = \overline{x} - \overline{y} \tag{6.1}$$

where $\bar{x}$ and $\bar{y}$ are the means of X and Y:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{6.2}$$

Mean Absolute Error (MAE)

MAE is the average of the absolute differences between observations and model predictions. It is less sensitive to outliers than RMSE and is always positive. Its unit is the same as X and Y units. The closer to 0, the better, meaning that X and Y are very similar.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{6.3}$$

Root Mean Square Error (RMSE)

RMSE measures the average magnitude of the squared errors, giving more weight to large errors. Its unit is the same as X and Y units. The closer to 0, the better, meaning that X and Y are very similar.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{6.4}$$

Scatter index (SI)

SI is the RMSE divided by the mean of X. It is unitless.

$$\text{SI} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}}{\overline{x}} \tag{6.5}$$

where $\overline{x}$ is the mean of X as defined in Eq. (6.2). The smaller the SI, the higher the agreement between X and Y.

Correlation Coefficient (CORR)
CORR quantifies the strength and direction of the linear relationship between observations and model predictions. It is unitless and varies between -1 and 1. A correlation of 1 means that X and Y vary in the same direction whereas a correlation of -1 means that X and Y vary in opposite directions. An absolute correlation of 1 also means that X and Y are strongly related whereas a correlation of 0 means that X and Y are very different.

$$\text{CORR} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{6.6}$$

where $\bar{x}$ and $\bar{y}$ are the means of X and Y defined in Eq. (6.2).

Taylor diagram
This diagram was imagined by Taylor (2001) to compare two datasets in a concise way by representing the correlation, the root-mean-square error, and the standard deviation of the two datasets on the same figure. One dataset is usually treated as the reference and the second dataset is compared to the reference dataset. Usually, the closer the compared point to the reference point, the better, meaning that the two datasets are similar.

6.2 Temporal trends in time series

As the time series used in this report cover a long period of time, trends in the data can be expected and must be acknowledged. For example, the surface air temperature is expected to increase with time. Concerning the wind, trends are shown to be quite uncertain (Lussana et al., 2022). Several methods exist to find out trends in time series. Here, we describe two methods: the linear regression and the Theil-Sen estimator.

6.2.1 Least-squares linear regression

The least-squares linear regression method draws a line by minimizing the sum of the squared distance between the estimated line and each point of the time series. The coefficient of determination R^2 gives a measure of the robustness of the trend. It varies between 0 and 1 and the closer to 1 the better.

6.2.2 Theil-Sen estimator

The Theil-Sen estimator (Theil, 1950; Sen, 1968) uses every possible pair of points within the time series to create a set of possible lines representing the trend. Among all lines' slopes, the linear trend is defined as the line with the median slope. A confidence interval can be obtained by for example looking at the 2.5th and 97.5th percentiles of the slopes for a 95% confidence interval. If both slopes have the same sign, it is likely that the trend is robust. Another measure of trend significance is the Kendall's tau method, which varies between -1 and 1, with -1 meaning that there is a perfectly decreasing relationship, 1 a perfectly increasing relationship, and 0 that there is no relationship. The Theil-Sen estimator is less sensitive to outliers than the least-squares regression (Wilks, 2019).

6.3 Extreme values analysis of univariate time series

When looking at the distribution of the values of one variable, the least frequent events or in other words the extreme events are found in the tails of the distribution (see Fig. 6.1). Most of the time, we are interested in the extremely high values, for example for heatwaves (high temperature), windstorms (strong wind), floods (heavy rain), storm surges (high water level/waves), among others.

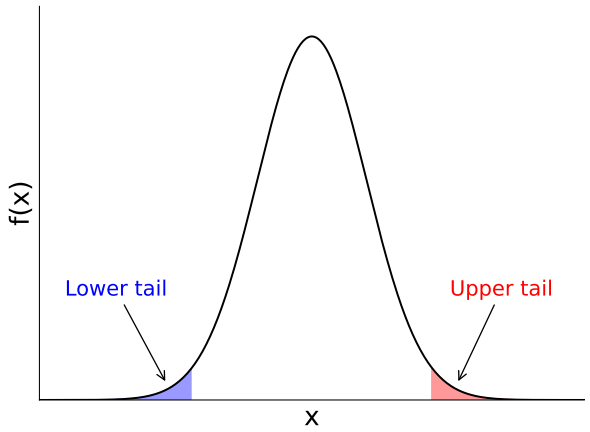

Figure 6.1: Example of a distribution and its tails. In metocean reports, we are most of the time interested in the upper tail.

The *Extremal Types Theorem* states that the highest values of a sample (the upper tail) taken from any distribution will follow a known distribution (Coles, 2001; Wilks, 2019). Depending on the definition of the extremes, the known distribution differs.

The extreme values are distributed following a *probability density function* f (see Fig. 6.1). Another important function is the *cumulative distribution function* F, which gives the probability of not exceeding a certain value, or its inverse named the *inverse cumulative distribution function* F^{-1}. The return values estimates (or return levels) are obtained by inverting the cumulative distribution function $F(x)$: $z_p = F^{-1}(x)$ where p is the probability of a return event within a given year.

6.3.1 Block maxima and Generalized extreme values

A method to define the extreme events is the Block-Maxima (BM) method where one picks the maximum value within a certain time interval (block). Most commonly, a block covers one year. Therefore, all yearly maxima will be retained. This means that if the dataset covers 20 years, only 20 values will be used for further analysis. A strong shortcoming of the BM method is that it only keeps the largest extreme for each year even if a second extreme occurred during the same year that was larger than extremes of other years.

The *Fisher–Tippett–Gnedenko theorem* states that extremes from block maxima are distributed along one distribution of the *Generalized Extreme Value* (GEV) distributions family.

The most general form of the probability density function of the GEV distribution is

$$f(x) = \frac{1}{\beta}\left[1 + \frac{\kappa(x-\zeta)}{\beta}\right]^{1-\frac{1}{\kappa}} \exp\left\{-\left[1 + \frac{\kappa(x-\zeta)}{\beta}\right]^{-\frac{1}{\kappa}}\right\}, \tag{6.7}$$

with $1 + \frac{\kappa(x-\zeta)}{\beta} > 0$. x represents the selected extreme values, β is the scale parameter, κ the shape parameter, and ζ the location parameter.

There are three main distributions in the GEV family: the Gumbel, the Fréchet, and the Weibull distributions, which are described below.

- GEV Type I: the Gumbel distribution

When the shape parameter $\kappa \to 0$, Eq. (6.7) becomes

$$f(x) = \frac{1}{\beta}\exp\left\{-\exp\left[-\frac{(x-\zeta)}{\beta}\right] - \frac{(x-\zeta)}{\beta}\right\}, \tag{6.8}$$

with $-\infty < x < +\infty$. The Gumbel distribution reaches its maximum at $x = \zeta$. The cumulative distribution function is

$$F(x) = \exp\left\{-\exp\left[-\frac{(x-\zeta)}{\beta}\right]\right\}. \tag{6.9}$$

The return level is given by

$$z_p = \zeta - \beta \ln\left(\ln(1-p)\right). \tag{6.10}$$

- GEV Type II: the Fréchet distribution

For $\kappa > 0$, Eq. (6.7) is called the Fréchet distribution or inverse Weibull distribution.

$$f(x) = \frac{1}{\kappa\beta}\left(\frac{x-\zeta}{\beta}\right)^{-\left(1+\frac{1}{\kappa}\right)} \exp\left[-\left(\frac{x-\zeta}{\beta}\right)^{-\frac{1}{\kappa}}\right]. \tag{6.11}$$

where the location parameter ζ is the minimum of the extreme values selected for the analysis (x_{min}). The cumulative distribution function takes the form:

$$F(x) = \exp\left[-\left(\frac{x-\zeta}{\beta}\right)^{-\frac{1}{\kappa}}\right]. \tag{6.12}$$

The return level is given by

$$z_p = \zeta + \beta\left[\ln\left(\frac{1}{1-p}\right)\right]^{-\kappa}. \tag{6.13}$$

This distribution will not be used in this report as it is mostly applicable to hydrology (Thompson and Emery, 2024).

- GEV Type III: the Weibull distribution

The 3-parameter Weibull distribution is

$$f(x) = \begin{cases} \frac{\kappa}{\lambda}\left(\frac{x-\zeta}{\lambda}\right)^{\kappa-1} \exp\left\{-\left(\frac{x-\zeta}{\lambda}\right)^{\kappa}\right\} & \text{for } x \geq \zeta \\ 0 & \text{for } x < \zeta, \end{cases} \tag{6.14}$$

where κ is the shape parameter (> 0), λ the scale parameter (> 0), and ζ the location parameter. Note that $f(x) \geq 0$. The cumulative distribution function of the 3-parameter Weibull distribution is

$$F(x) = \begin{cases} 1 - \exp\left[-\left(\frac{x-\zeta}{\lambda}\right)^{\kappa}\right] & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases} \tag{6.15}$$

The return level is

$$z_p = \zeta + \lambda\left(-\ln p\right)^{\frac{1}{\kappa}}. \tag{6.16}$$

When the location parameter $\zeta = 0$, Eq. (6.14) reduces to the 2-parameter Weibull distribution:

$$f(x) = \begin{cases} \frac{\kappa}{\lambda}\left(\frac{x}{\lambda}\right)^{\kappa-1} \exp\left\{-\left(\frac{x}{\lambda}\right)^{\kappa}\right\} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0, \end{cases} \tag{6.17}$$

where λ is the scale parameter and κ the shape parameter. The associated cumulative distribution function is

$$F(x) = \begin{cases} 1 - \exp\left[-\left(\frac{x}{\lambda}\right)^{\kappa}\right] & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases} \tag{6.18}$$

The return level is

$$z_p = \lambda\left(-\ln p\right)^{\frac{1}{\kappa}}. \tag{6.19}$$

6.3.2 Peak Over Thresholds and Generalized Pareto distributions

The Peak-Over-Threshold (POT) method defines the extremes as all values above a certain threshold and separated by a certain period of time to ensure that the extremes are independent from each other. A shortcoming of the POT method is that the choice of the threshold is relatively subjective and the results will be sensitive to the threshold chosen. Moreover, if there is a strong trend in the data, for example temperatures strongly increasing with time, the POT method will only select values in the last period of the data. However, an advantage of the POT method over the BM method described in the previous section is that it usually selects more extreme values, rendering the fit of the distribution more robust. Various methods can be employed to find the most appropriate threshold:

- Coles (2001) suggests to use a range of thresholds and choose the threshold just before which both the scale and shape parameters start to change (see Fig. 4.2 in Coles, 2001),
- when using daily data, u should be chosen such that the number of extreme values is equal to $1.65\, n\_years$ where n_years is the number of years in the time series (see Coles, 2001, and references therein),

- Outten and Sobolowski (2021) choose the minimum of the yearly maxima (BM) as threshold.

From the *Pickands-Balkema-De Haan theorem*, the extremes selected with the POT method are supposed to follow one of the distributions of the Generalized Pareto (GP) family. The type I has the following probability density function

$$f(x) = \frac{1}{\sigma}\left[1 + \frac{\kappa(x-u)}{\sigma}\right]^{-\frac{1}{\kappa}-1}, \tag{6.20}$$

where u represents the threshold above which the extreme values are selected (or the location parameter), κ the shape parameter, and σ the scale parameter. Differently from the other distributions, due to the presence of the threshold u, the GP distribution model the threshold exceedances $(x-u)$ and not the extreme values (x).
The cumulative distribution function is

$$F(x) = 1 - \left[1 + \frac{\kappa(x-u)}{\sigma}\right]^{-\frac{1}{\kappa}} \tag{6.21}$$

and the return value is given by

$$z_p = u + \frac{\sigma}{\kappa}\left(p^{-\kappa} - 1\right). \tag{6.22}$$

When the shape parameter $\kappa \to 0$, Eq. (6.20) reduces to the probability density function of the exponential distribution

$$f(x) = \frac{1}{\lambda}\exp\left(-\frac{x}{\lambda}\right), \tag{6.23}$$

where $\lambda > 0$ is the scale parameter. The associated cumulative distribution function is

$$F(x) = 1 - \exp\left(-\frac{x}{\lambda}\right) \tag{6.24}$$

and the return level is

$$z_p = -\lambda \ln(p). \tag{6.25}$$

6.3.3 NORSOK adjustment

When calculating the return levels for different directional sectors, the NORSOK adjustment consists to increase the return periods by the factor $N/2$, where N is the number of different sectors, meaning that a 100-yr return period will correspond to a $100N$-yr return period. However, the resulting directional return level must be reduced to the return level obtained for all wind directions at once if it exceeds it.

6.3.4 Fitting techniques

The probability density functions or cumulative distribution functions presented above have various parameters whose values depend on the selected extremes. To obtain these parameters, we have to fit the distribution to the data. There exists various methods to perform the fitting. The choice of the method depends on both the distribution to fit and the amount of data available. We here below present three different methods.

Method of moments
The distribution parameters are estimated from the statistical moments of the data sample, such as mean, variance, skewness, among others. Equating the analytical formulas of the moments to their empirical values yields the estimation of the parameters. This method usually provides a good fit of the mode of the distribution, but not always of the extremes.

Maximum likelihood estimation
The maximum likelihood method finds the values of the parameters that maximize the likelihood function L or rather its natural logarithm l, which is the function of the parameter θ knowing x

$$l(\theta|x) = \ln L(\theta|x) = \ln f_\theta(x), \tag{6.26}$$

θ being either the scale, the shape, or the location parameter. To maximize l for every parameter, the derivative of l relative to the parameter considered is calculated and set equal to 0. Solving the resulting equation yields the most probable value of the parameter. The maximizing is solved iteratively with numerical techniques.

6.3.5 Return periods

Modeled - from the distributions
The cumulative distribution function $F(x)$ represents the probability of non-exceedance $F(x) = 1 - p$ of the value x, with p being the probability of exceedance of x $(= 1 - F(x))$. The inverse of p corresponds to the return period of x. More precisely, the averaged return period $R(x)$ of a value x is

$$R(x) = \frac{1}{\omega\,[1 - F(x)]} = \frac{1}{\omega\,p} \tag{6.27}$$

where ω is the average sampling frequency. If BM is used to obtain the extremes, $\omega = 1$ because there is one event per year. For the POT method, ω is the expected rate of occurrence or the mean number of occurrences per year. In this case, the duration of an extreme event is considered to be the time resolution of the data. When the initial data is used, one must take into account the extreme event duration in the return value estimates. Therefore, $\omega = \frac{PT}{\tau}$ where τ is the extreme event duration (in hours), T is the number of hours in one year $(= 365.25 * 24)$, and P is the 'sector' probability $(= 1/12$ for monthly return values and 1 for annual return values).

For stationary time series, the probability p of one event occurring within a time period T exceeding the return period R is determined by the Poisson distribution (see e.g., Thompson and Emery, 2024) as follows:

$$p = 1 - \exp\left(-\frac{T}{R}\right). \tag{6.28}$$

The probability of a 1-yr return period event $(R = 1)$ within a given year $(T = 1)$ is thus $p \simeq 0.632$. The probability is the same for a 10-yr return event within a 10-yr period $(R = T = 10)$. However, the probability of a 5-yr return event $(R = 5)$ within a given year $(T = 1)$ is $p \simeq 0.181 \sim \frac{1}{5}$. Similarly, the probability of a 10-yr return event $(R = 10)$ within a given year $(T = 1)$ is $p = 0.095 \simeq \frac{1}{10}$ and the probability of a 100-yr return event is $p = 0.00995 \simeq \frac{1}{100}$.

Therefore, the 1, 5, 10, 100, 1000 and 10000-year return value estimates correspond to p equal to 0.63, 5^{-1}, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, respectively.

Empirical - from the data

The probability of non-exceedance of each value of the dataset can be calculated as follows

$$p_{emp} = \frac{r - \alpha}{n + 1 - \alpha - \beta} \tag{6.29}$$

where r represent the ranks of the selected extreme values x, n is the number of extreme values. α and β are parameters for the Weibull plotting position and are here set to 0 following Makkonen (2006). Therefore, the empirical probability of non-exceedance reduces to

$$p_{emp} = \frac{r}{n + 1} \tag{6.30}$$

Note that p_{emp} does not depend on the data distribution $f(x)$. Ideally, the empirical distribution function lines up perfectly with the theoretical distribution.

The empirical return period is the inverse of the probability of exceedance

$$R_{emp} = \frac{1}{1 - p_{emp}} = \frac{n + 1}{n - r + 1} \tag{6.31}$$

6.3.6 Inclusion of climate change

In the extreme value analysis, we consider that the climate is stationary. This may not be true for variables such as air temperature and locations where sea ice used to be, meaning that return periods are overestimated and the return values are underestimated if the trend (or non-stationarity of the climate) is not taken into account. For variables such as wind speed and wave height, the annual variability is so high that any trend is usually not very significant. One method to include the non-stationary features would be to make the location parameter vary linearly with time and keep the scale and shape parameters constant (Cheng et al., 2014).

6.4 Extreme values analysis of multivariate time series

When at least two variables are considered together, a joint model is built using all original data and not only the extremes. The conditional modeling approach is a method for creating a multivariable probability distribution by combining marginal and conditional probability distributions. For example, the joint distribution of two variables Var_1 and Var_2, named $f_{jd}(Var_1, Var_2)$, is given by the product of the marginal distribution of Var_1, named $f(Var_1)$, and the conditional distribution of Var_2 given Var_1 ($g(Var_2|Var_1)$) as follows

$$f_{jd}(Var_1, Var_2) = f(Var_1)\, g(Var_2|Var_1). \tag{6.32}$$

Practically, this is done by dividing the first variable into fixed-width intervals. The conditional distribution is then fitted to the data within each of these intervals producing a set of conditional distribution parameters corresponding to each interval. To obtain a smooth joint distribution, *dependency functions* are fitted to the parameters found for each interval.

In general, the parameters of the marginal distributions are obtained using either the maximum likelihood or the method of moments. The parameters of the joint distributions are obtained using the method of moments. The dependency functions are fitted using the maximum likelihood.

6.4.1 The LoNoWe model for the joint distribution of significant wave height and wave period

The LogNormal-Weibull model (or LoNoWe-model) is used to represent the long-term joint distribution of the significant wave height (H_s) and the peak wave period (T_p). Most of this analysis follows the work by Moan et al. (2005).

The LoNoWe joint density function is expressed by

$$f_{jd}(H_s, T_p) = f(H_s)\, g(T_p|H_s) \tag{6.33}$$

where the marginal distribution of H_s is given by a LogNormal distribution for the lowest part of the distribution

$$f(H_s) = \frac{1}{\sqrt{2\pi}\sigma h} \exp\left[-\frac{(\ln h - \mu)^2}{2\sigma^2}\right] \quad \text{for} \quad h \leq \eta, \tag{6.34}$$

where μ and σ are the mean and standard deviation of lnH_s, and by a 2-parameter Weibull distribution for the high-end of the distribution

$$f(Hs) = \beta \frac{H_s^{\beta-1}}{\alpha^\beta} \exp\left[-\left(\frac{H_s}{\alpha}\right)^\beta\right] \quad \text{for} \quad h > \eta \tag{6.35}$$

where α and β are the scale and shape parameters of the Weibull distribution, respectively. η represents the shifting point between the two distributions, which is selected using a χ^2 test, fulfilling the principal of continuity between the two distributions, both for the probability density functions and the cumulative distribution functions at the shifting point.

The distribution of the peak period T_p conditional on the significant wave height H_s is assumed to follow a pure LogNormal distribution:

$$f(T_p|H_s) = \frac{1}{\sqrt{2\pi}\sigma T_p} \exp\left[-\frac{(\ln T_p - \mu)^2}{2\sigma^2}\right] \tag{6.36}$$

where

$$\mu = a_1 + a_2 H_s^{a_3} \quad \text{with} \quad a_3 \leq 1 \tag{6.37}$$

and

$$\sigma^2 = b_1 + b_2\, e^{-b_3 H_s} \tag{6.38}$$

represent the mean and variance of $\ln T_p$, respectively, which are functions of H_s. The coefficients a_1, a_2, a_3, b_1, b_2, and b_3 are obtained by curve fitting Eqs. (6.37) and (6.38) to the mean and variance of $\ln T_p$ within each H_s-bin.

6.4.2 Joint distribution of significant wave height and wave period

As in Eq. (6.32), the joint distribution of H_s and T_p is given by the product of the marginal distribution of H_s (Eq. (6.40)) and the conditional distribution of T_p given H_s (Eq. (6.41)):

$$f_{jd}(H_s, T_p) = f(H_s)\, g(T_p|H_s). \tag{6.39}$$

Following the recommendation by DNV GL (2017), the joint H_S/T_p distribution consists of a marginal 3-parameter Weibull distribution of the significant wave height H_s:

$$f(H_s) = \frac{\kappa}{\lambda}\left(\frac{H_s-\zeta}{\lambda}\right)^{\kappa-1}\exp\left\{-\left(\frac{H_s-\zeta}{\lambda}\right)^{\kappa}\right\}, \tag{6.40}$$

where κ is the shape parameter, λ the scale parameter, and ζ the location parameter (see Eq. (6.14)), and of a conditional log-normal distribution of the wave period T_p:

$$g(T_p|H_s) = \frac{1}{\sqrt{2\pi}\,T_p\,\sigma_{T_p|H_s}}\exp\left\{-\frac{(\ln T_p-\mu_{T_p|H_s})^2}{2\,\sigma^2_{T_p|H_s}}\right\} \tag{6.41}$$

where the parameters $\mu_{T_p|H_s}$ and $\sigma_{T_p|H_s}$ depend on H_s and are defined (with dependency functions) as follows

$$\mu_{T_p|H_s} = a_1 + a_2\,H_s^{a_3} \tag{6.42}$$

$$\sigma_{T_p|H_s} = b_1 + b_2\,e^{b_3 H_s} \tag{6.43}$$

In practice, the probability density functions (Eqs. (6.40) and (6.41)) are fitted using the method of moments (see Sect. 6.3.4) for each H_s interval while the three a_i and b_i $(i = 1, 2, 3)$ parameters in Eqs. (6.42) and (6.43) are obtained using the maximum likelihood estimation (see Sect. 6.3.4).

6.4.3 Joint distribution of wind speed, significant wave height, and peak wave period

This model is based on the joint model for wind and wave described in Li et al. (2015). It consists of a 2-parameter Weibull distribution for wind speed U, a 2-parameter Weibull distribution for the significant wave height H_s dependent on U, and a log-normal distribution for the peak wave period T_p dependent on H_s, such that:

$$f_{jd}(U, H_s, T_p) = f(U)\,g(H_s|U)\,h(T_p|H_s), \tag{6.44}$$

with

$$f(U) = \frac{\kappa}{\lambda}\left(\frac{U}{\lambda}\right)^{\kappa-1}\exp\left\{-\left(\frac{U}{\lambda}\right)^{\kappa}\right\}, \tag{6.45}$$

$$g(H_s|U) = \frac{\kappa_{H_s|U}}{\lambda_{H_s|U}}\left(\frac{U}{\lambda_{H_s|U}}\right)^{\kappa_{H_s|U}-1}\exp\left\{-\left(\frac{U}{\lambda_{H_s|U}}\right)^{\kappa_{H_s|U}}\right\}, \tag{6.46}$$

and

$$h(T_p|H_s) = \frac{1}{\sqrt{2\pi}\,T_p\,\sigma_{T_p|H_s}}\exp\left\{-\frac{(\ln T_p-\mu_{T_p|H_s})^2}{2\,\sigma^2_{T_p|H_s}}\right\}. \tag{6.47}$$

The scale ($\lambda_{H_s|U}$) and shape ($\kappa_{H_s|U}$) parameters for the Weibull distribution of H_s conditional on U are modeled using power laws (meaning that they depend on U to the power of some unknown number) as follows

$$\lambda_{H_s|U} = c_1 + c_2 U^{c_3} \tag{6.48}$$

$$\kappa_{H_s|U} = d_1 + d_2 U^{d_3}, \tag{6.49}$$

while the parameters of the log-normal distribution of T_p conditional on H_s are as follows

$$\mu_{T_p|H_s} = e_1 + e_2 {H_s}^{e_3} \tag{6.50}$$

$$\sigma_{T_p|H_s} = f_1 + f_2 \exp\left(f_3 H_s\right) \tag{6.51}$$

The parameters of the probability density functions (Eqs. (6.45)-(6.47)) are estimated using the method of moments while the parameters of the dependency functions (Eqs. (6.48-6.51)) are obtained using the maximum likelihood method.

6.4.4 Joint distribution of significant wave height and wind speed

Following DNV GL (2017), the joint distribution of significant wave height H_s and wind speed U is given by

$$f_{jd}(H_s, U) = f(H_s)\, g(U|H_s) \tag{6.52}$$

where $f(H_s)$ is a 3-parameter Weibull distribution given by Eq. (6.40) and the conditional distribution of the wind speed U conditional on H_s follows a 2-parameter Weibull distribution (Eq. 6.17) such as

$$f(U|H_s) = \kappa \frac{U^{\kappa_{U|H_s}-1}}{\lambda_{U|H_s}^{\kappa_{U|H_s}}} \exp\left\{-\left(\frac{U}{\lambda_{U|H_s}}\right)^{\kappa_{U|H_s}}\right\}, \tag{6.53}$$

where $\kappa_{U|H_s}$ is the shape parameter, $\lambda_{U|H_s}$ is the scale parameter and u the wind speed. The dependency functions of the two parameters are power laws of H_s as follows

$$\kappa_{U|H_s} = c_1 + c_2 H_s^{c_3}, \tag{6.54}$$

$$\lambda_{U|H_s} = c_4 + c_5 H_s^{c_6}. \tag{6.55}$$

6.4.5 Obtaining environmental contours

Environmental contours are combinations of environmental variables, which represent the most extreme conditions expected within the given return period. There are two common methods of obtaining the contour as described in (DNV GL, 2017, 3.7.2).

IFORM The Inverse First-Order Reliability Method (IFORM), introduced by Winterstein et al. (1993), works by transforming the joint distribution of environmental variables into a standard normal space. In this space, a circle (or sphere for 3D joint distributions) is drawn, with its radius corresponding to a specified exceedance probability. The points are then transformed back to the physical space to form the contour. All points on the resulting contour therefore have a constant probability of being exceeded.

Iso-density The iso-density approach constructs a contour defined by a constant value of the joint probability density function of the model. It is created by selecting a marginal return value (e.g., estimated 100-year significant wave height H_s), finding the most probable conditional values of the other variables, and obtaining the corresponding joint probability density function value. This value is used to create the contour. While all points on the contour are equally likely to occur, their probability of being exceeded varies. This method is mainly used to obtain contours that correspond to specific marginal return values, obtained by any method described in section 6.3.

6.4.6 Steepness criterion

In the H_s/T_p joint distribution, we include the steepness criterion as defined by DNV GL (2017):

$$S_p = \frac{2\pi}{g}\frac{H_s}{T_p^2} \tag{6.56}$$

depending on the significant wave height H_s and the peak wave period T_p. The line actually displayed follows the equation:

$$H_s = \frac{S_p\, g}{2\pi} T_p^2, \tag{6.57}$$

where $S_p = 1/15$ for $T_p \leq 8$ s and $S_p = 1/25$ for $T_p \geq 15$ s. Between $T_p = 8$ and $T_p = 15$ s, the line is linearly interpolated. g is the gravity of Earth ($\simeq 9.81$ m s^{-2}).

Although this criterion does not represent a physical limit, it provides a lower limit for the peak wave period (T_p) for a given significant wave height (H_s) where H_s/T_p combinations become unrealistic. Therefore, the environmental contours located left of this line should be considered as artificial.

6.5 Reproducibility of the results

Note that almost all calculations are performed using the functions from the *metocean-stats* Python module (`https://github.com/MET-OM/metocean-stats`), which builds upon *pyextremes* for the marginal return value estimates (Bocharov, 2023), and *virocon* for the joint return value estimates (Haselsteiner et al., 2019). Most of the data are obtained using the *metocean-api* Python module (`https://github.com/MET-OM/metocean-api`). Some data are not publicly available, but most of them can be found at `https://thredds.met.no` or at `https://cds.climate.copernicus.eu/`.

References

Bocharov, G., 2023: pyextremes: Extreme Value Analysis (EVA) in Python. https://github.com/georgebv/pyextremes.

Cheng, L., A. AghaKouchak, E. Gilleland, and R. W. Katz, 2014: Non-stationary extreme value analysis in a changing climate. *Climatic Change*, **127**, 353–369, https://doi.org/10.1007/s10584-014-1254-5.

Coles, S., 2001: *An introduction to statistical modeling of extreme values*. Springer series in statistics, Springer, London, https://doi.org/10.1007/978-1-4471-3675-0.

DNV GL, 2017: DNVGL-RP-C205 Environmental conditions and environmental loads, Recommended practice. *Det Norske Veritas GL AS*.

Haselsteiner, A. F., J. Lehmkuhl, T. Pape, K.-L. Windmeier, and K.-D. Thoben, 2019: ViroCon: A software to compute multivariate extremes using the environmental contour method. *SoftwareX*, **9**, 95–101, https://doi.org/10.1016/j.softx.2019.01.003.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Li, L., Z. Gao, and T. Moan, 2015: Joint distribution of environmental condition at five European offshore sites for design of combined wind and wave energy devices. *Journal of Offshore Mechanics and Arctic Engineering*, **137**, 031 901, https://doi.org/10.1115/1.4029842.

Lussana, C., I. Hanssen-Bauer, J. E. Haugen, A. Dobler, H. O. Hygen, A. V. Dyrrdal, and H. Haakenstad, 2022: Notes on 1991–2020 wind speed climatology based on NORA3 near-surface data. Tech. rep., Norwegian Meteorological Institute.

Makkonen, L., 2006: Plotting positions in extreme value analysis. *Journal of Applied Meteorology and Climatology*, **45 (2)**, 334–340, https://doi.org/10.1175/JAM2349.1.

Moan, T., Z. Gao, and E. Ayala-Uraga, 2005: Uncertainty of wave-induced response of marine structures due to long-term variation of extratropical wave conditions. *Marine Structures*, **18 (4)**, 359–382, https://doi.org/10.1016/j.marstruc.2005.11.001.

NORSOKSTANDARD-N-003, 2017: Actions and action effects. Technical report, NORSOK STANDARD.

Outten, S., and S. Sobolowski, 2021: Extreme wind projections over Europe from the Euro-CORDEX regional climate models. *Weather and Climate Extremes*, **33**, 100 363, https://doi.org/10.1016/j.wace.2021.100363.

Sen, P. K., 1968: Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, **63**, 1379–1389, https://doi.org/10.1080/01621459.1968.10480934.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, **106**, 7183–7192, https://doi.org/10.1029/2000JD900719.

Theil, H., 1950: A rank-invariant method of linear and polynomial regression analysis I, II, and III. *Nederl. Akad. Wetensch. Proc.*, **53**, 386–392, 521–525, and 1397–1412.

Thompson, R. E., and W. J. Emery, 2024: *Data Analysis Methods in Physical Oceanography*, Vol. Fourth Edition. Elsevier.

Wilks, D. S., 2019: *Statistical Methods in the Atmospheric Sciences, Fourth Edition*. Elsevier.

Winterstein, S. R., T. C. Ude, C. A. Cornell, P. Bjerager, and S. Haver, 1993: Environmental parameters for extreme response: Inverse FORM with omission factors. *Proceedings ICOSSAR-93, Innsbruck, Austria*, https://doi.org/10.1029/2011JC007465.