Fundamentals and examples of **Marine Data Analysis**

Instrumentation and sampling

Time series analysis

Statistical methods

Spatial analysis of data fields

Climatological assessments

Carlos Román Cascón

University of Cádiz

carlos.roman@uca.es

Marine Data Literacy Course

European University of the Seas

7th November 2023

Outline

1. Instrumentation and sampling characteristics.

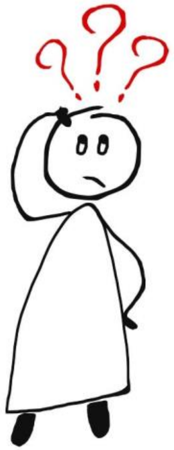
2. Data analysis. How to deal with data…
   - Basic statistics, climatology, anomalies, correlation…

3. Spatial analysis of data.

4. Time series analysis.
   - Fourier analysis, Harmonic analysis, Spectral analysis, Wavelet…

5. Some examples!

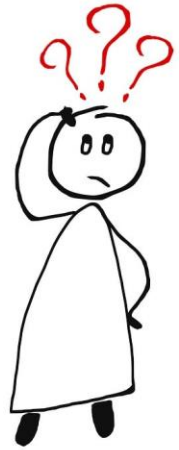
WHY DATA ANALYSIS?



WHY DATA ANALYSIS?

We have our hypothesis and we want to check it to obtain **scientific conclusions.**

We need to design the experiment, to obtain data, and to **analyse** them.



Typical Scientific Procedure (of course, after your hypothesis & experimental design!)

READ → REPRESENT → PROCESS → CALCULATE → THINK → ADVANCED STUDIES → **SCIENTIFIC CONCLUSIONS**

0. **Read** the data → Fight against the files you receive!

1. Just **Represent** the variable(s) you are interested in → see their evolution in time, their spatial distribution (horizontal, vertical), etc.

2. **Processing** → edit the data in files to correct errors, transform the data, quality control, etc.

3. **Calculate** basic statistics → mean, standard deviation, anomalies (if we know the climatology), etc.

4. **Think** → Correlation with other variables? Try to understand the physical processes.

5. **Advanced** studies → Spectral analysis (Fourier), Harmonic analysis, etc.

6. Obtain **scientific conclusions.**

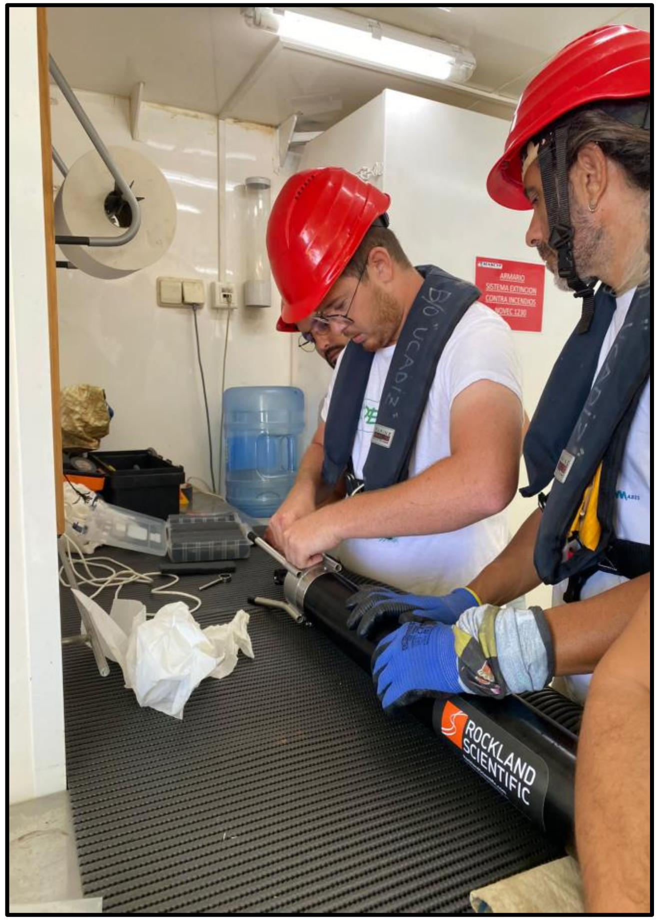
1. Instrumentation and sampling characteristics

1. Instrumentation and sampling characteristics

BEFORE THE DATA ANALYSIS, we need to know details about the instrumentation and the experiment.

Instrument calibration

(can be done several times, for example before each field campaign)



Instrument factory…

Comparison of two devices in the same place…

Always NOTE type of calibration and date!

Absolute accuracy

(related to systematic errors)

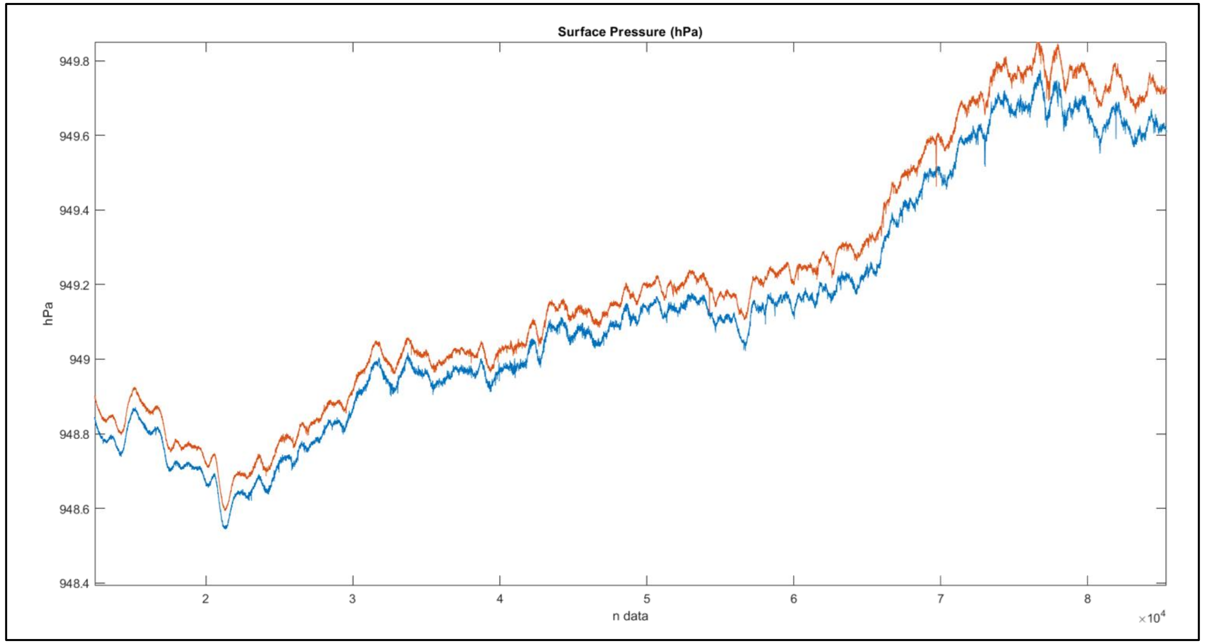
Shifts in records (calibration needed). Deviation from TRUE value (**bias**).

Oceanographical example

2 scientific groups are measuring the same (nutrients in some area).

Their obtained values are different (mean) due to differences in the methods/instrumentation used.

These are systematic errors. To check which one is more accurate we need reference measurements.

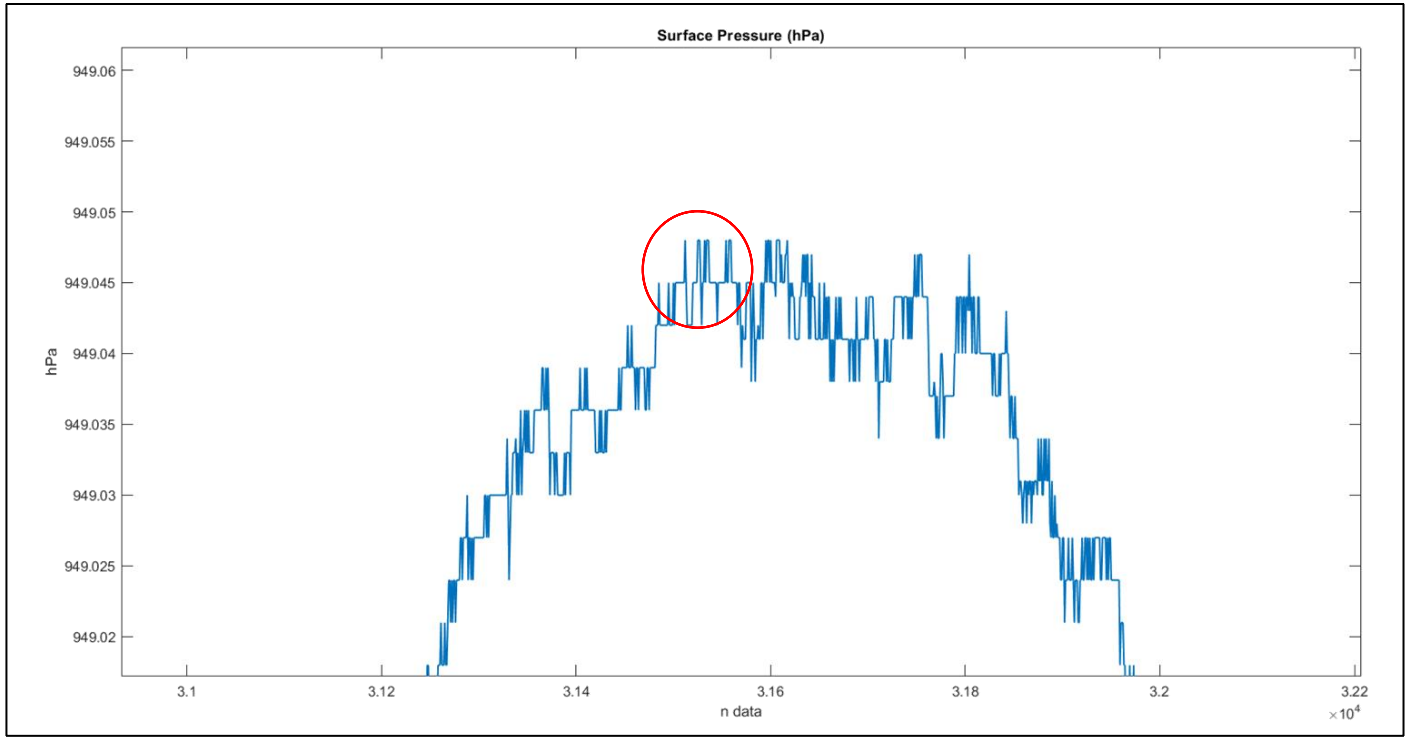


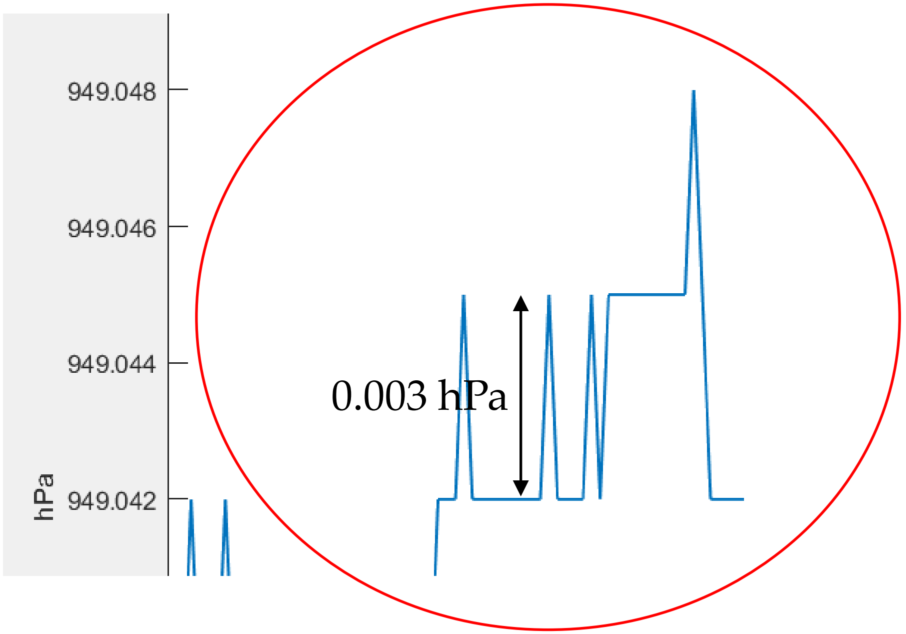
Precision

(related to random errors)

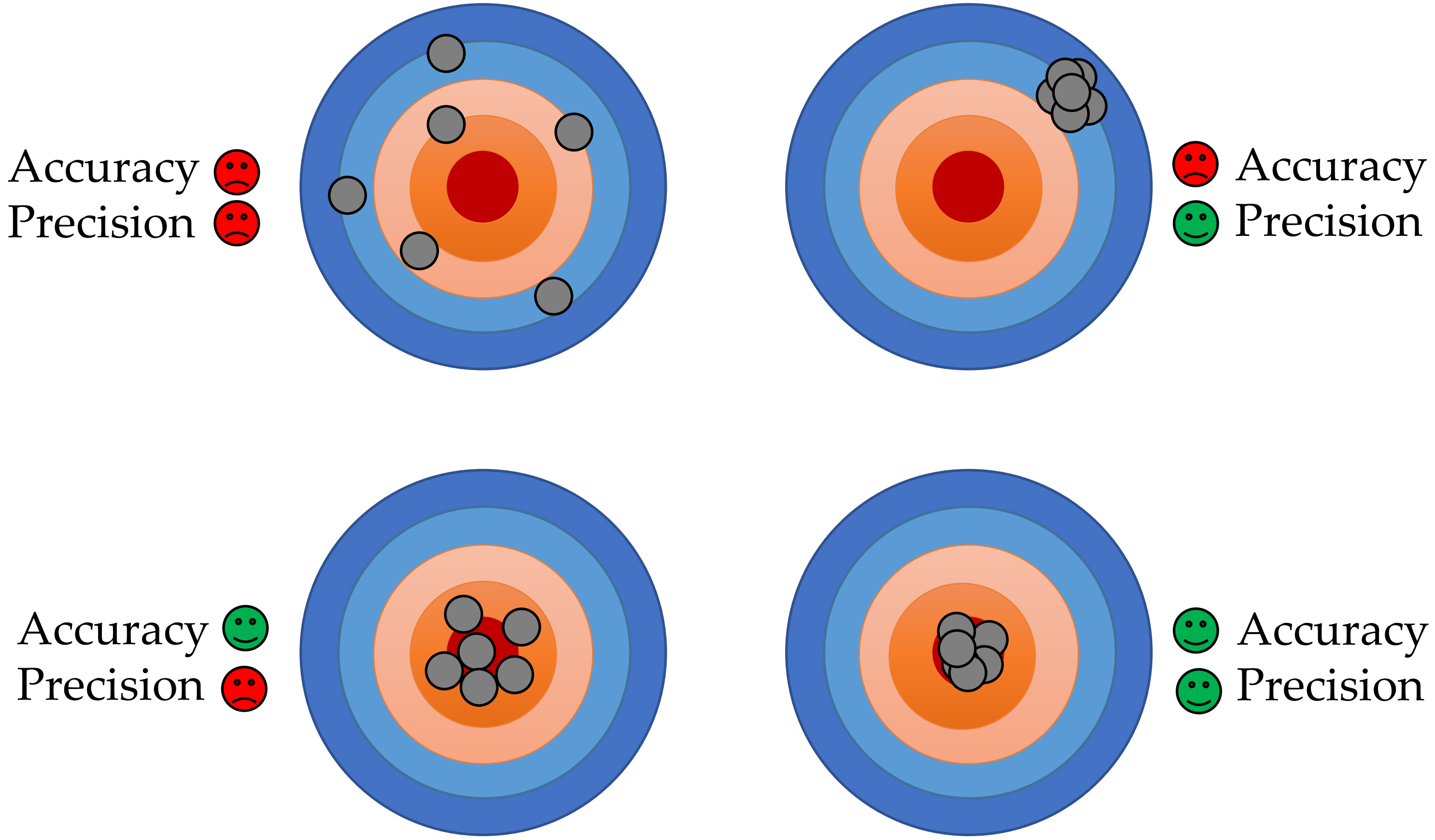
Ability to repeat the observation without deviation. Deviation of one measurement from another.

Fluctuations in the measurement due to limitations in the precision of the instruments
It is quantified through the **variance or standard deviation**.





Accuracy vs. Precision



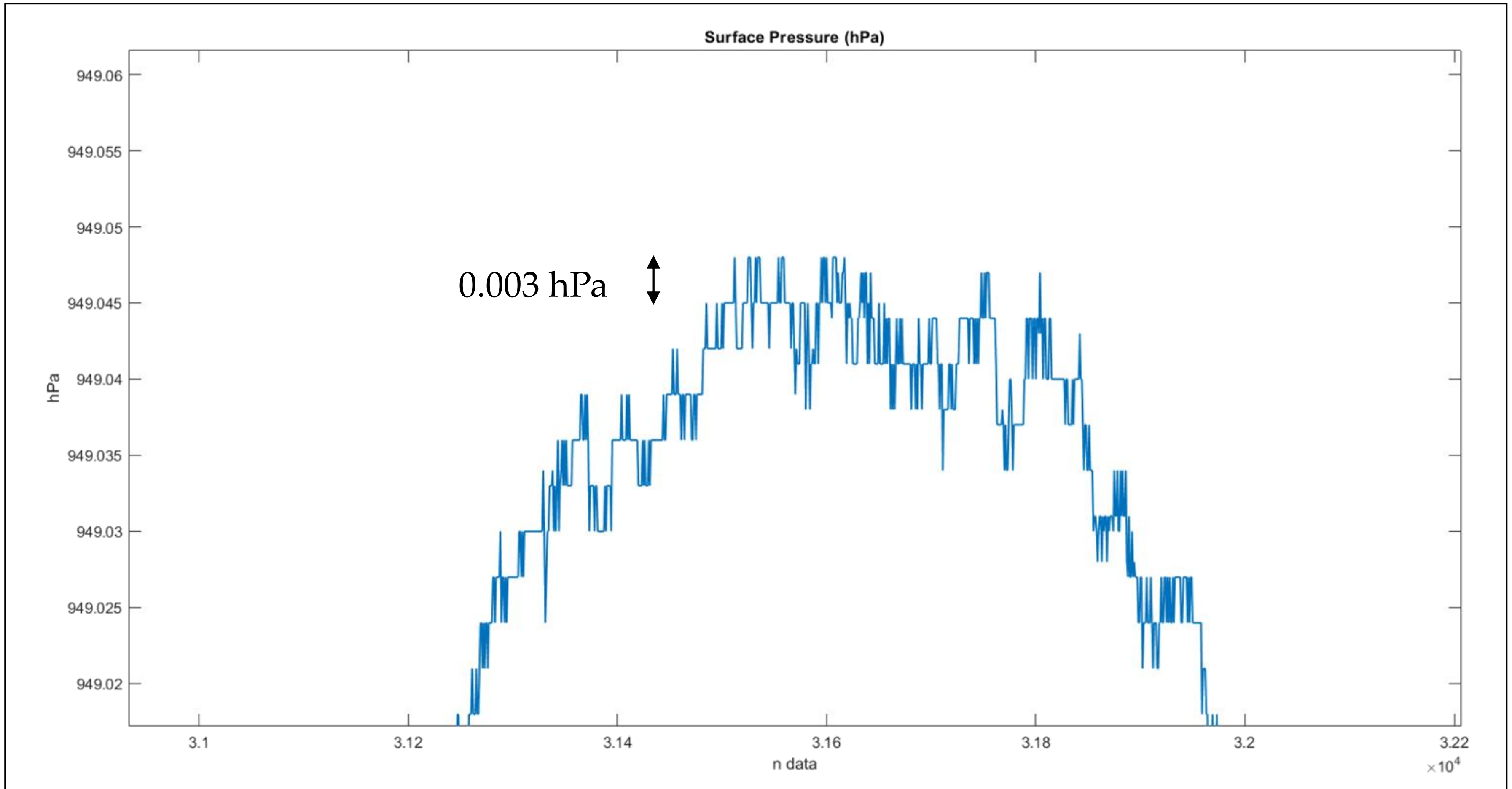
Resolution

(related to the readability of the measurement system)

Smallest change the sensor can capture and display.

Data =

949.2720000000000
949.2720000000000
949.2720000000000
949.2720000000000
949.2749999999999
949.2749999999999
949.2720000000000
949.2720000000000
949.2720000000000
949.2749999999999



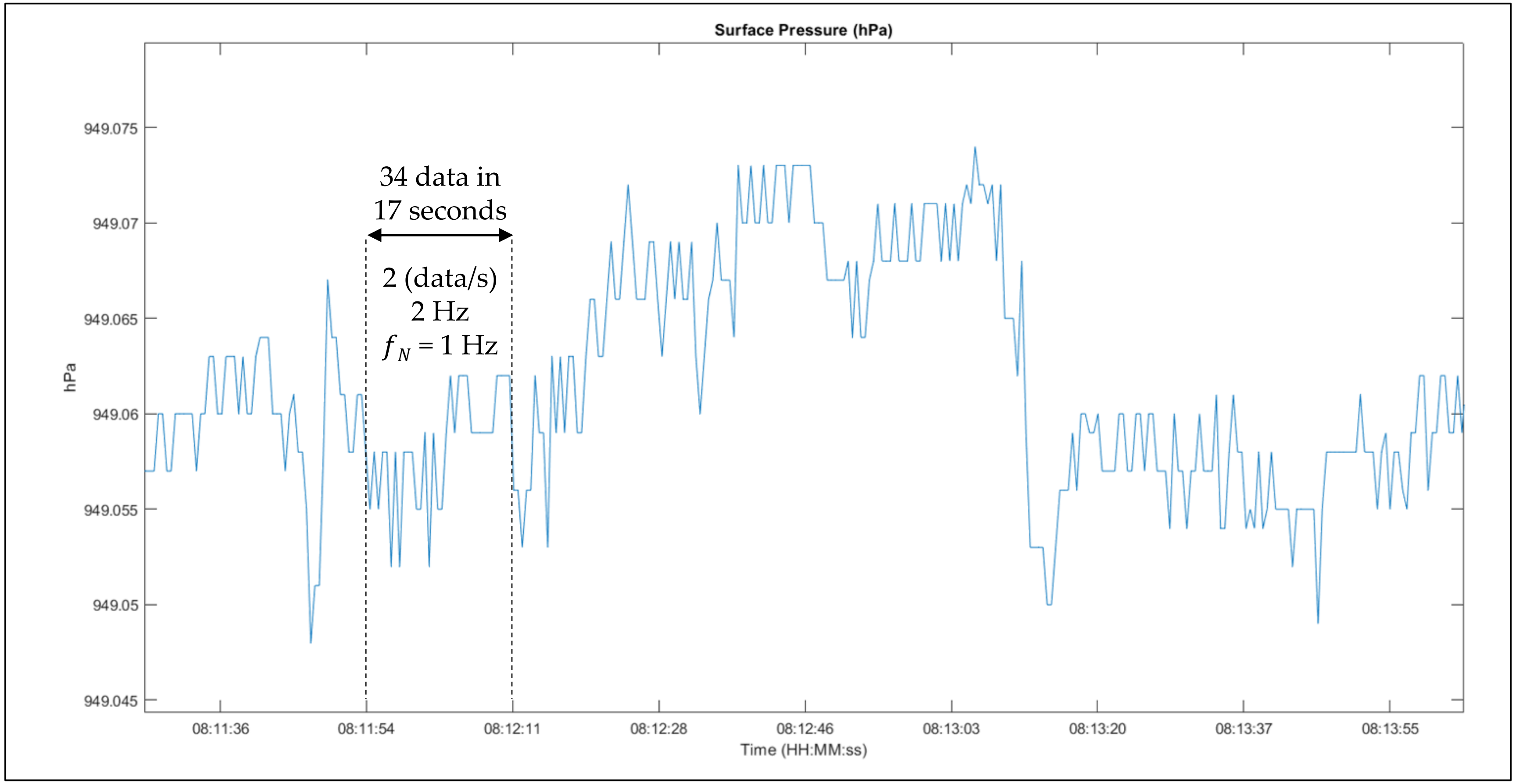
Sampling interval (Δt)
(also time resolution)

Time between measurements.
It should be often enough to detect the processes we are interested in.

Nyquist frequency

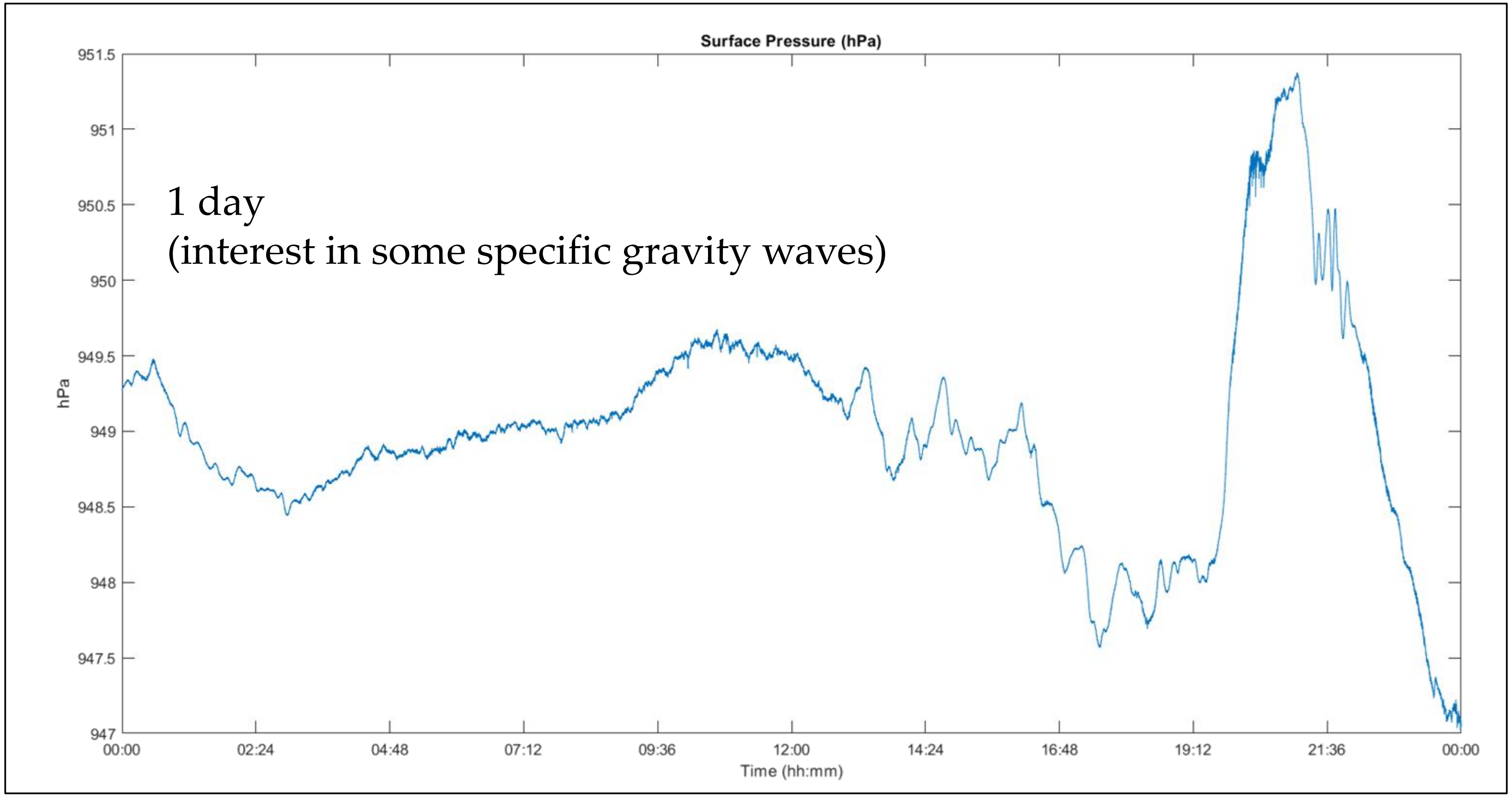
$$f_N = 1/(2\Delta t)$$

(it will be half of the instrument frequency)



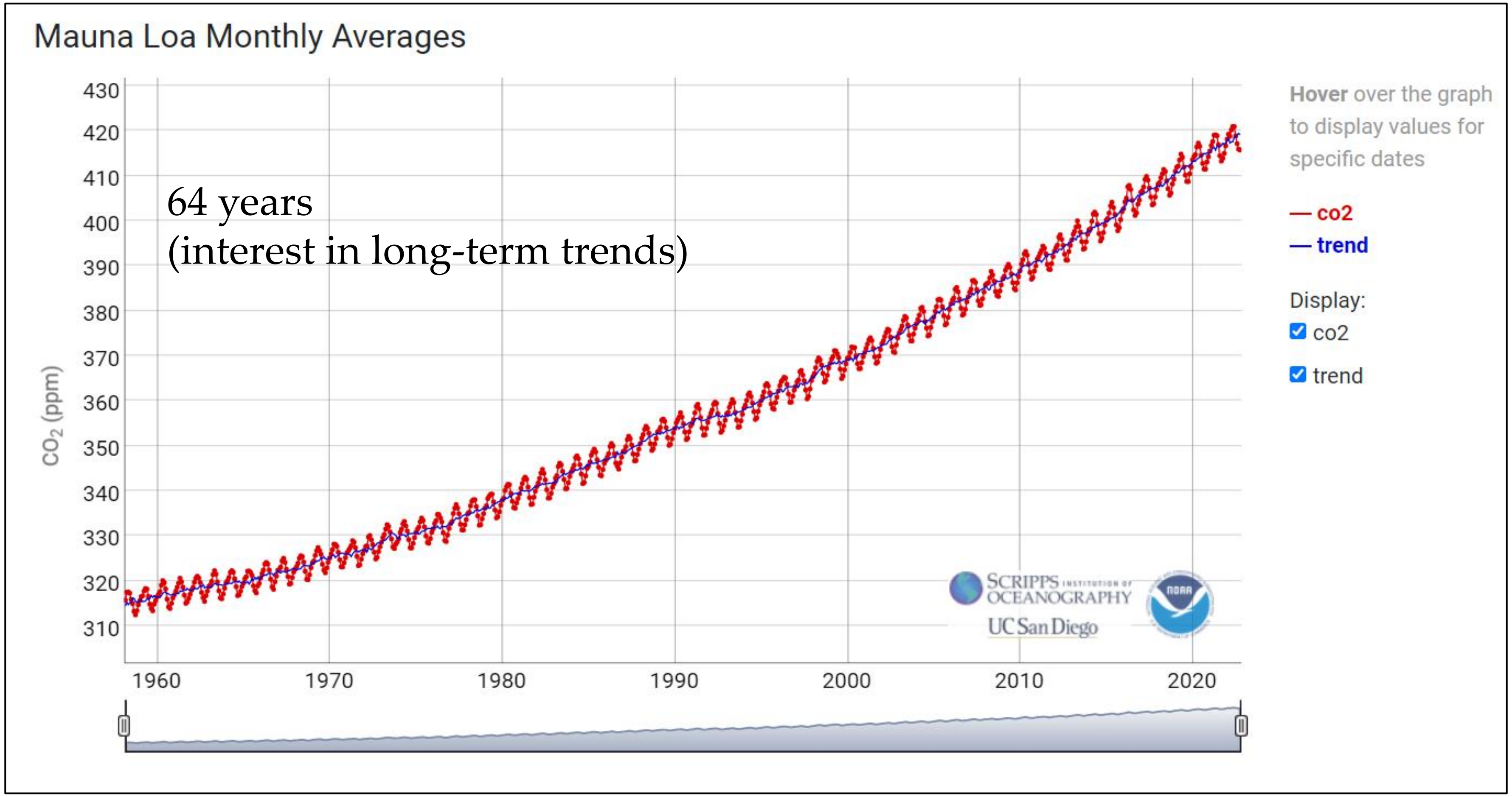
Sampling duration

It should allow for a statistically significant picture of the processes studied.



Sampling duration

It should allow for a statistically significant picture of the processes studied.



https://gml.noaa.gov/ccgg/trends/graph.html

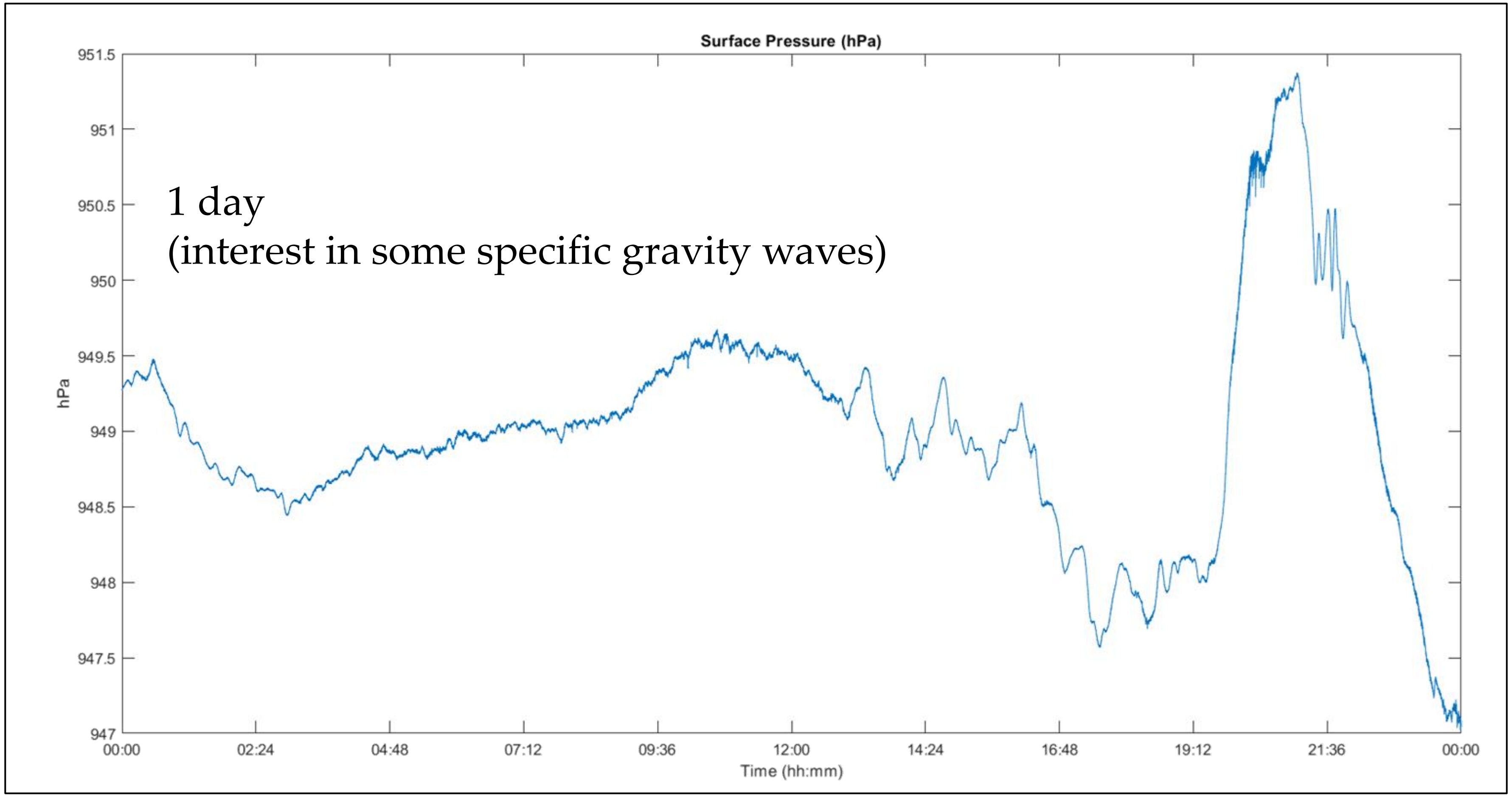
Sampling duration

It should allow for a statistically significant picture of the processes studied.

Fundamental frequency
$f_0 = 1/(T)$,
where T = N Δt

24*60*60*2 = 172800
(data in 1 day)
T = 172800 * 0.5
T= 86400
f_0 = 1/86400 =
0.0000115 Hz
(cycles/s)

(x60x60x24) =
(1 cycle/day)



Sampling duration

It should allow for a statistically significant picture of the processes studied.

Fundamental frequency

$f_0 = 1/(T)$,

where T = N Δt

24*60*60*2 = 172800

(data in 1 day)

T = 172800 * 0.5

T= 86400

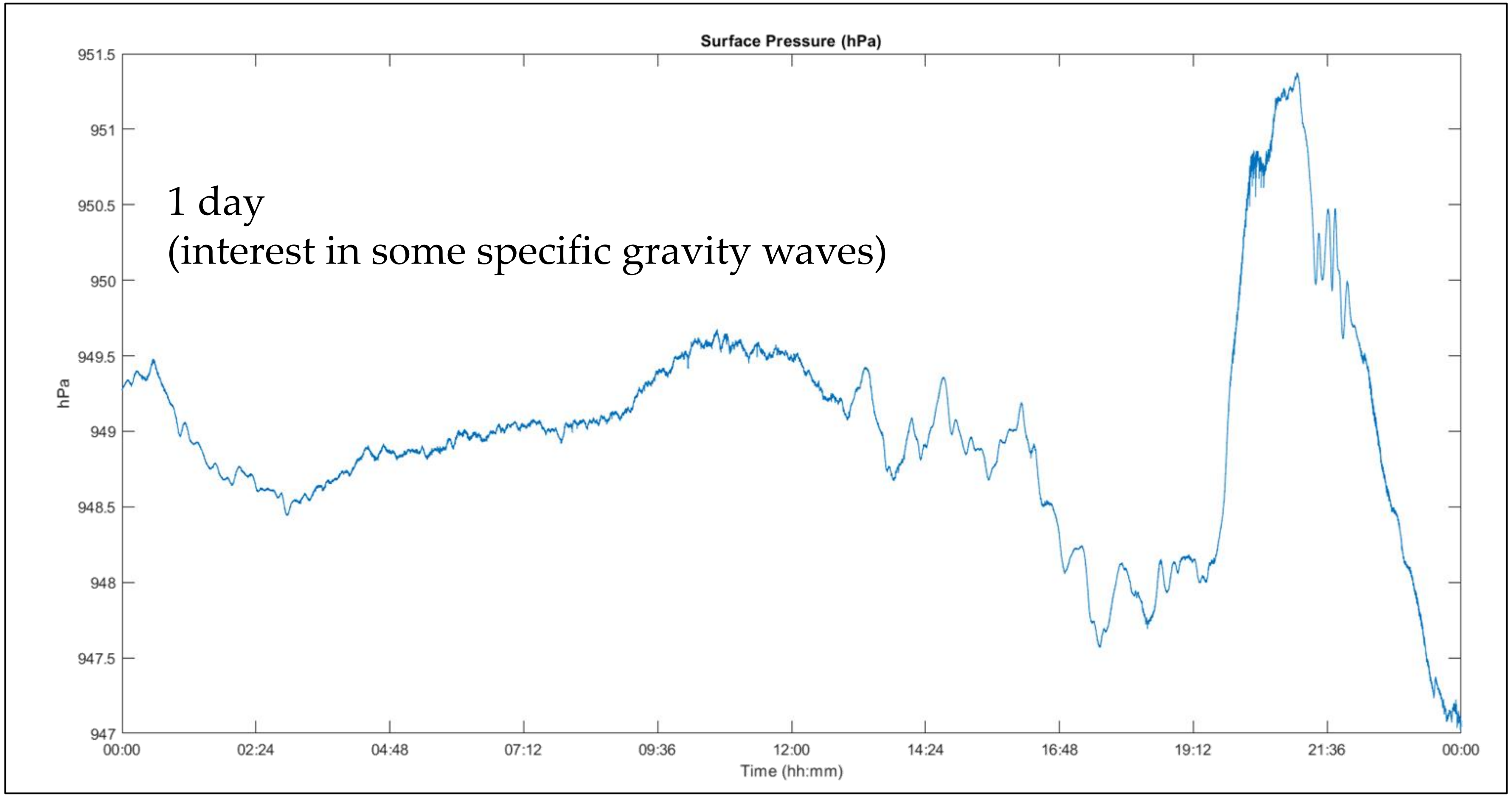
f_0 = 1/86400 =

0.0000115 Hz

(cycles/s)

(x60x60x24) =

(1 cycle/day)



Nyquist frequency

$f_N = 1/(2\Delta t)$

Δt = 0.5 s

f = 2 Hz

f_N = 1 / (2*0.5)

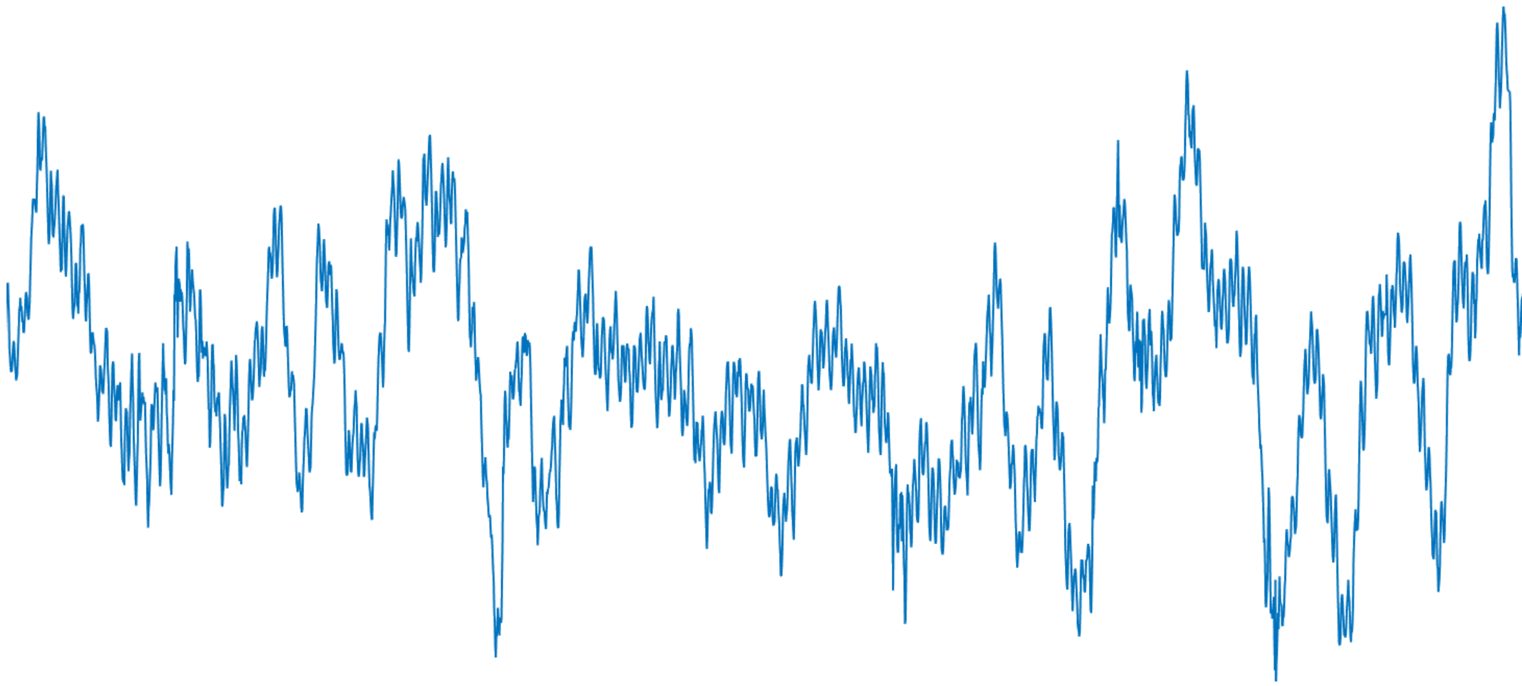
f_N = 1 Hz

(1 cycle each s)

(x60x60x24) =

(86400 cycles/day)

2. Data Analysis



Typical Scientific Procedure

READ → REPRESENT → PROCESS → CALCULATE → THINK → ADVANCED STUDIES → **SCIENTIFIC CONCLUSIONS**

0. **Read** the data → Fight against the files you receive!

1. Just **Represent** the variable(s) you are interested in → see their evolution in time, their spatial distribution (horizontal, vertical), etc.

2. **Processing** → edit the data in files to correct errors, transform the data, quality control, etc.

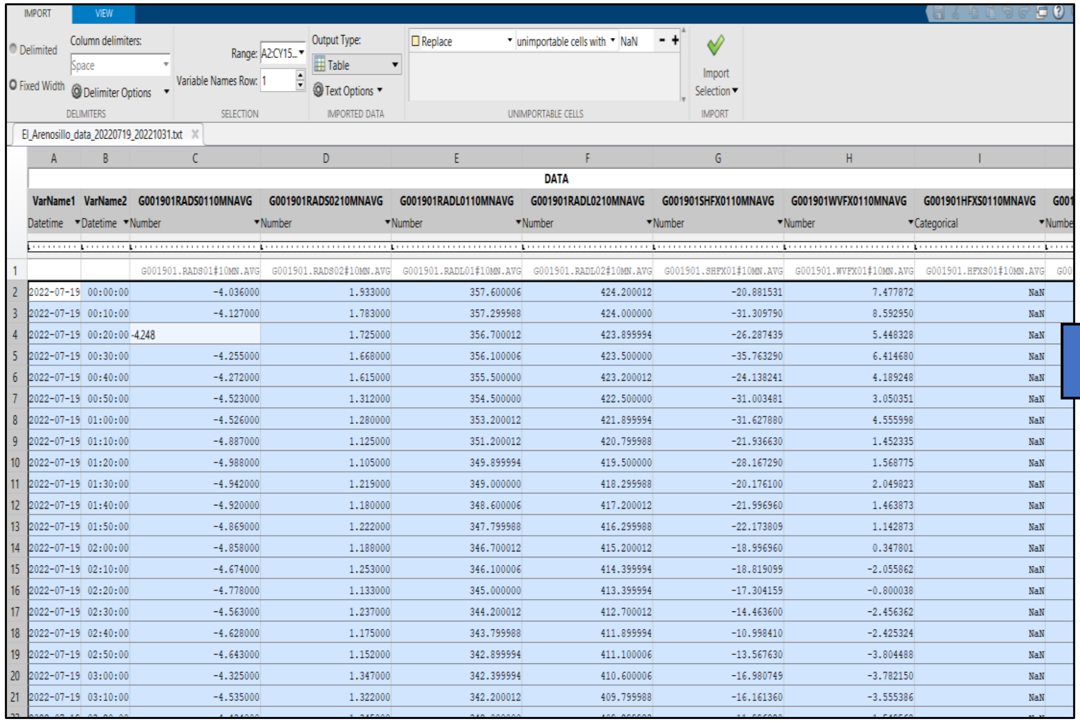
3. **Calculate** basic statistics → mean, standard deviation, anomalies (if we know the climatology), etc.

4. **Think** → Correlation with other variables? Try to understand the physical processes.

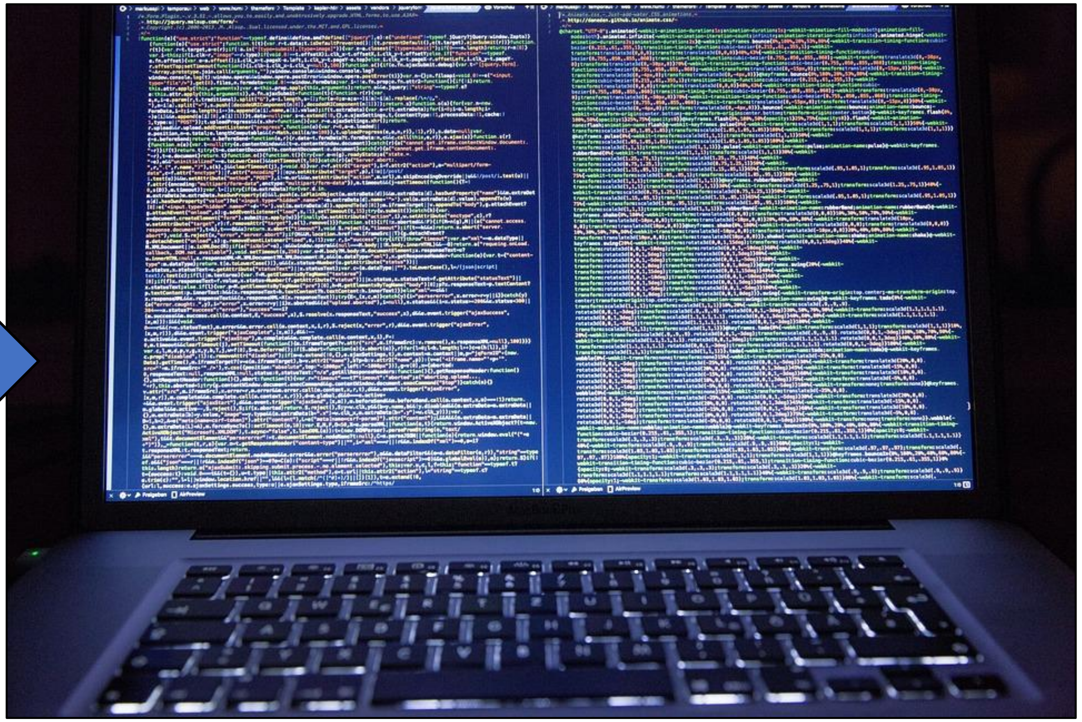
5. **Advanced** studies → Spectral analysis (Fourier), Harmonic analysis, etc.

6. Obtain **scientific conclusions.**

ascii files, netcdf files, excel files, HDF5, binary...



	A	B	C	D	E	F	G	H	I
	VarName1	VarName2	G001901RADS0110MNAVG	G001901RADS0210MNAVG	G001901RADL0110MNAVG	G001901RADL0210MNAVG	G001901SHFX0110MNAVG	G001901WVFX0110MNAVG	G001901HFXS0110MNAVG
	Datetime	Datetime	Number	Number	Number	Number	Number	Number	Categorical
1			G001901.RADS01‡10MN.AVG	G001901.RADS02‡10MN.AVG	G001901.RADL01‡10MN.AVG	G001901.RADL02‡10MN.AVG	G001901.SHFX01‡10MN.AVG	G001901.WVFX01‡10MN.AVG	G001901.HFXS01‡10MN.AVG
2	2022-07-19	00:00:00	-4.036000	1.933000	357.600006	424.200012	-20.881531	7.477872	NaN
3	2022-07-19	00:10:00	-4.127000	1.783000	357.299988	424.000000	-31.309790	8.592950	NaN
4	2022-07-19	00:20:00	-4248	1.725000	356.700012	423.899994	-26.287439	5.448328	NaN
5	2022-07-19	00:30:00	-4.255000	1.668000	356.100006	423.500000	-35.763290	6.414680	NaN
6	2022-07-19	00:40:00	-4.272000	1.615000	355.500000	423.200012	-24.138241	4.189248	NaN
7	2022-07-19	00:50:00	-4.523000	1.312000	354.500000	422.500000	-31.003481	3.050351	NaN
8	2022-07-19	01:00:00	-4.526000	1.280000	353.200012	421.899994	-31.627880	4.555998	NaN
9	2022-07-19	01:10:00	-4.887000	1.125000	351.200012	420.799988	-21.936630	1.452335	NaN
10	2022-07-19	01:20:00	-4.988000	1.105000	349.899994	419.500000	-28.167290	1.568775	NaN
11	2022-07-19	01:30:00	-4.942000	1.219000	349.000000	418.299988	-20.176100	2.049823	NaN
12	2022-07-19	01:40:00	-4.920000	1.180000	348.600006	417.200012	-21.996960	1.463873	NaN
13	2022-07-19	01:50:00	-4.869000	1.222000	347.799988	416.299988	-22.173809	1.142873	NaN
14	2022-07-19	02:00:00	-4.858000	1.188000	346.700012	415.200012	-18.996960	0.347801	NaN
15	2022-07-19	02:10:00	-4.674000	1.253000	346.100006	414.399994	-18.819099	-2.055862	NaN
16	2022-07-19	02:20:00	-4.778000	1.133000	345.000000	413.399994	-17.304159	-0.800038	NaN
17	2022-07-19	02:30:00	-4.563000	1.237000	344.200012	412.700012	-14.463600	-2.456362	NaN
18	2022-07-19	02:40:00	-4.628000	1.175000	343.799988	411.899994	-10.998410	-2.425324	NaN
19	2022-07-19	02:50:00	-4.643000	1.152000	342.899994	411.100006	-13.567630	-3.804488	NaN
20	2022-07-19	03:00:00	-4.325000	1.347000	342.399994	410.600006	-16.980749	-3.782150	NaN
21	2022-07-19	03:10:00	-4.535000	1.322000	342.200012	409.799988	-16.161360	-3.555386	NaN



FIRST QUICKLOOKS!

Examples:

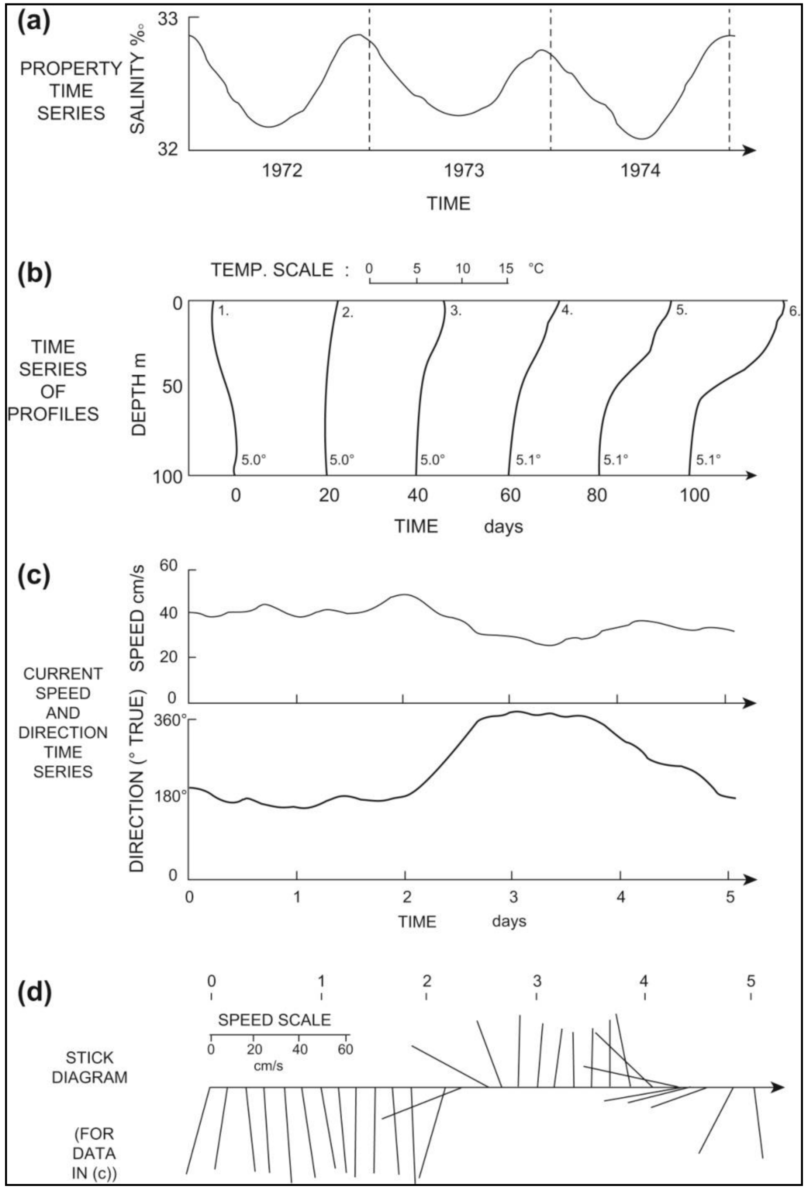
a) Variable along the time

b) Profiles time series

c) Current speed and direction along the time

d) Stick diagram (same than in c)

…



From Emery and Thomson (2014)

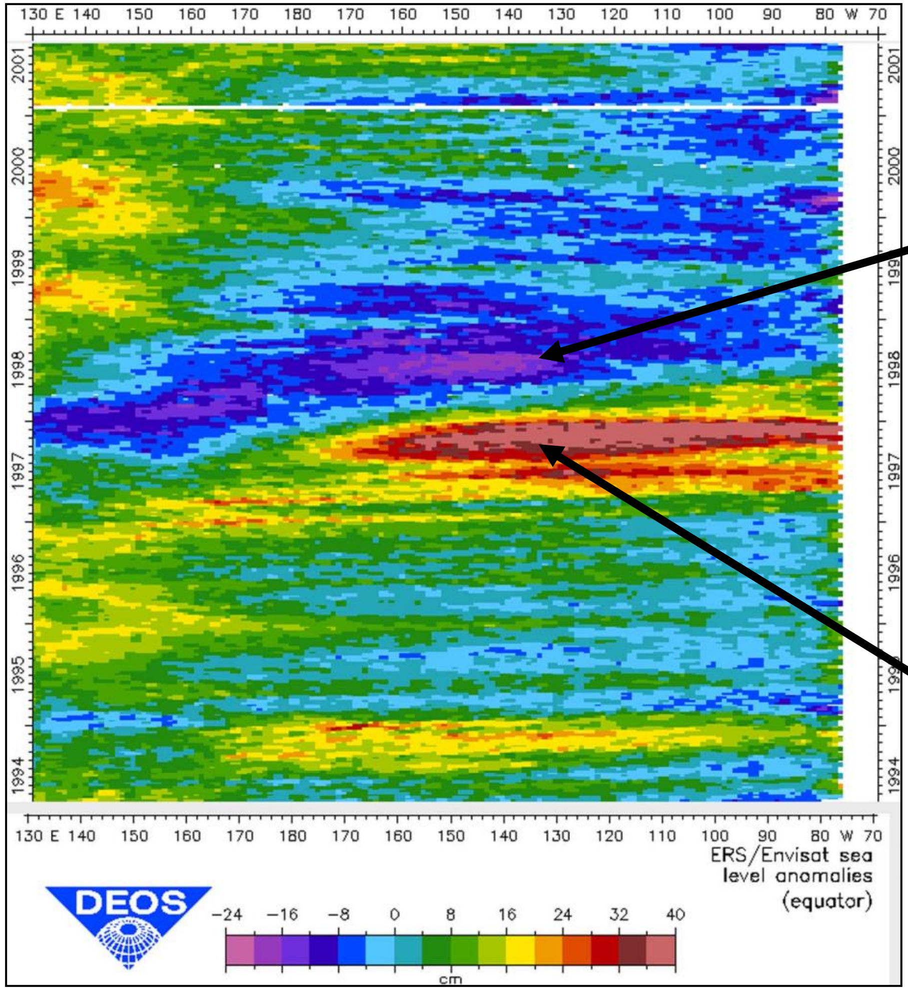
Hovmöller diagram

Time in y-axis

Space in x-axis

Variable in colors

Sea level height (SLH) anomaly in the Equatorial Pacific



Negative anomaly!

(propagating from W to E)

Positive anomaly!

(El Niño 1997)

Depth in y-axis
Latitude in x-axis
Variable in contours

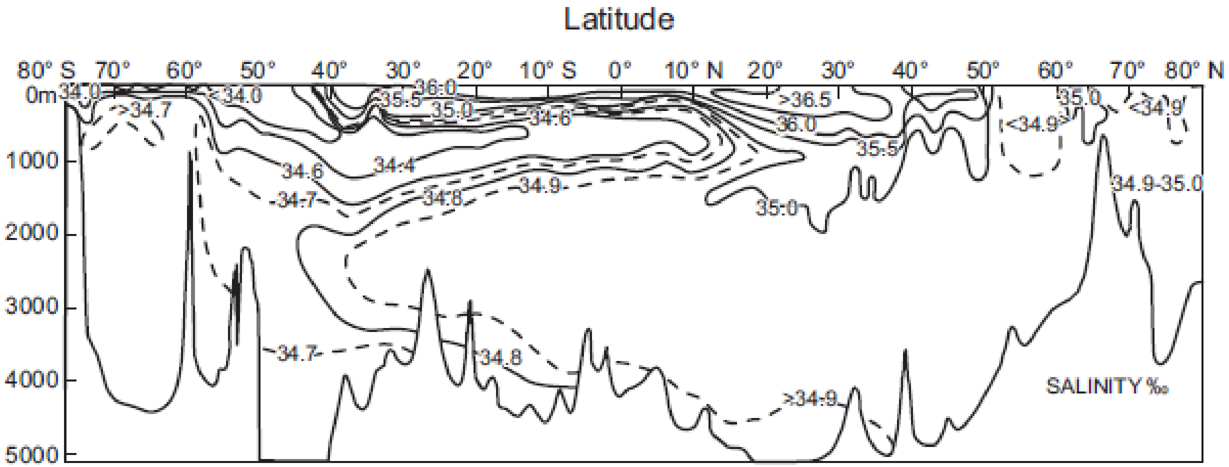
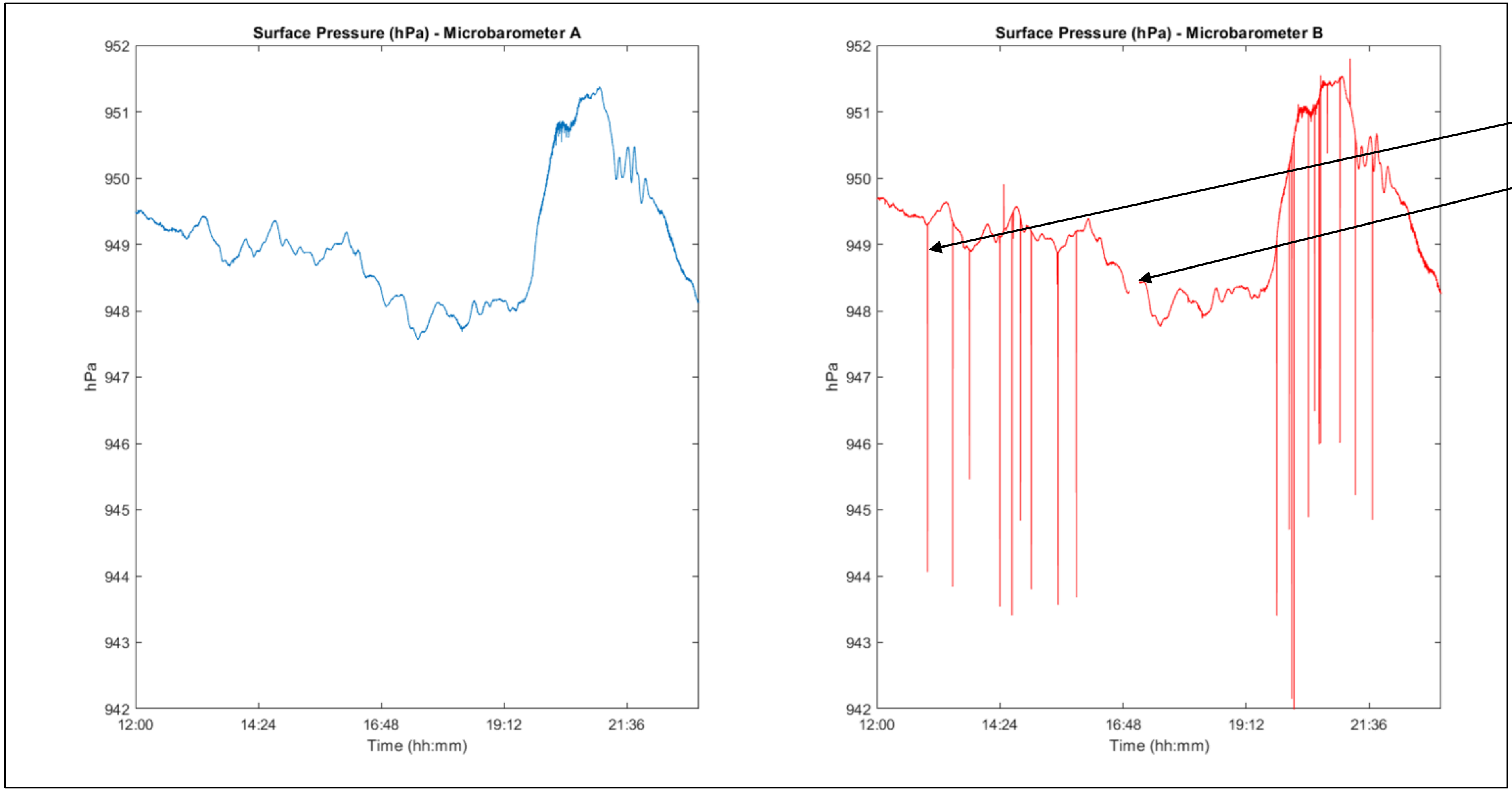


FIGURE 2.1 Latitudinal section of salinity in the western basin of the Atlantic Ocean. *(After Spiess (1928).)*
From Emery and Thomson (2014)

Exaggeration of the y-axis in comparison with the x-axis needed!

ALWAYS SAVE THE ORIGINAL DATA FILE! (raw data)
(security copy)



Remove anomalous data

Fill gaps

Correct known systematic errors (known bias)

Define a reasonable range of variation

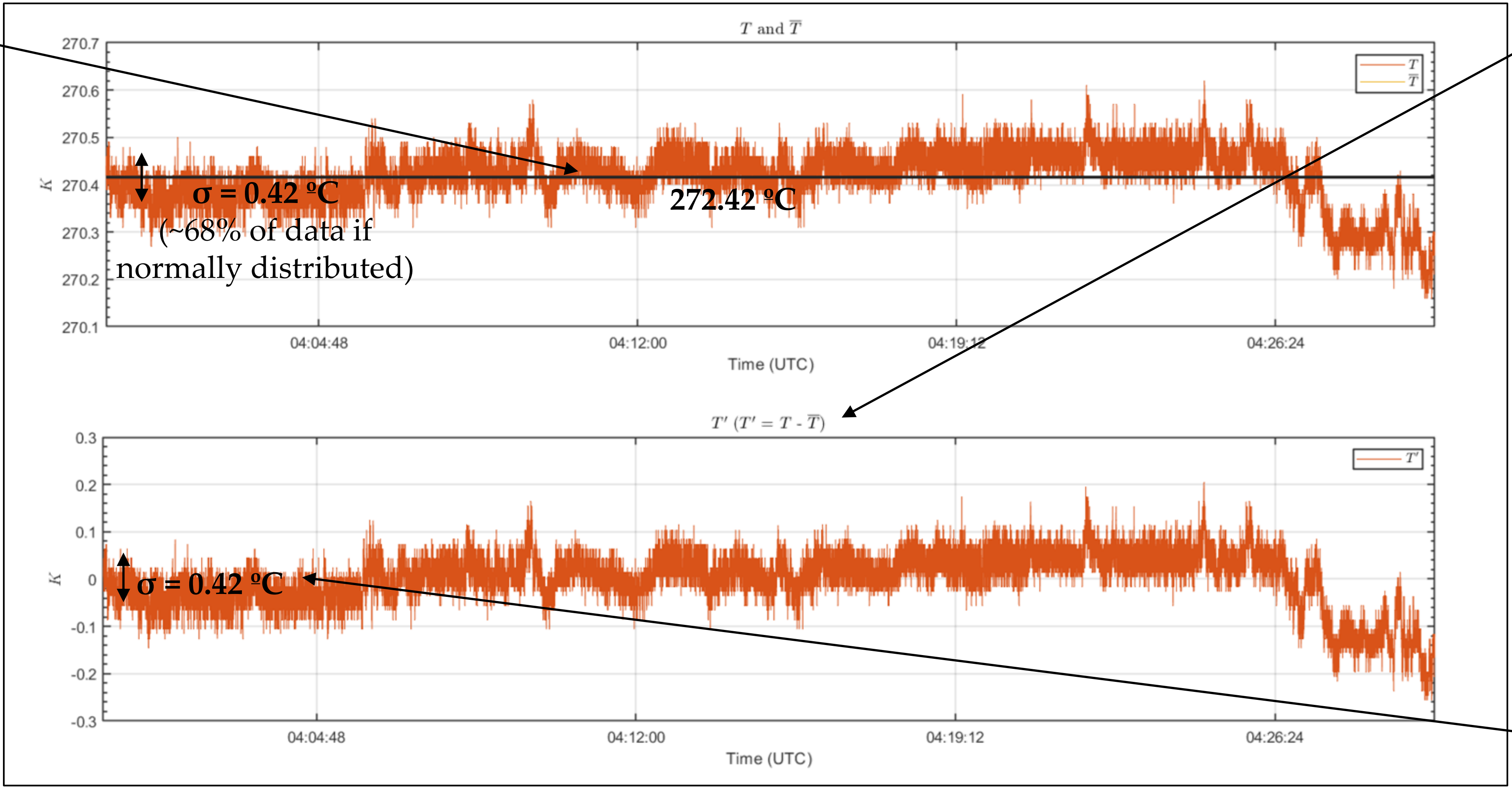
Quality control flags

…

Mean, standard deviation, anomaly (if we know the climatology), …

Mean

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$



Anomaly (or fluctuation)

$$x' = x - \bar{x}$$

Mean fluctuation in the example is ~0 ºC, but note the short sampling period (30 min)

Variance

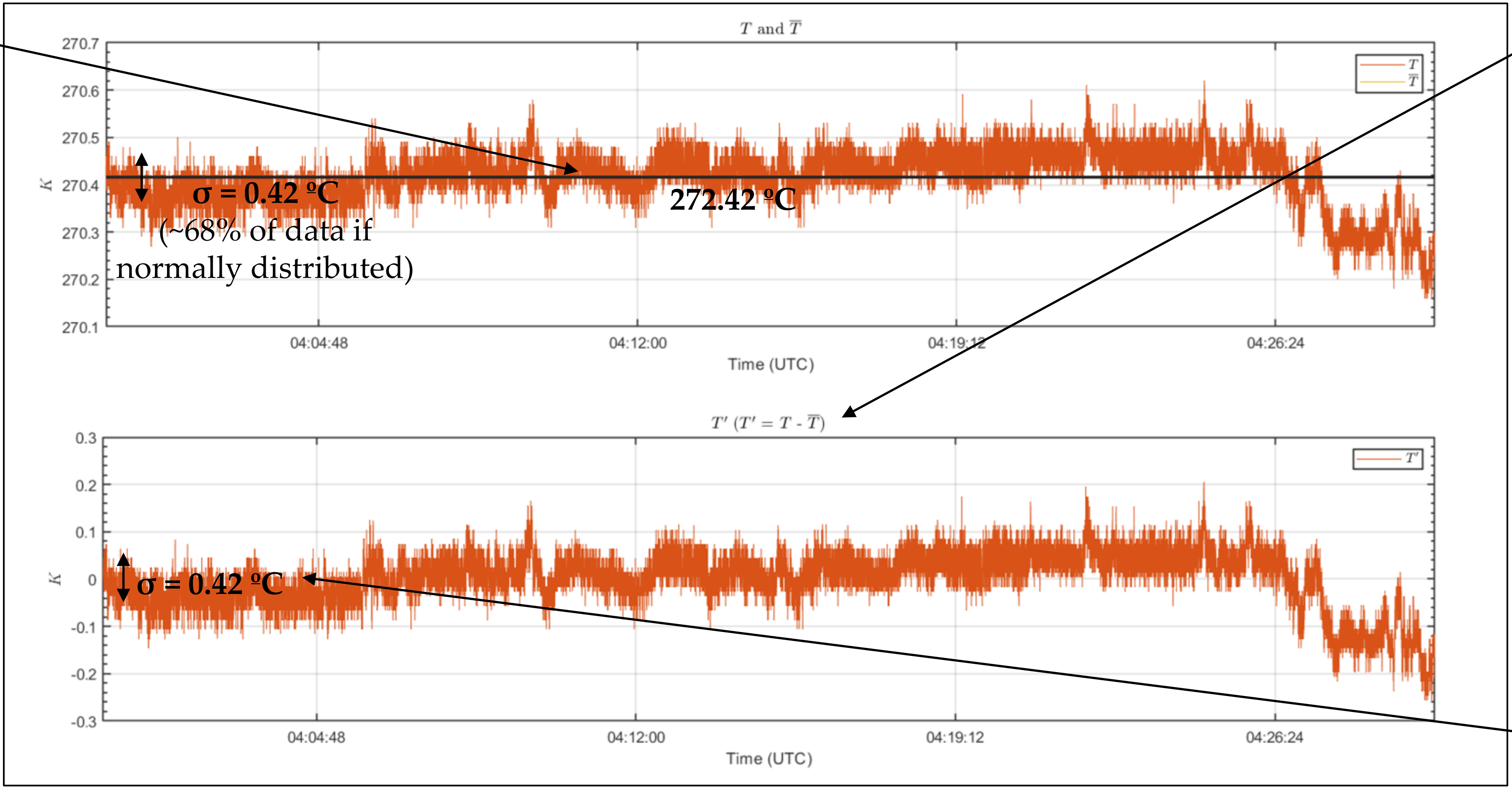
$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i-\bar{x})^2$$

Standard deviation

$$\sigma$$

Mean

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$



Anomaly (or fluctuation)

$$x' = x - \overline{x}$$

Mean fluctuation in the example is ~0 ºC, but note the short sampling period (30 min)

Variance

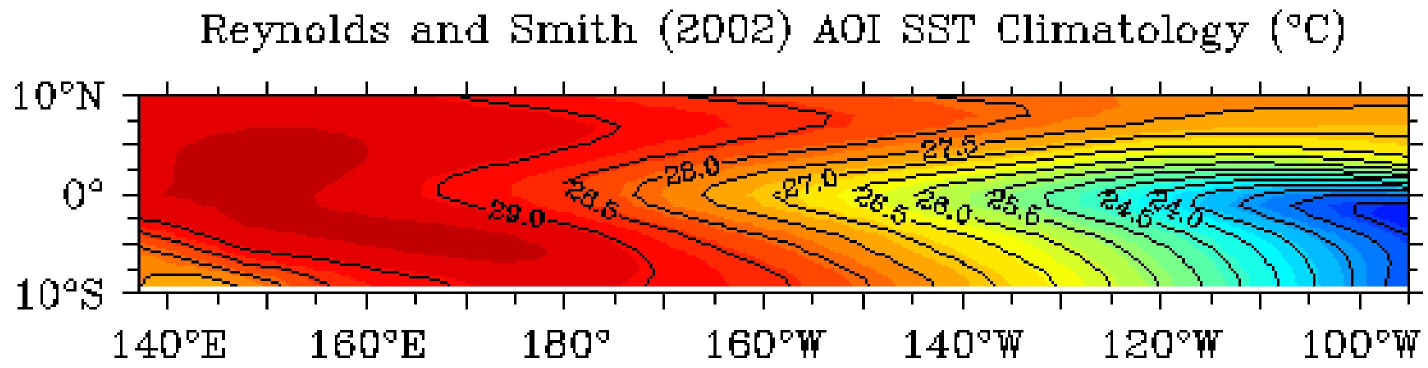
$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i-\overline{x})^2$$

Standard deviation

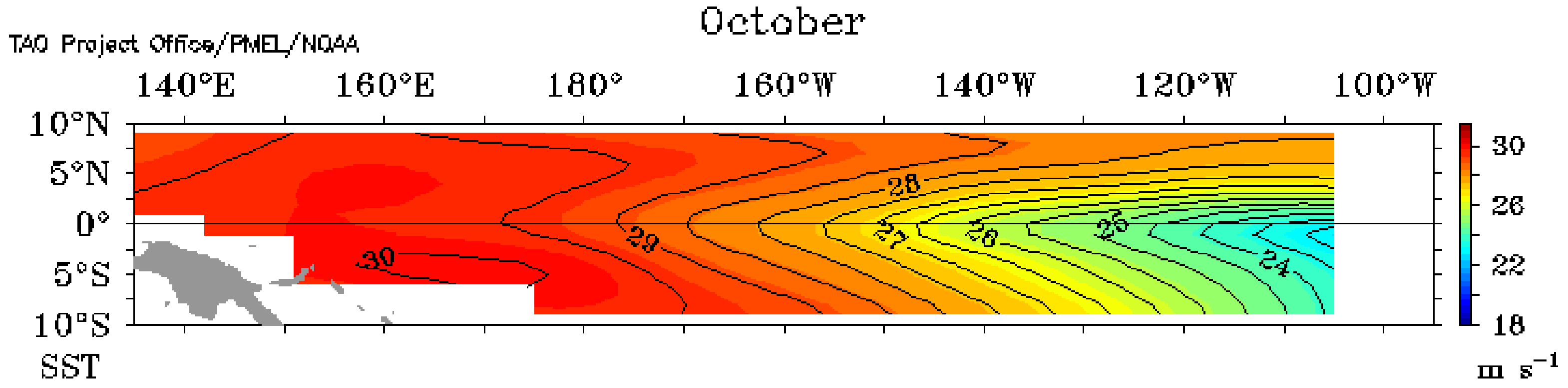
σ

Caution when working with circular variables (wind direction, currents direction…)

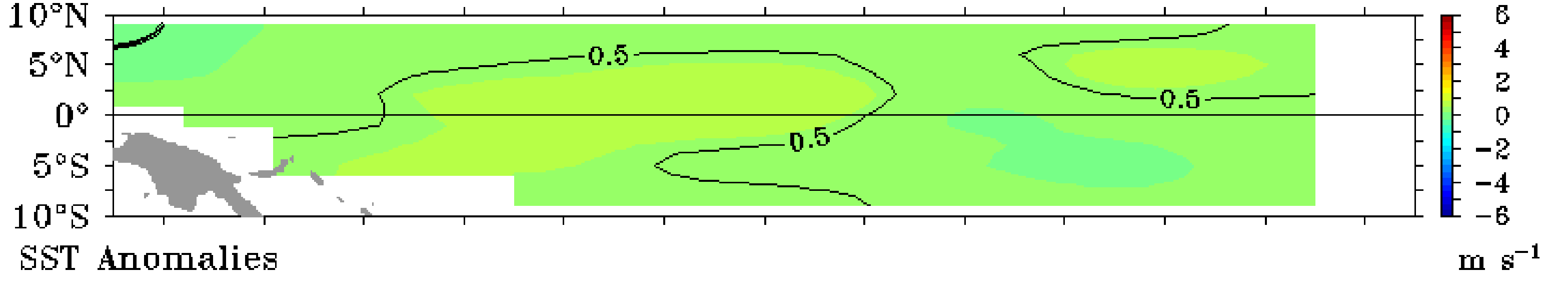
SST climatology*
*(*long timeseries ~30 years)*



Observations in Oct. 2012

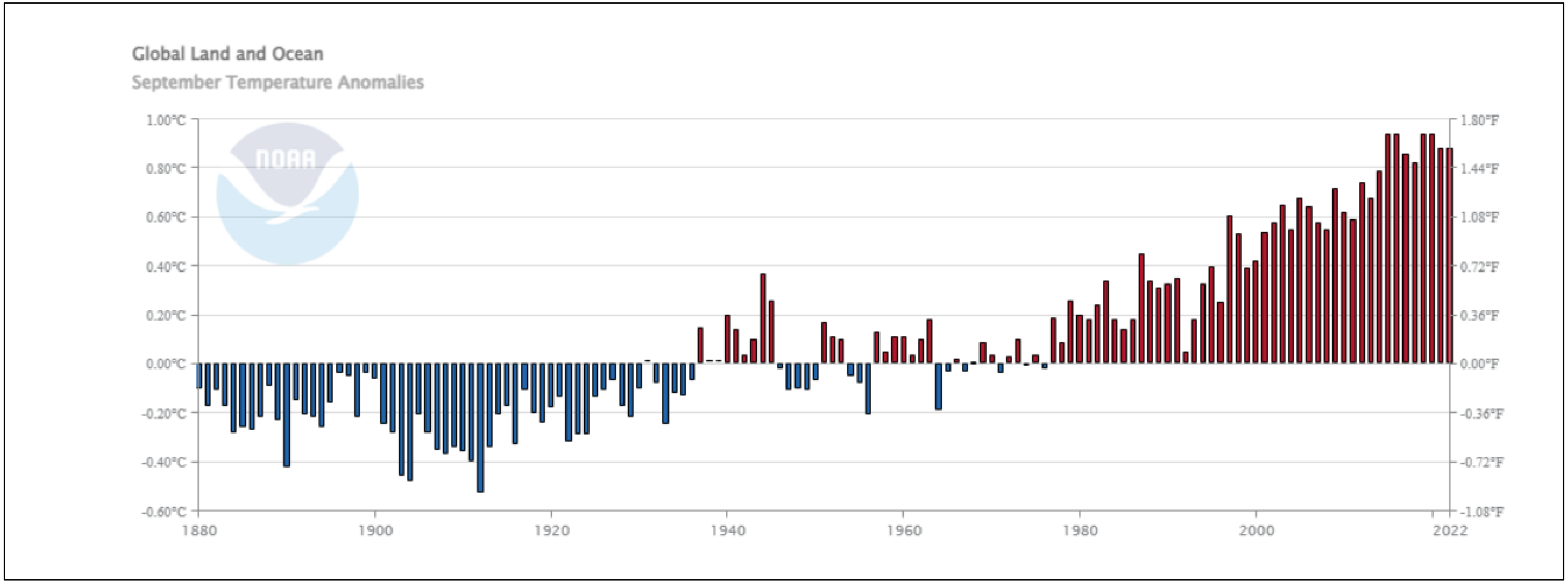


SST anomaly in Oct. 2012
(obs - climat)



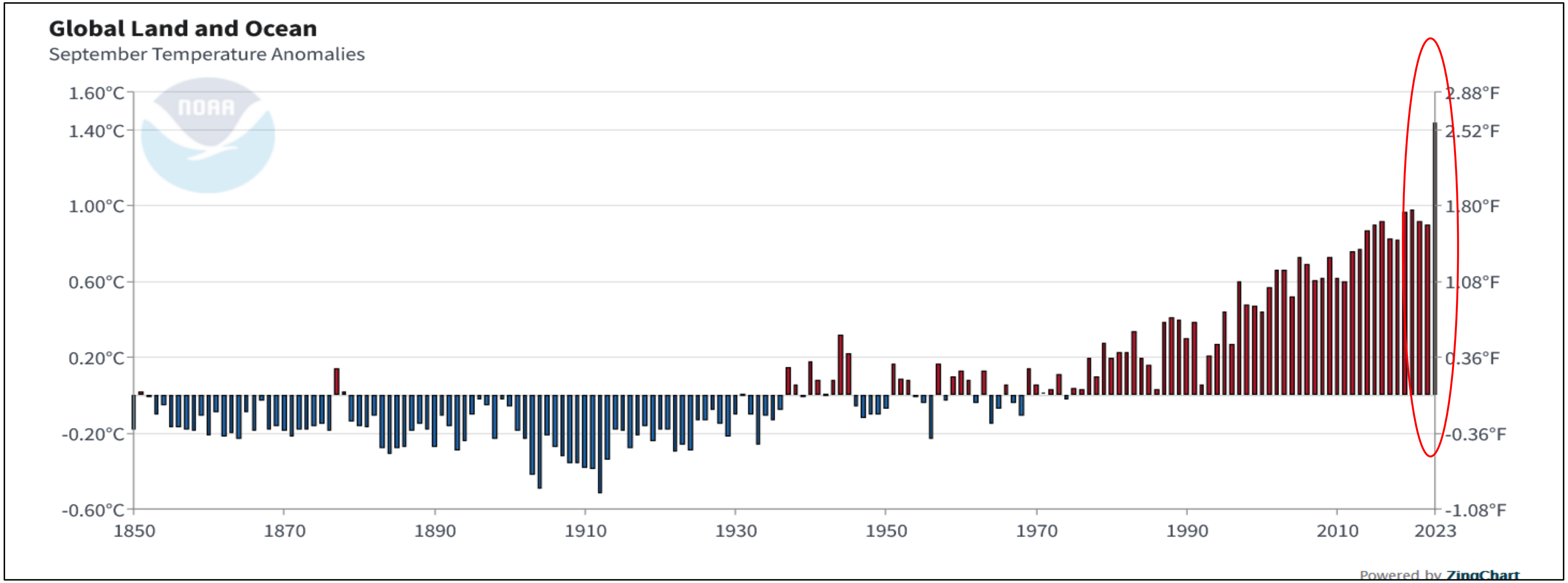
http://www.pmel.noaa.gov/tao/

The September 2022 global surface temperature departure tied September 2021 as the fifth highest for September in the 143-year record at 0.88°C (1.58°F) above the 20th century average of 15.0°C (59.0°F). The ten warmest Septembers on record have all occurred since 2012. September 2022 also marked the 46th consecutive September and the 453rd consecutive month with temperatures, at least nominally, above the 20th century average.

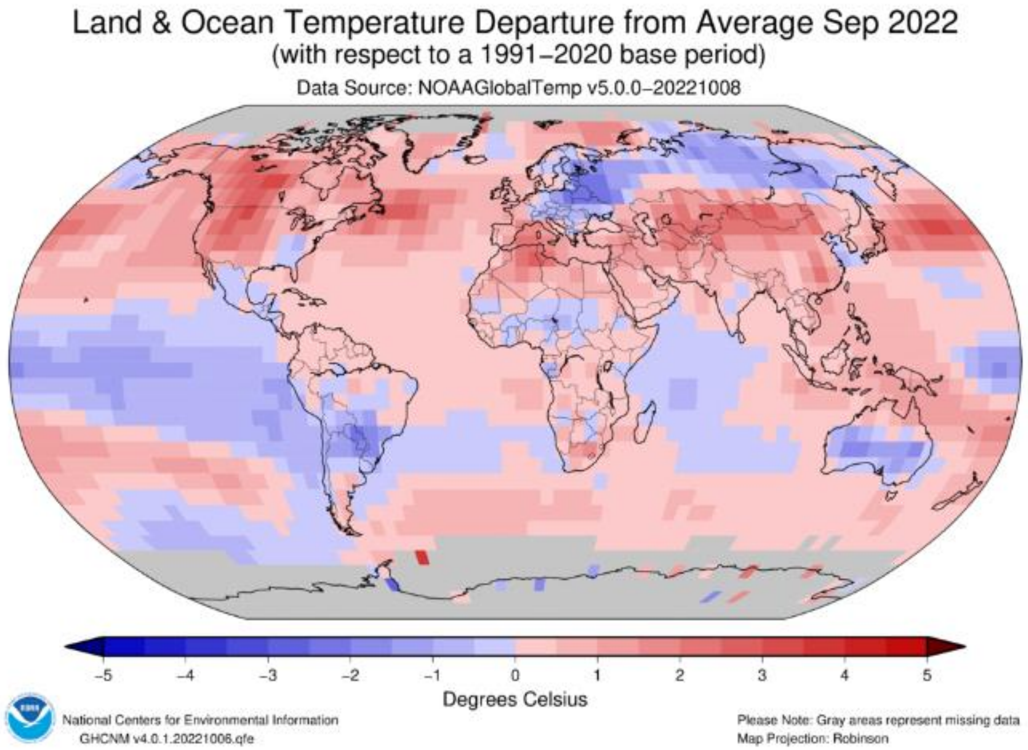


https://www.ncdc.noaa.gov/sotc/global/202209

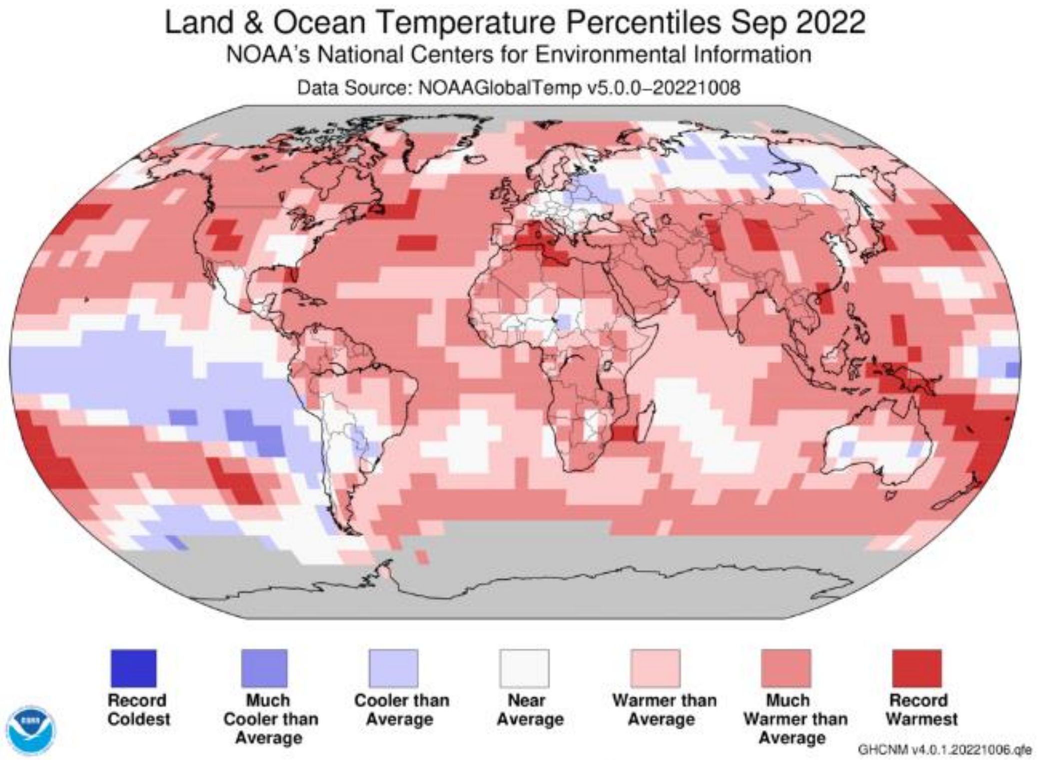
The September global surface temperature was 1.44°C (2.59°F) above the 20th-century average of 15.0°C (59.0°F), making it the warmest September on record. September 2023 marked the 49th-consecutive September and the 535th-consecutive month with temperatures at least nominally above the 20th-century average. September 2023 was 0.46°C (0.83°F) above the previous record from September 2020, and marks the largest positive monthly global temperature anomaly of any month on record. The September 2023 global temperature anomaly surpassed the previous record-high monthly anomaly from March 2016 by 0.09°C (0.16°F). The past ten Septembers (2014–2023) have been the warmest Septembers on record.



https://www.ncdc.noaa.gov/sotc/global/202209

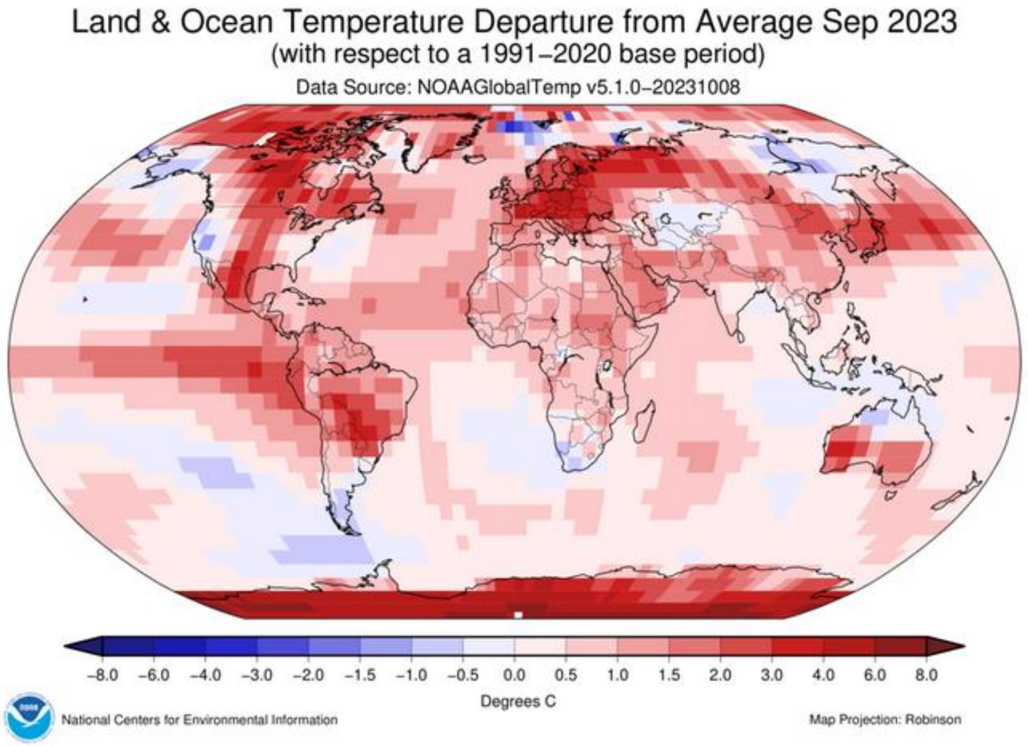


[September 2022 Blended Land and Sea Surface Temperature Anomalies in degrees Celsius]

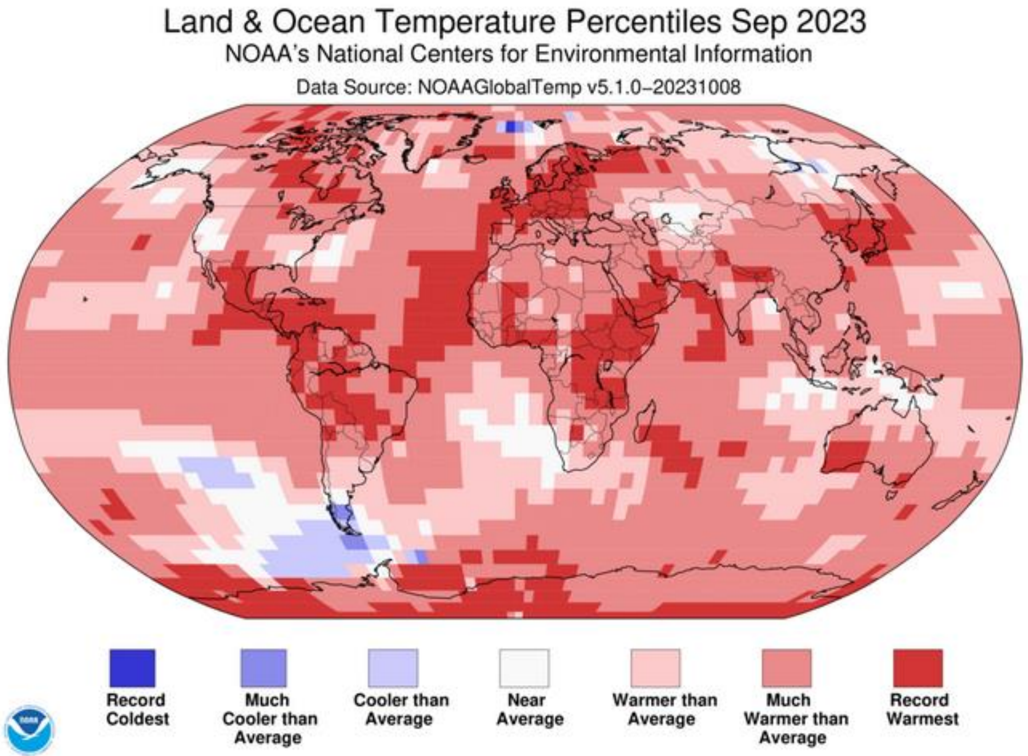


[September 2022 Blended Land and Sea Surface Temperature Percentiles]

https://www.ncdc.noaa.gov/sotc/global/202209



[September 2023 Blended Land and Sea Surface Temperature Anomalies in degrees Celsius]

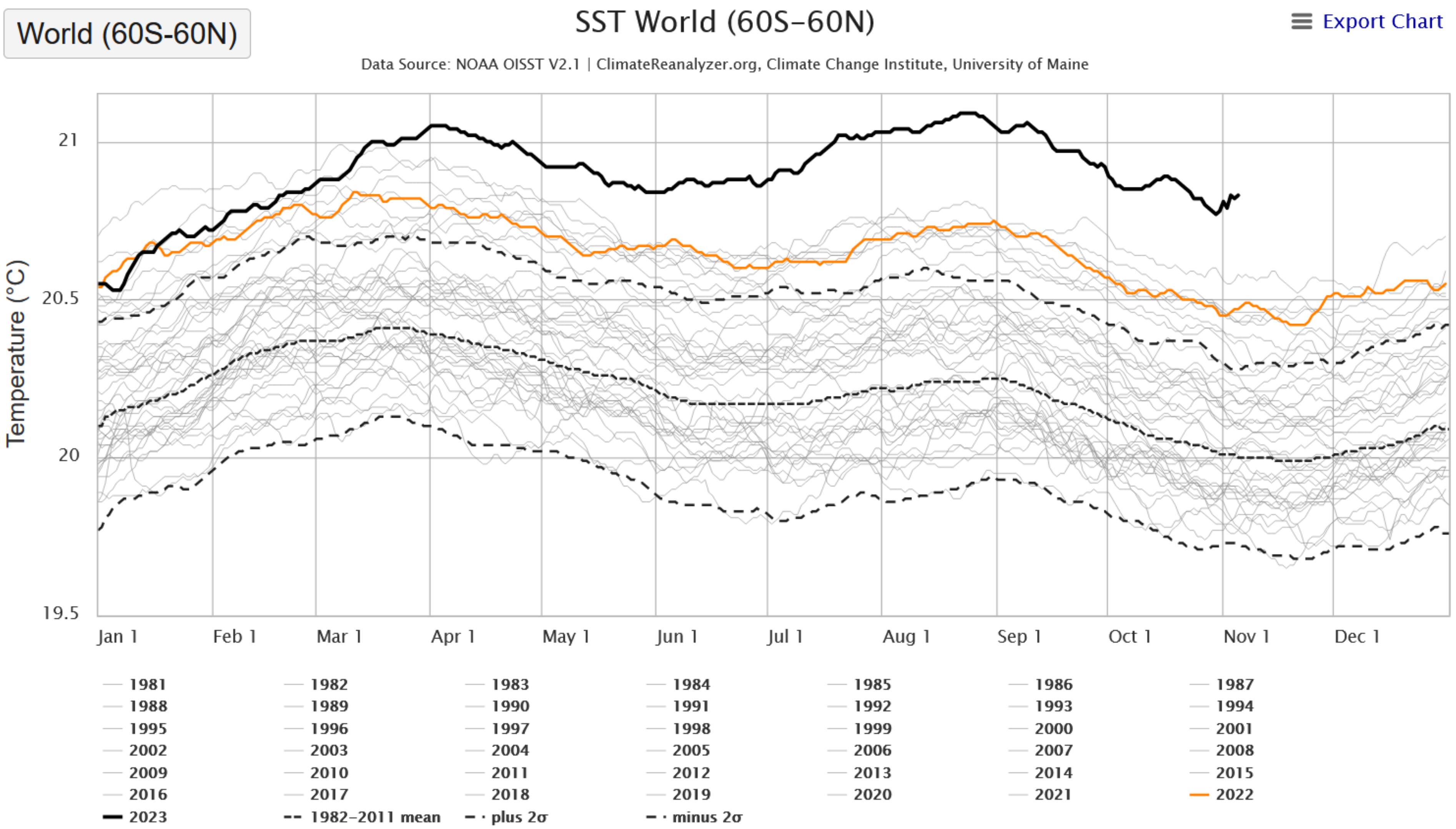


[September 2023 Blended Land and Sea Surface Temperature Percentiles]

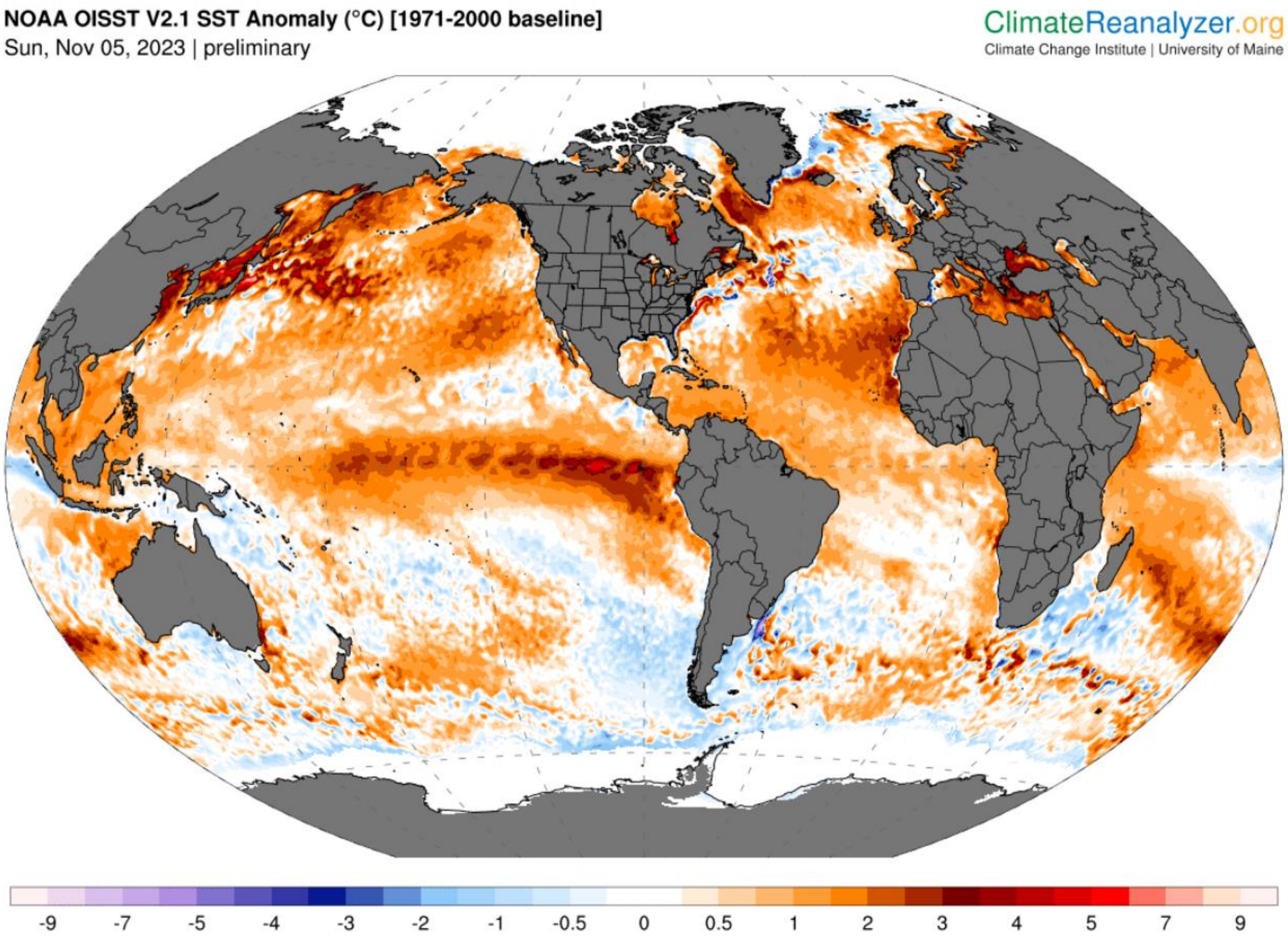
https://www.ncdc.noaa.gov/sotc/global/202209

Choose Area: °C/°F

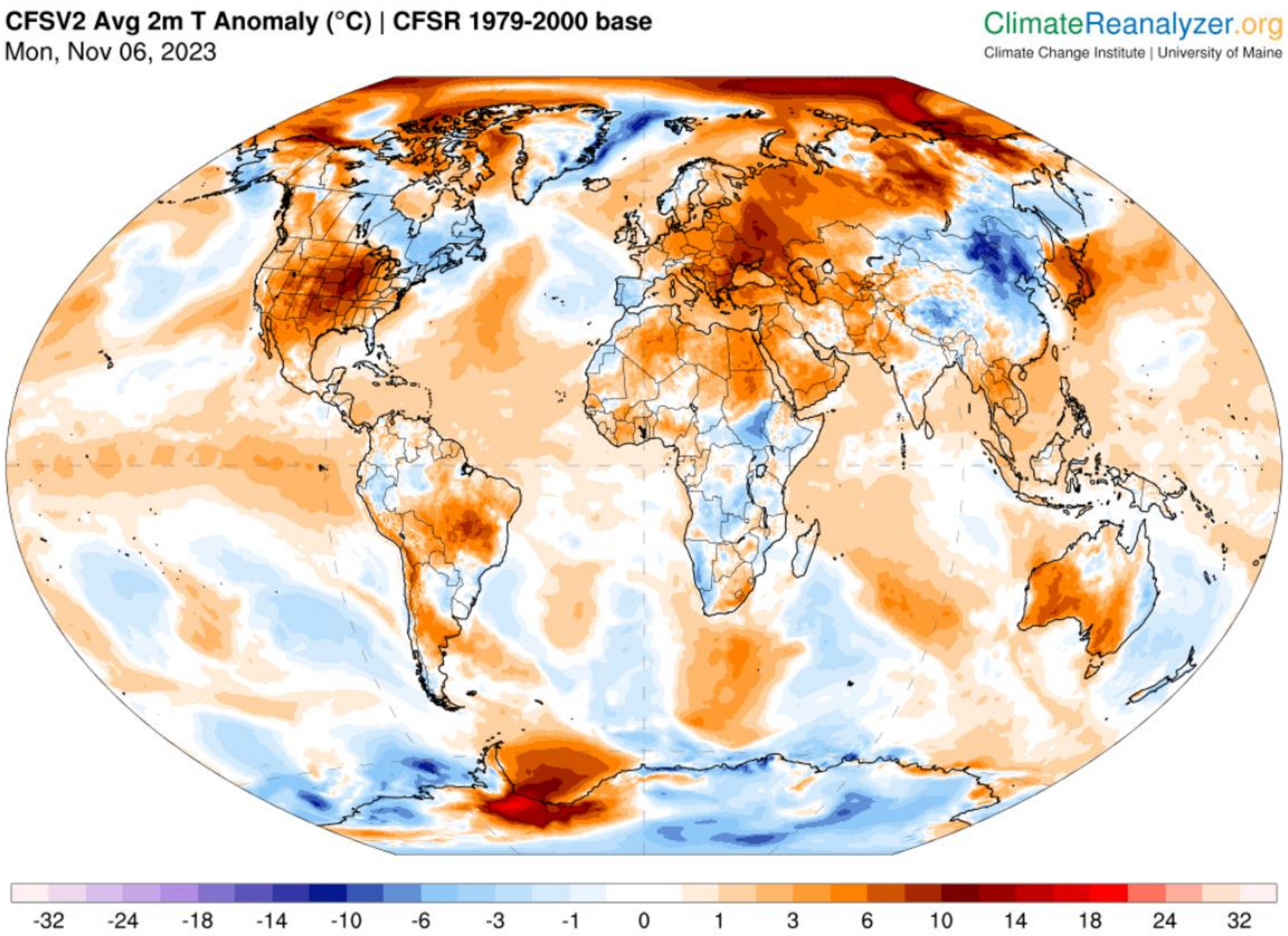
World (60S-60N)



https://climatereanalyzer.org/clim/sst_daily/

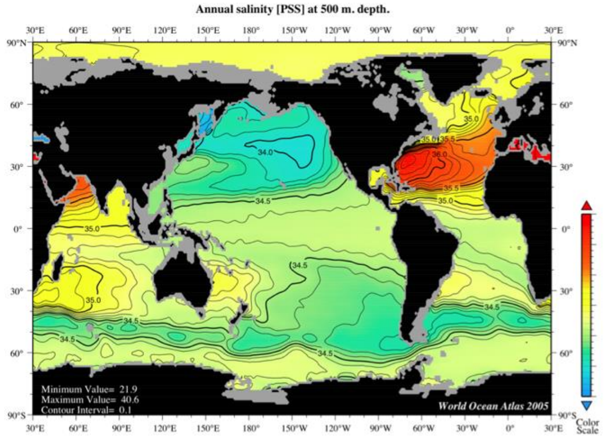


https://climatereanalyzer.org/wx/DailySummary/#sstanom

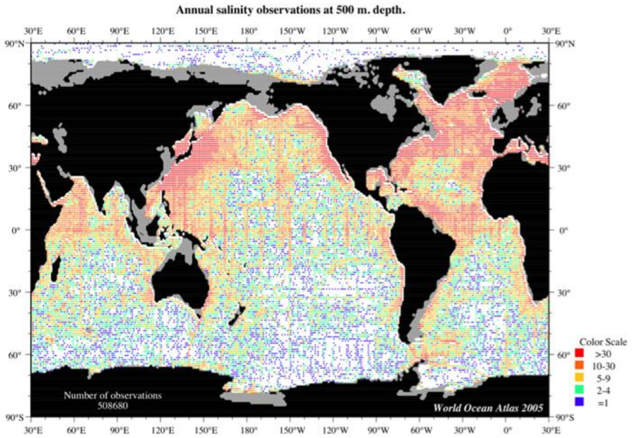


https://climatereanalyzer.org/wx/todays-weather/?var_id=t2anom&ortho=1&wt=1

Salinity (PSS) at 500 m depth (climatology)



(a)

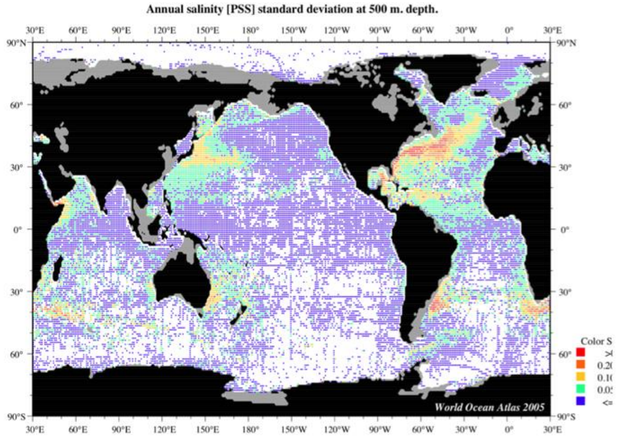


(b)

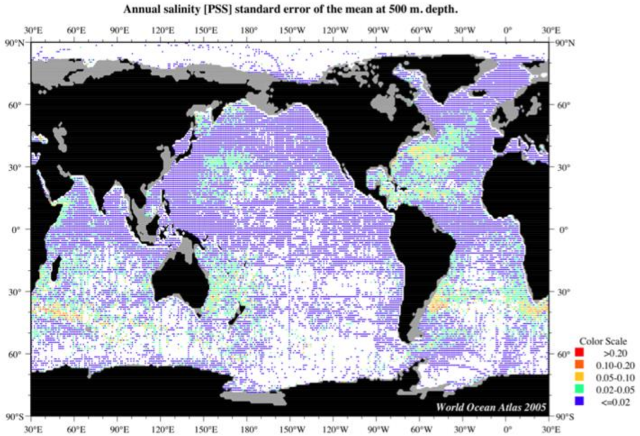
Salinity (PSS) at 500 m depth (observations)

Salinity (PSS) Standard deviation

Variability of the variable



(c)



(d)

Salinity (PSS) Standard error

Indicator of how well the variable was measured

Is my variable related to other variables?

Is this relationship interesting to analyse?

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})^2$$

Variance

Covariance & Correlation

$$\mathrm{cov}(x,y) = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) =$$

$$r_{x,y} = \frac{\mathrm{cov}(x,y)}{s_x s_y}$$

Covariance

Statistical relationship between two variables

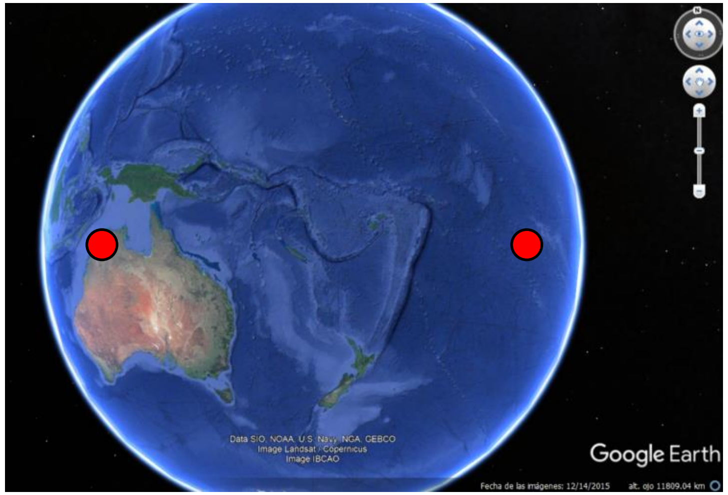
Correlation

Covariance divided by std product of both vars

[-1 to 1]

3. Spatial Analysis of Data Fields

Spatial analysis

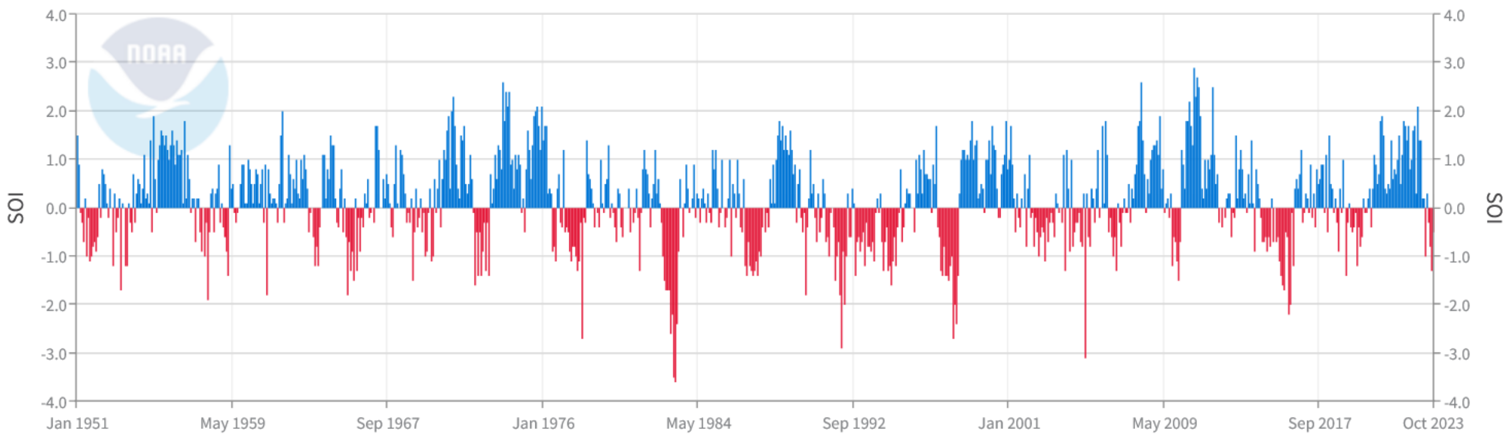


$$SOI = \frac{\left(sSLP_{Tahiti} - sSLP_{Darwin}\right)}{\sigma_{monthly}}$$

Southern Oscillation Index (SOI)

The Southern Oscillation Index (SOI) is a standardized index based on the observed sea level pressure (SLP) differences between Tahiti and Darwin, Australia. The SOI is one measure of the large-scale fluctuations in air pressure occurring between the western and eastern tropical Pacific (i.e., the state of the Southern Oscillation) during El Niño and La Niña episodes. In general, smoothed time series of the SOI correspond very well with changes in ocean temperatures across the eastern tropical Pacific. The negative phase of the SOI represents below-normal air pressure at Tahiti and above-normal air pressure at Darwin. Prolonged periods of negative (positive) SOI values coincide with abnormally warm (cold) ocean waters across the eastern tropical Pacific typical of El Niño (La Niña) episodes. The methodology used to calculate SOI is available below. More information can be found at the Climate Prediction Center SOI page.

Southern Oscillation Index (SOI)



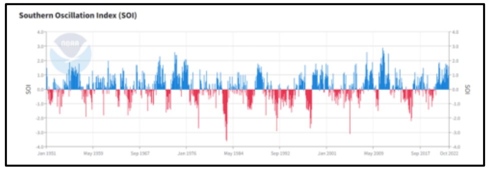
Source: https://www.cpc.ncep.noaa.gov/data/indices/soi

Powered by **ZingChart**

https://www.ncei.noaa.gov/access/monitoring/enso/soi

Spatial analysis

We can correlate the evolution of this index



With the time evolution of several variables defined spatially

For example, SST…

Southern Oscillation Index (SOI)

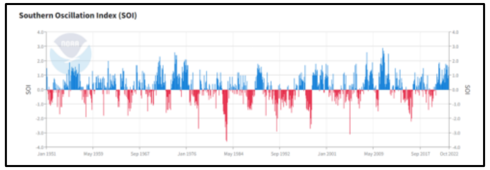
The Southern Oscillation Index (SOI) is a standardized index based on the observed sea level pressure (SLP) differences between Tahiti and Darwin, Australia. The SOI is one measure of the large-scale fluctuations in air pressure occurring between the western and eastern tropical Pacific (i.e., the state of the Southern Oscillation) during El Niño and La Niña episodes. In general, smoothed time series of the SOI correspond very well with changes in ocean temperatures across the eastern tropical Pacific. The negative phase of the SOI represents below-normal air pressure at Tahiti and above-normal air pressure at Darwin. Prolonged periods of negative (positive) SOI values coincide with abnormally warm (cold) ocean waters across the eastern tropical Pacific typical of El Niño (La Niña) episodes. The methodology used to calculate SOI is available below. More information can be found at the Climate Prediction Center SOI page.

https://www.ncei.noaa.gov/access/monitoring/enso/soi

https://psl.noaa.gov/data/correlation/

Spatial analysis

We can correlate the evolution of this index (timeseries)



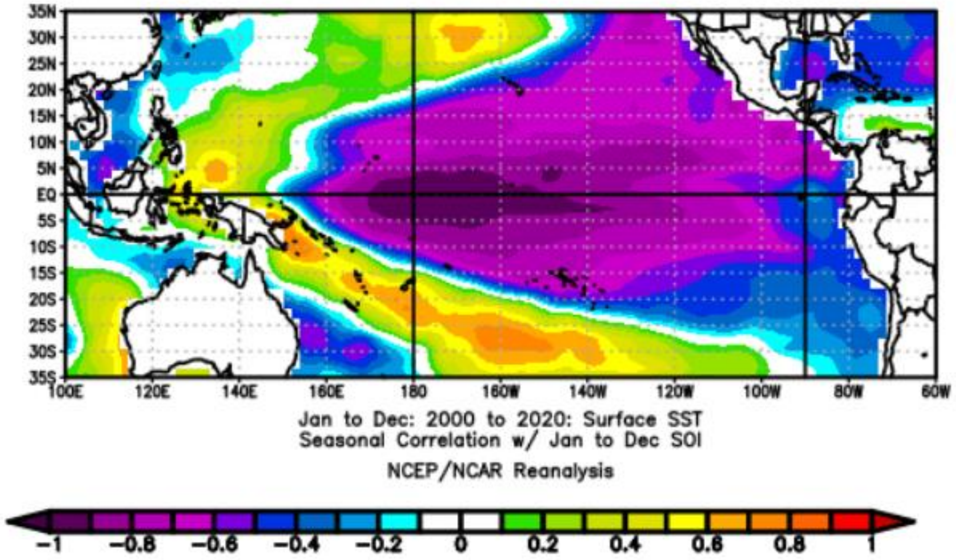
With the time evolution of several variables defined spatially

For example, SST defined spatially

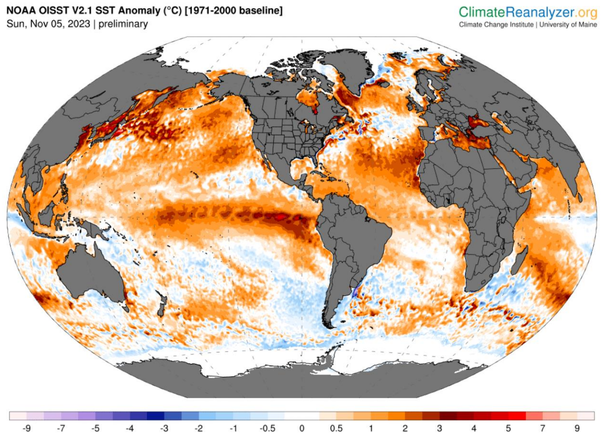
Southern Oscillation Index (SOI)

The Southern Oscillation Index (SOI) is a standardized index based on the observed sea level pressure (SLP) differences between Tahiti and Darwin, Australia. The SOI is one measure of the large-scale fluctuations in air pressure occurring between the western and eastern tropical Pacific (i.e., the state of the Southern Oscillation) during El Niño and La Niña episodes. In general, smoothed time series of the SOI correspond very well with changes in ocean temperatures across the eastern tropical Pacific. The negative phase of the SOI represents below-normal air pressure at Tahiti and above-normal air pressure at Darwin. Prolonged periods of negative (positive) SOI values coincide with abnormally warm (cold) ocean waters across the eastern tropical Pacific typical of El Niño (La Niña) episodes. The methodology used to calculate SOI is available below. More information can be found at the Climate Prediction Center SOI page.

https://www.ncei.noaa.gov/access/monitoring/enso/soi



NOAA Physical Sciences Laboratory



https://psl.noaa.gov/data/correlation/

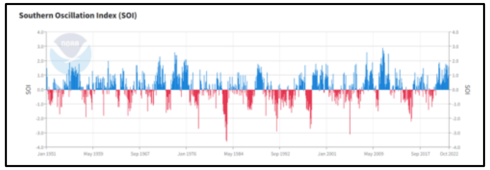
Spatial analysis

Southern Oscillation Index (SOI)

The Southern Oscillation Index (SOI) is a standardized index based on the observed sea level pressure (SLP) differences between Tahiti and Darwin, Australia. The SOI is one measure of the large-scale fluctuations in air pressure occurring between the western and eastern tropical Pacific (i.e., the state of the Southern Oscillation) during El Niño and La Niña episodes. In general, smoothed time series of the SOI correspond very well with changes in ocean temperatures across the eastern tropical Pacific. The negative phase of the SOI represents below-normal air pressure at Tahiti and above-normal air pressure at Darwin. Prolonged periods of negative (positive) SOI values coincide with abnormally warm (cold) ocean waters across the eastern tropical Pacific typical of El Niño (La Niña) episodes. The methodology used to calculate SOI is available below. More information can be found at the Climate Prediction Center SOI page.

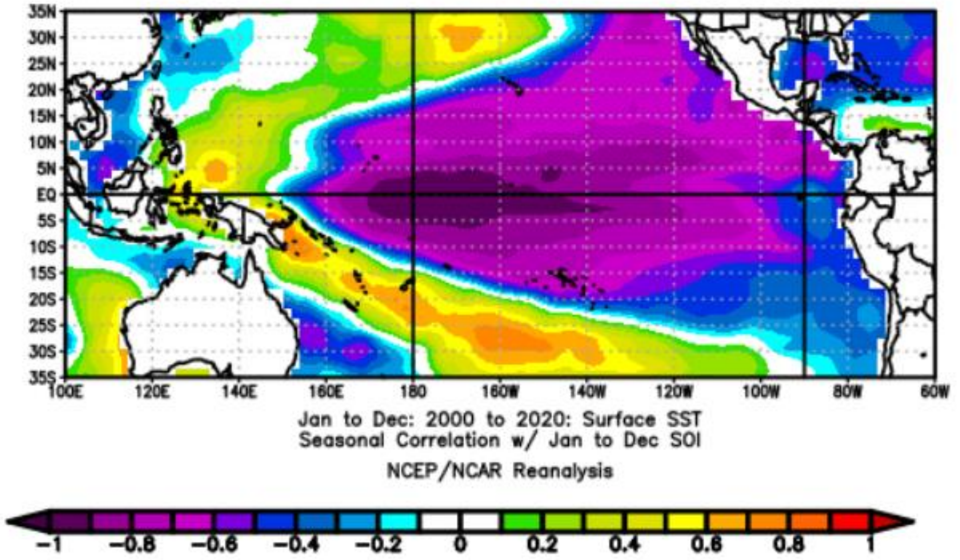
https://www.ncei.noaa.gov/access/monitoring/enso/soi

We can correlate the evolution of this index (timeseries)

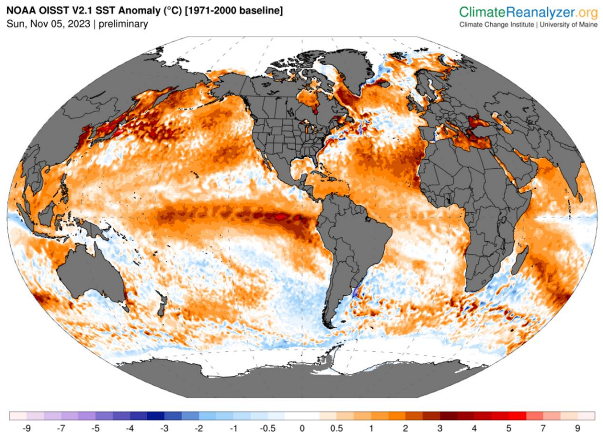


With the time evolution of several variables defined spatially

For example, SST defined spatially



NOAA Physical Sciences Laboratory



Interesting for the study of teleconnections!

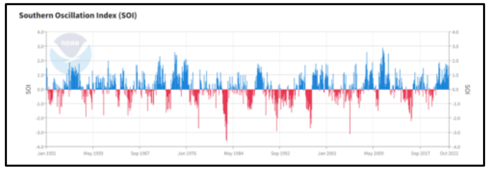
https://psl.noaa.gov/data/correlation/

Spatial analysis

Southern Oscillation Index (SOI)

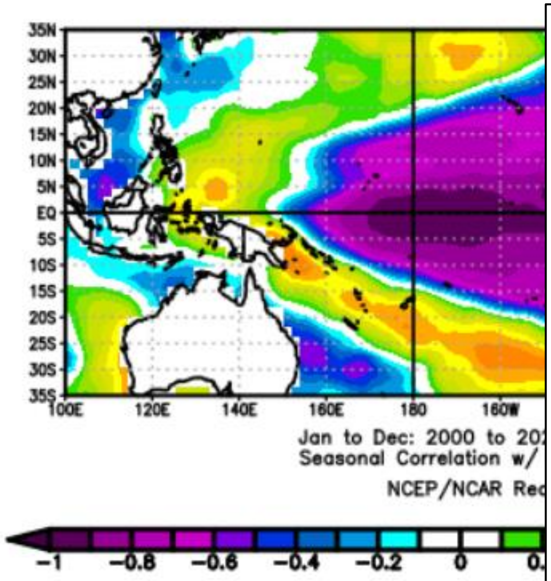
The Southern Oscillation Index (SOI) is a standardized index based on the observed sea level pressure (SLP) differences between Tahiti and Darwin, Australia. The SOI is one measure of the large-scale fluctuations in air pressure occurring between the western and eastern tropical Pacific (i.e., the state of the Southern Oscillation) during El Niño and La Niña episodes. In general, smoothed time series of the SOI correspond very well with changes in ocean temperatures across the eastern tropical Pacific. The negative phase of the SOI represents below-normal air pressure at Tahiti and above-normal air pressure at Darwin. Prolonged periods of negative (positive) SOI values coincide with abnormally warm (cold) ocean waters across the eastern tropical Pacific typical of El Niño (La Niña) episodes. The methodology used to calculate SOI is available below. More information can be found at the Climate Prediction Center SOI page.

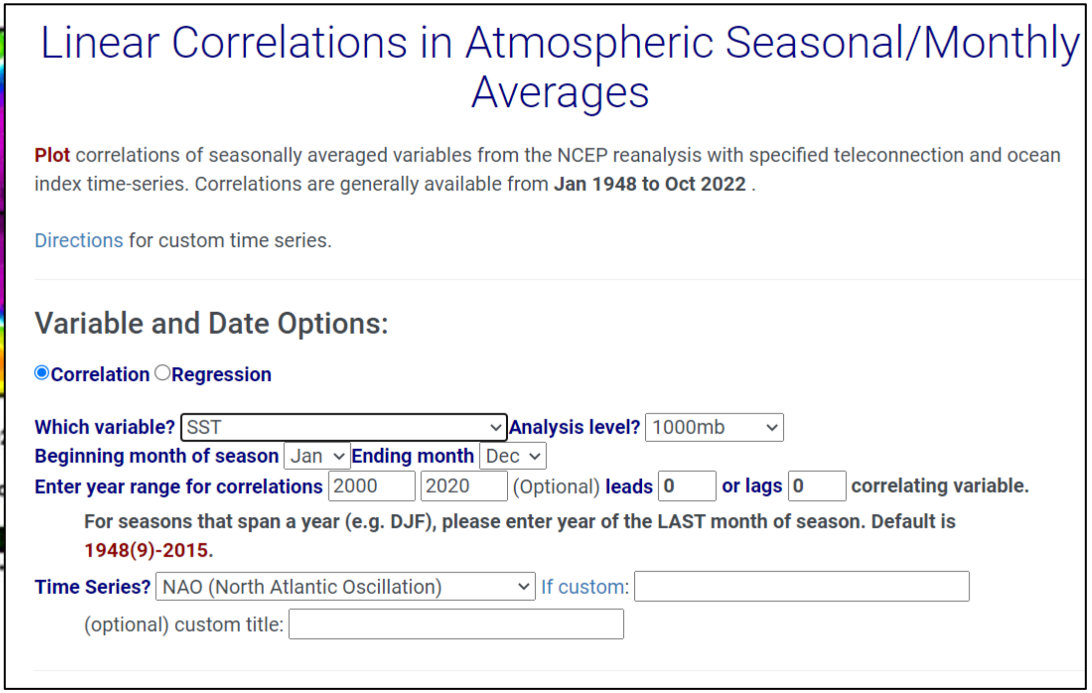
We can correlate the evolution of this index



With the time evolution of several variables defined spatially

For example SST…



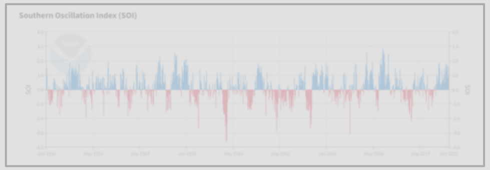


https://psl.noaa.gov/data/correlation/ ⇐ very nice website to try these correlations…

Spatial analysis

Southern Oscillation Index (SOI)

We can correlate the evolution of this index



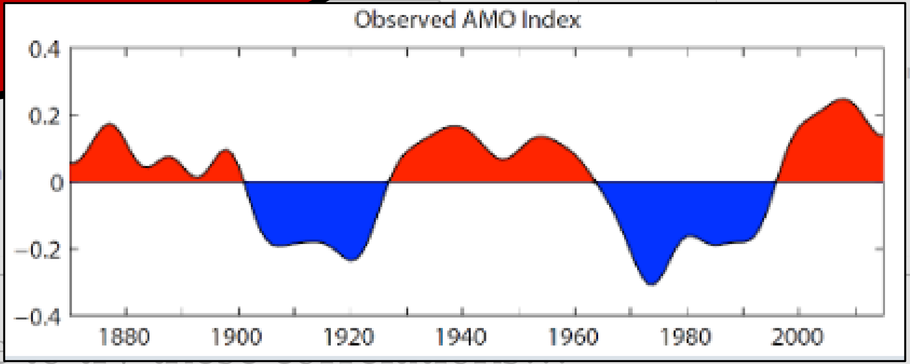
With the time evolution of several variables defined spatially

For example SST…

TAKE CARE WITH CORRELATIONS!!

~~***cum hoc ergo propter hoc***~~
~~('with this, therefore because of this')~~

Correlation does not imply causation!



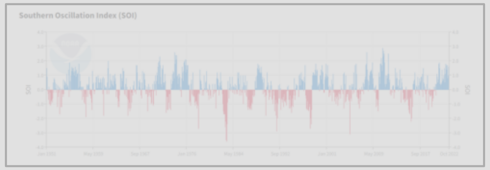
https://psl.noaa.gov… nice website…

Spatial analysis

Southern Oscillation Index (SOI)

…tandardized index based on the observed sea level pressure (SLP) differences between …e of the large-scale fluctuations in air pressure occurring between the western and …llation) during El Niño and La Niña episodes. In general, smoothed time …mperatures across the eastern tropical Pacific. The negative phase of …ormal air pressure at Darwin. Prolonged periods of negative …s across the eastern tropical Pacific typical of El Niño (La Niña) …information can be found at the Climate Prediction Center

We can correlate the evolution of this index



With the time evolution of several variables defined spatially

For example SST…

TAKE CARE WITH CORRELATIONS!!

cum hoc ergo propter hoc

('with this, therefore because of this')

Correlation does not imply causation!

…eric Seasonal/Monthly

…nalysis with specified teleconnection and ocean …3 to Oct 2022 .

Analysis level? 1000mb

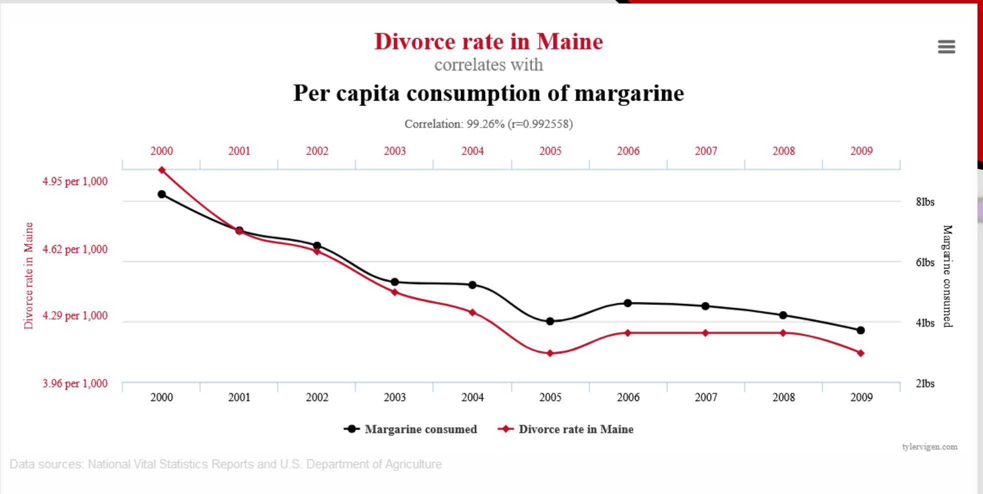
Ending month Dec

…correlations 2000 2020 (Optional) leads 0 or lags 0 correlating variable.

For seasons that span a year (e.g. DJF), please enter year of the LAST month of season. Default is 1948(9)-2015.

Time Series? NAO (North Atlantic Oscillation) If custom:

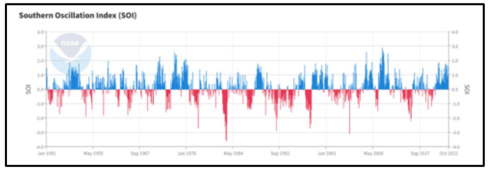
(optional) custom title:

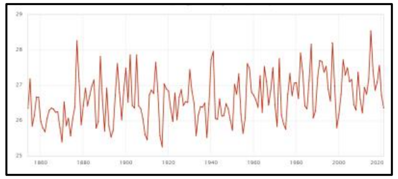


very r…

https://www.tylervigen.com/spurious-correlations

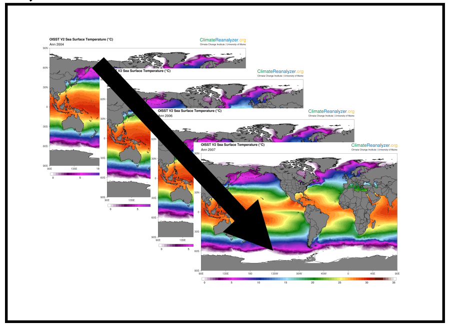
Spatial analysis





We were correlating a vector (SOI evolution) with another vector (SST evolution)…

but we had this last vector defined for each grid point of the map



Maybe we want to correlate two spatially-distributed variables
(like SST … and another one (for example the wind speed timeseries in each point))

….

Multivariate analysis

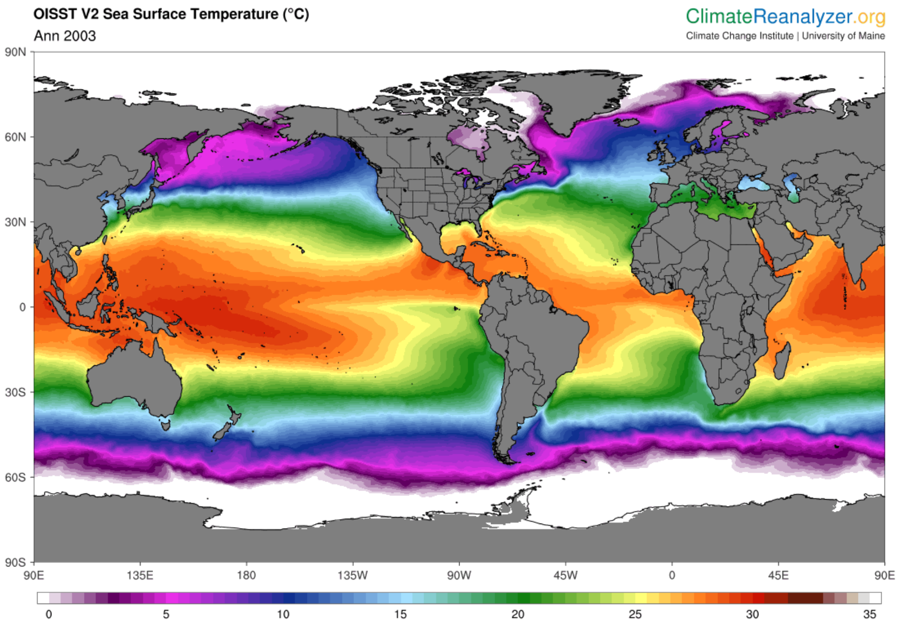
How to deal with the spatial + temporal variability?

Empirical Orthogonal Functions (EOFs)

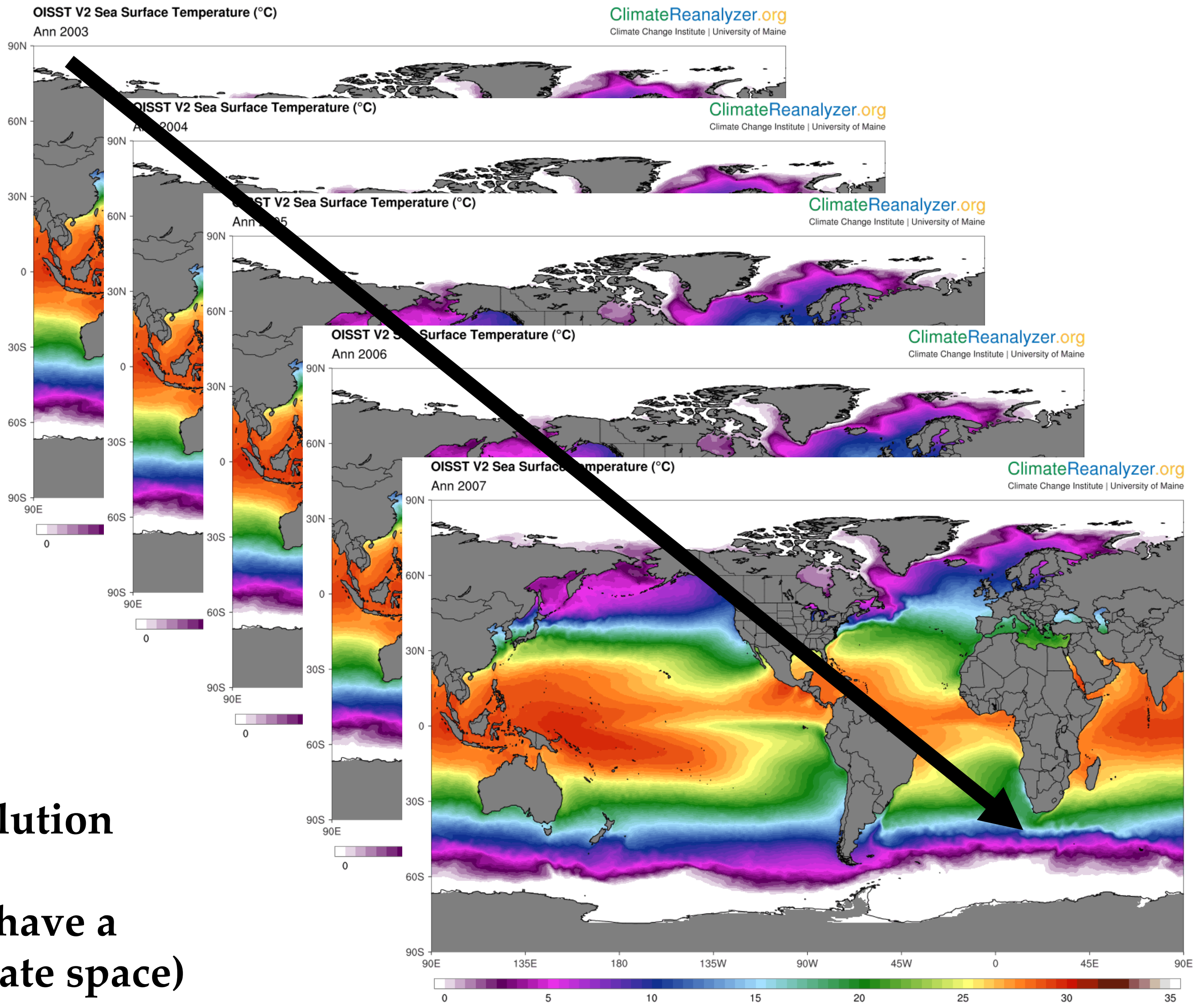
- also known as Principal Component Analysis (PCA) -

Compact description of the spatial and temporal variability (modes)

Defined by the covariance


OISST V2 Sea Surface Temperature (°C)
Ann 2003
ClimateReanalyzer.org
Climate Change Institute | University of Maine
90N
60N
30N
0
30S
60S
90S
90E
135E
180
135W
90W
45W
0
45E
90E
0
5
10
15
20
25
30
35

Spatial analysis — EOFs (Empirical Ortogonal Functions)



Climate is too complex and can vary in several ways….

Studying all the possible variations is too complicated…

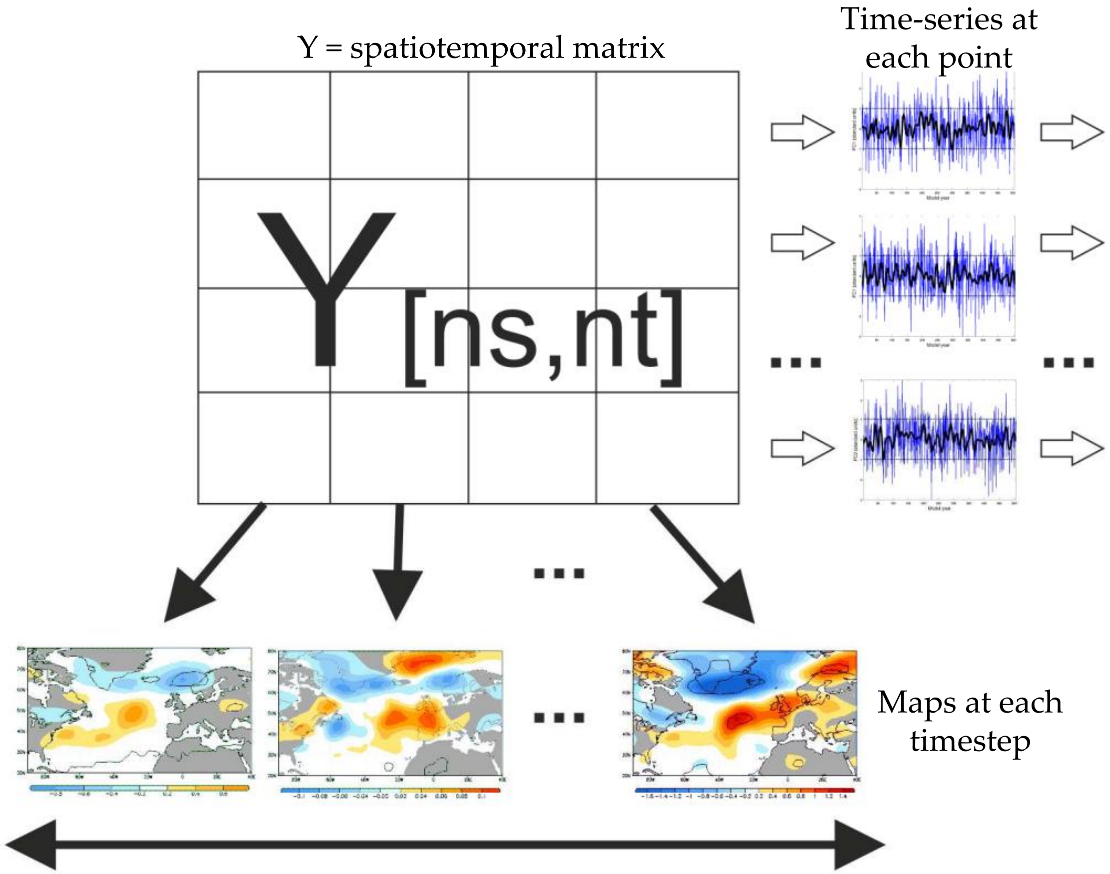
Never too years are going to be the same…

But they can be similar…

Time evolution

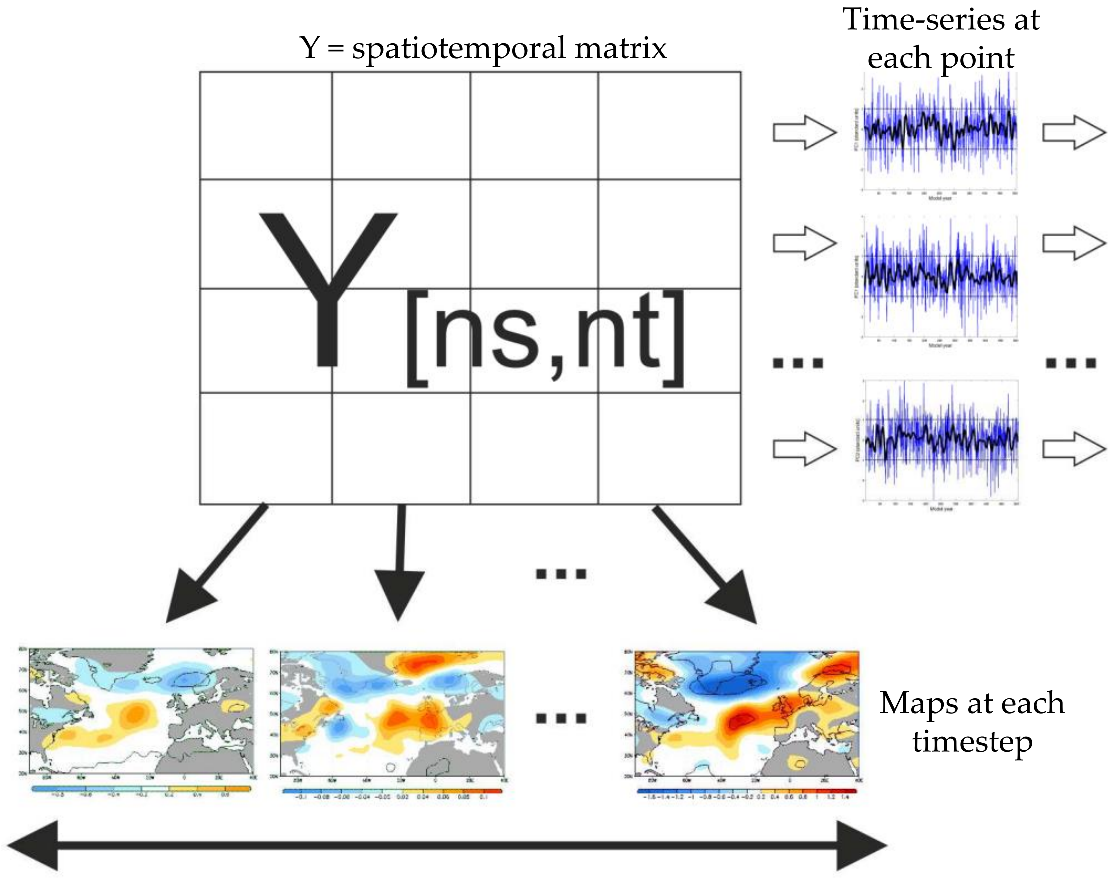
(now we have a multivariate space)

There are statistical methods to deal with this complexity (EOF)



From Gómara (2015), adapted from Rodríguez-Fonseca (2001)

There are statistical methods to deal with this complexity (EOF)



From Gómara (2015), adapted from Rodríguez-Fonseca (2001)

Mathematical calculation related to diagonalisation of covariance matrix….

$$C = \frac{1}{nt} Y'[ns,nt] \cdot Y'^{T}[nt,ns]$$

In this case, Y'[ns,nt] is the anomaly of the two-dimensional (space) variable
And Y'^{T}[ns,nt] is the same at other time

$$(C - \lambda_i I)\vec{e_i} = 0$$

Diagonalization of the matrix of covariances, with eigenvectors (e_i) and eigenvalues (λ_i)

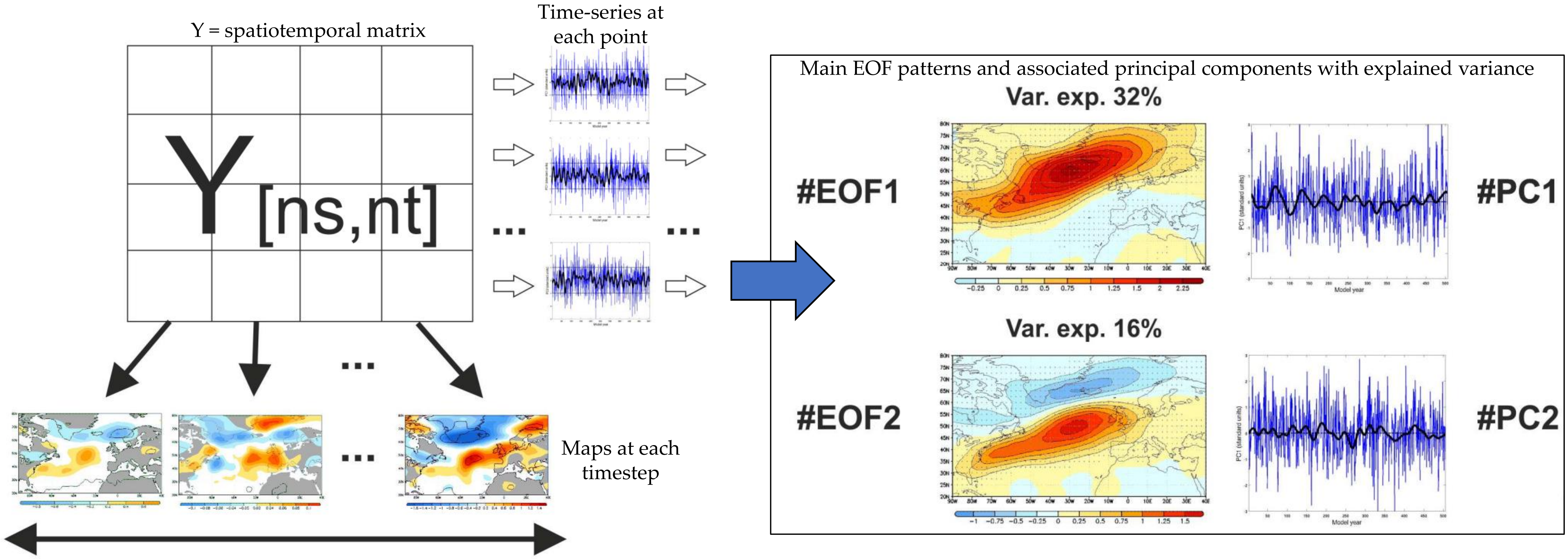
$$\overrightarrow{\alpha_i[nt,1]} = Y'^{T}[nt,ns] \cdot \overrightarrow{e_i[ns,1]}$$

The PC are calculated as the projection of the anomalous field (Y') onto the corresponding eigenvector (EOF)

$$\overrightarrow{\alpha_i[nt,1]} = Y'^{T}[nt,ns] \cdot \overrightarrow{e_i[ns,1]}$$

This is a clean way to analyse the variability. With a few "modes" we can describe almost the whole system. Maybe we want to correlate the evolution of one of the main modes with something… (as we did for AIO but using spatial structures).

We <u>reduce</u> one of the dimensions of the data
(removing redundant information in the time-evolving fields)



From Gómara (2015), adapted from Rodríguez-Fonseca (2001)

4. Time series analysis

We see a cyclic evolution…

What kind of signal is in my data?

How can we analyse their frequency content?

WHAT DO WE WANT?

Periodic variability of the timeseries – separate it from other aperiodic fluctuations

How is its spectra?

Frequency (time) - Wavelength (space)

WHAT DO WE WANT?

Periodic variability of the timeseries – separate it from other aperiodic fluctuations

How is its spectra?

Frequency (time) - Wavelength (space)

WHAT DO WE NEED?

Long timeseries with the appropriate time resolution (Sampling theorem)

"the highest detectable frequency or wavenumber (Nyquist) is determined by the data separation (frequency)"

10 min data → one cycle each 20 min (highest detectable freq.)

WHAT DO WE WANT?

Periodic variability of the timeseries – separate it from other aperiodic fluctuations

How is its spectra?

Frequency (time) - Wavelength (space)

WHAT DO WE NEED?

Long timeseries with the appropriate time resolution (Sampling theorem)

"the highest detectable frequency or wavenumber (Nyquist) is determined by the data separation (frequency)"

10 min data → one cycle each 20 min (highest detectable freq.)

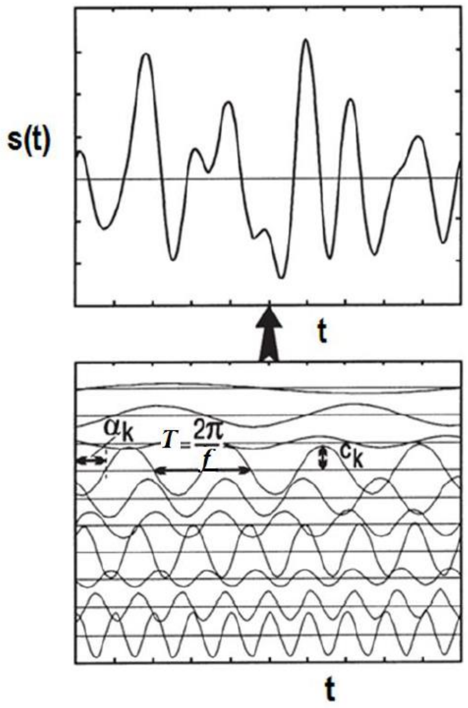
WHAT SHOULD WE USE?

Fourier analysis
Harmonic analysis
Spectral analysis
Wavelet…

A time series is like a combination of periodic (or almost) components + noise (+ trends)

Periodic components → almost fixed amplitudes and phases

Fourier analysis → Identification of these components of a timeseries

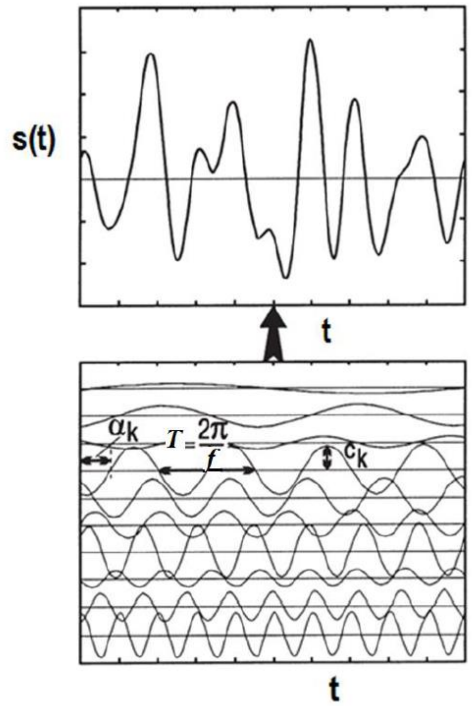


A time series is like a combination of periodic (or almost) components + noise (+ trends)

Periodic components → almost fixed amplitudes and phases

Fourier analysis → Identification of these components of a timeseries

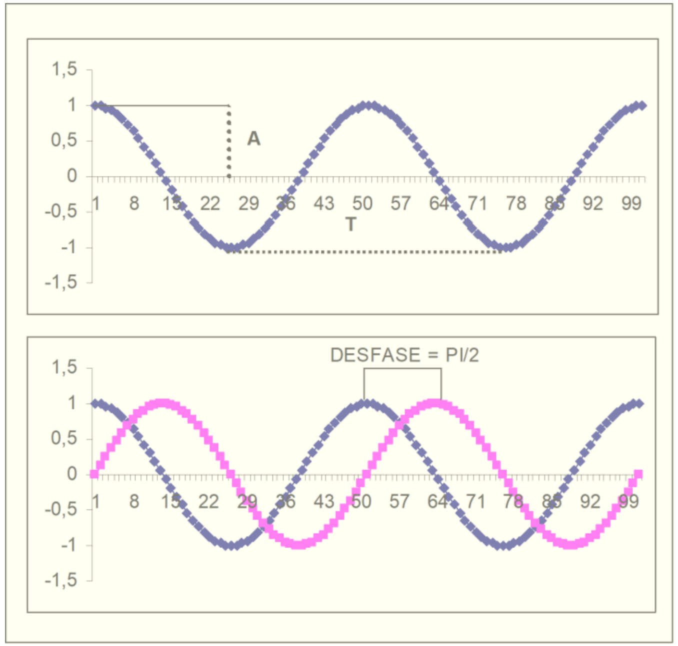
- A, amplitude.
- T, period.
- θ, lag.



If Y is periodic → Y(t) = Y(t + T)

$$Y_t = A\cos(\frac{2\pi t}{T} + \theta)$$

$$Y_t = a\cos(\omega t) + b\,\mathrm{sen}(\omega t)$$

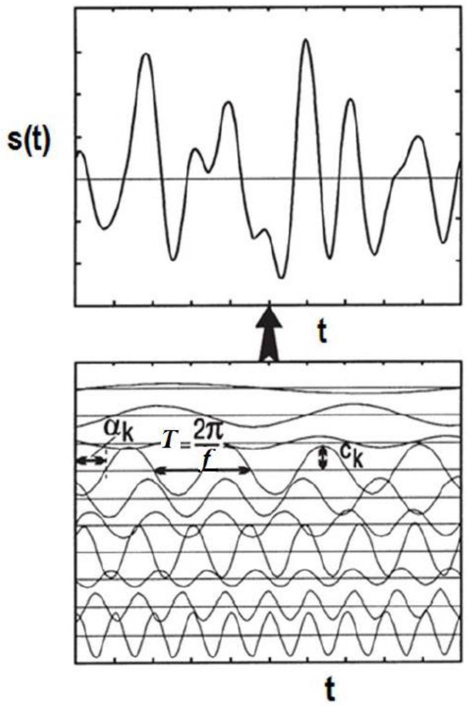


A time series is like a combination of periodic (or almost) components + noise (+ trends)

Periodic components → almost fixed amplitudes and phases

Fourier analysis → Identification of these components of a timeseries

- A, amplitude.
- T, period.
- θ, lag.

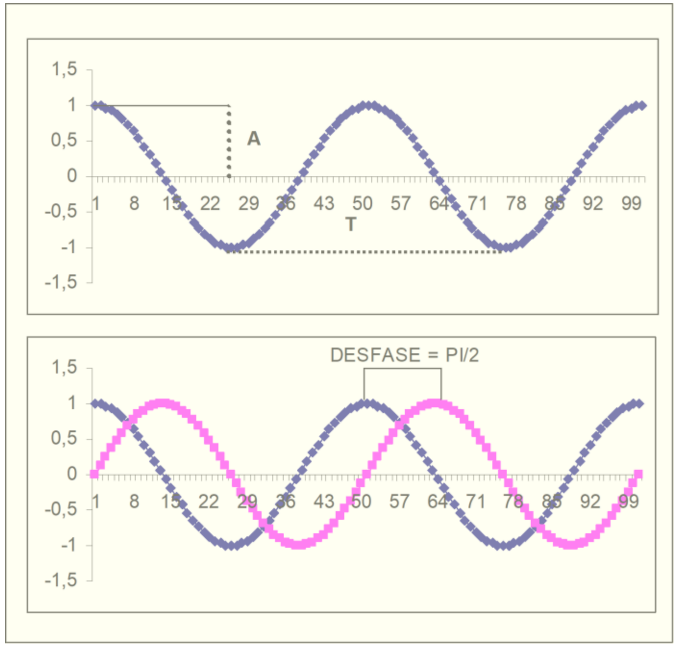


If Y is periodic → Y(t) = Y(t + T)

$$Y_t = A\cos\left(\frac{2\pi t}{T} + \theta\right)$$

$$Y_t = a\cos(\omega t) + b\,\mathrm{sen}(\omega t)$$

CALCULATION OF COEFFICIENTS!



CALCULATION OF COEFFICIENTS!

$$Y_t = a\cos(\omega t) + b\,\mathrm{sen}(\omega t)$$

$$a_k = \frac{2}{T}\int_0^T s(t)\cos\left(\frac{k2\pi t}{T}\right)dt$$

$$b_k = \frac{2}{T}\int_0^T s(t)\,\mathrm{sen}\left(\frac{k2\pi t}{T}\right)dt$$

T = (N-1)Δt is the sampling length… k = 0, 1, 2 N/2…

$$a_k = \frac{2}{(N-1)}\sum_{i=0}^{N} s(i\Delta t)\cos\left(\frac{k}{(N-1)\Delta t}2\pi i\Delta t\right)$$

$$b_k = \frac{2}{(N-1)}\sum_{i=0}^{N} s(i\Delta t)\,\mathrm{sen}\left(\frac{k}{(N-1)\Delta t}2\pi i\Delta t\right)$$

See very clear example in Emery and Thomson (2014)

TABLE 5.11 Monthly Mean Sea Surface Temperature (SST) (°C) at Amphitrite Point (48° 55.16′ N, 125° 32.17′ W) on the West Coast of Canada for January 1982 through December 1983

YEAR 1982												
n	1	2	3	4	5	6	7	8	9	10	11	12
SST	7.6	7.4	8.2	9.2	10.2	11.5	12.4	13.4	13.7	11.8	10.1	9.0
YEAR 1983												
n	13	14	15	16	17	18	19	20	21	22	23	24
SST	8.9	9.5	10.6	11.4	12.9	12.7	13.9	14.2	13.5	11.4	10.9	8.1

TABLE 5.12 Fourier Coefficients and Frequencies for the Amphitrite Point Monthly Mean Temperature Data

p	Frequency (cpmo)	Period (month)	Coefficient A_p (°C)	Coefficient B_p (°C)	Coefficient C_p (°C)	Phase θ_p (degrees)
0	0	–	21.89	0	21.89	0
1	0.042	24	−0.55	−0.90	1.05	−121.4
2	0.083	12	−1.77	−1.99	2.67	−131.7
3	0.125	8	0.22	−0.04	0.23	−10.3
4	0.167	6	−0.44	−0.06	0.45	−172.2
5	0.208	4.8	0.09	−0.07	0.11	−37.9
6	0.250	4	0.08	−0.04	0.09	−26.6
7	0.292	3.4	0.01	−0.16	0.16	−58.0
8	0.333	3	−0.03	−0.16	0.16	−100.6
9	0.375	2.7	−0.14	0.05	0.15	160.3
10	0.417	2.4	−0.09	−0.07	0.11	−142.1
11	0.458	2.2	−0.08	−0.12	0.14	−123.7
12	0.500	2	−0.15	0	0.15	0

Frequency is in cycles per month (cpmo). $A_0/2$ is the mean temperature and θ_p is the phase lag for the pth component taken counterclockwise from the positive Ap axis.

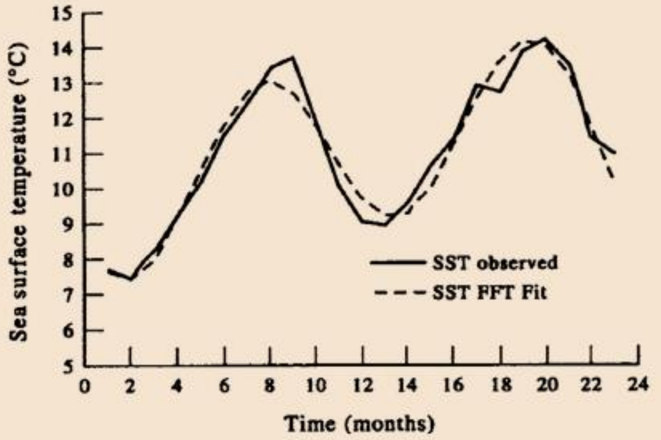
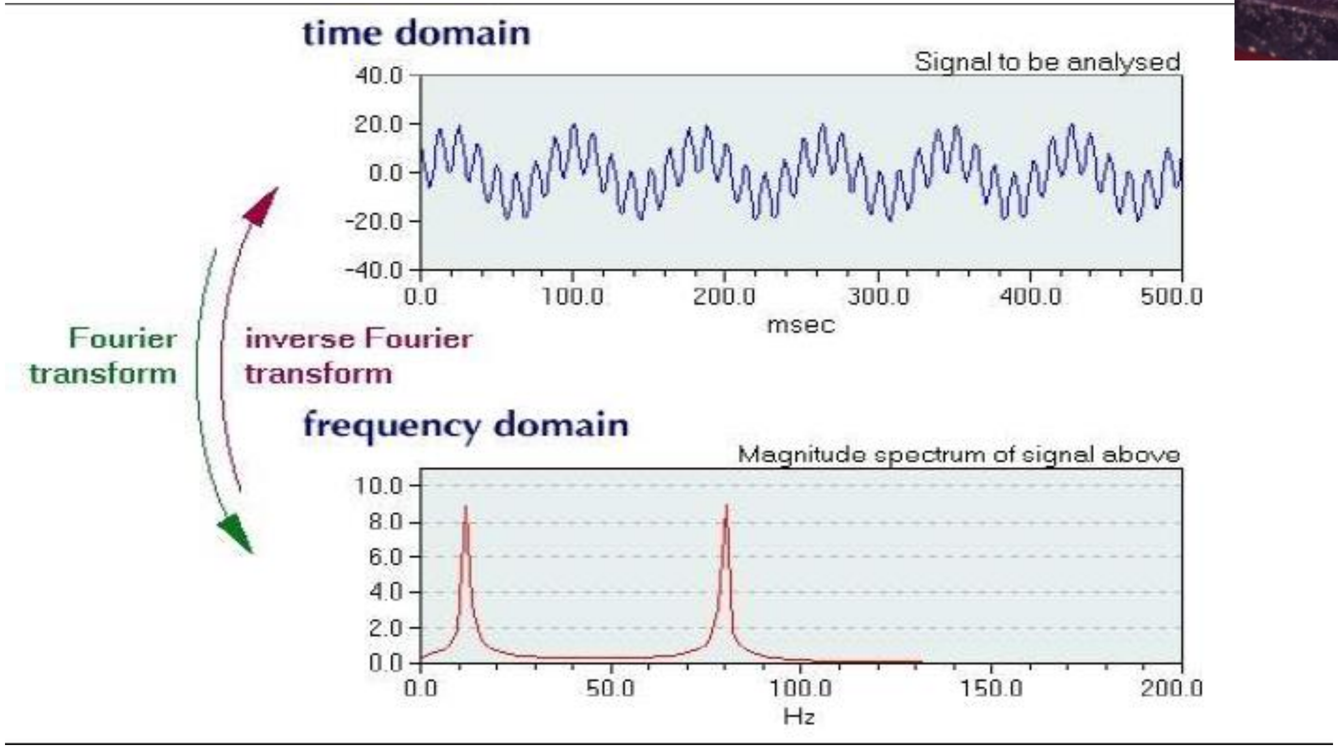


FIGURE 5.57 Monthly mean sea surface temperature (SST) record for Amphitrite Point on the west coast of Vancouver Island (see Table 5.11). The bold line is the original 24-month series; the dashed line is the SST time series generated using the first three Fourier components, f_p, $p = 0, 1, 2$, corresponding to the mean, 24-month, and 12-month cycles (Fourier components appear in Table 5.12). FFT, fast Fourier transform.

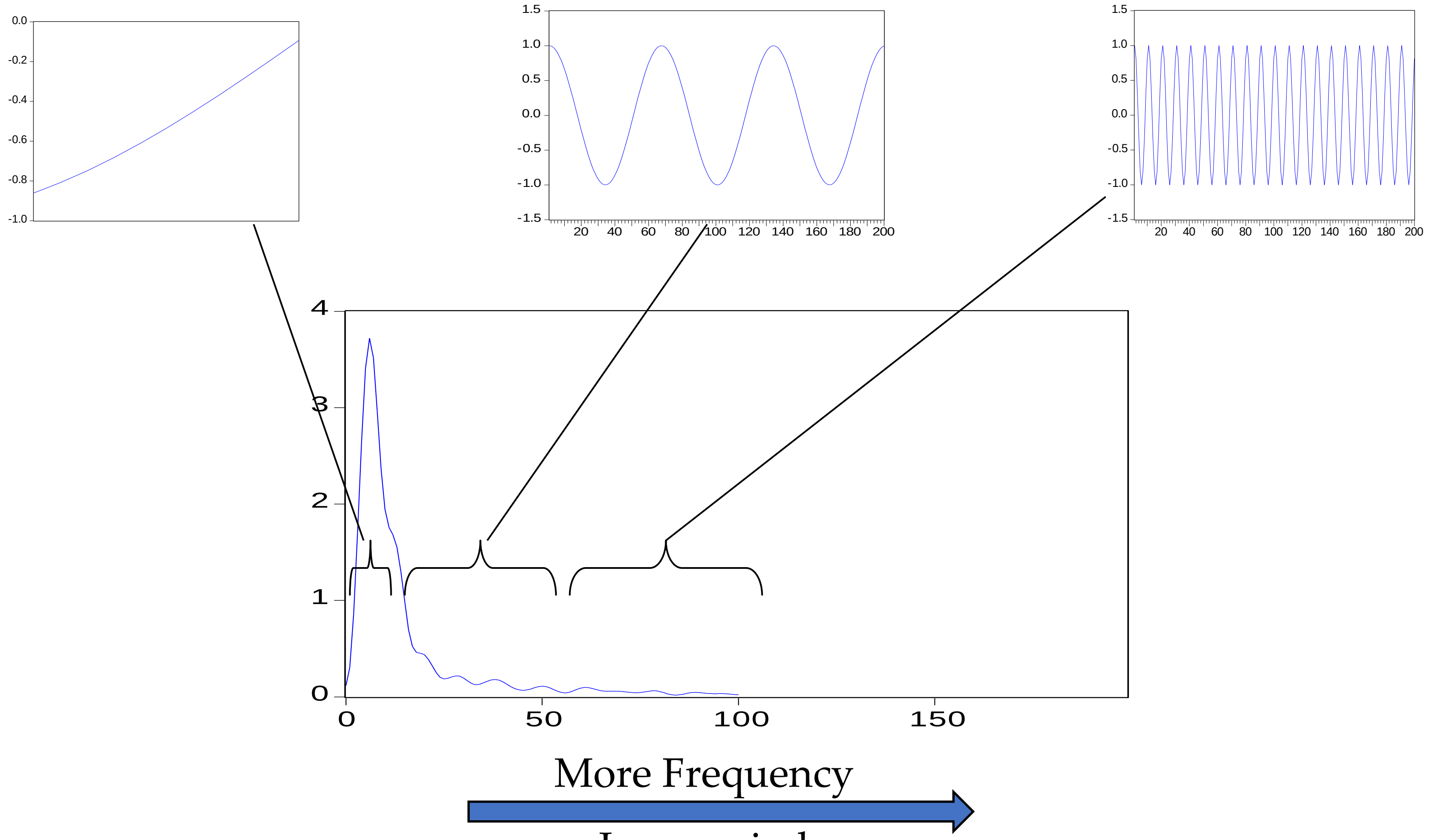
Nyquist freq. = $fN = 1/(2\Delta t)$ = 1/(2*1) = 0.5 (cycles/month) (p=12) (remember we have 24 data)

Fundamental freq. = f1 = 1/24 = 0.042 (p=1)

Periodogram



Periodogram



Periodogram: formulation…

Our timeseries seems to follow this "model"…

$$Y_t = \sum_{i=1}^{k} (a_p \cos \omega_i t + b_i \operatorname{sen} \omega_i t) + \varepsilon_t$$

We assume that our frequencies are…

$$\omega_i = \frac{2\pi p_i}{N} \qquad p_i = 1,\ldots,k$$

The parameters a and b are determined as:

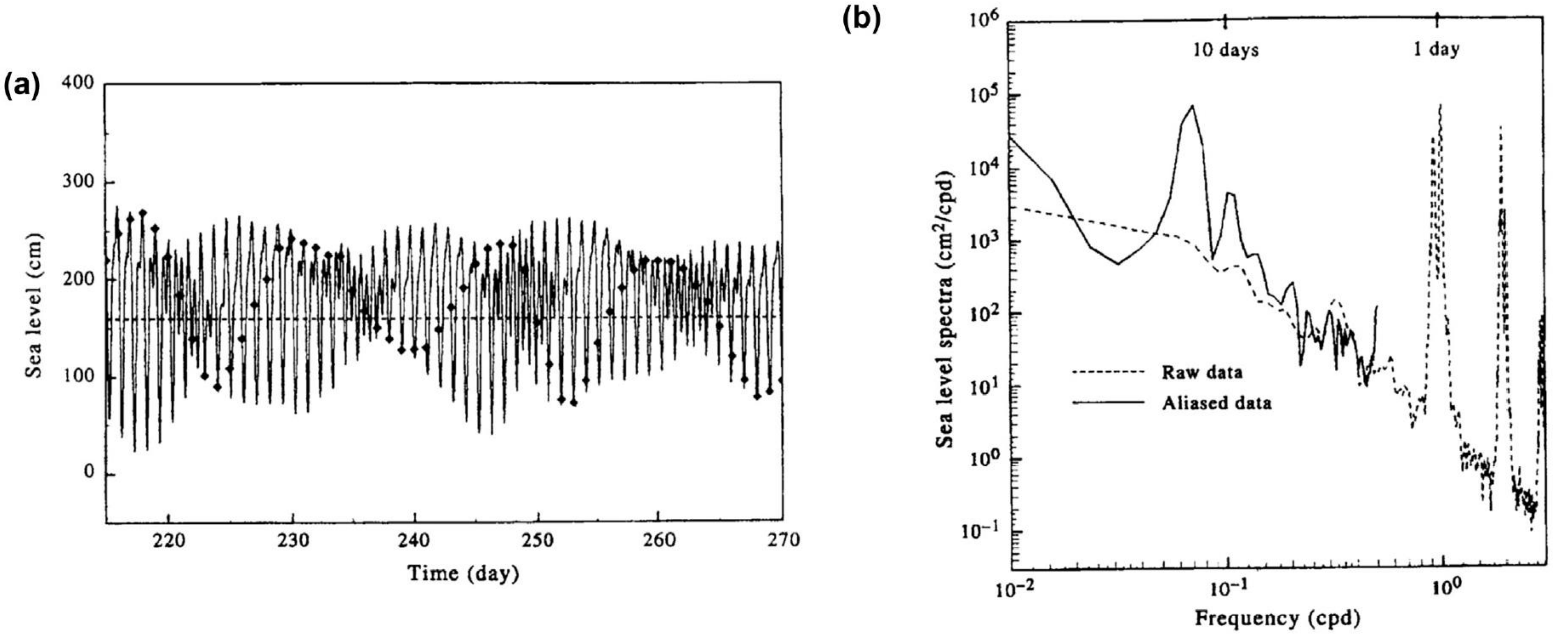
$$\hat{a}_p = \frac{2}{N} \sum_{t=1}^{N} Y_t \cos p\omega_o t \qquad \hat{a}_0 = \sum_{t=1}^{N} \frac{Y_t}{N} \qquad \hat{a}_{N/2} = \frac{1}{N} \sum_{t=1}^{N} Y_t \cos \pi t$$

$$\hat{b}_p = \frac{2}{N} \sum_{t=1}^{N} Y_t \operatorname{sen} p\omega_o t$$

And finally, we calculate our periodogram:

$$I(\omega_p) = \frac{(a_p^2 + b_p^2)}{2\omega_0}$$

Aliasing



We lose frequencies higher than Nyquist freq. and their energy is integrated in larger frequencies…

FIGURE 5.18 The origin of aliasing. (a) The solid line is the tide height recorded at Victoria, British Columbia over a 60-day period from July 29 to September 27, 1975 (time in Julian days). The diamonds are the sea-level values one would obtain by only sampling once per day. (b) The power spectrum obtained from the two data series in (a). In this case, the high frequency energy (dashed curve) gets folded back into the spectrum at lower (aliased) frequencies (solid curve).

From Emery and Thomson (2014)

We need to adjust the record length to match the desired Fourier components

Tidal frequencies → Integer multiples of the fundamental freq. (1/T)
Use Fourier analysis to find constituent amplitudes and phases →

TABLE 5.12 Fourier Coefficients and Frequencies for the Amphitrite Point Monthly Mean Temperature Data

p	Frequency (cpmo)	Period (month)	Coefficient A_p (°C)	Coefficient B_p (°C)	Coefficient C_p (°C)	Phase θ_p (degrees)
0	0	–	21.89	0	21.89	0
1	0.042	24	−0.55	−0.90	1.05	−121.4
2	0.083	12	−1.77	−1.99	2.67	−131.7
3	0.125	8	0.22	−0.04	0.23	−10.3
4	0.167	6	−0.44	−0.06	0.45	−172.2
5	0.208	4.8	0.09	−0.07	0.11	−37.9
6	0.250	4	0.08	−0.04	0.09	−26.6
7	0.292	3.4	0.01	−0.16	0.16	−58.0
8	0.333	3	−0.03	−0.16	0.16	−100.6
9	0.375	2.7	−0.14	0.05	0.15	160.3
10	0.417	2.4	−0.09	−0.07	0.11	−142.1
11	0.458	2.2	−0.08	−0.12	0.14	−123.7
12	0.500	2	−0.15	0	0.15	0

Frequency is in cycles per month (cpmo). $A_0/2$ is the mean temperature and θ_p is the phase lag for the pth component taken counterclockwise from the positive Ap axis.

TIDAL CONSTITUENTS (M_2, K_1, S_2…)
The letter indicates the different types of tides in each frequency band.
The number, the cycles per lunar day

We need to adjust the record length to match the desired Fourier components

Tidal frequencies → Integer multiples of the fundamental freq. (1/T)
Use Fourier analysis to find constituent amplitudes and phases

TIDAL CONSTITUENTS (M_2, K_1, S_2…)
The letter indicates the different types of tides in each frequency band.
The number, the cycles per lunar day

↓

HARMONIC ANALYSIS
(the user specifies the frequencies to be examined!)
Application of least-square techniques to obtain the constituents
Find constituent amplitudes and phases → Use them for tidal predictions

Length of the record – IMPORTANT!

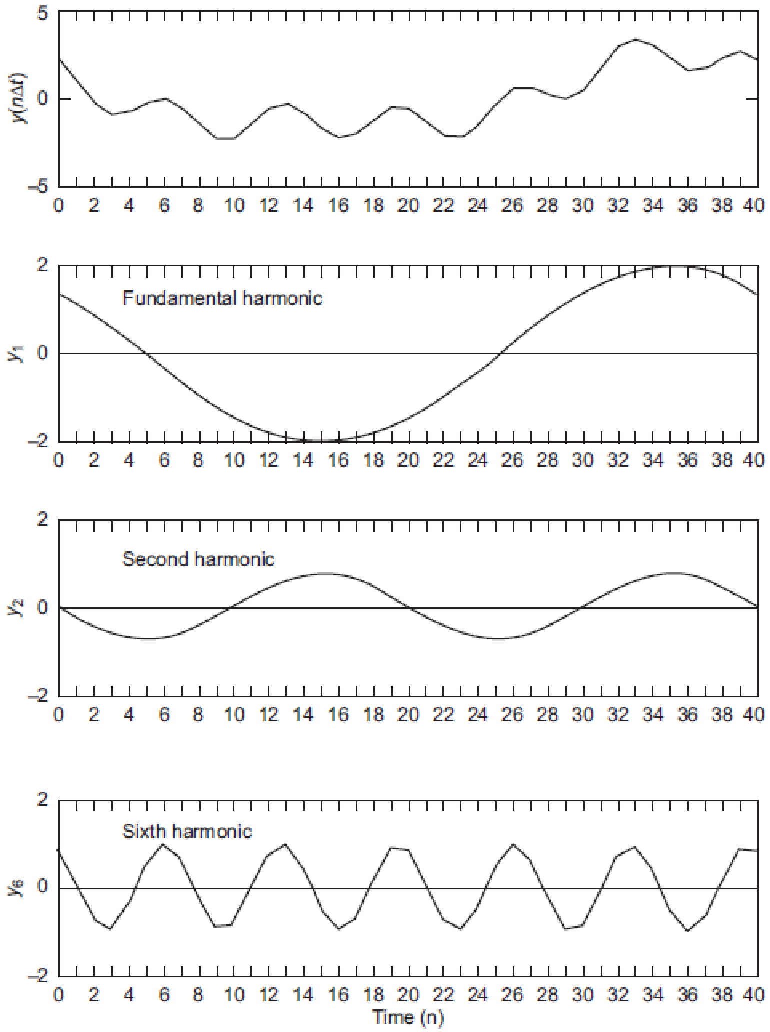
Principal lunar - one cycle of M_2 = 12.42 h (0.0805 cycles/hour = 1.93 cycles/day)

But we want to solve also the other constituents

ΔT = 1

N = 40

T = N ΔT = 40 (total lenght)



1/T harmonic (fundamental harmonic)

1 cycle every 40 n

2/T harmonic

2 cycles every 40 n

6/T harmonic

FIGURE 5.55 Discrete subsampling of a continuous signal, $y(t)$. The sampling interval is $\Delta t = 1$ time unit and the fundamental frequency is $f_1 = 1/T$, where $T = N\Delta t$ is the total record length and $N = 40$. The signal $y(t)$ is the sum of the first, second, and sixth harmonics which have the form $y_k(n\Delta t) = C_k \cos[(2\pi kn/N) + \phi_k]$; $k = 1, 2, 6$; $n = 0, 1, \ldots, 40$.

From Emery and Thomson (2014)

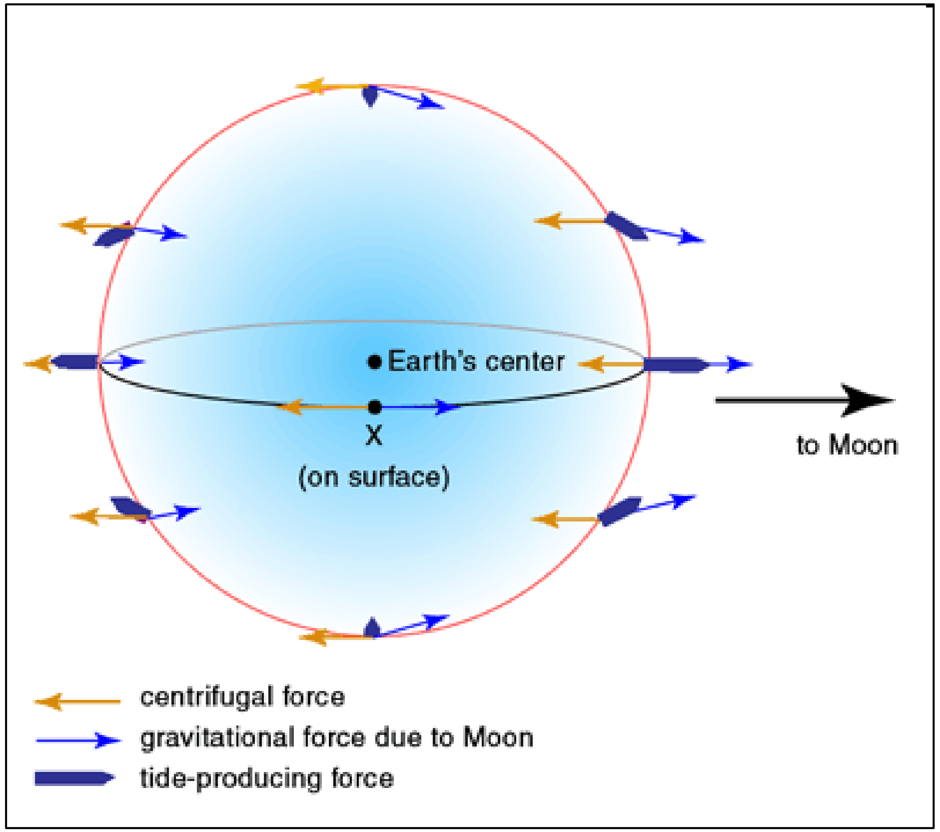
Newton Equilibrium Theory Hypothesis

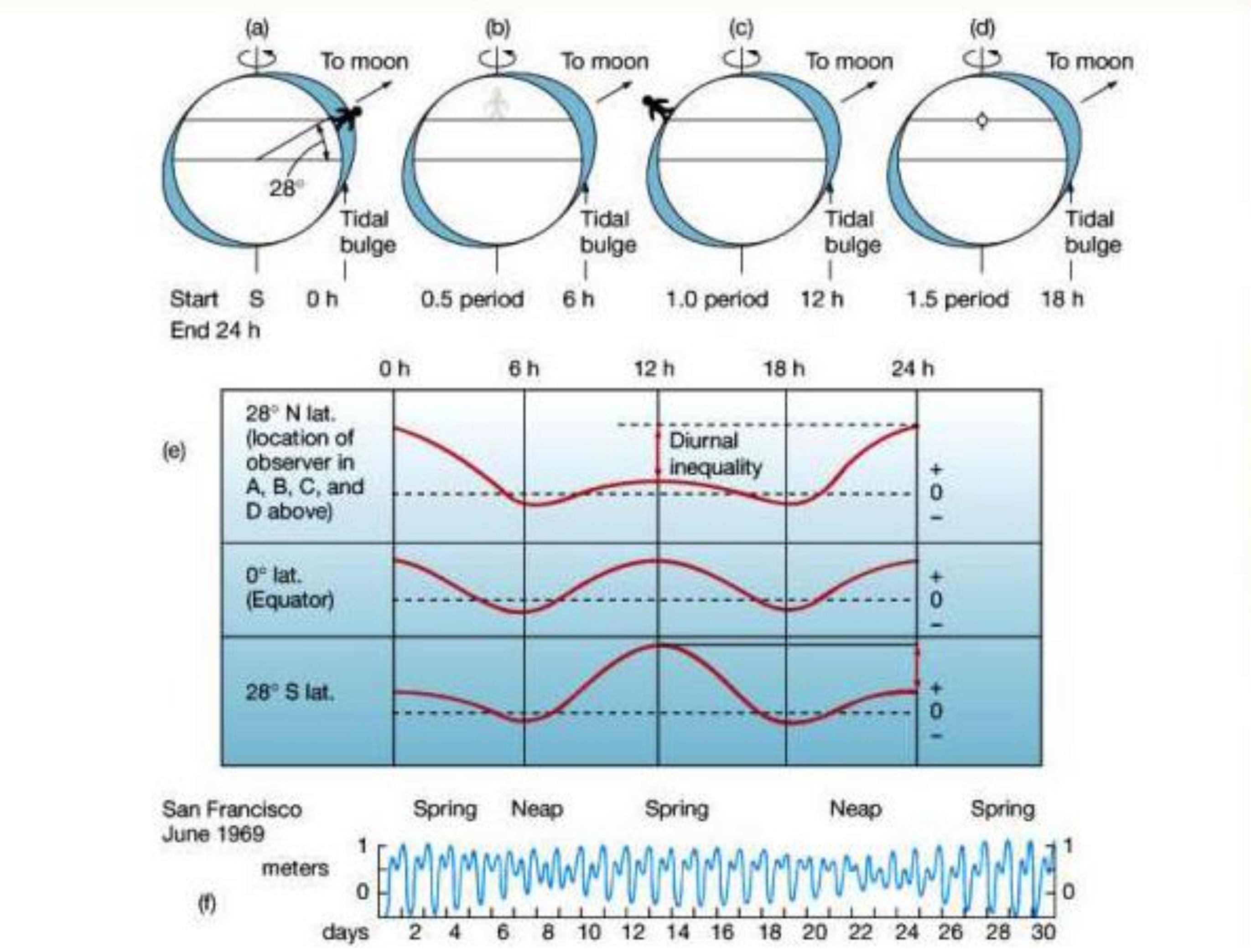
If the Earth is completely covered by water (no continents), and if the depth of the sea is enough to eliminate friction with the bottom surface, there should be an instantaneous response to the tidal forces.

TIDES
Gravitational + centrifugal forces

Earth rotation (24 hours)
Orbit of the Moon around Earth (27.32 days)
Orbit of Earth around Sun (365.25 days)

Tides effects:
- Periodic sea level changes
- Tidal currents

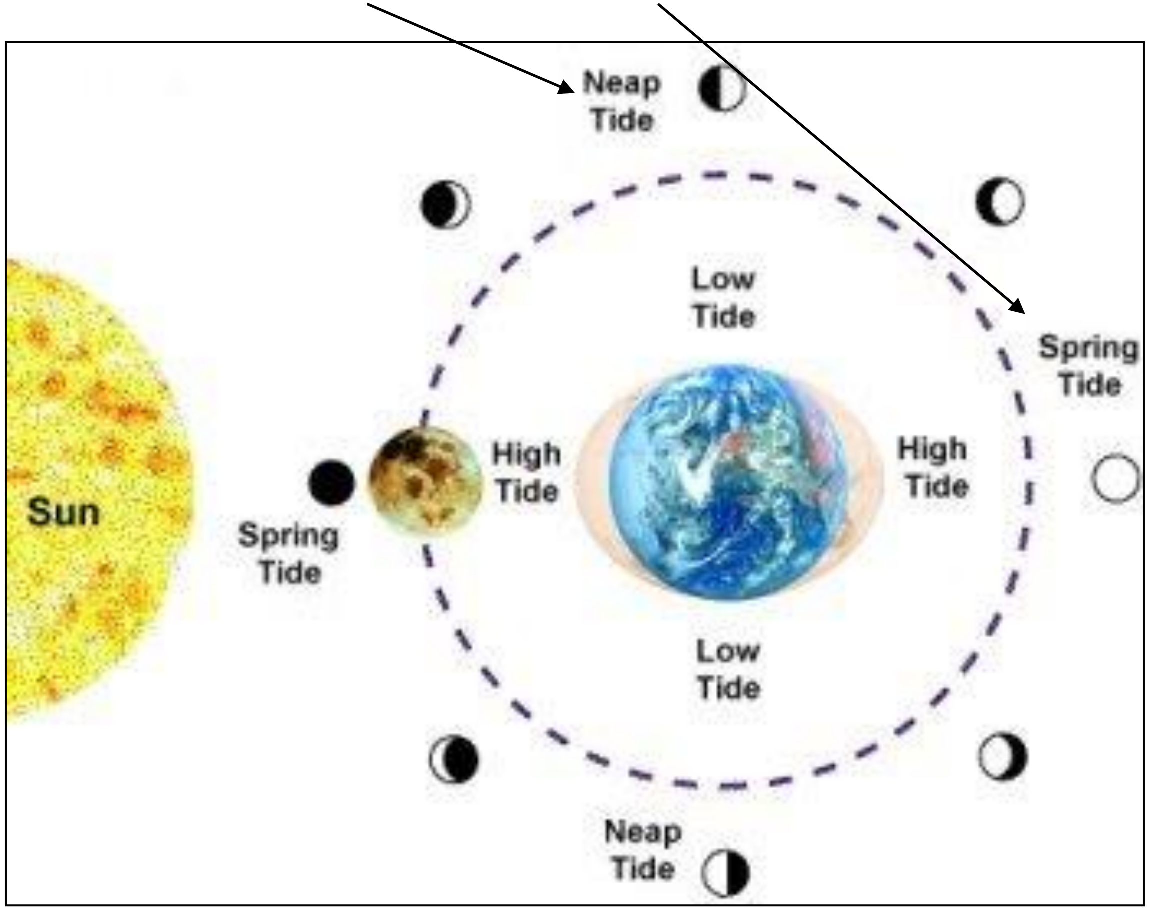




Earth rotation
Semidiurnal tide (12.4 h)

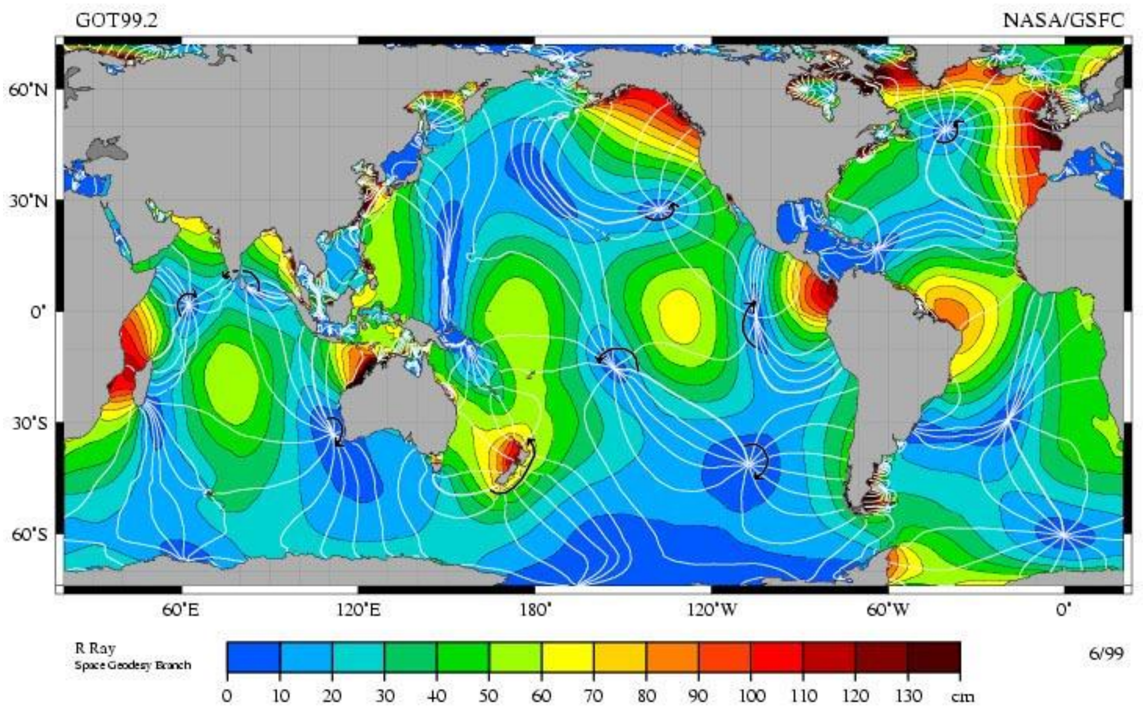
Moon-orbit inclination
Mixed and diurnal tide (24 h)

Sun + Moon effect
Neap and spring tides



Laplace (improvement of the Equilibrium Theory Hypothesis)

- **Finite depth**: obstacle for the immediate response
- **Coriolis effect**: wave rotated anticlockwise in the Northern Hemisphere.
- **Continents**: obstacle for wave propagation: reflection, diffraction, refraction.
- **Coast shallow waters**: waves amplification.



- Wavelength of thousands of kilometers: shallow water waves.
- Characteristics of progressive and stationary waves.
- Analysis through **harmonic decomposition** (simple waves of different periods).
- Each harmonic is characterized by its phase and amplitude.
- Represented in tidal maps.

The harmonic analysis from tides observational data allows for:

- Tidal prediction.
- Understand the tides at specific areas
- Interpretation of results in terms of hydrodynamics in the area

$$\sum_{i=1}^{M} A_i \cos(\omega_i t - \varphi_i)$$

A_i = Amplitude

ω_i = Angular velocity= $2\pi/T_i$

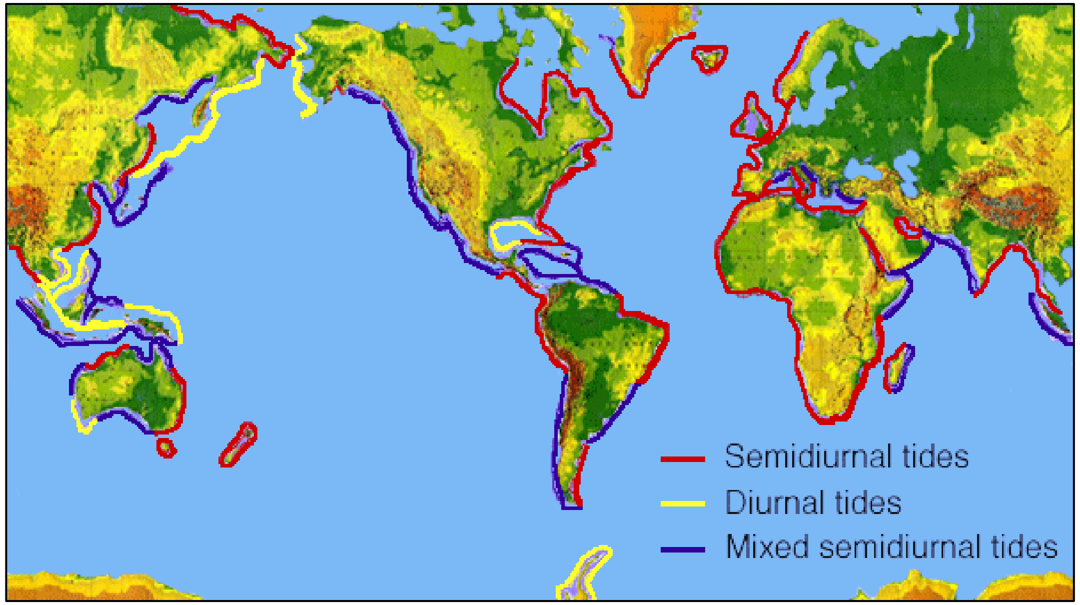
T_i = Constituent period

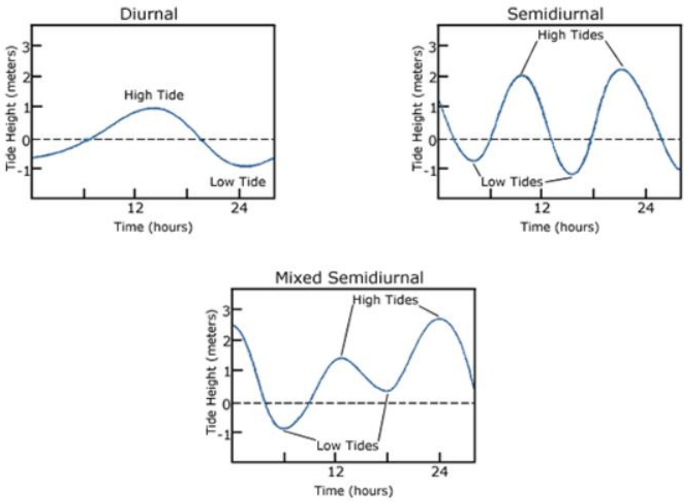
φ_i = Constituent phase

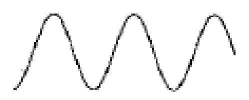
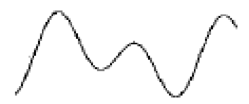
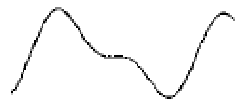
The tidal variations in a determined point can be represented as the sum of a finite number of harmonics since the angular frequency of these harmonics is known.

	Argument					Period (hsm)	Relative amplitude
	i_b	i_c	i_d	i_e	i_f		
Principal semi-diurnal							
M2	0	0	0	0	0	12,42	1,0000
S2	2	-2	0	0	0	12,00	0,4652
N2	-1	0	1	0	0	12,66	0,1915
Principal diurnal							
O1	-1	0	0	0	0	25,82	0,4151
K1	1	0	0	0	0	23,93	0,2921
P1	1	-2	0	0	0	24,06	0,1932
Long-period							
Mm	1	0	-1	0	0	27,55 dsm	0,0909
Msf	2	0	0	0	0	14,76 dsm	0,1723

Principal harmonic constituents of tides





Tidal Type	Form Number	Typical Form
Semidiurnal Tides	Less than 0.25	
Mixed, Semidiurnal	0.25 – 1.5	
Mixed, Diurnal	1.5 – 3.0	
Diurnal Tides	More than 3.0	

$$F = \frac{K_1 + O_1}{M_2 + S_2}$$

Input

Data!

→ Least-square adjustment →

Output

Amplitude and phase of the harmonic constituents

(Z_0, A_i, φ_i)

The number of the astronomic constituents depends on the length of the timeseries.

The least-square adjustment is based on the minimization of the sum of the squared residual terms.

$$R^2 = \left\{ \sum_{n=1}^{N} \eta_{obs}(t_n) - \left[C.\cos(\omega t_n) + S.\sin(\omega t_n) \right] \right\}^2$$

Input

Amplitude and phase of the harmonic constituents

→ Sum of waves →

Output

Sea-level evolution due to tides (prediction)

The more constituents are used, the better the prediction

Input

Amplitude and phase of the harmonic constituents → Sum of waves →

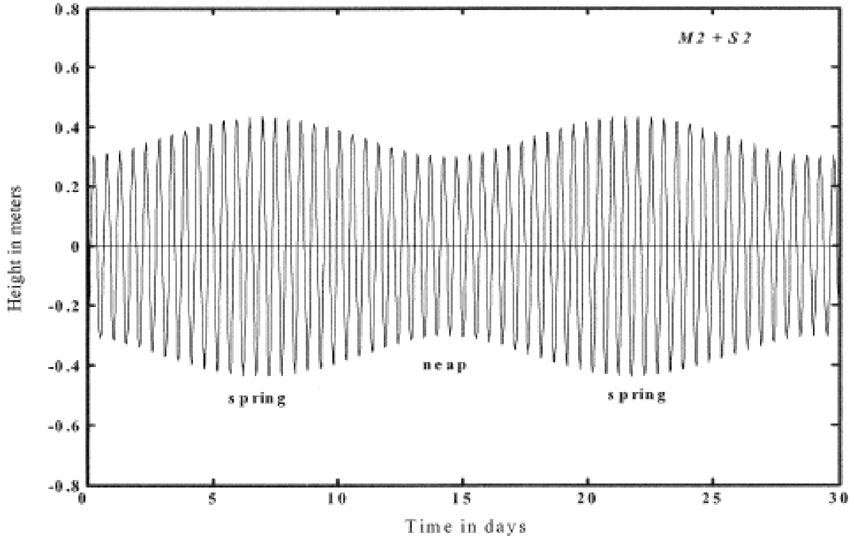
Output

Sea-level evolution due to tides (prediction)

The more constituents are used, the better the prediction

Moon and Solar largest constituents

$M_2 + S_2$



Input

Amplitude and phase of the harmonic constituents

→ Sum of waves →

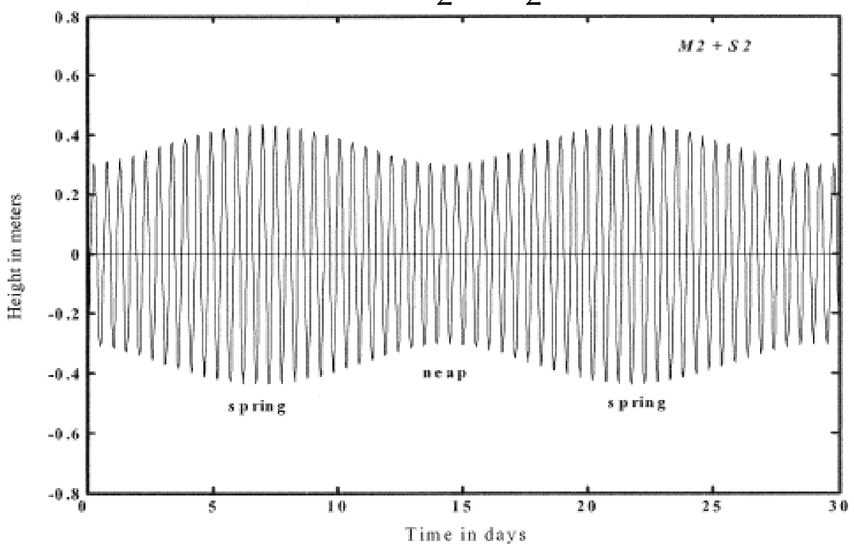
Output

Sea-level evolution due to tides (prediction)

The more constituents are used, the better the prediction

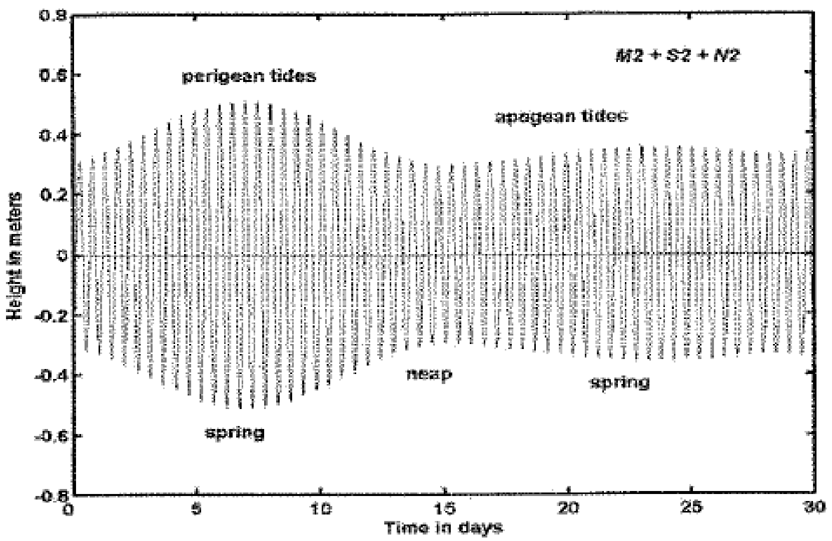
Moon and Solar largest constituents

$M_2 + S_2$



Effect of apogee and perigee

$M_2 + S_2 + N_2$



Input

Amplitude and phase of the harmonic constituents

→ Sum of waves →

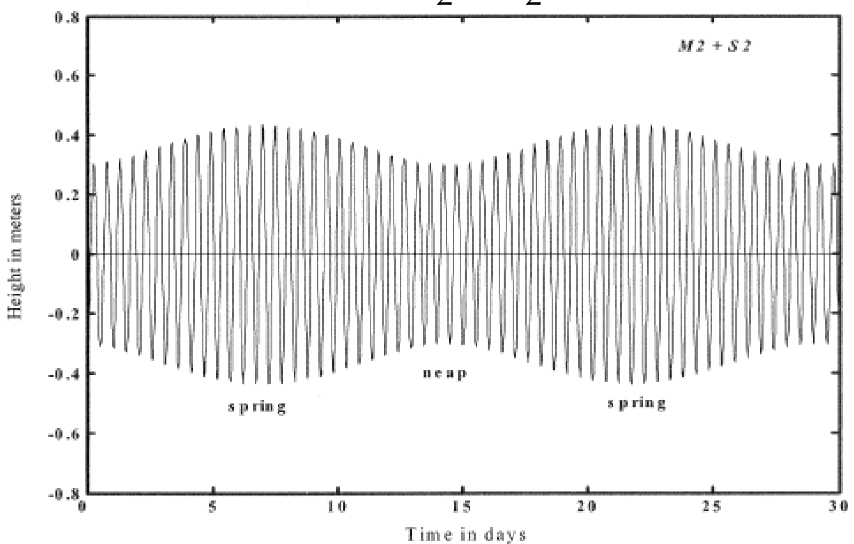
The more constituents are used, the better the prediction

Output

Sea-level evolution due to tides (prediction)

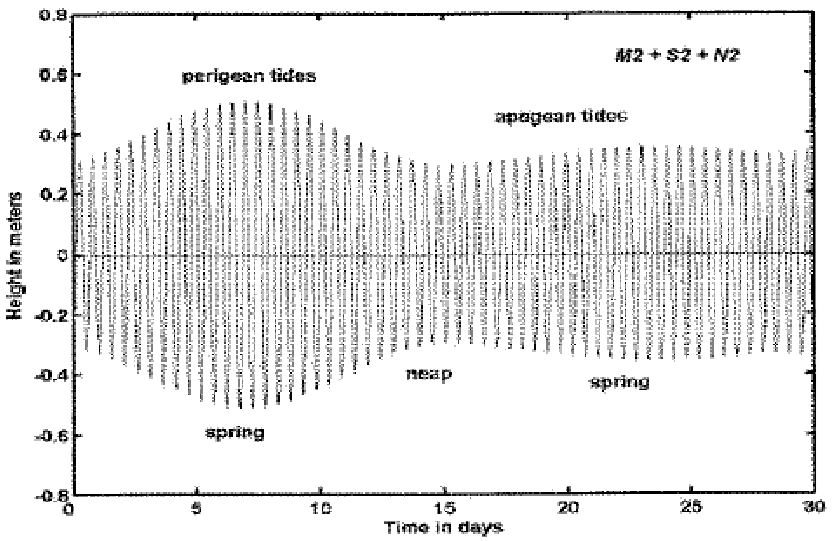
Moon and Solar largest constituents

$M_2 + S_2$



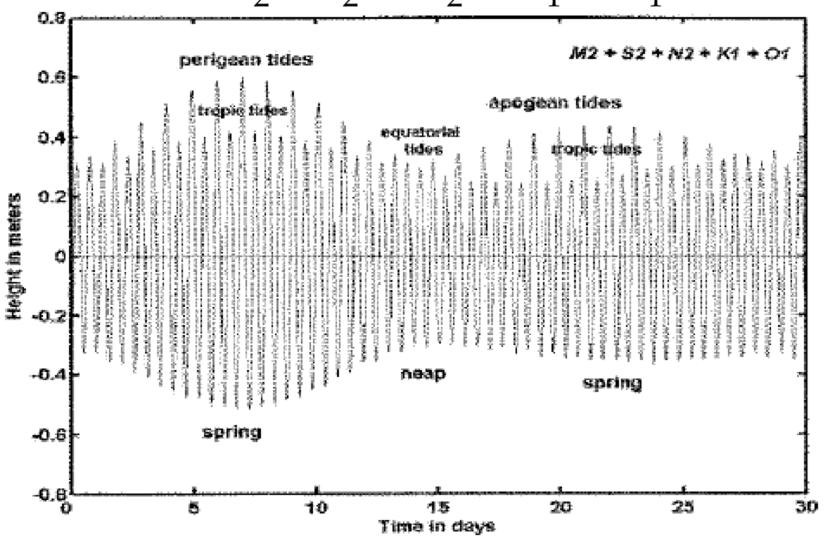
Effect of apogee and perigee

$M_2 + S_2 + N_2$

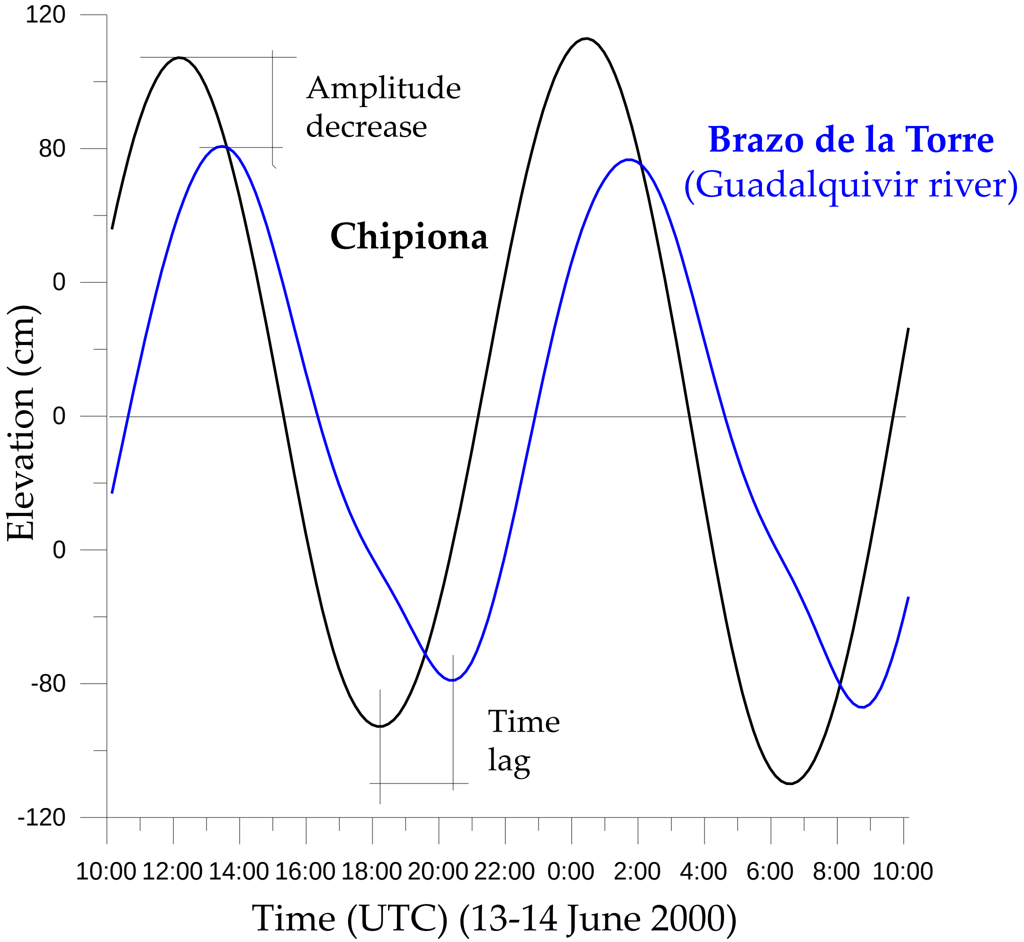


Diurnal inequality (relative sun and moon position and specific location)

$M_2 + S_2 + N_2 + K_1 + O_1$



Further local complication… in shallow waters the tides are modified by the friction with the bottom surface, originating distortions in the tidal wave…



These effects can be also described through certain harmonic constituents through calculations with the main astronomical constituents.

We have sea level data of Ceuta

8760 data … (8760/24 = 365) …

we have hourly data for a full year

We have sea level data of Ceuta

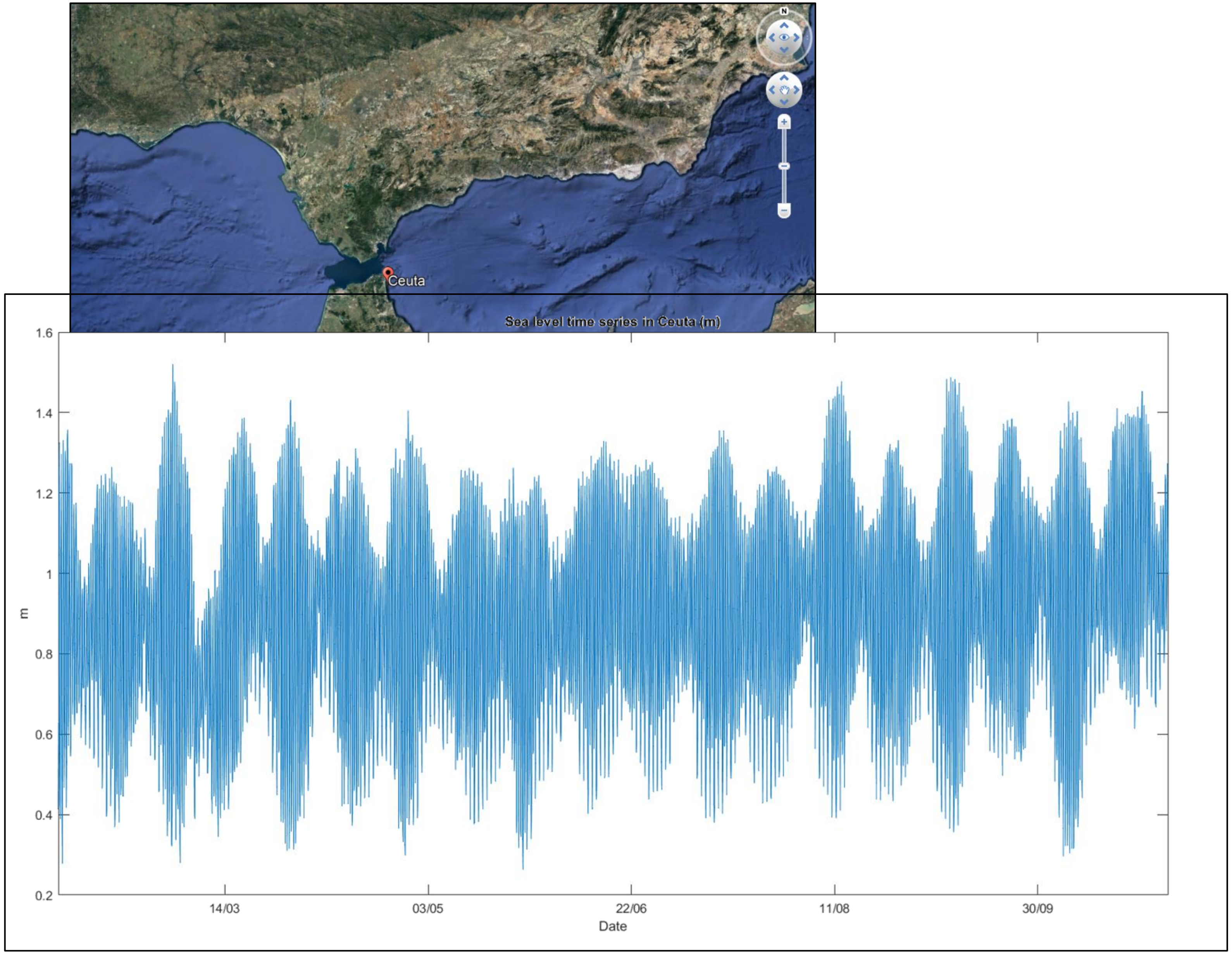
8760 data … (8760/24 = 365) …

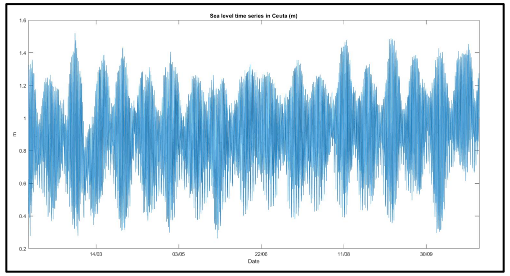
we have hourly data for a full year

From a quick overview we see

~20 “large” cycles

… and other things (a lot of cycles)

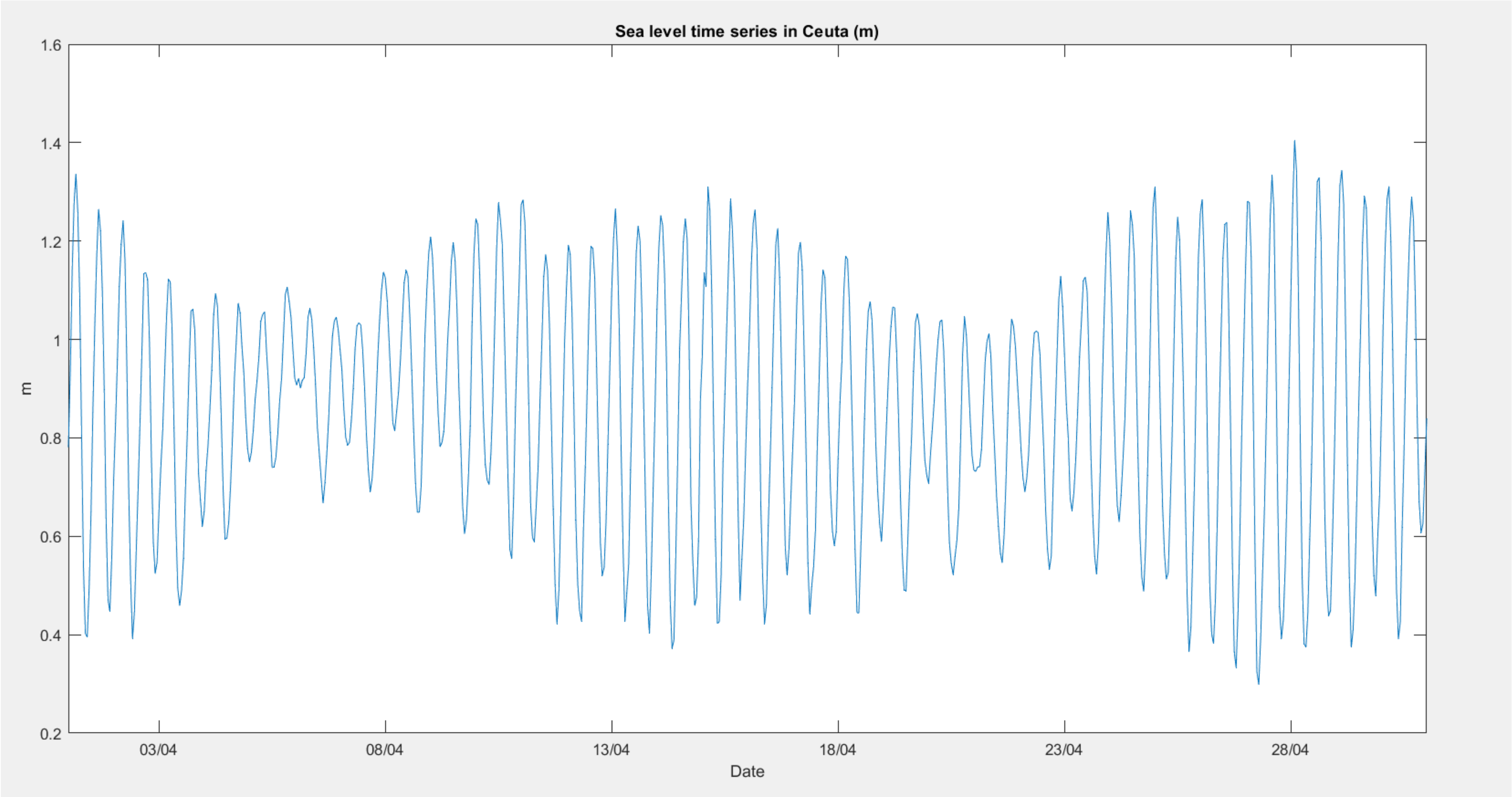


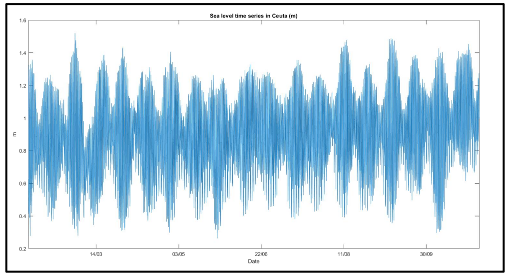


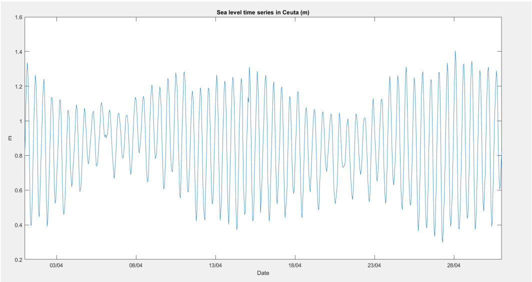
1-month zoom

We see almost 2 cycles

And many others…

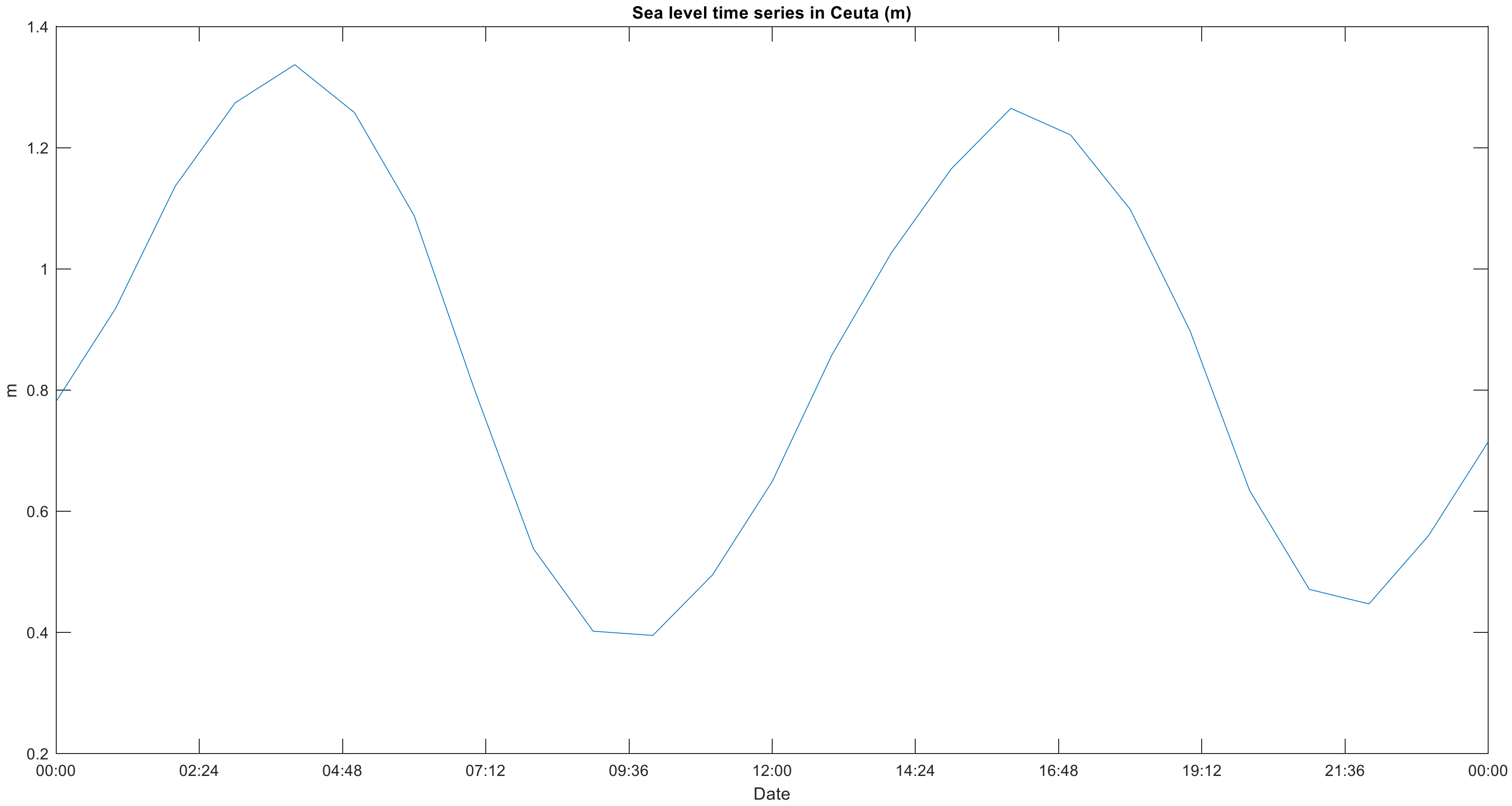


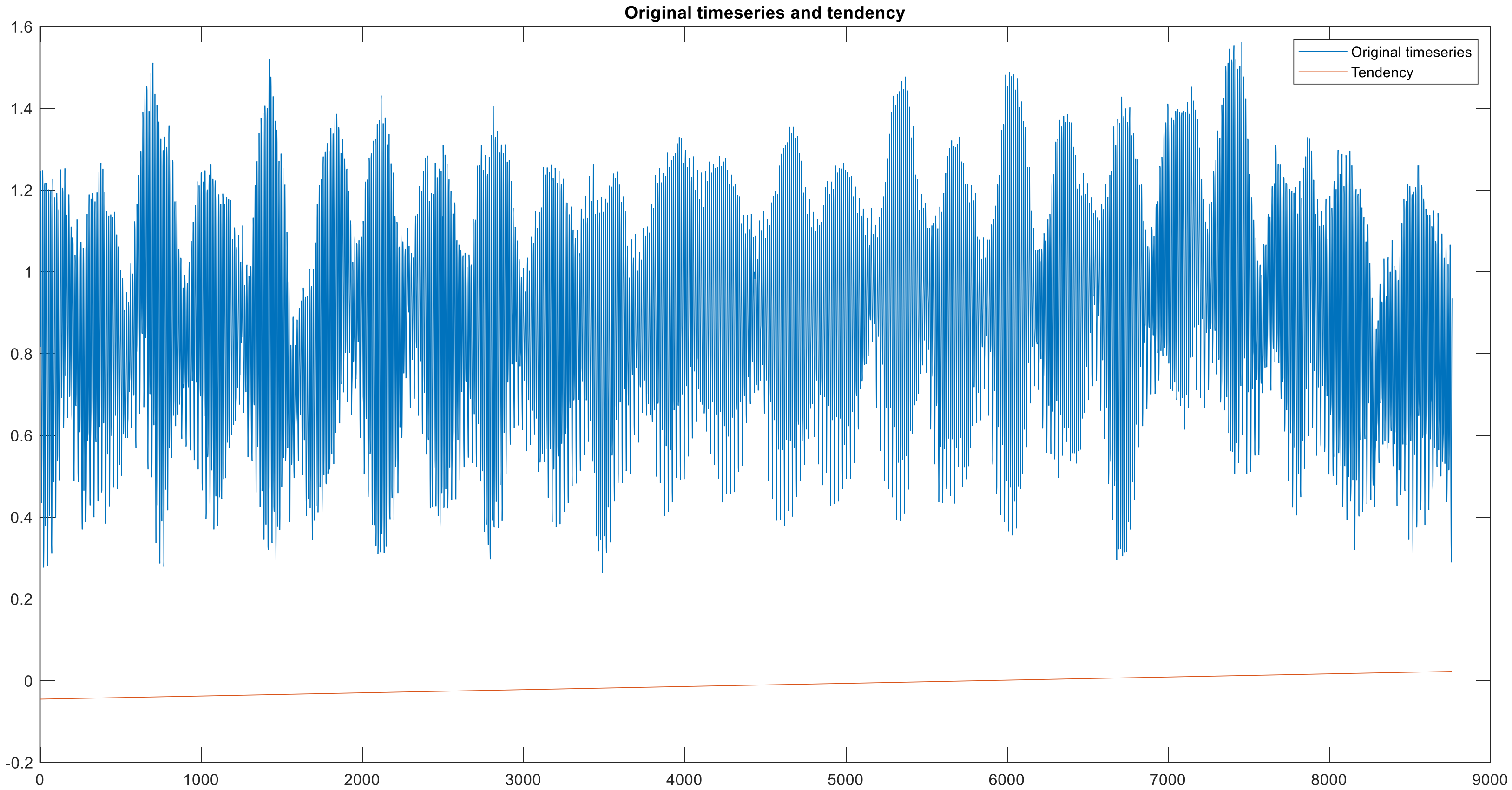




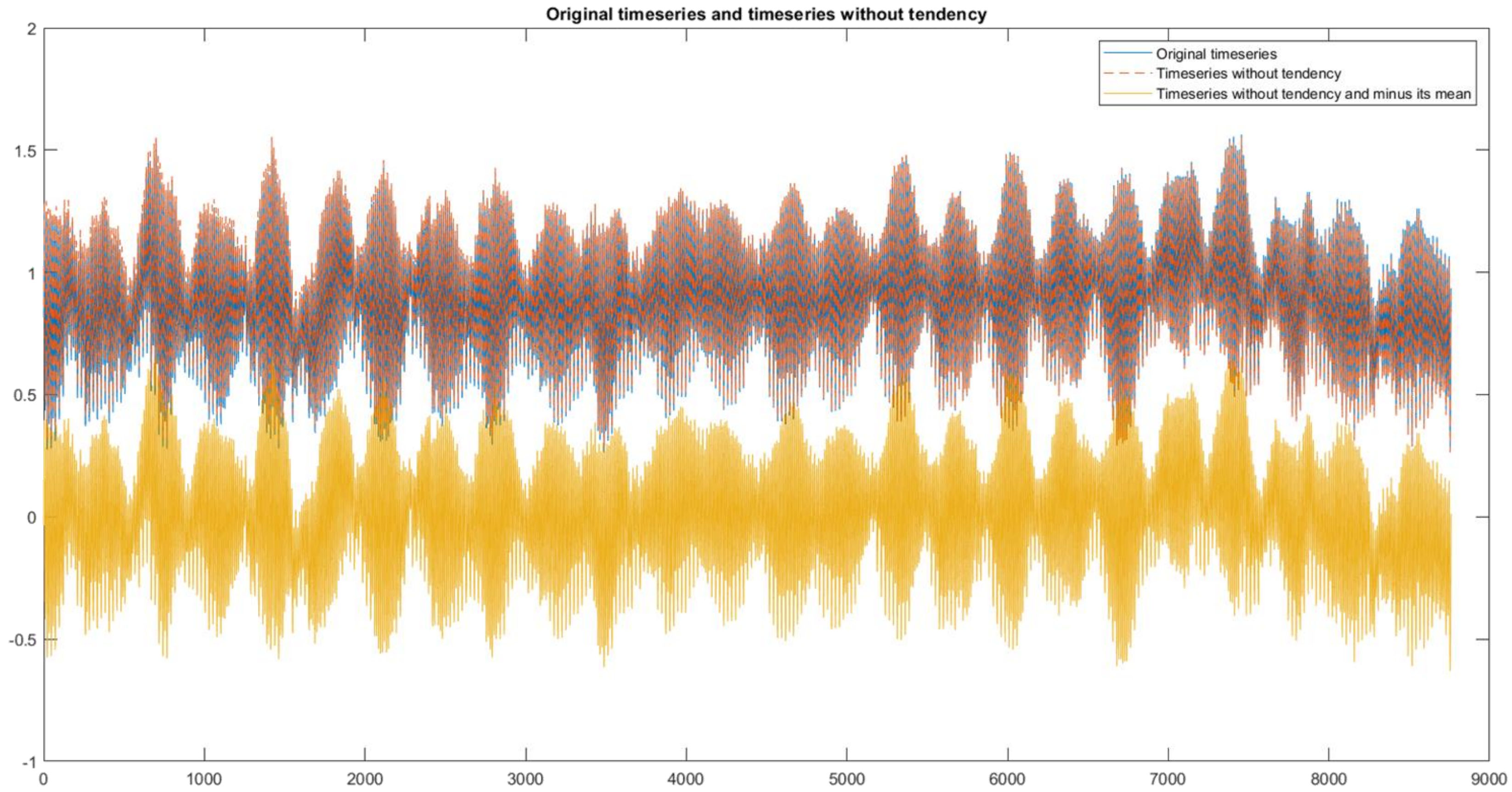
1-day zoom

We see ~2 cycles



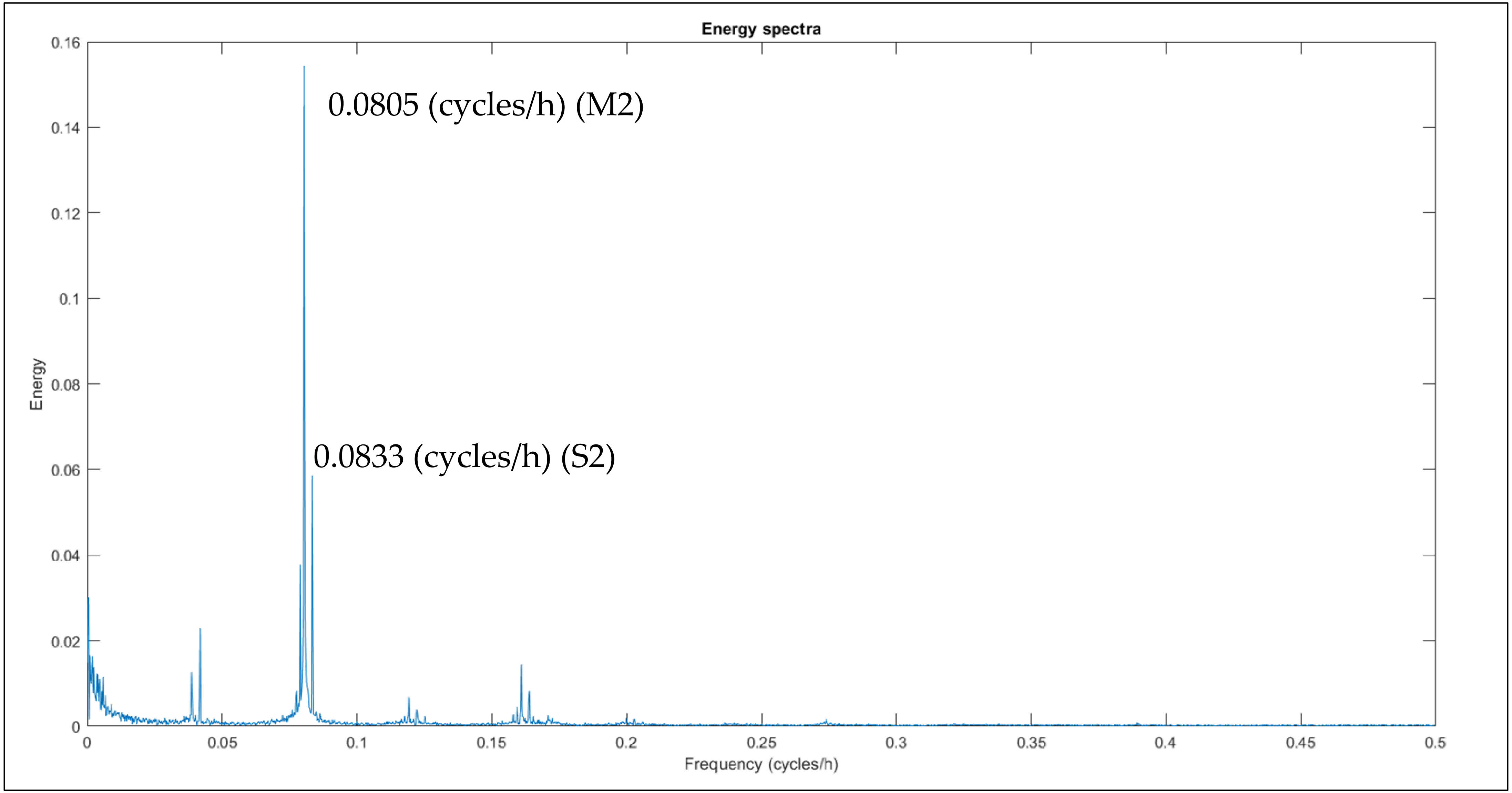


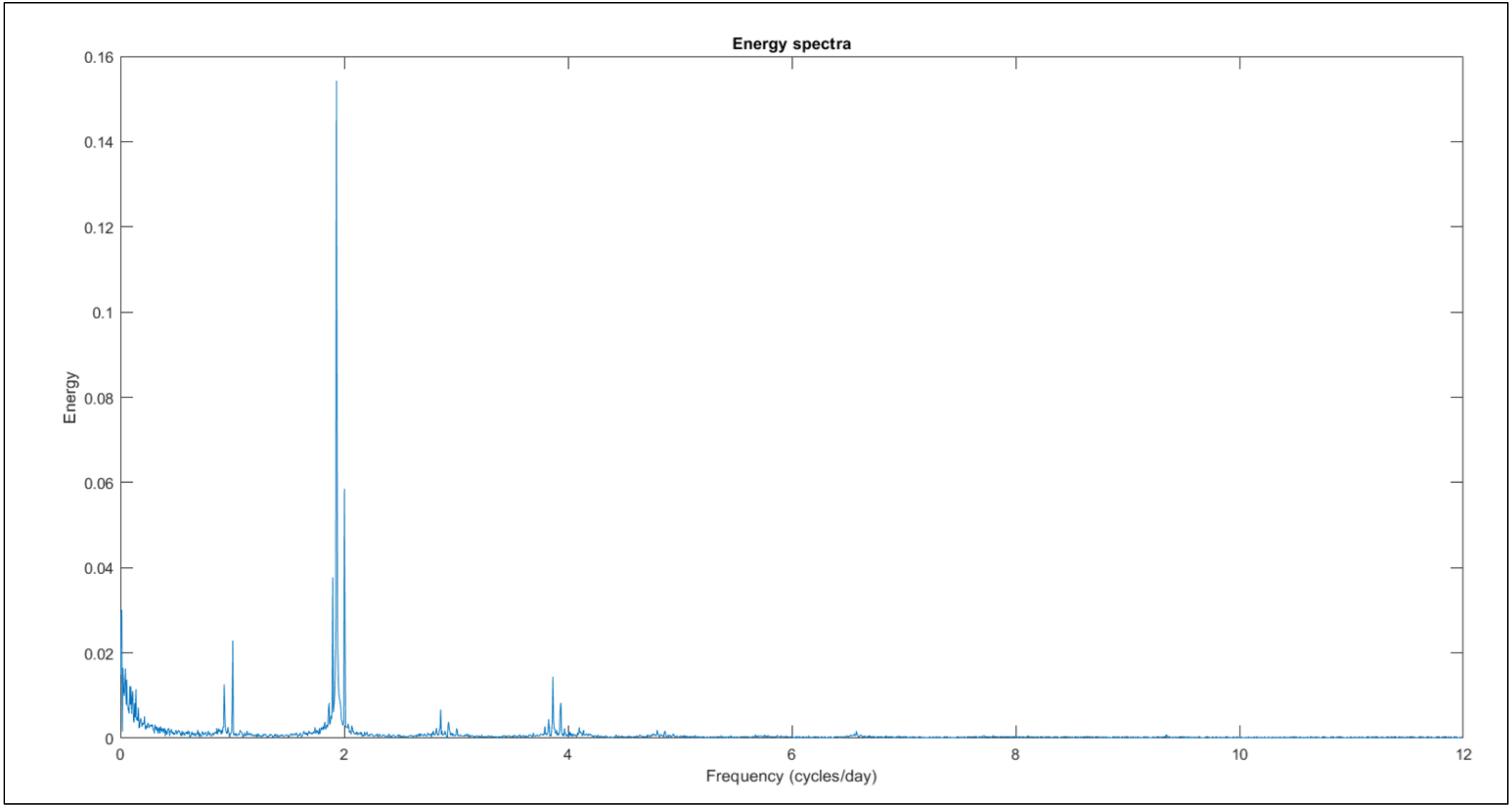
Removal of trend

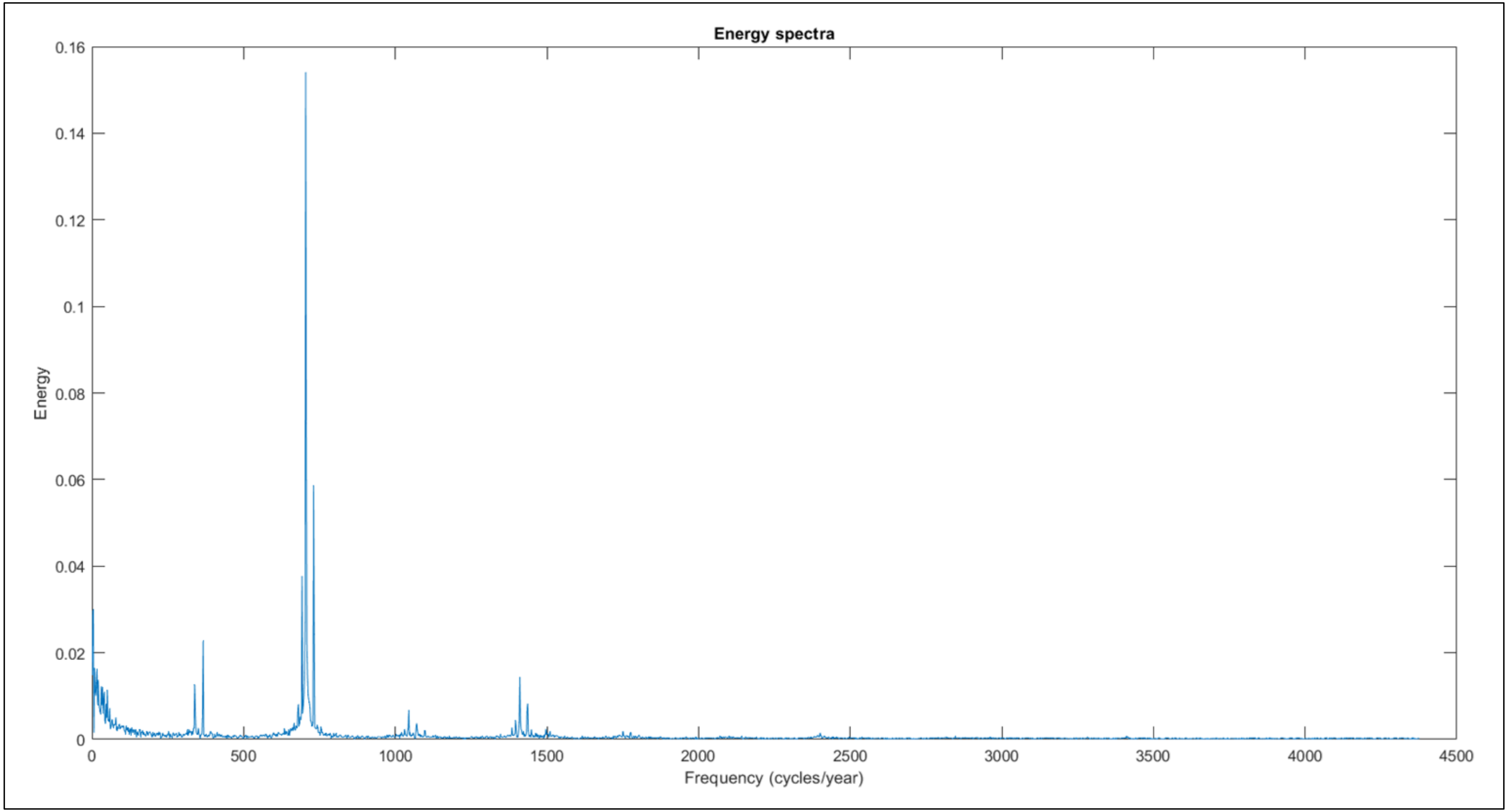


Removal of trend & Removal of mean

FT analysis → FFT

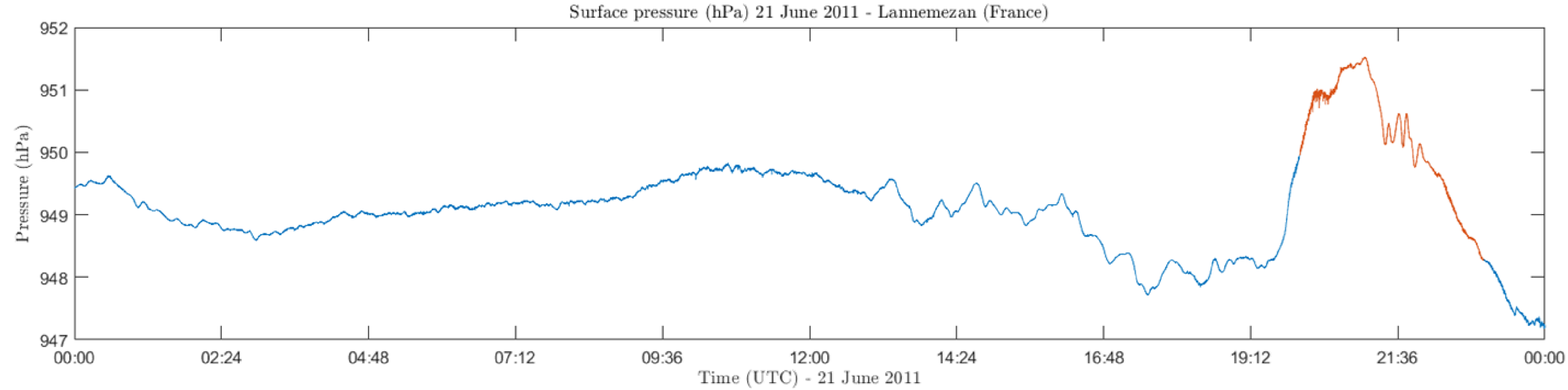

Energy spectra
0.0805 (cycles/h) (M2)
0.0833 (cycles/h) (S2)
Energy
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Frequency (cycles/h)

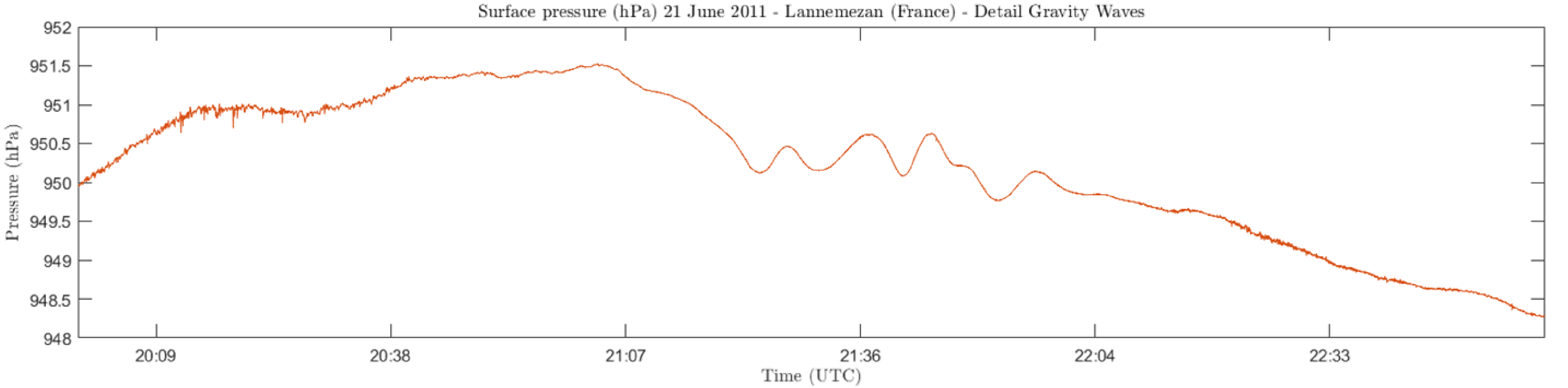

Energy spectra
Energy
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
2
4
6
8
10
12
Frequency (cycles/day)


Energy spectra
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Energy
0
500
1000
1500
2000
2500
3000
3500
4000
4500
Frequency (cycles/year)

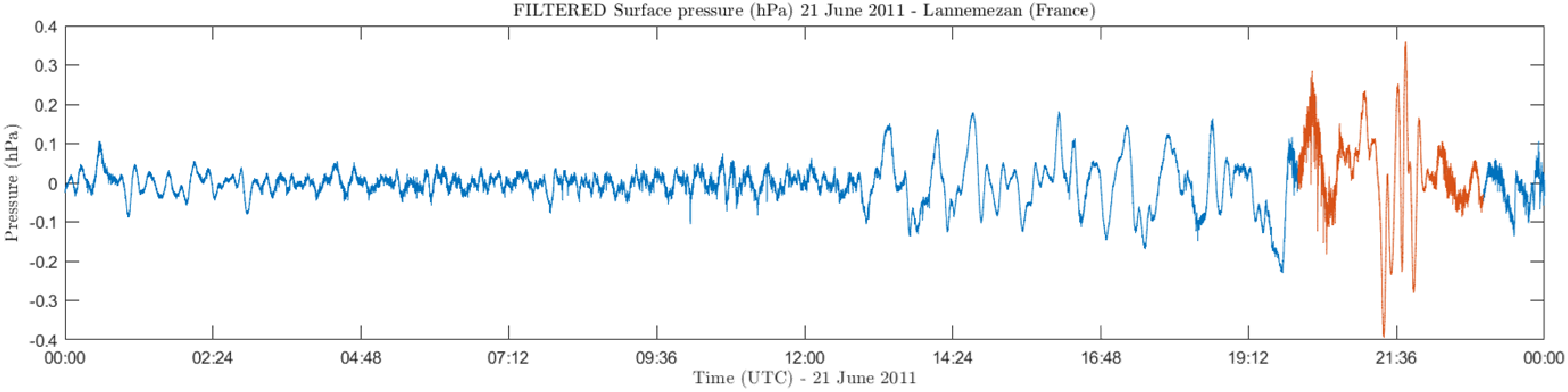
Wavelet to analyse an interesting pressure signal…

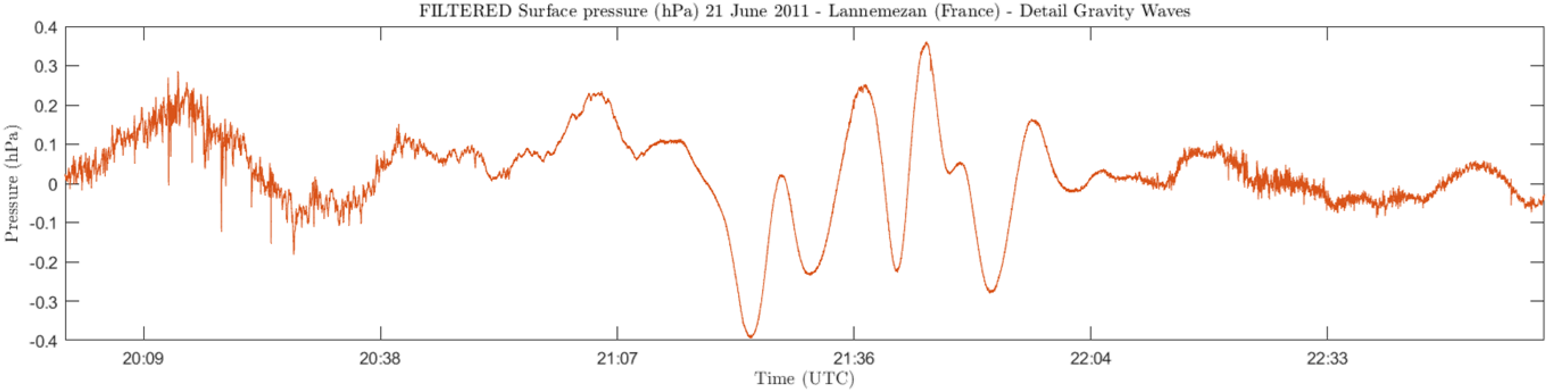
WAVELETS → When the timeseries are not stationary (event-like signal)





Interesting pressure signal…

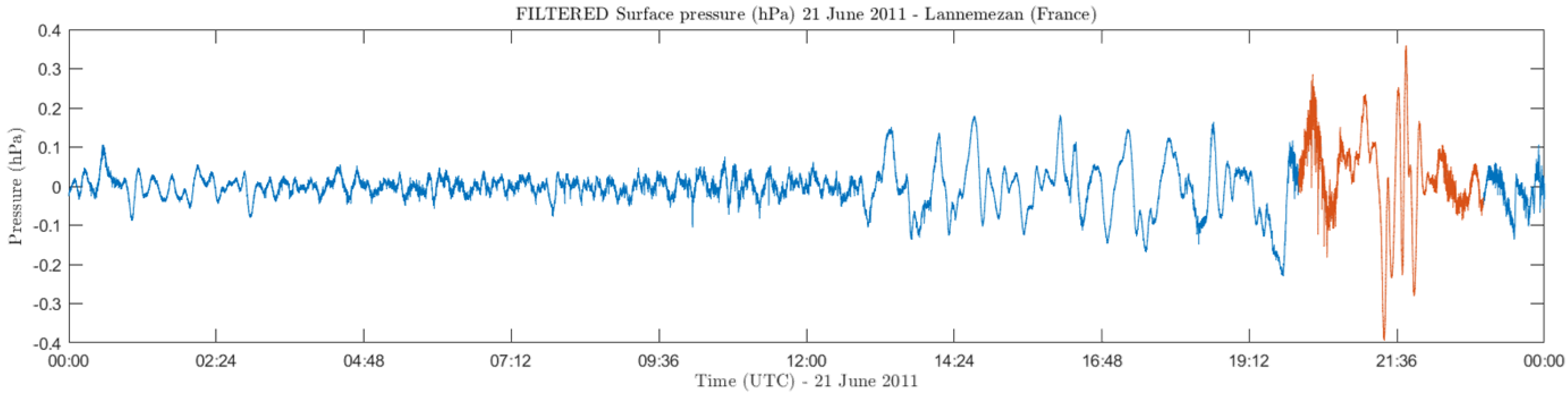


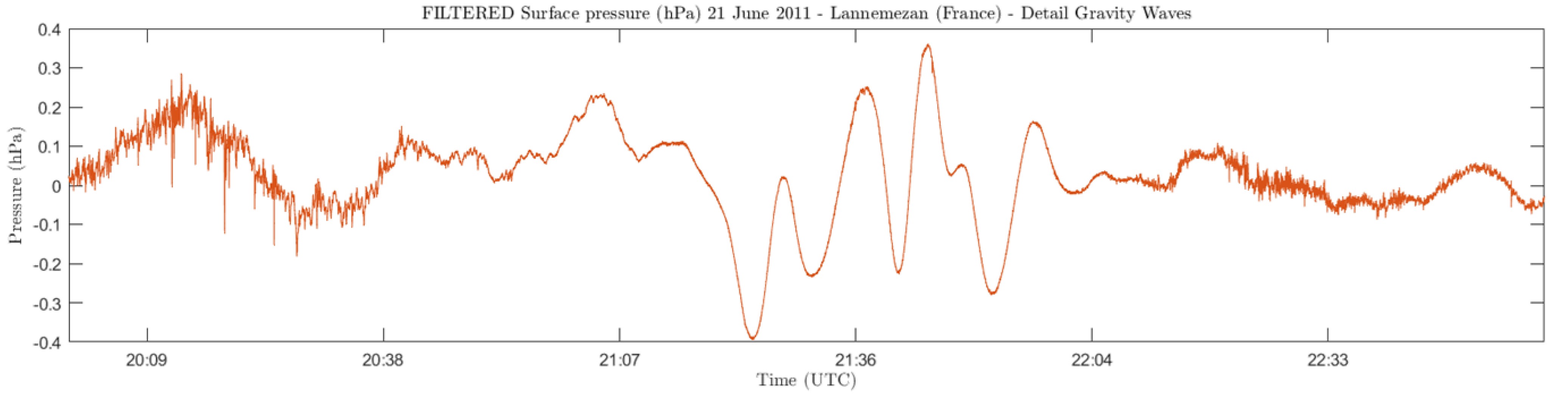


Interesting pressure signal…

<u>Always caution with the filters, use them for specific applications, but not for timeseries analysis type Fourier analysis</u>

<u>High-pass</u>

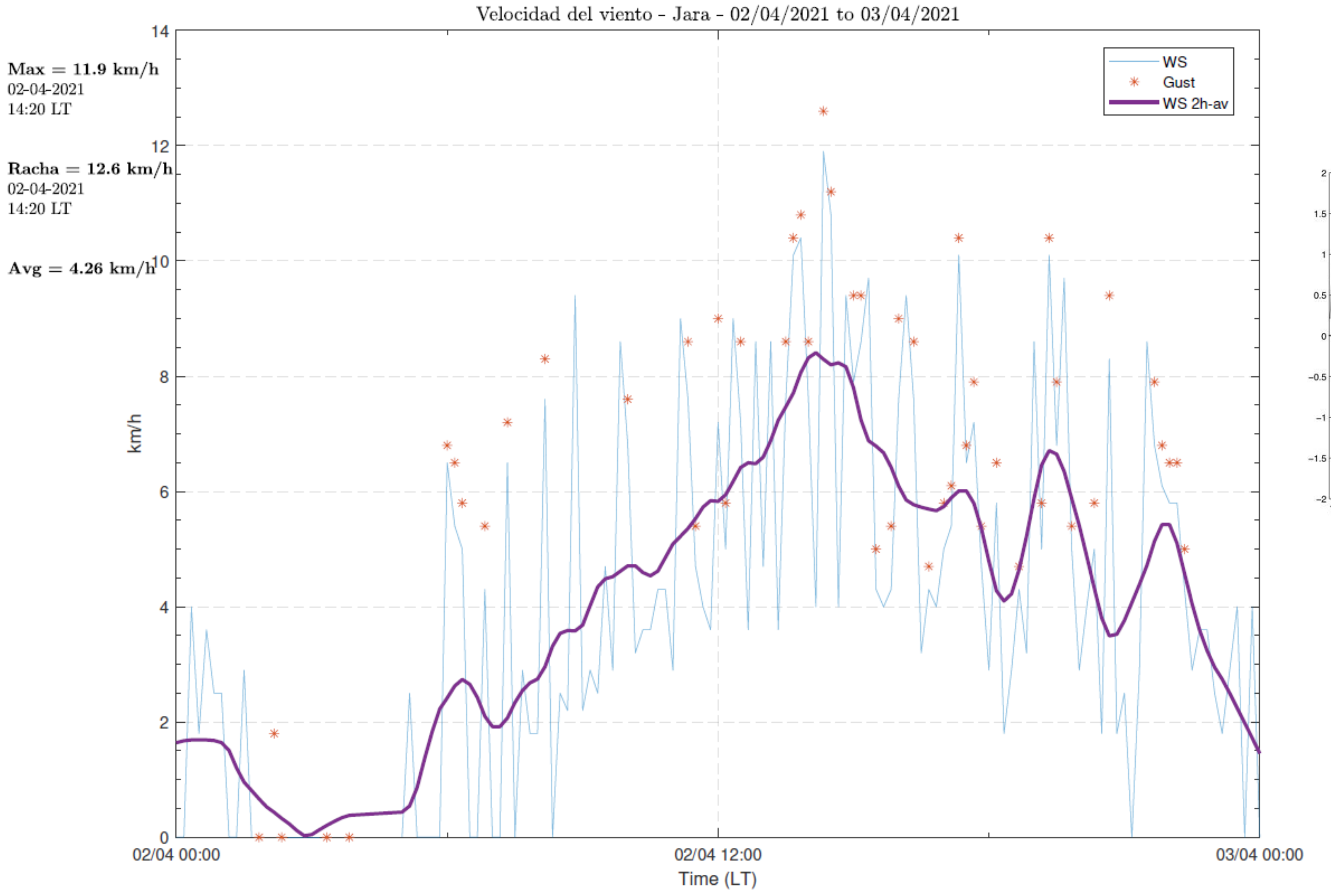


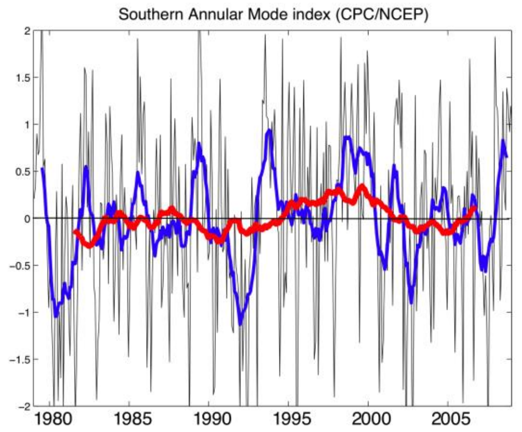


Interesting pressure signal…

Always caution with the filters, use them for specific applications, but not for timeseries analysis type Fourier analysis

Low-pass

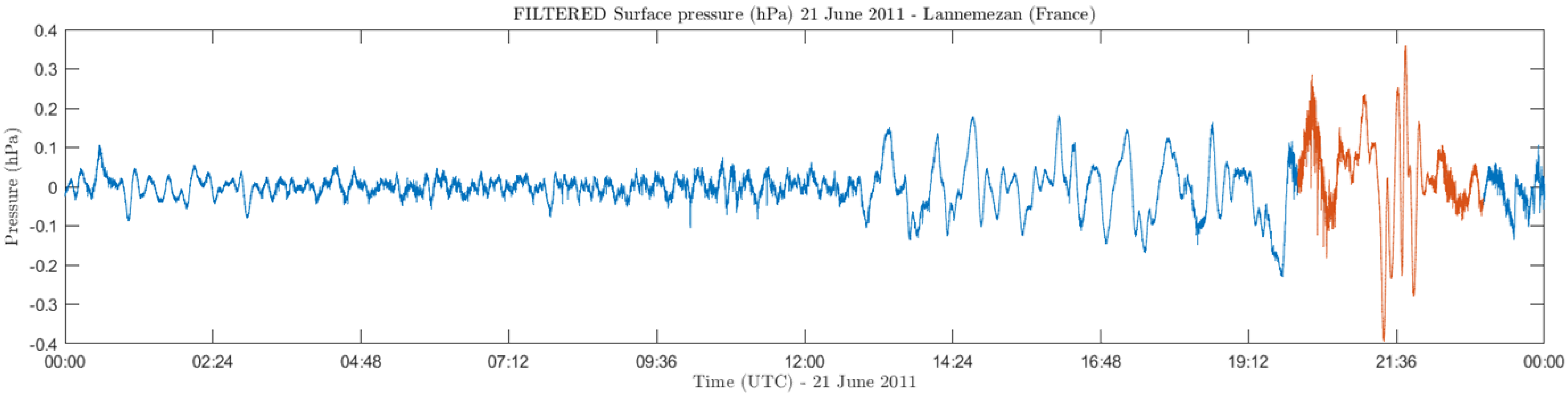


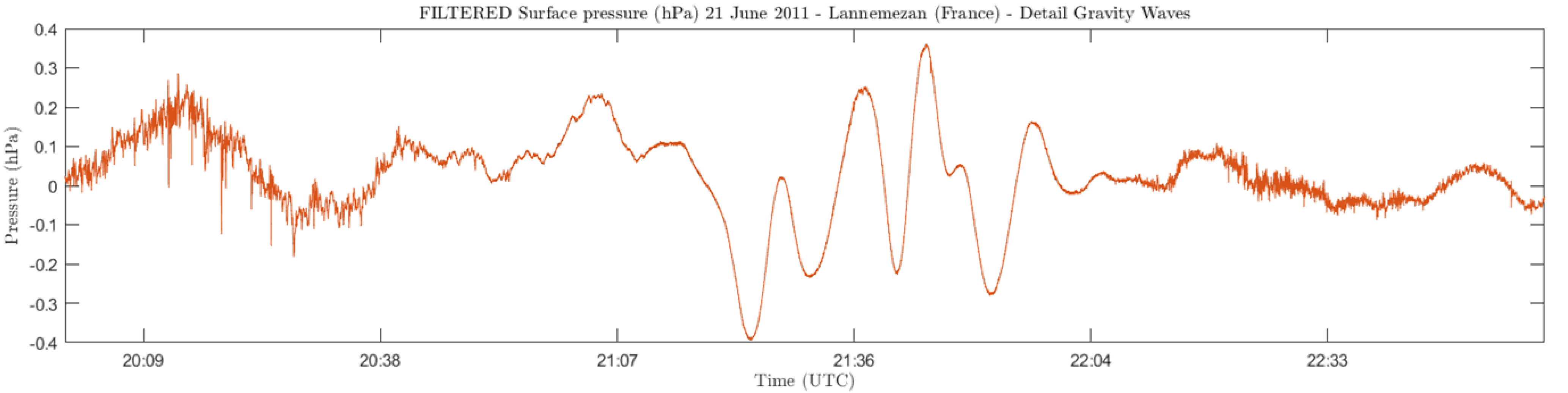


Interesting pressure signal…

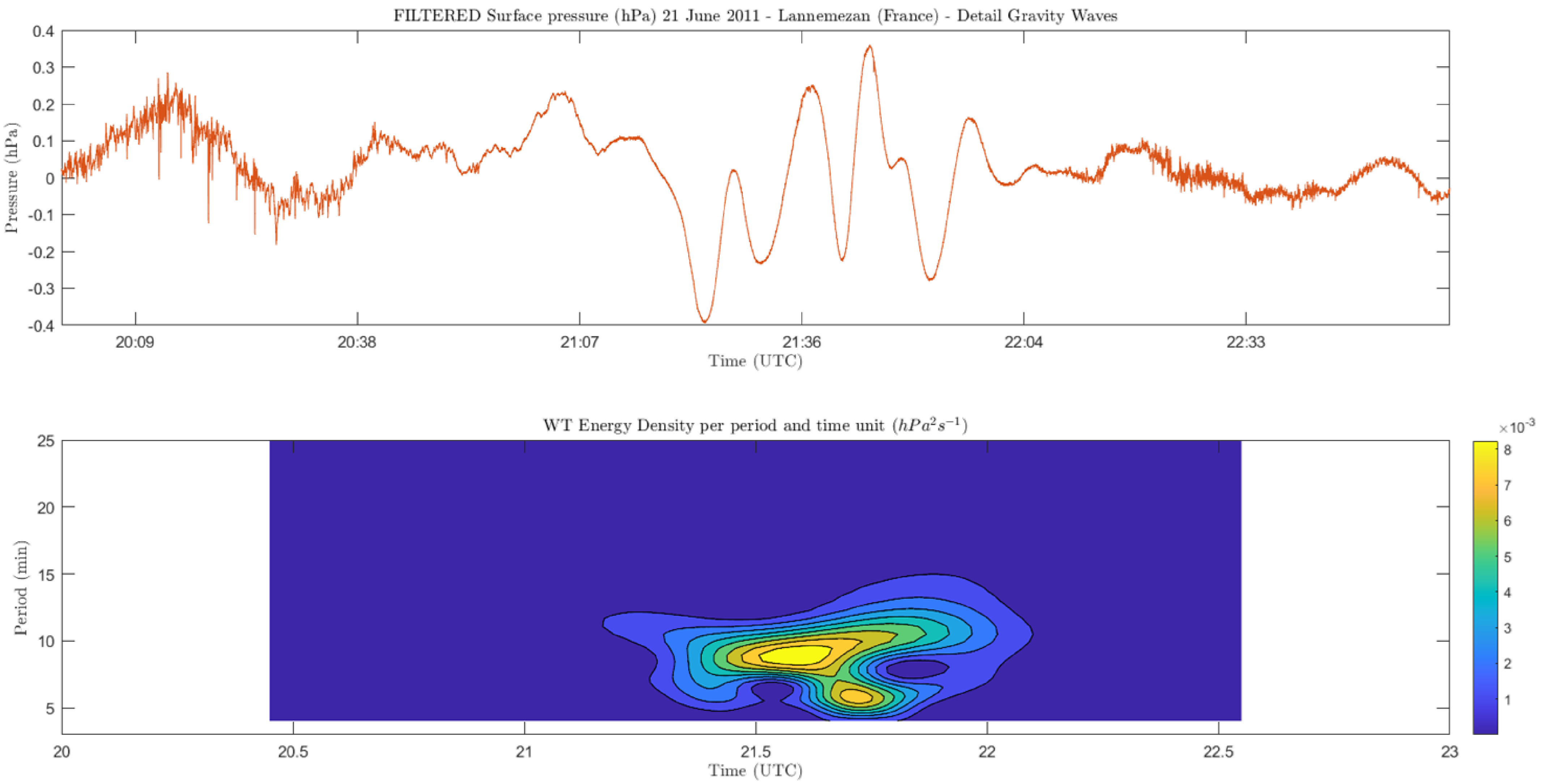
<u>**Always caution with the filters, use them for specific applications, but not for timeseries analysis type Fourier analysis**</u>

<u>**High-pass**</u>





Interesting pressure signal…



5. A practical example of data analysis

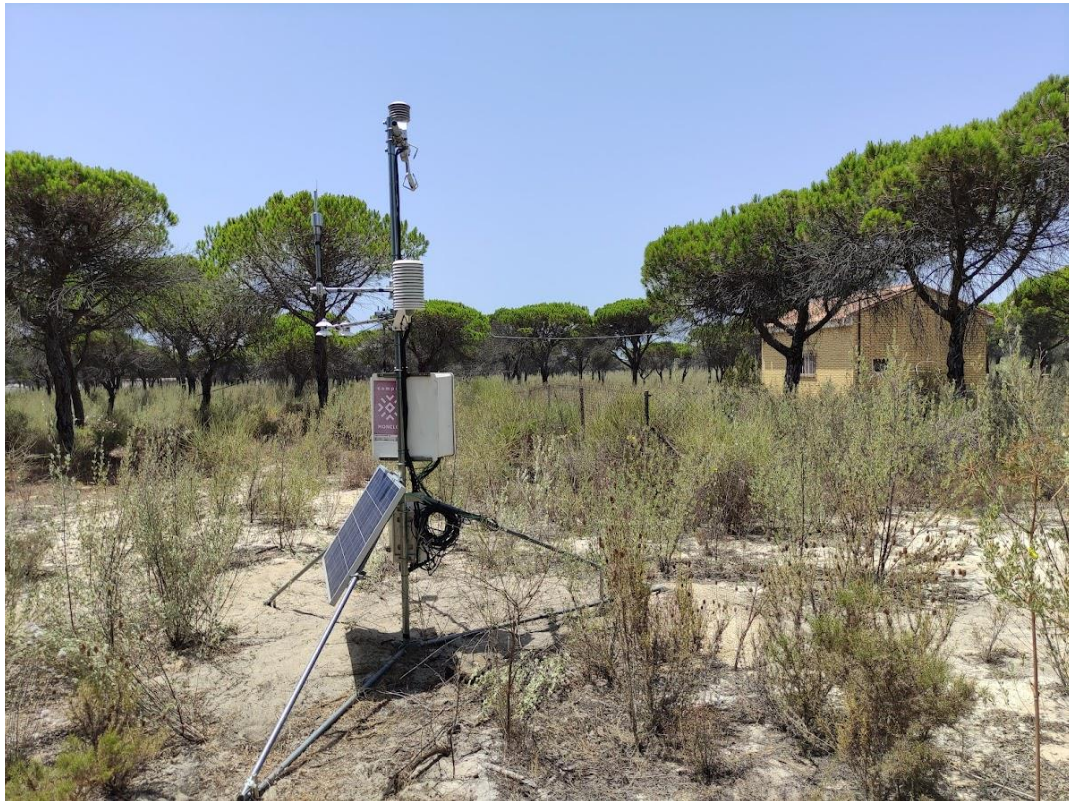
Example of data analysis

We are interested in the **coastal breezes** in the area

We have model simulations (not shown now) and we want to evaluate the model…
How well is the model working?

Observational data!

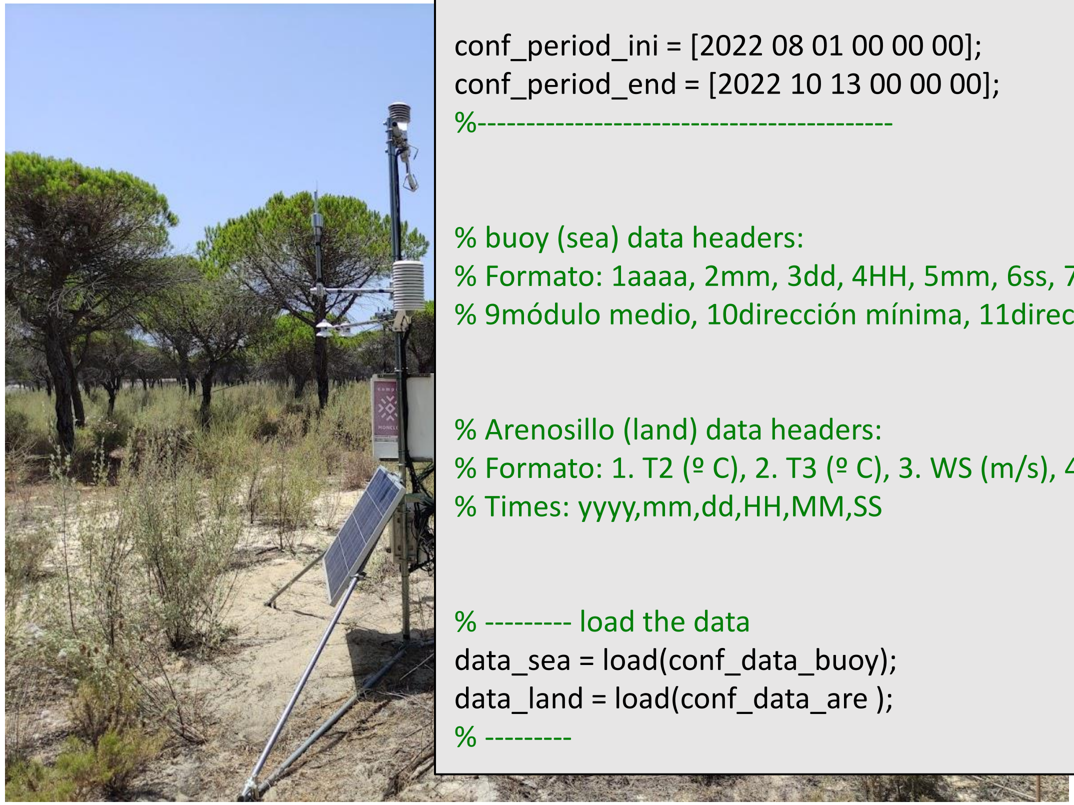
Know what you have!



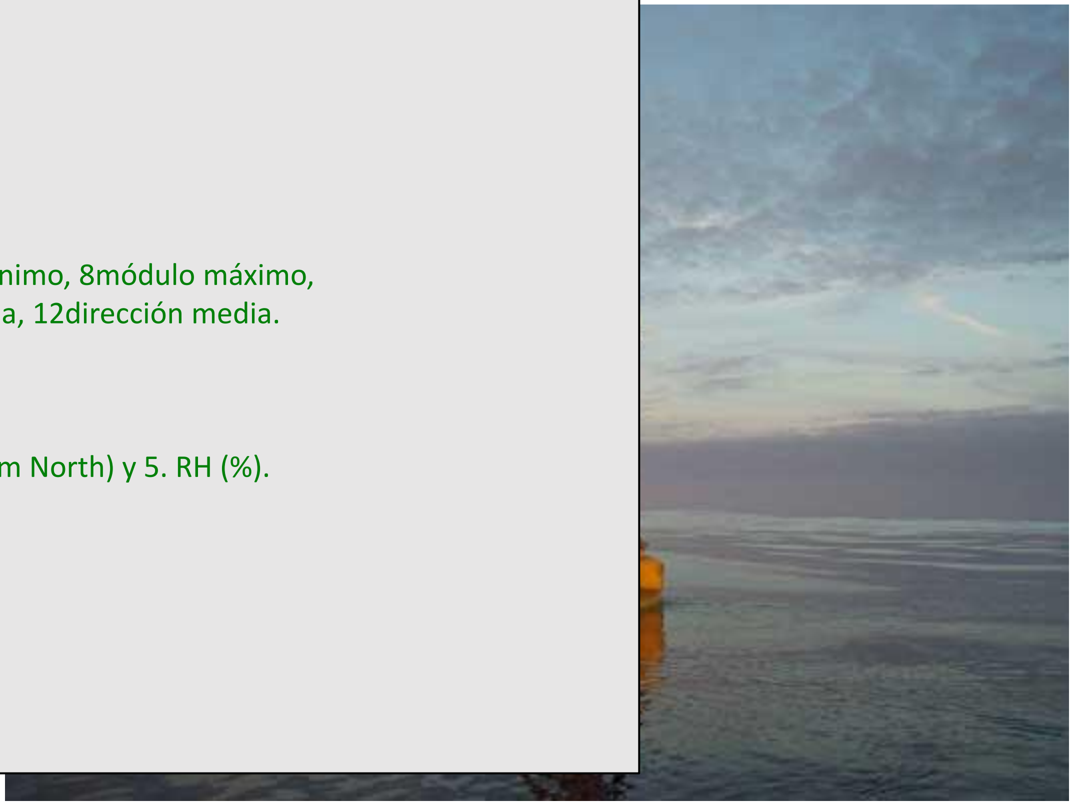
https://www.ndbc.noaa.gov/station_page.php?station=62001

Example of data analysis

Load the data!

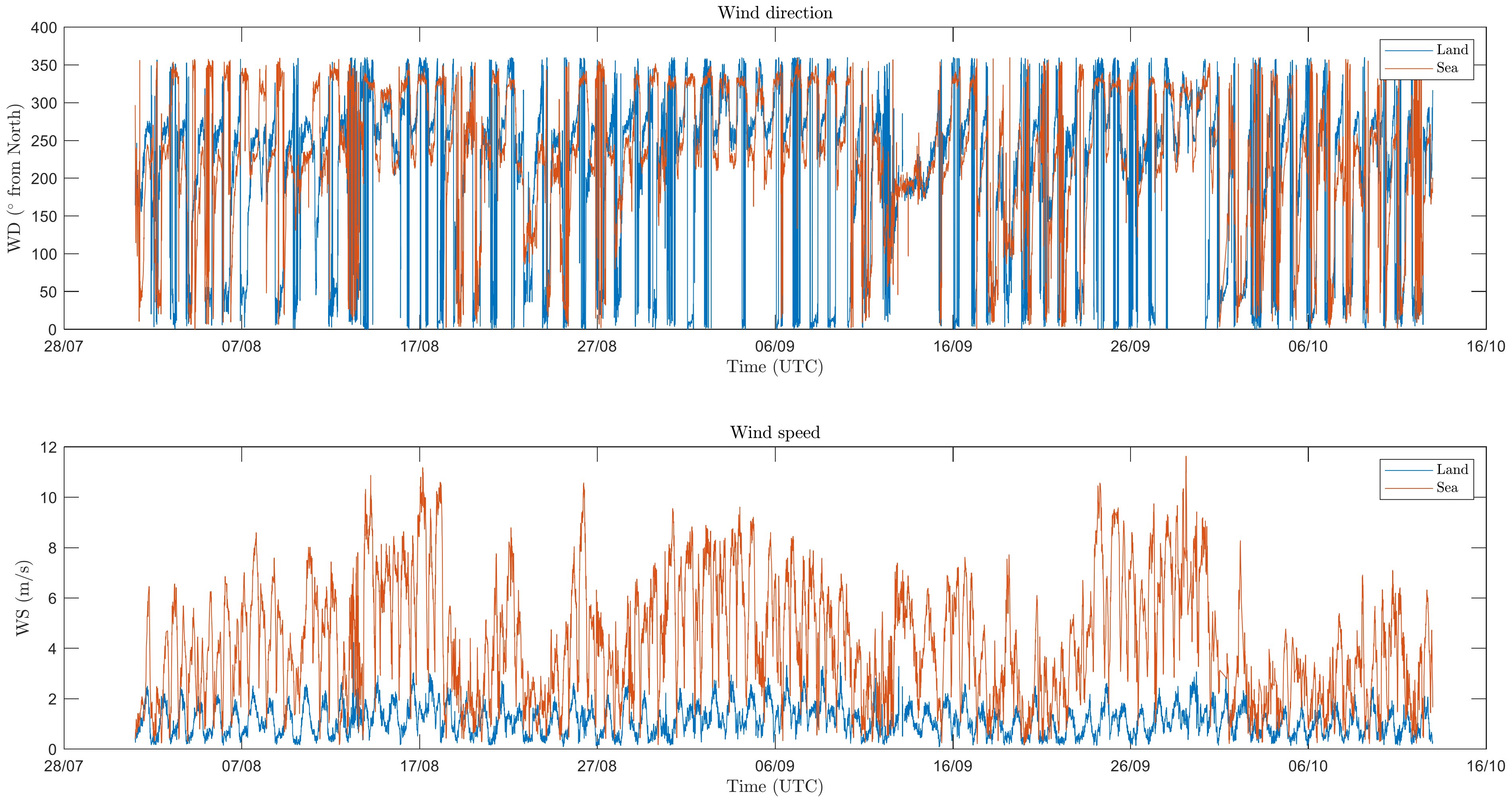


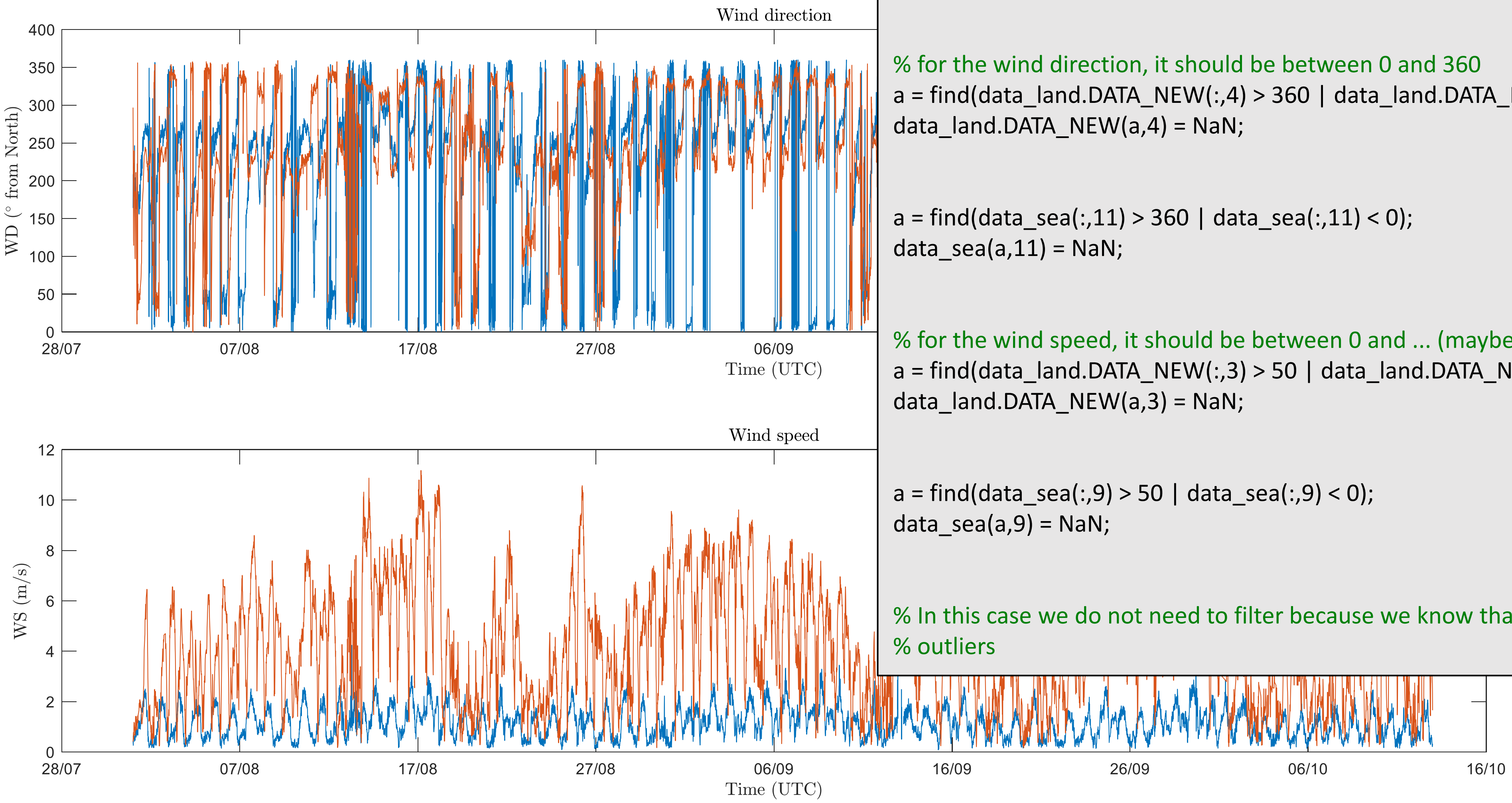
```
%% Load the data
%-------------------------------------------
conf_data_buoy = 'data_buoy.dat';
conf_data_are = 'DATA_ARENOSILLO_20220719_20221031.mat';


conf_period_ini = [2022 08 01 00 00 00];
conf_period_end = [2022 10 13 00 00 00];
%-------------------------------------------


% buoy (sea) data headers:
% Formato: 1aaaa, 2mm, 3dd, 4HH, 5mm, 6ss, 7modulo mínimo, 8módulo máximo,
% 9módulo medio, 10dirección mínima, 11dirección máxima, 12dirección media.


% Arenosillo (land) data headers:
% Formato: 1. T2 (º C), 2. T3 (º C), 3. WS (m/s), 4. WD (º from North) y 5. RH (%).
% Times: yyyy,mm,dd,HH,MM,SS


% --------- load the data
data_sea = load(conf_data_buoy);
data_land = load(conf_data_are );
% ---------
```



https://www.ndbc.noaa.gov/station_page.php?station=62001

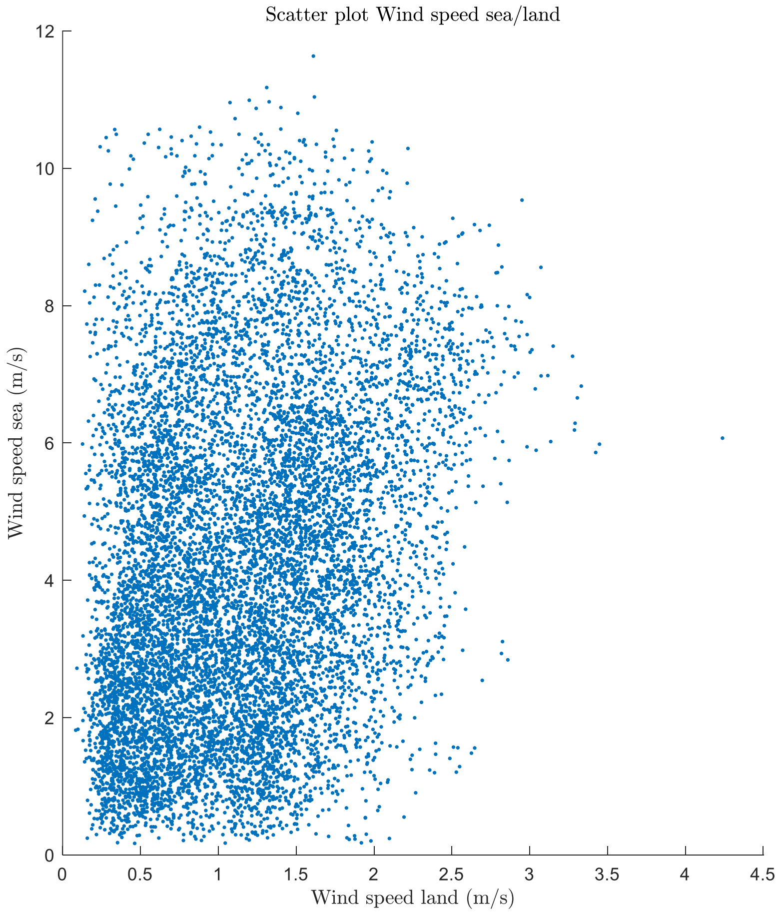
First quick looks!

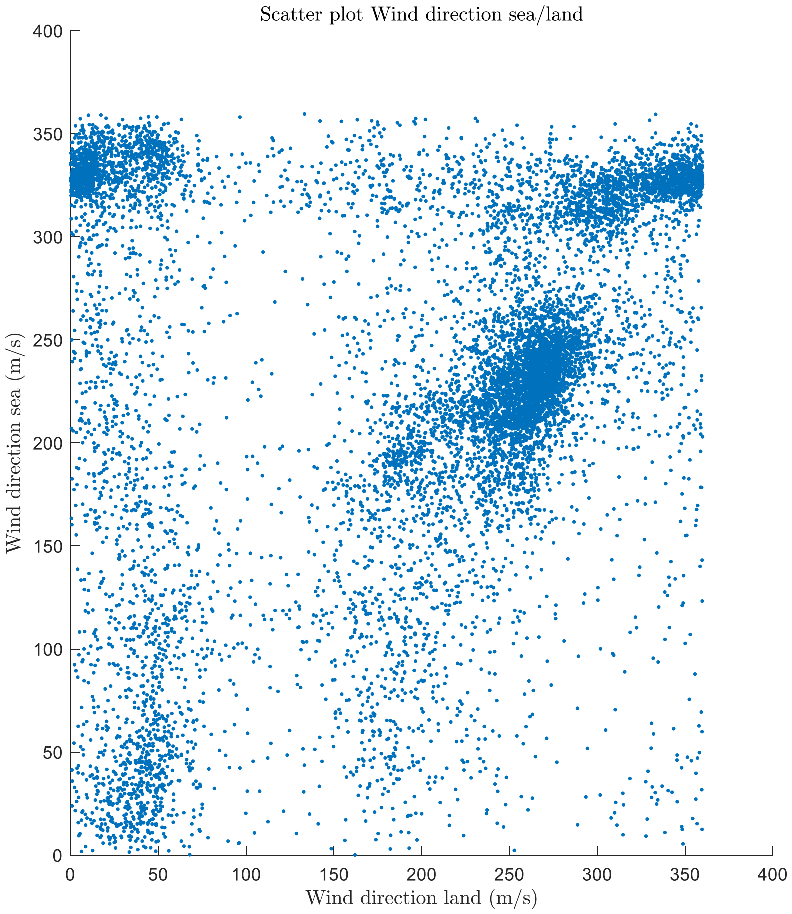




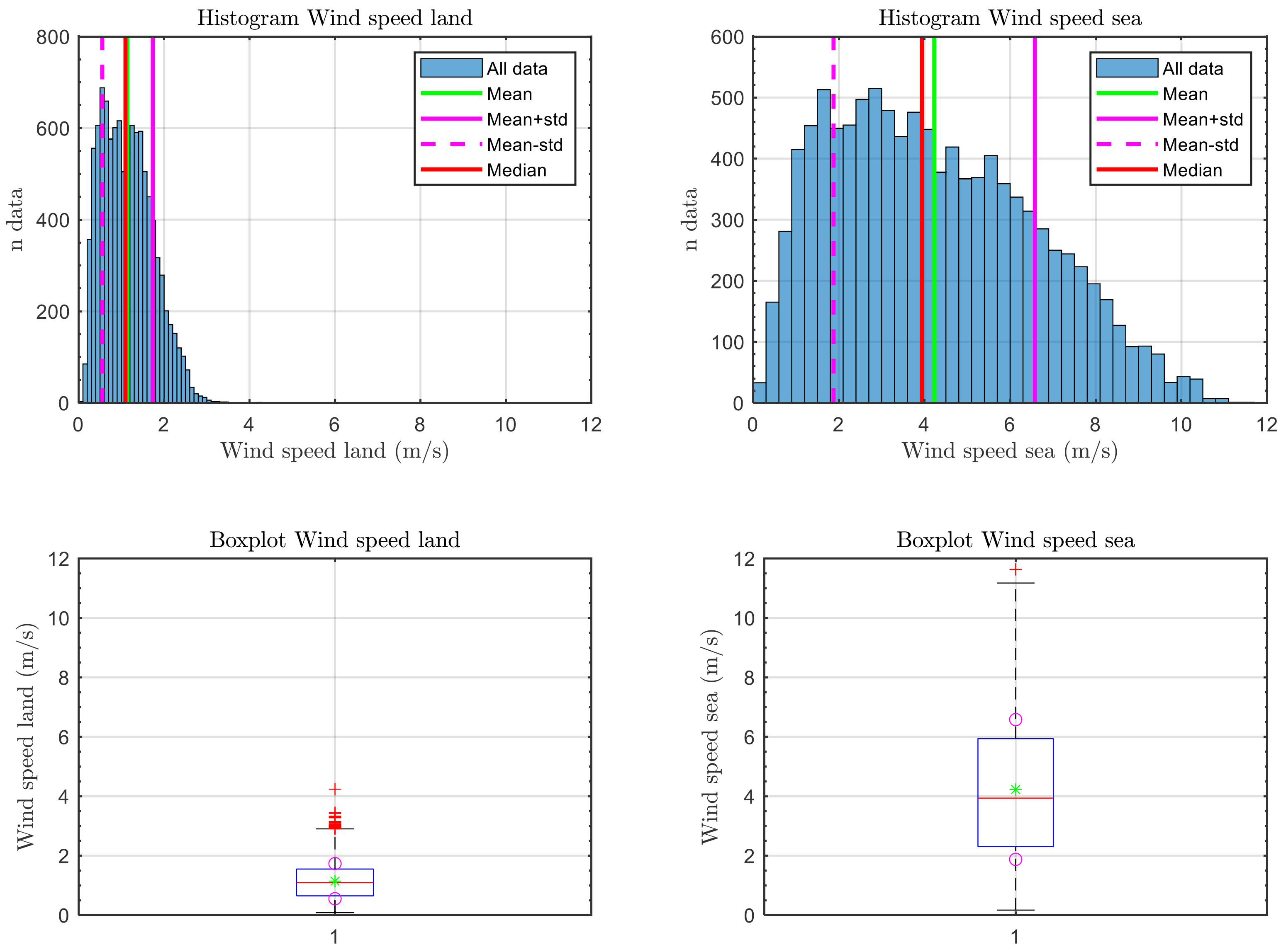

```
%% Fast quality control (just remove values outside reasonable limits)

% for the wind direction, it should be between 0 and 360
a = find(data_land.DATA_NEW(:,4) > 360 | data_land.DATA_NEW(:,4) < 0);
data_land.DATA_NEW(a,4) = NaN;

a = find(data_sea(:,11) > 360 | data_sea(:,11) < 0);
data_sea(a,11) = NaN;

% for the wind speed, it should be between 0 and ... (maybe 40 m/s?)
a = find(data_land.DATA_NEW(:,3) > 50 | data_land.DATA_NEW(:,3) < 0);
data_land.DATA_NEW(a,3) = NaN;

a = find(data_sea(:,9) > 50 | data_sea(:,9) < 0);
data_sea(a,9) = NaN;

% In this case we do not need to filter because we know that there are not
% outliers
```

Some other quick looks!



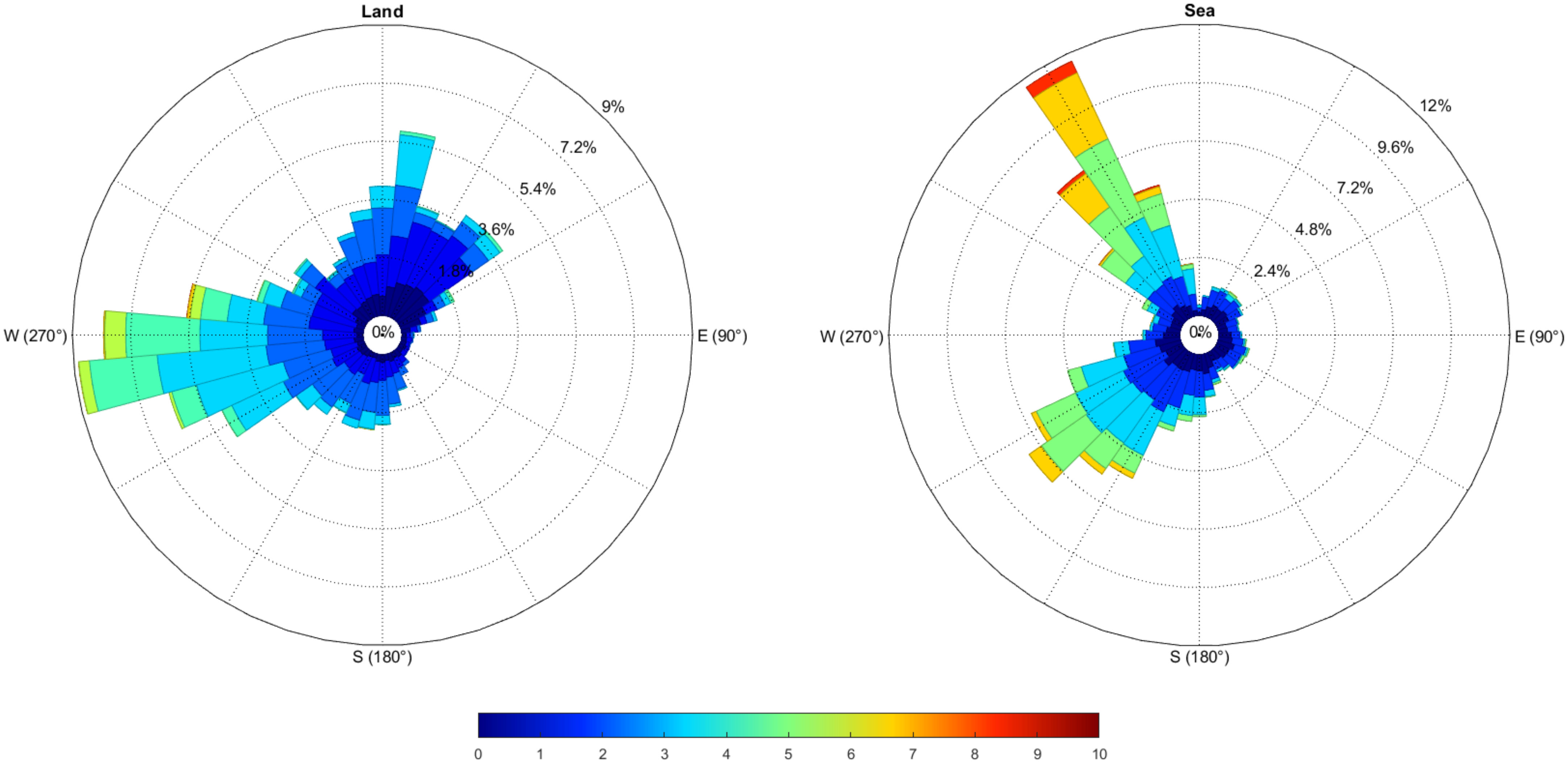


Some calculations to see the data distribution… and compare



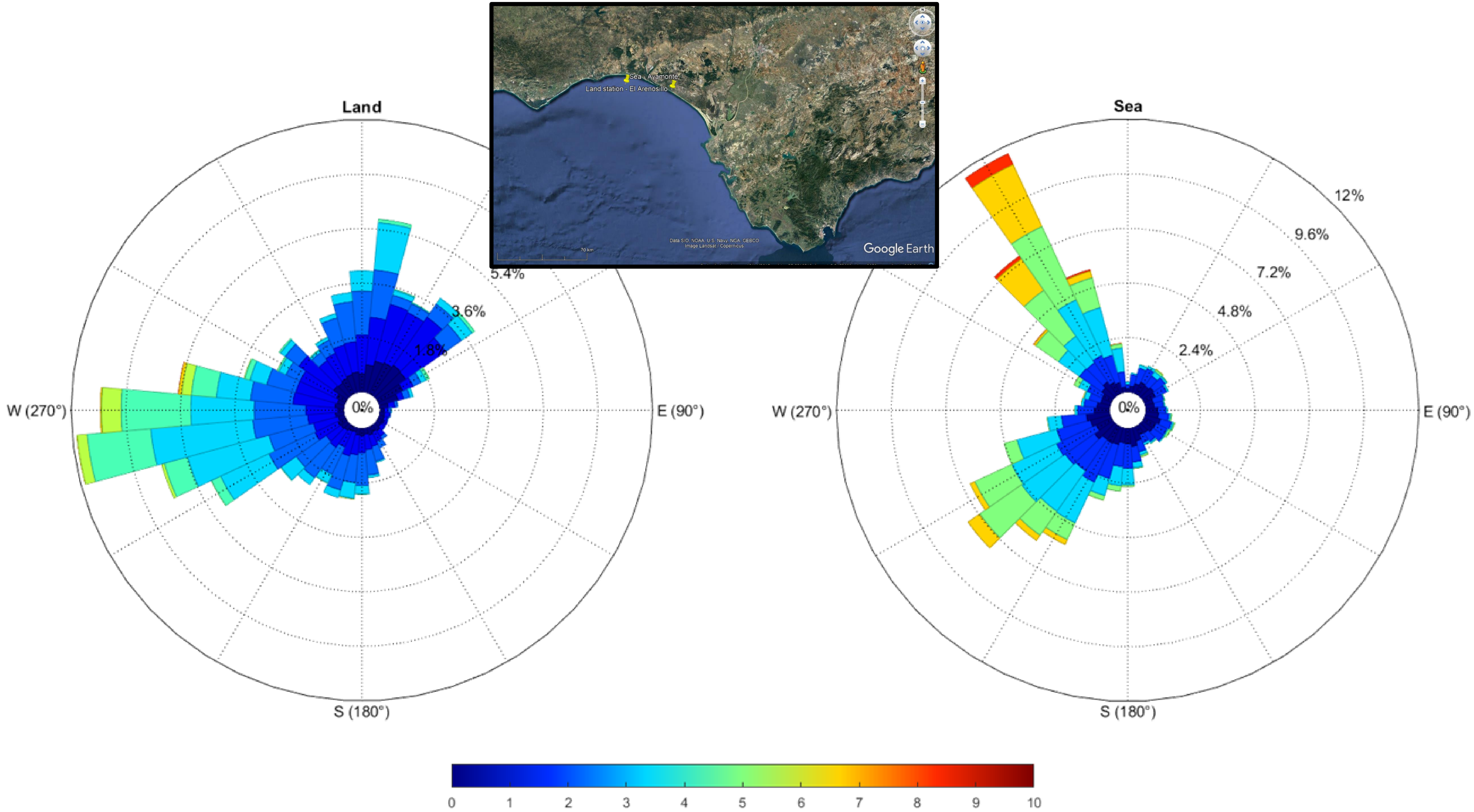
Some calculations to see the data distribution… and compare

Wind Roses



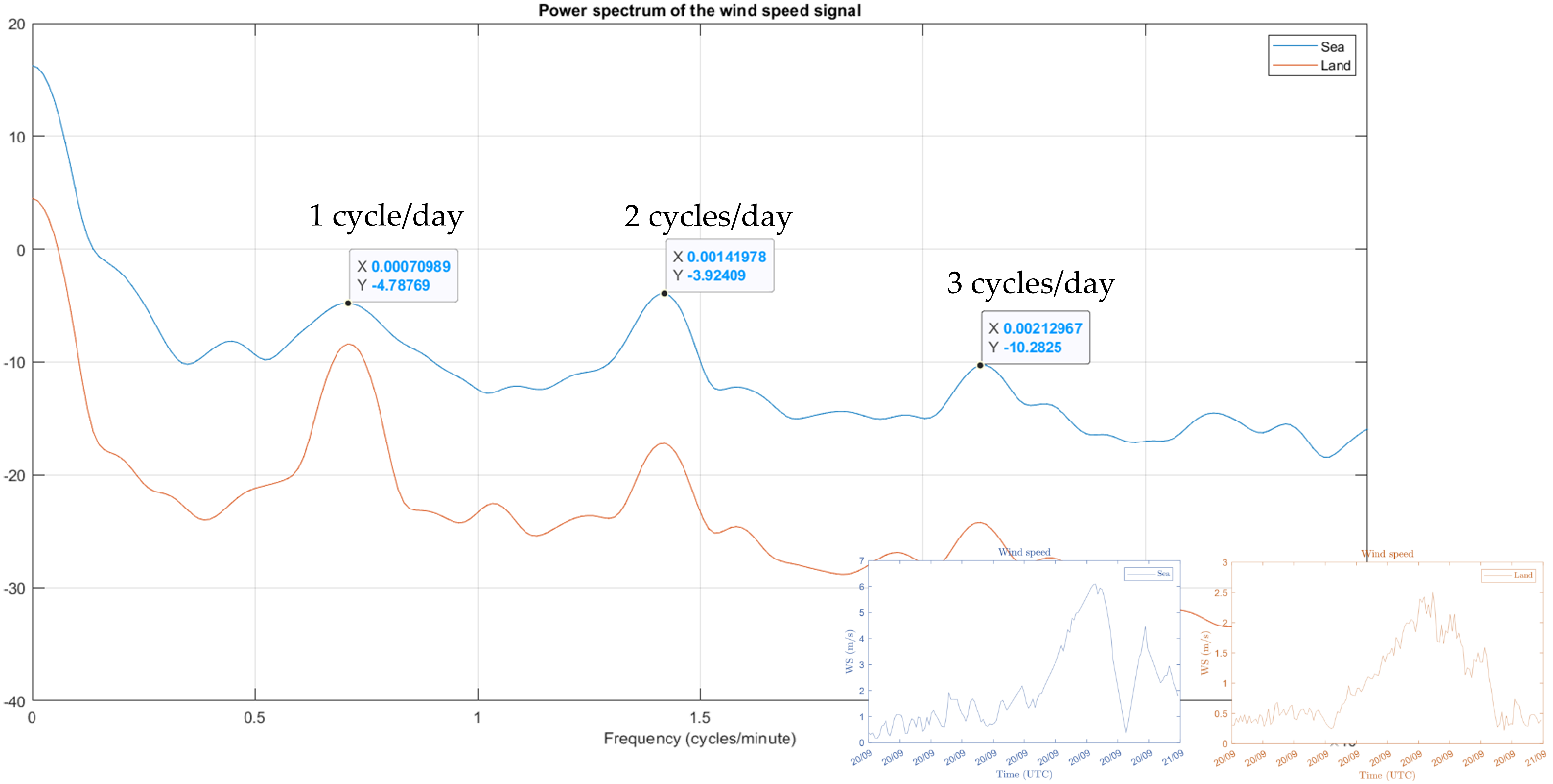
Example of data analysis

Some calculations to see the data distribution… and compare



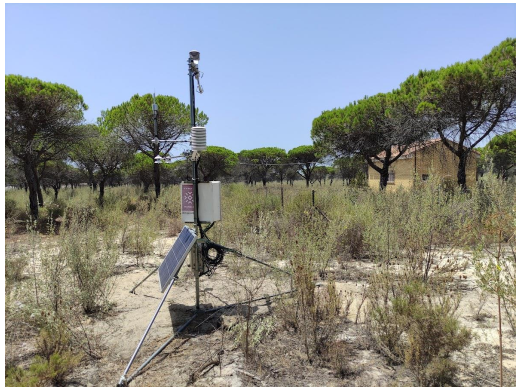
Example of data analysis

Further calculations…

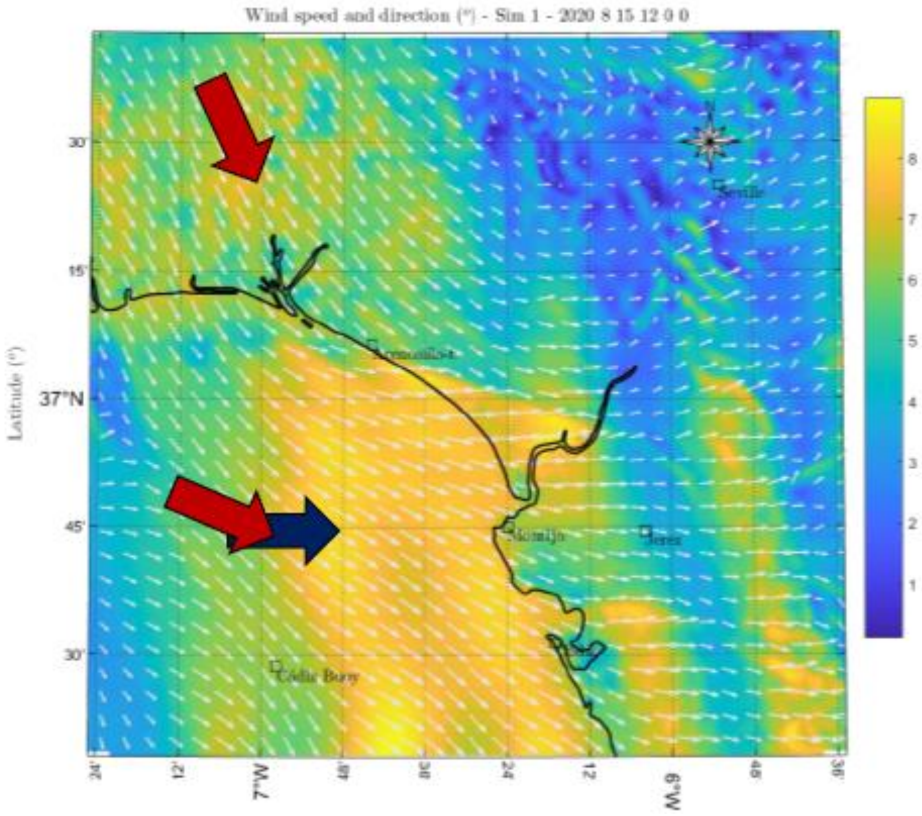


And then… Comparison with model output…

Analysis of case studies…



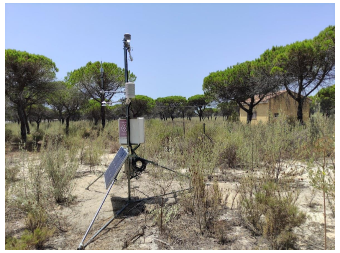
vs

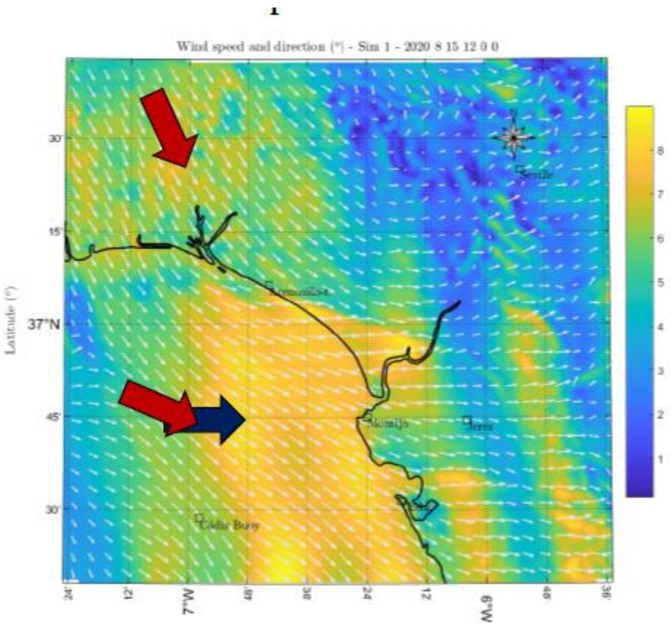


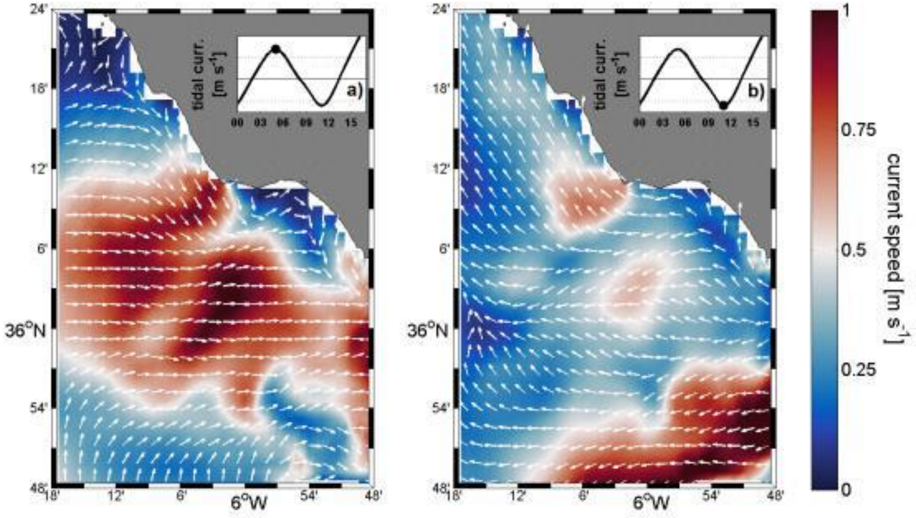
Some calculations to see the data distribution… and compare

Relation with other variables… & processes (heat dumping, surface oceanic currents…)

(what are the main factors affecting the coastal breezes?)



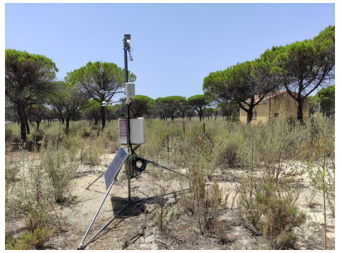


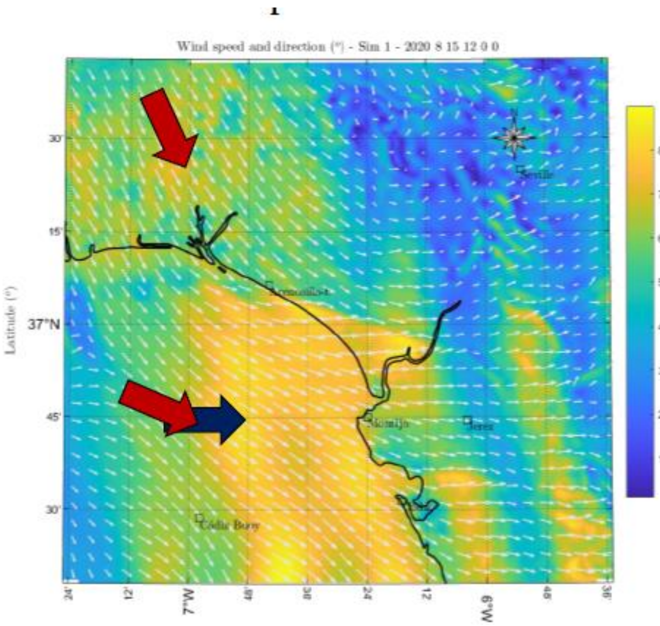


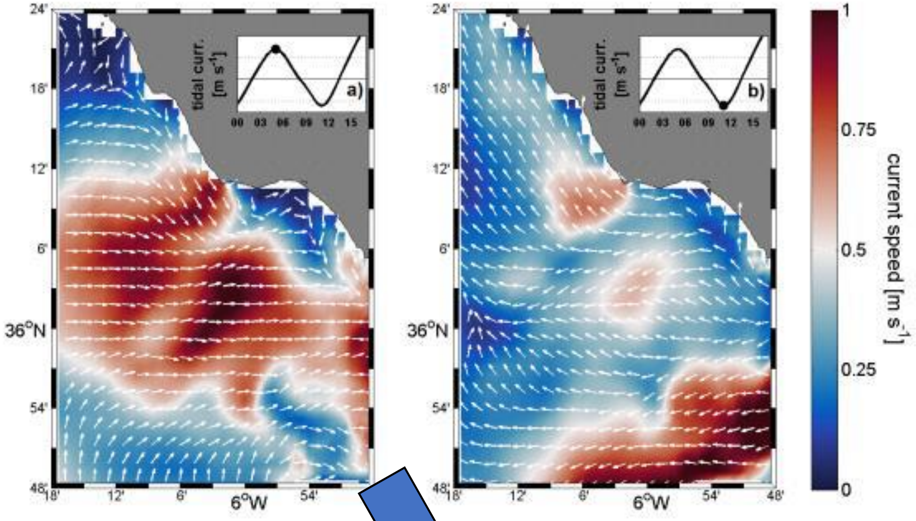
Bolado-Penagos et al., 2020
Progress in Oceanography 181, 102219.
https://doi.org/https://doi.org/10.1016/j.pocean.2019.102219

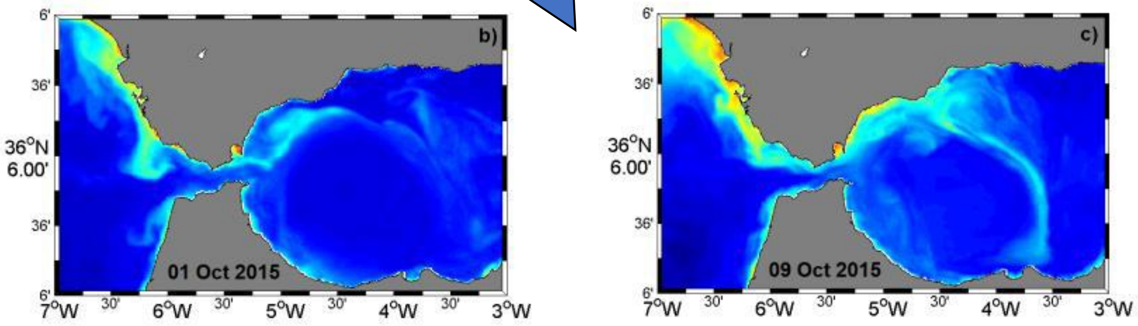
Relation with other variables… & processes (heat dumping, surface oceanic currents…)

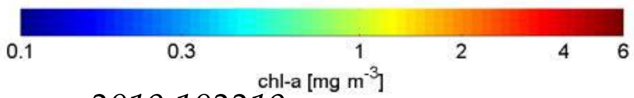
(what are the main factors affecting the coastal breezes?)







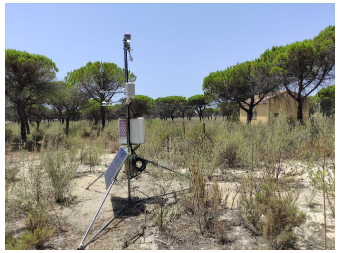


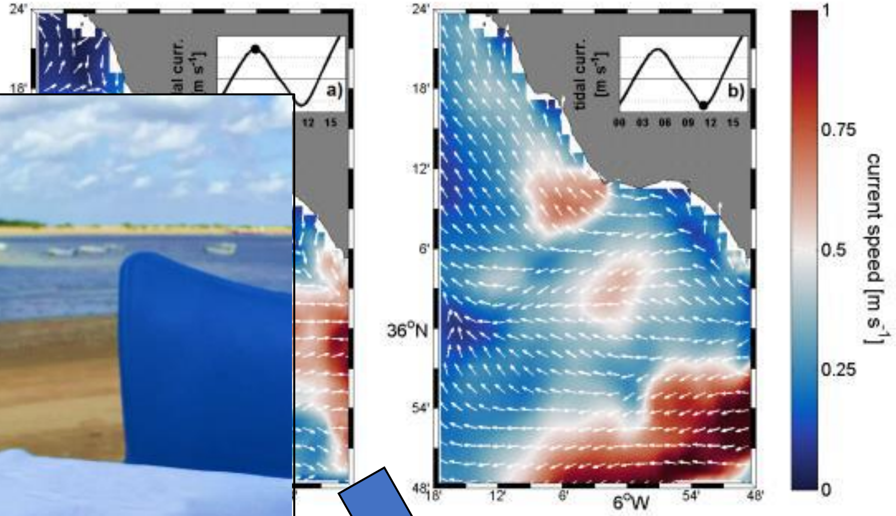


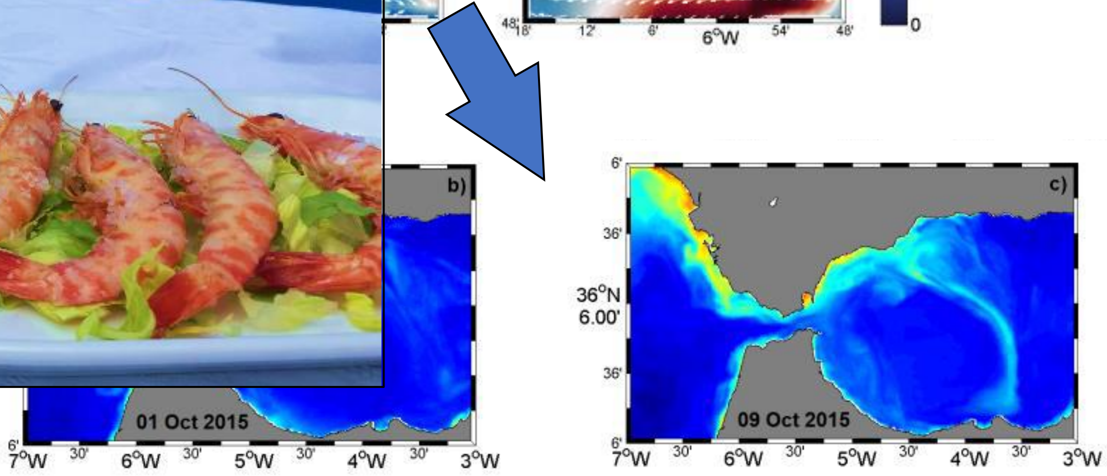
Bolado-Penagos et al., 2020
Progress in Oceanography 181, 102219.
https://doi.org/https://doi.org/10.1016/j.pocean.2019.102219

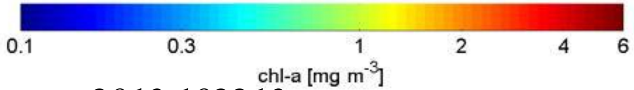
Relation with other variables… & processes (heat dumping, surface oceanic currents…)

(what are the main factors affecting the coastal breezes?)









Bolado-Penagos et al., 2020
Progress in Oceanography 181, 102219.
https://doi.org/https://doi.org/10.1016/j.pocean.2019.102219

https://climexp.knmi.nl/start.cgi

https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/history/method.shtml

https://ajdawson.github.io/eofs/latest/examples/elnino_standard.html

https://climatereanalyzer.org/

https://psl.noaa.gov/data/correlation/

Recommended Bibliography

Book: Data analysis methods in physical oceanography. *Thomson and Emery*

Book: Descriptive physical Oceanography: An Introduction. *Talley, Pickard, Emery & Swift*. (Chapter 6)

Book: An Introduction to Boundary Layer Meteorology. *Stull*

Book: Introduction to Micrometeorology. *Arya*

waves…

Thank you for your attention!

waves…

carlos.roman@uca.es

waves…

Caused by waves…

waves…

waves…