Erasmus+

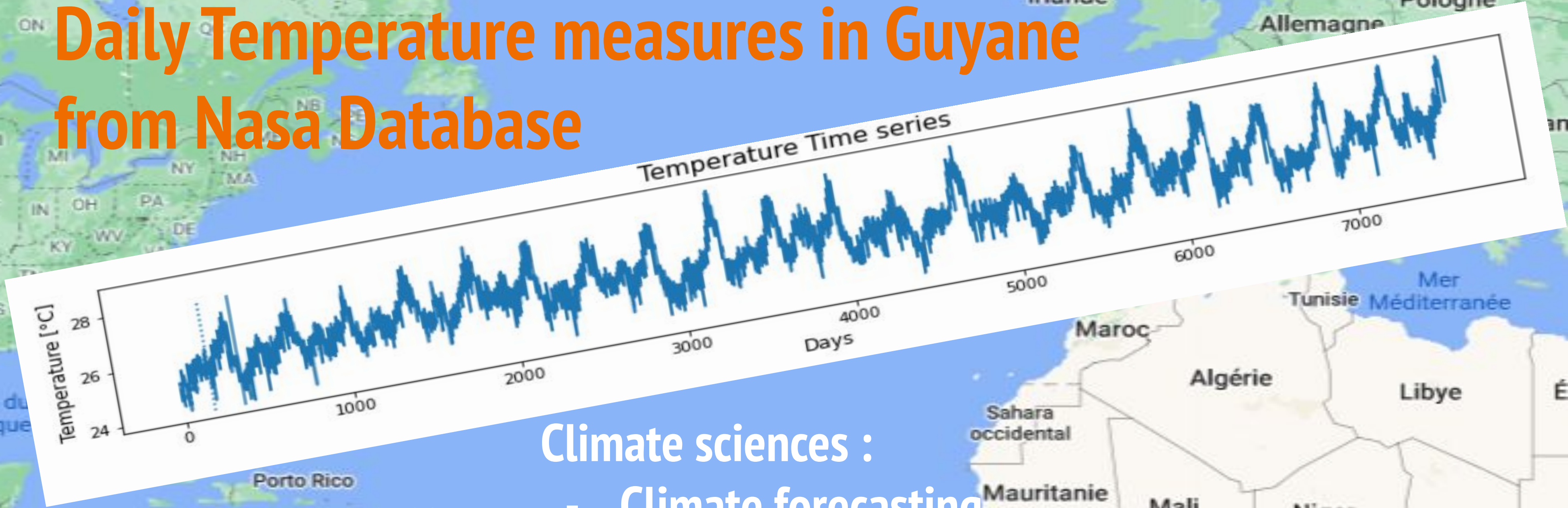
TU Graz

Time-series Analysis and Filtering

Lab 01

Aline Gauthier - Dominik Brandstetter - Fabian Veider

Daily Temperature measures in Guyane from Nasa Database



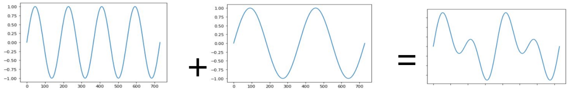
Climate sciences :

- Climate forecasting
- Climate change

Lat : 4.95091
Lon : -52.30179

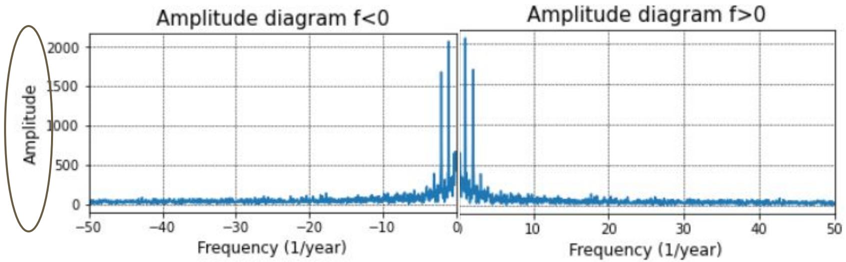
Amplitude Diagram

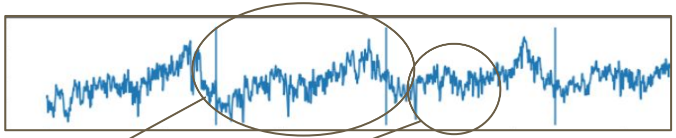
Amplitude spectrum is symmetric
⟶ Amplitude is **/ 2** and **x Nb of samples** ⟶ Amplitude Unit = (°C x Nb samples) / 2



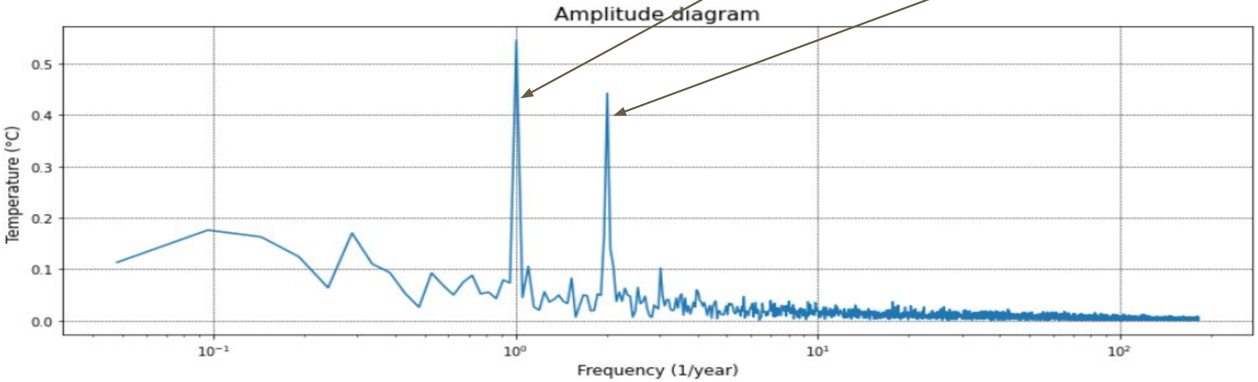
$f_0 = 1$ ⟶ one cycle per year
$2 \times f_0 = 2$ ⟶ quicker component

?



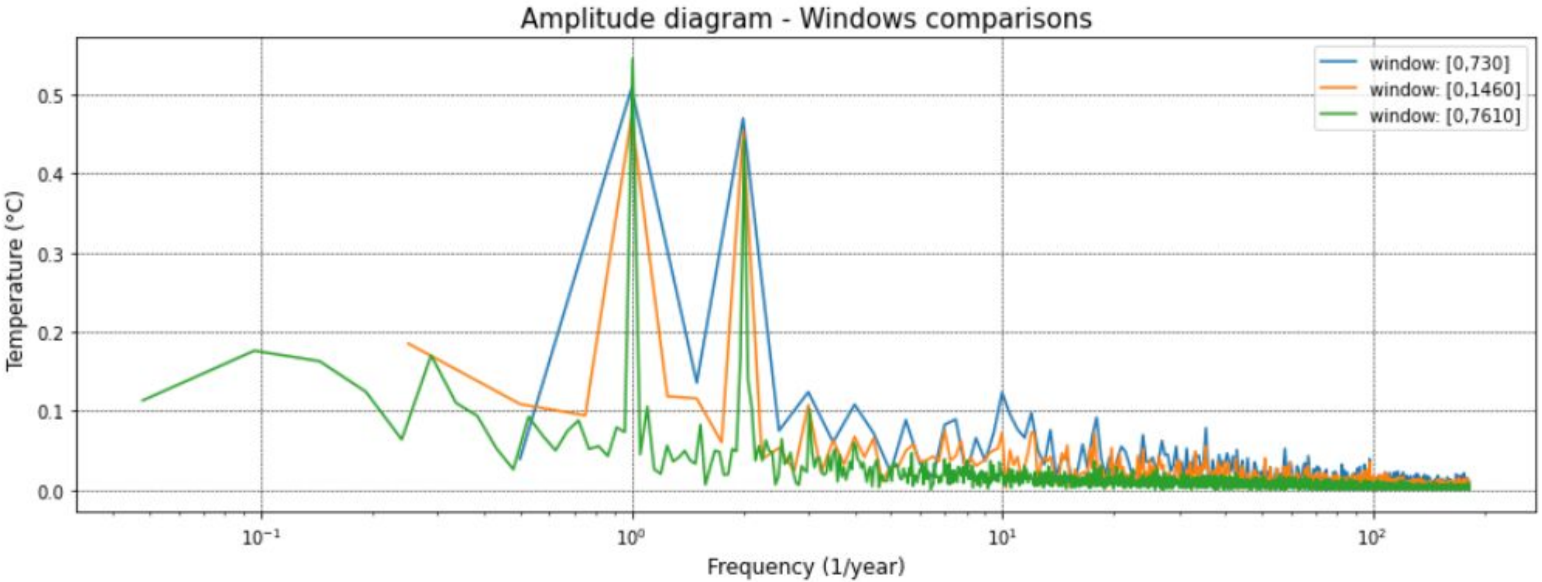


$$\sum_{n=1}^{N} |f(n)| = 27.5\,°C$$



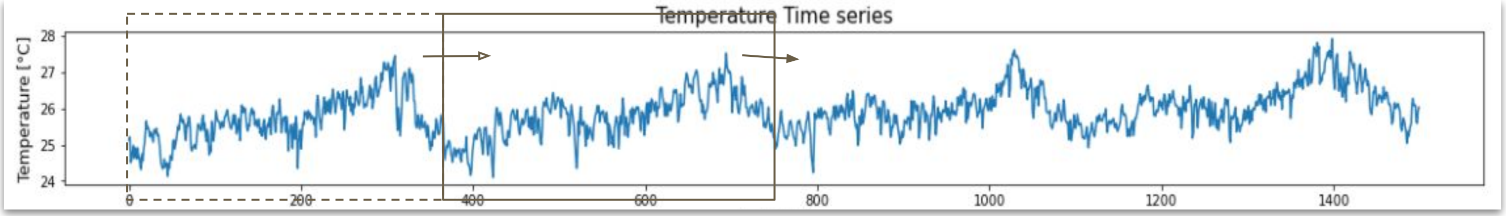
Importance of Windowing

small window ⟶ larger peaks
large window ⟶ smaller peaks

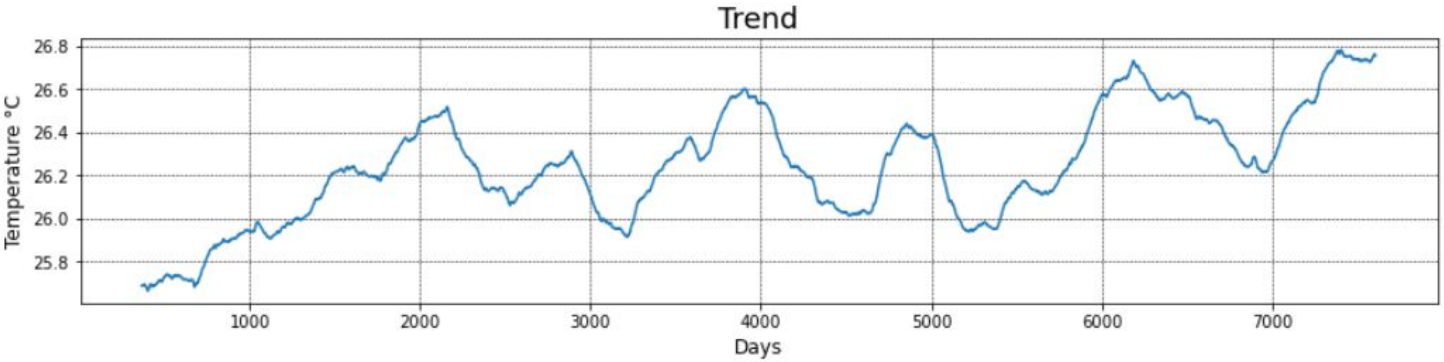


Trend (Moving average)

We compute the mean for a 365 days span ...



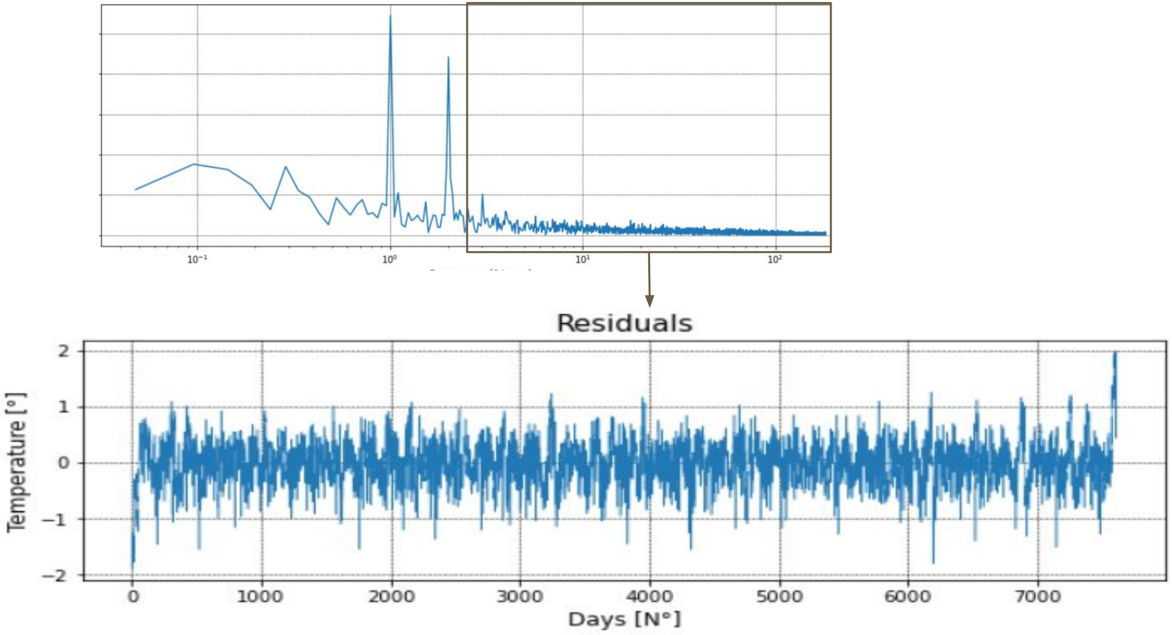
365 days

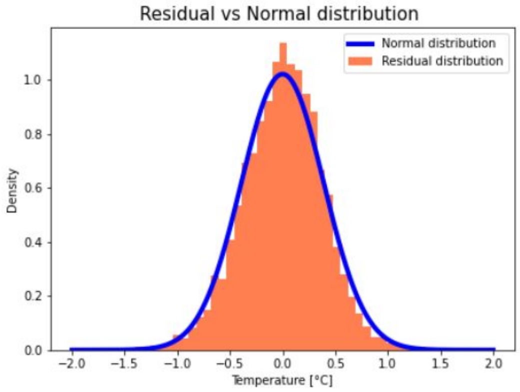


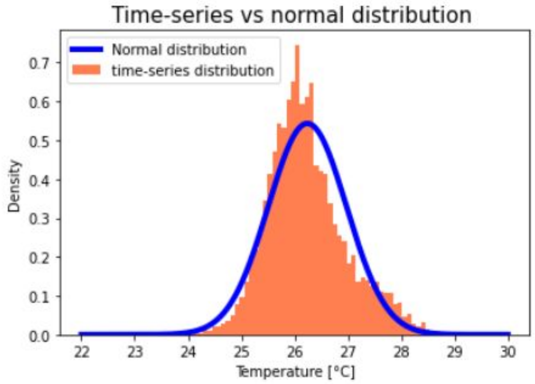
→ Hypothesis : temperature is increasing ? → Global warming?

Residuals

One the Amplitude diagram, we saw that two frequencies are higher (f=1 and f=2).
We keep only the frequencies above 2 to get the residuals.







SARIMA model ARIMA(p,d,q) (P,D,Q)s

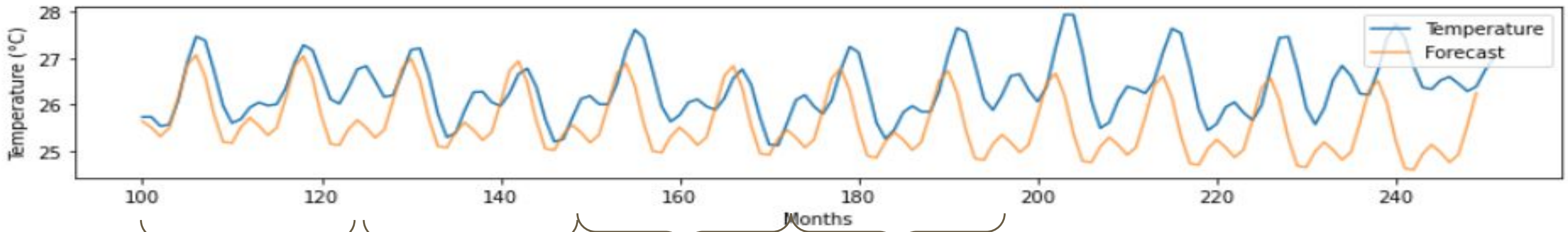
ARIMA is one of the most widely used forecasting methods for data forecasting.
S: Seasonality AR: autoregressive I: Integrated MA: Moving Average

Autoregressive specifies that the output variable depends on its own previous values (p,P)
Moving-average specifies that the output variable depends on past errors. (q,Q)
Integrate differencing serie to have a stationary serie (d,D)

$$y_t = c + \phi_1 y_{t-1}$$

Example with $\longrightarrow$ **ARIMA(1,1,1) (0,1,1)$_{12}$**

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} \qquad y'_t = y_t - y_{t-1} \qquad y'_t = y_t - y_{t-m}$$



2 ans : $r^2 = 0.85$ $r^2 = 0.66$ $r^2 = 0.54$ $r^2 = 0.40$

http://ijarcs.info/index.php/Ijarcs/article/download/4590/4172
8.9 Seasonal ARIMA models | Forecasting: Principles and Practice (otexts.com)

Autocorrelation

$$\mu_x = E[x_i] = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma_x^2 = E[(x_i - \mu_x)^2]$$

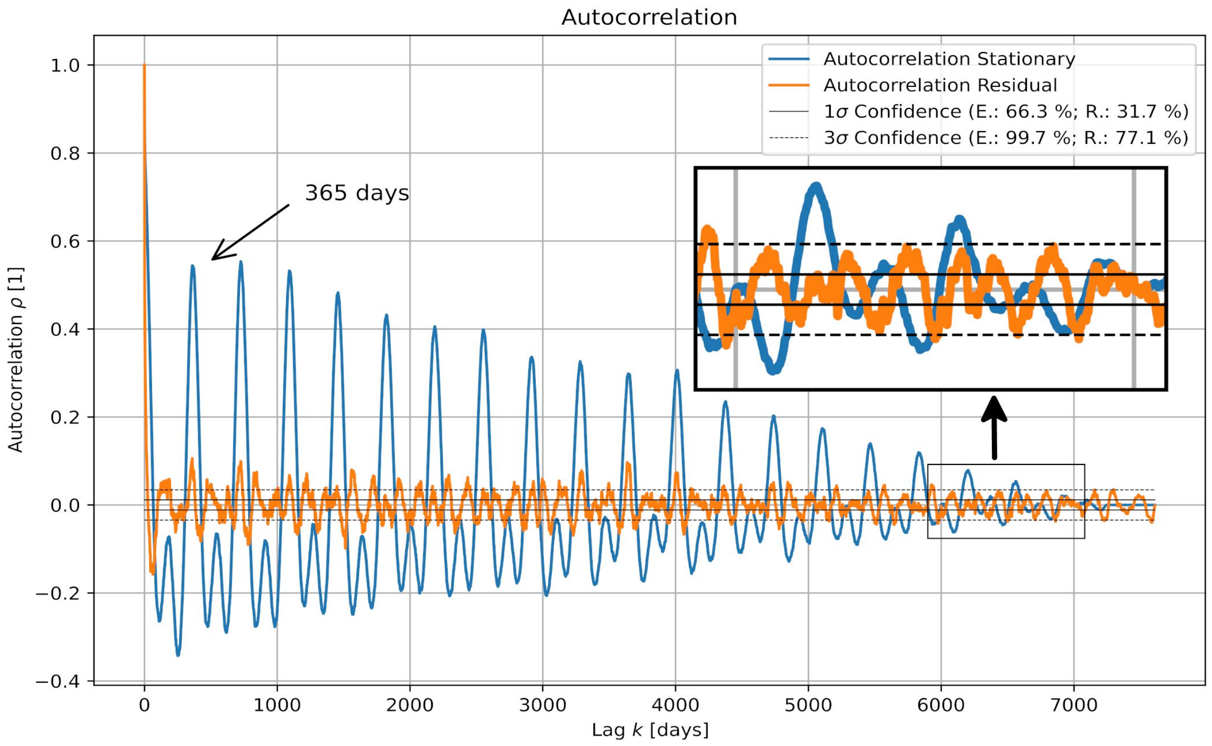
- (pearson) covariance between to discrete time series: $$cov(x_i, y_i) = \frac{E[(x_i - \mu_x) * (y_i - \mu_y)]}{\sigma_x * \sigma_y}$$
- autocorrelation is the covariance between a series and a shifted version of itself:

$$\rho(k) = cov(x_i, x_{i+k}) = \frac{E[(x_i - \mu_x)(x_{i+k} - \mu_x)]}{\sigma_x^2} \qquad k = 1, 2, ..., N-1$$

- additional information:
  - per definition: $\rho(0) = 1$
  - “decay” due to decreasing overlap of finite series (slightly different definition of ρ(k)

Source: Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics)* (5th ed.). Wiley.

Applied Autocorrelation



- clear positive correlation for k = n * (365 days) ⇒ annual cycle
- inverted peak for k = n * 1 year + 185 days ⇒ second warm period
- decay of peak heights is to be expected
- too many values outside confidence intervals ⇒ still signal left

Autocorrelation & Sample Mean

- average temperature per year
  ⇒ variance of the sample mean?
- "naïve" approach (only for uncorrelated measurements):

$$\sigma_\mu^2 = \frac{1}{N}\sigma_x^2$$

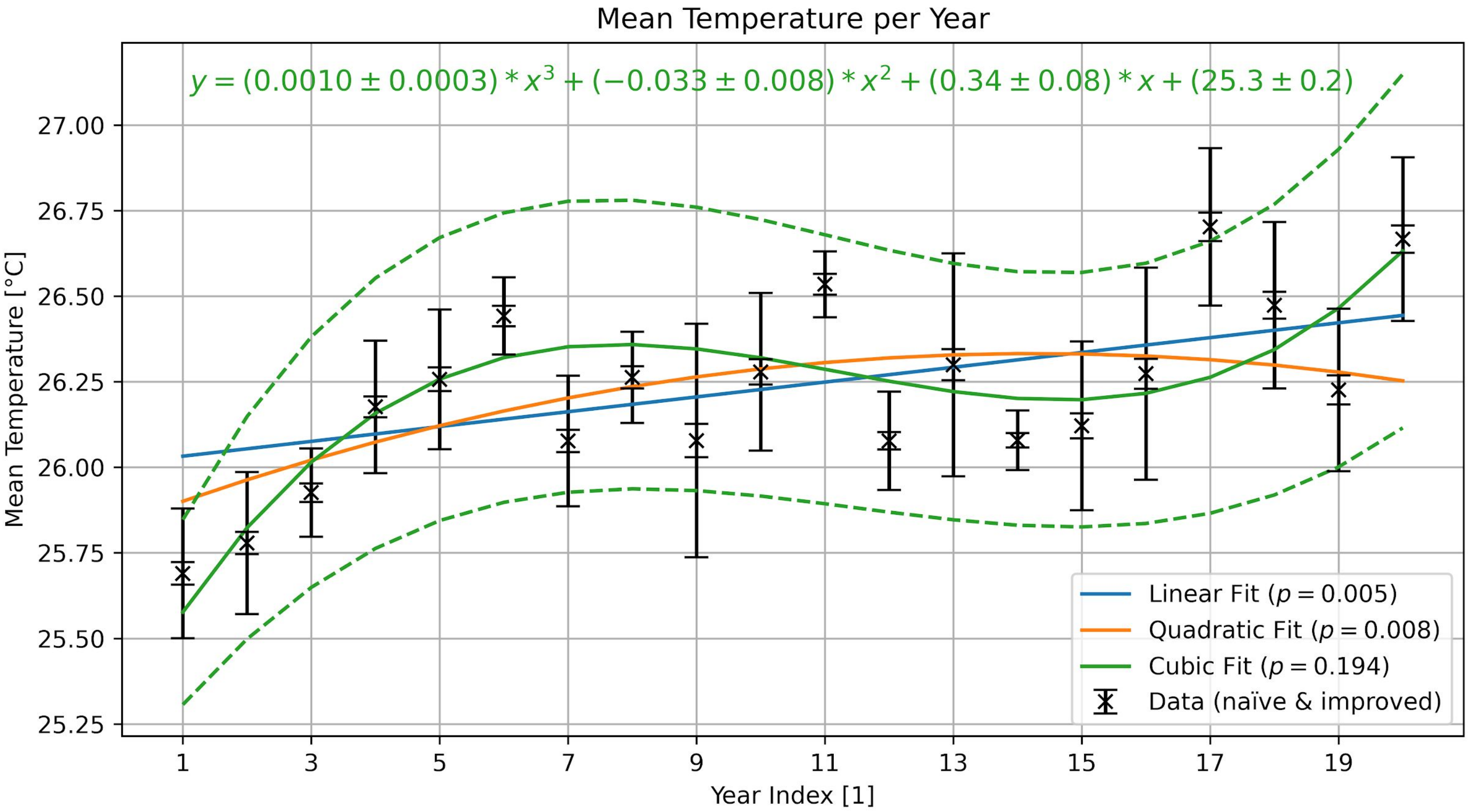
- improved version by taking autocorrelation into account:

$$\tilde{\sigma}_\mu^2 = \frac{\sigma_\mu^2}{N}\left[1 + \frac{2}{N}\sum_{k=1}^{N-1}(N-k)\rho(k)\right]$$

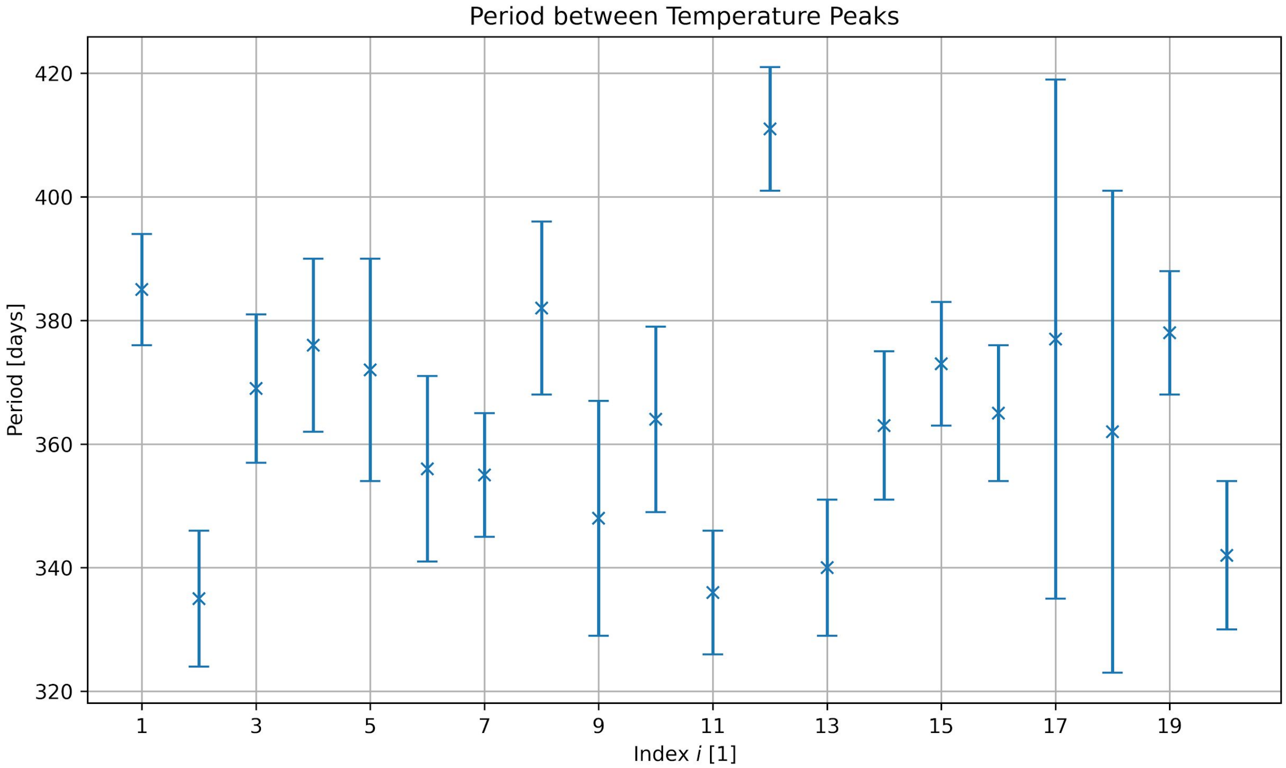
- autocorrelation reduces "number of independent measurements"
- requires: stationary process (= no trend)

Source: Witt, T. J., & Fletcher, N. E. (2010). Standard deviation of the mean and other time series properties of voltages measured with a digital lock-in amplifier. Metrologia, 47(5), 616–630. https://doi.org/10.1088/0026-1394/47/5/012

Mean Temperature



Annual Period



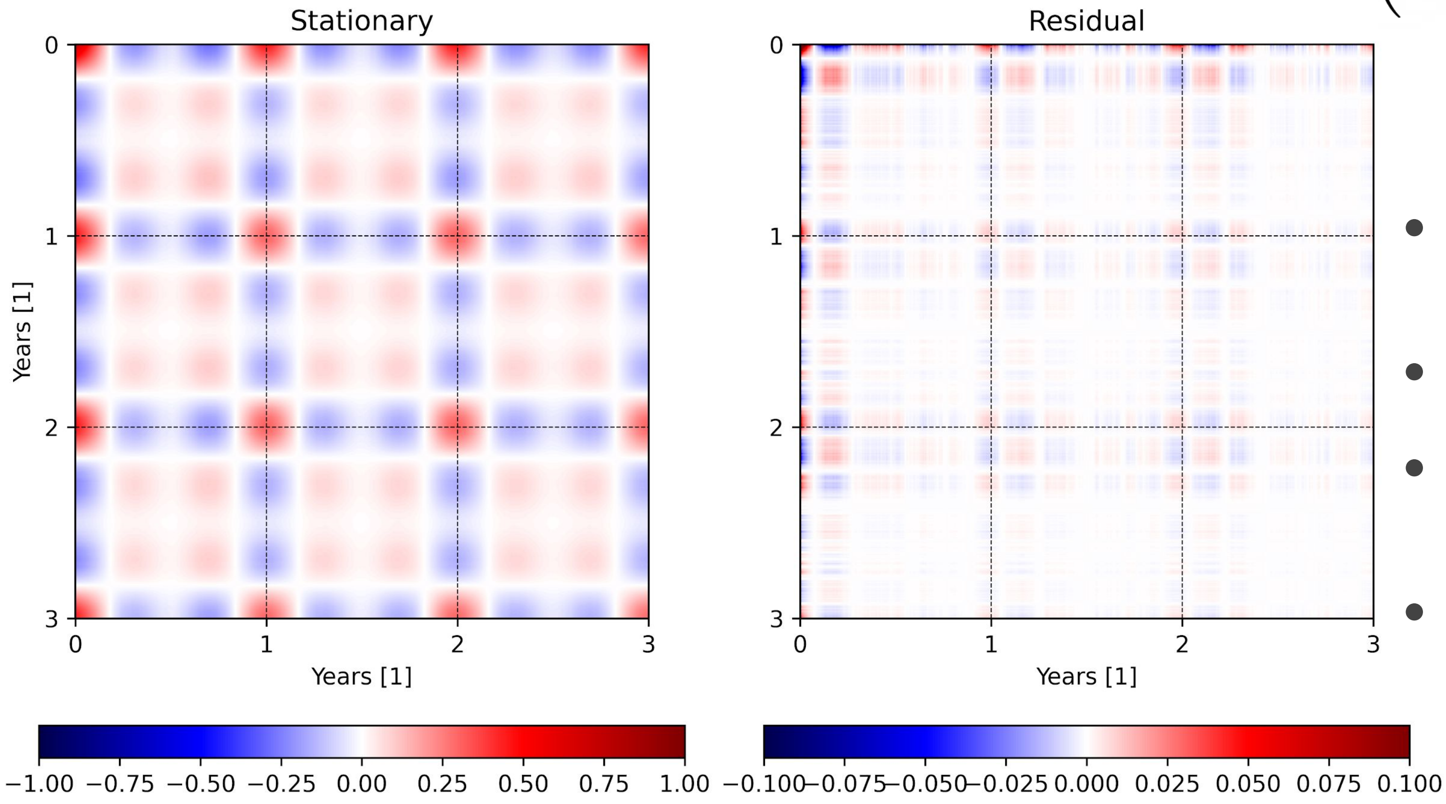
- Period T = (364 ± 4) days
- no clear correlation or trend
- large uncertainty due to dull peaks
- outlier at i = 12

  (Grubb's test for p = 0.15)

Bonus: Autocorrelation matrix

$$\left(\hat{P}\right)_{ij} := \rho_i * \rho_j$$

$$\rho(k) \to \rho_k$$

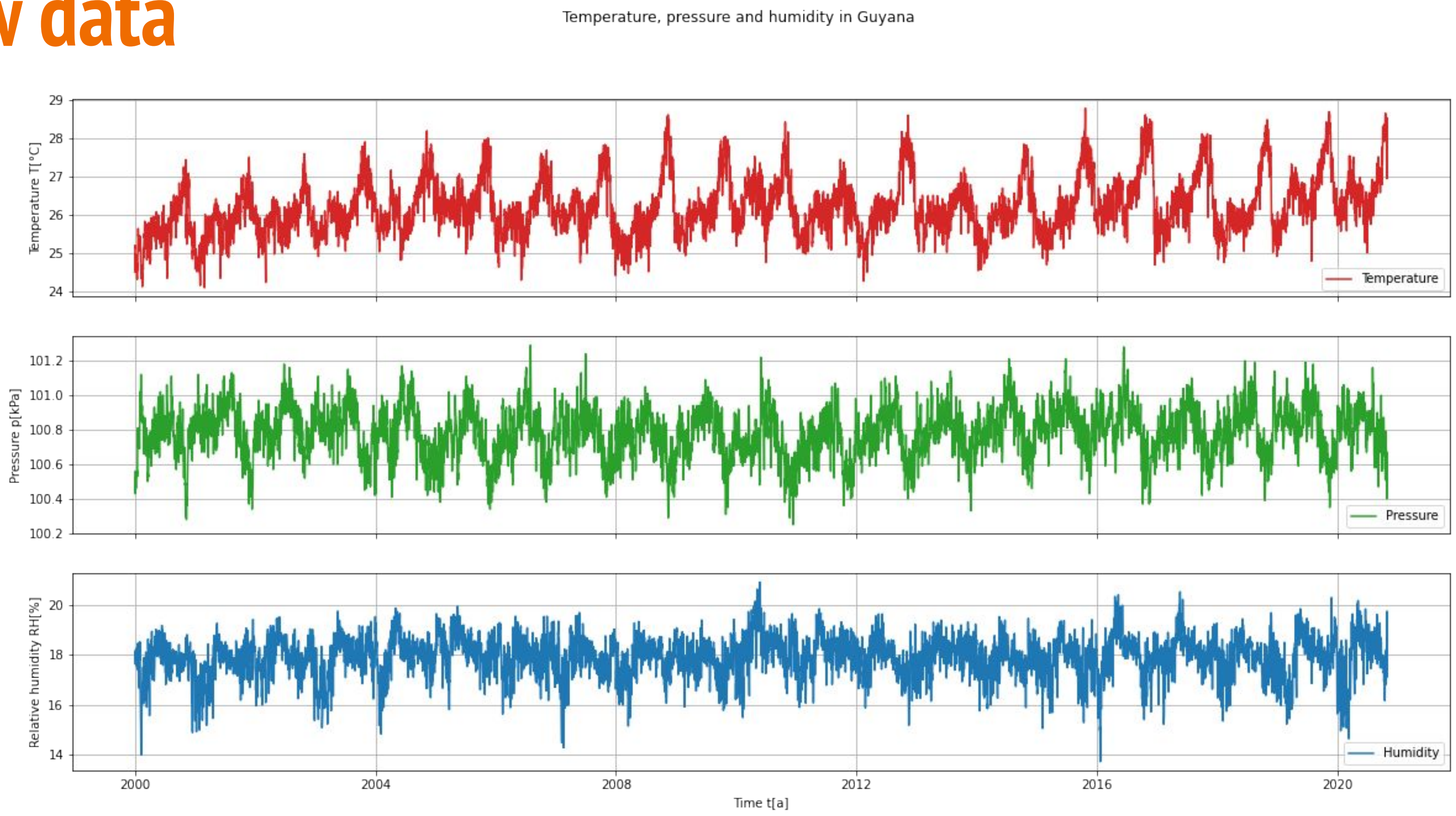


- visualization of autocorrelation
- only first 3 years
- symmetrix per definition
- different scales!

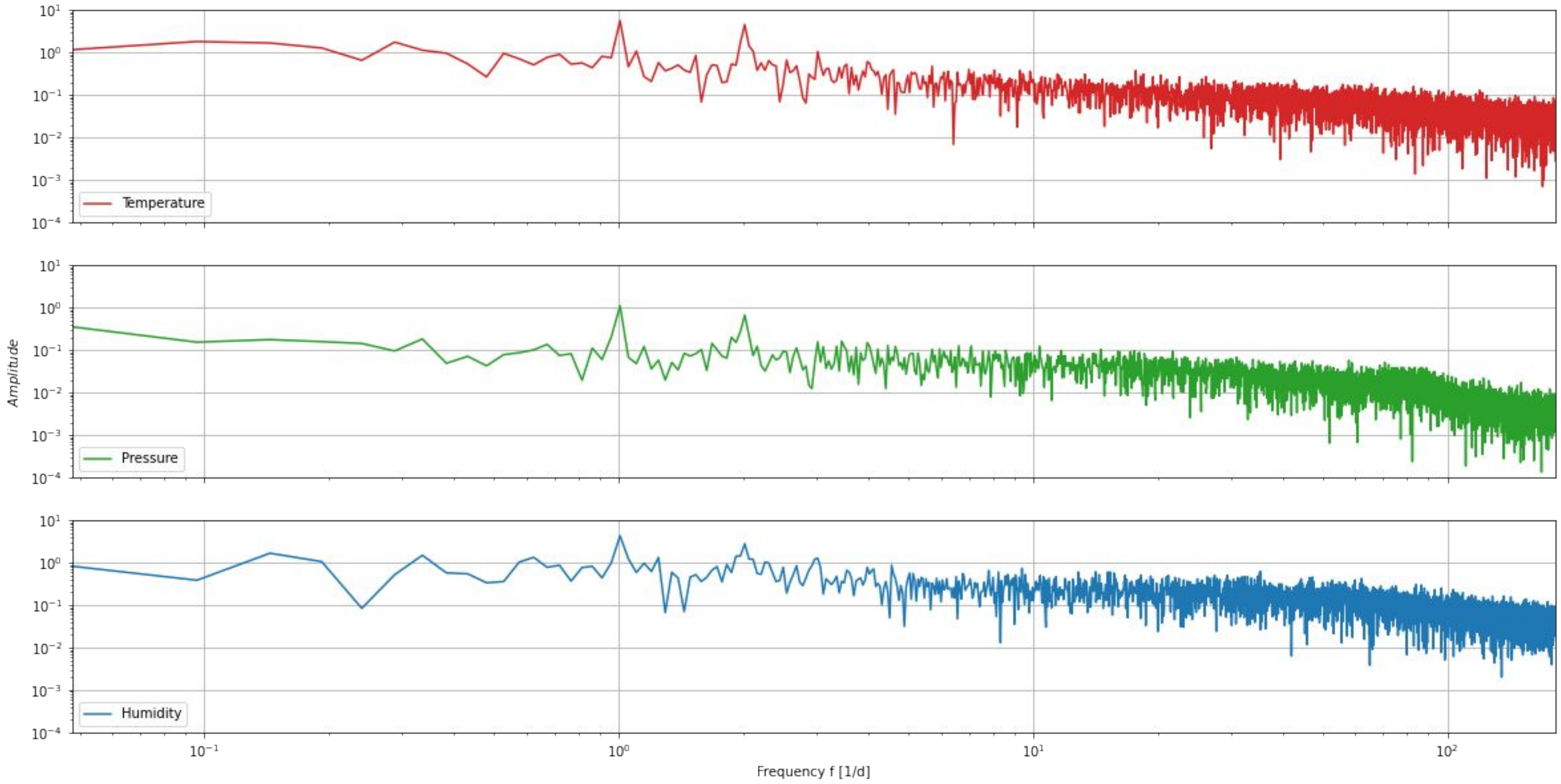
Comparison to pressure and humidity

- How does the temperature compare to pressure and humidity?
- Dominant frequencies
- Trend, seasonality, noise

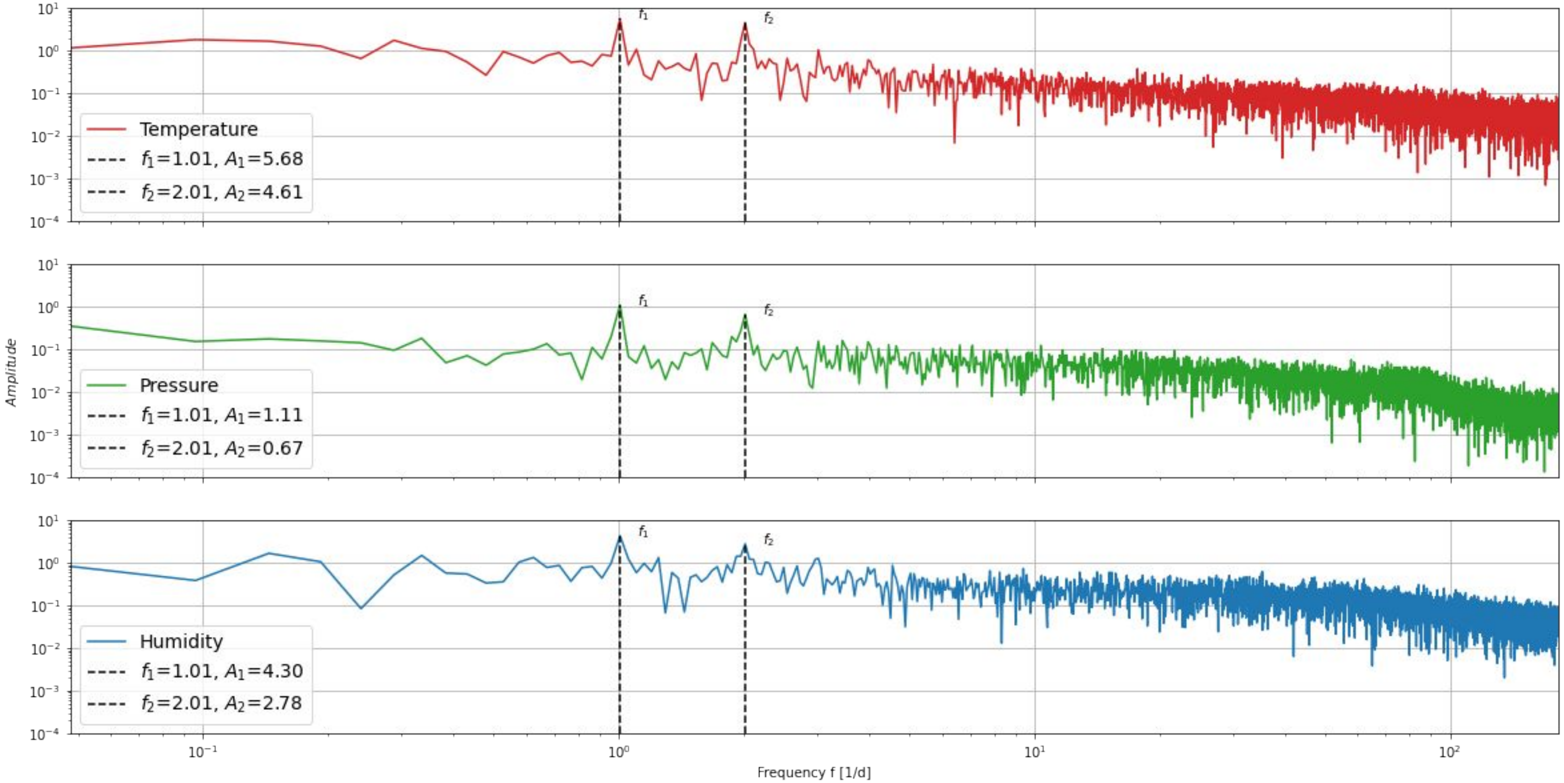
Raw data


Temperature, pressure and humidity in Guyana
Temperature T[°C]
29
28
27
26
25
24
Temperature
Pressure p[kPa]
101.2
101.0
100.8
100.6
100.4
100.2
Pressure
Relative humidity RH[%]
20
18
16
14
Humidity
2000
2004
2008
2012
2016
2020
Time t[a]

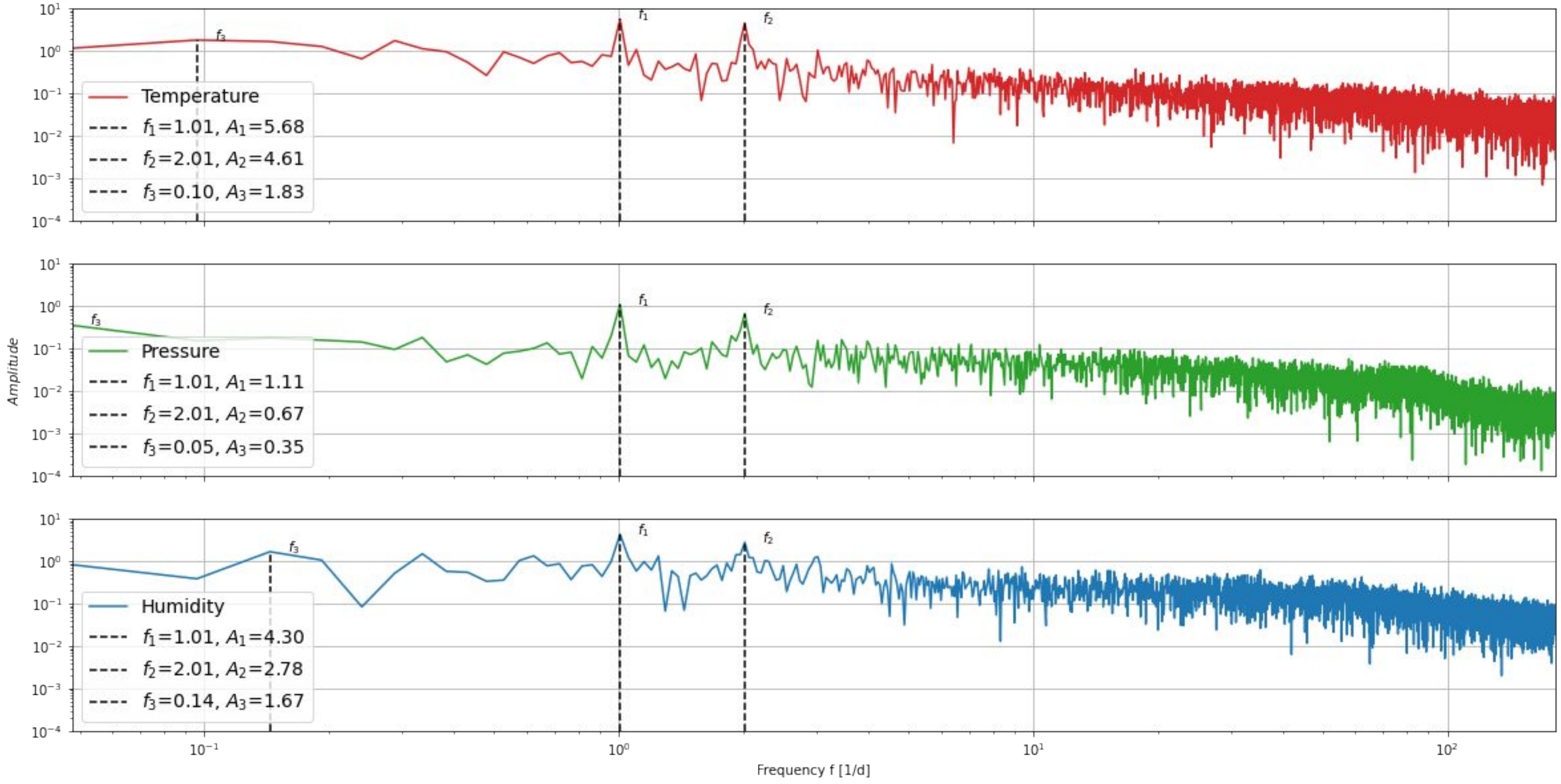
FFT - Amplitude



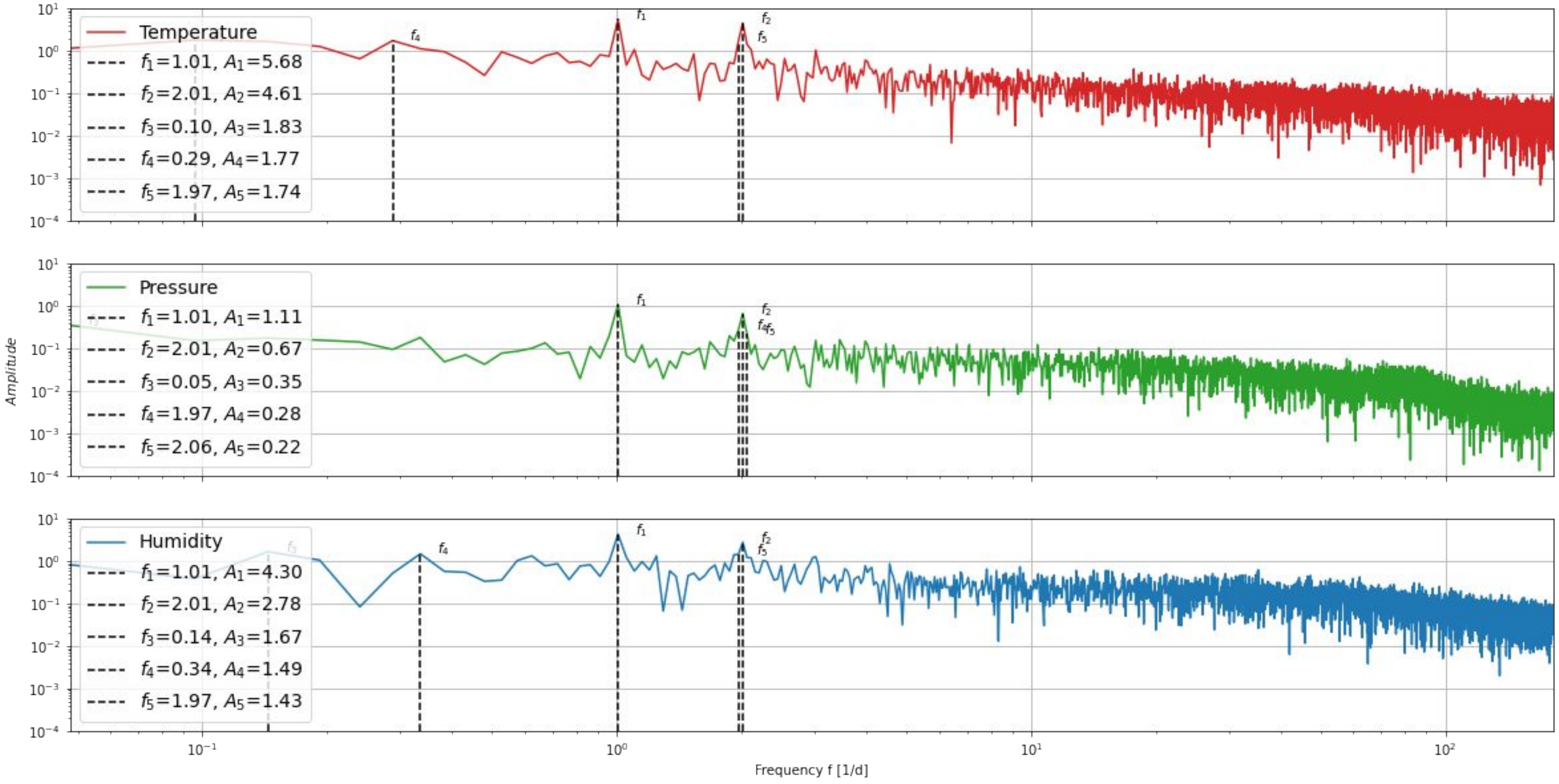
Dominant frequencies



Dominant frequencies



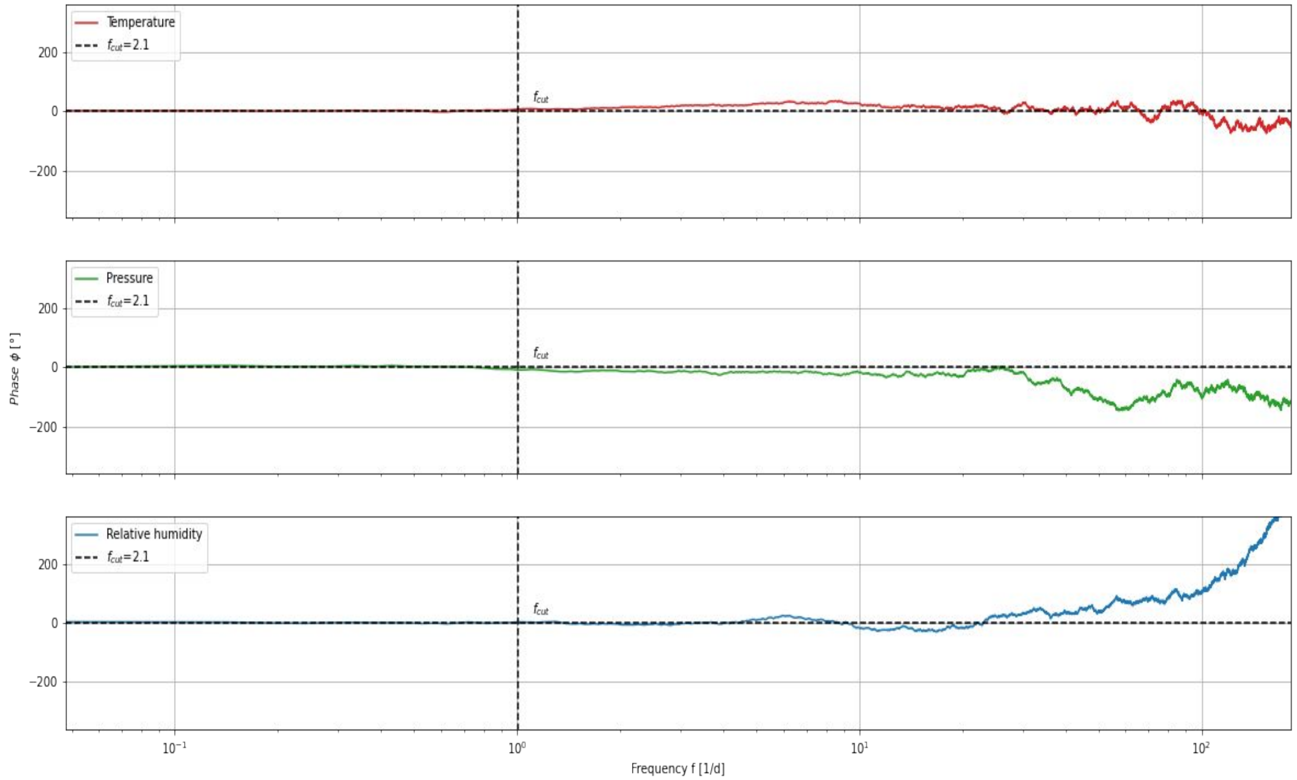
Dominant frequencies



Dominant frequencies - Interpretation

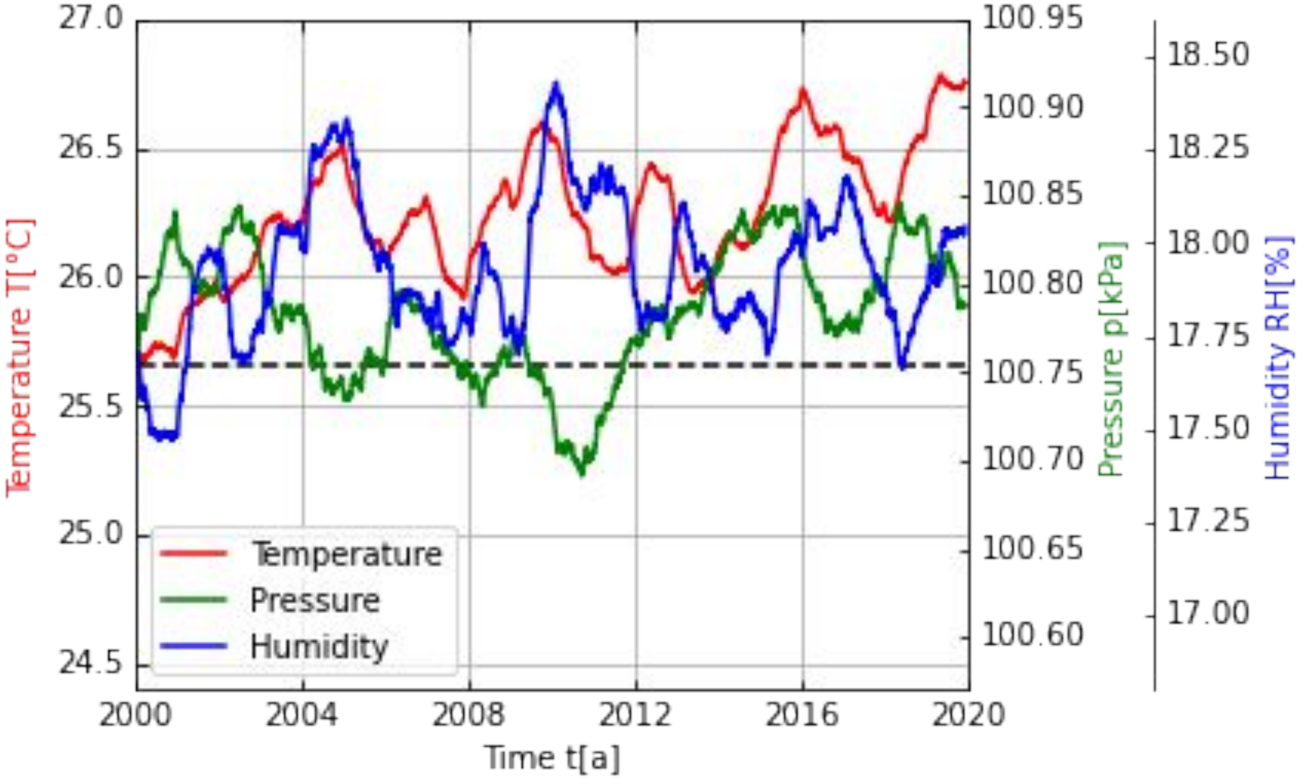
- Same two dominant frequencies
  - Two seasons visible in pressure and humidity
- f > 2.1 only small amplitudes
  - Apply lowpass filter

FFT - Phase



- Small phase shift when f < 2.1
- Strong increase for humidity

Trend comparison



- Visible increase in temperature
- Pressure and humidity oscillate

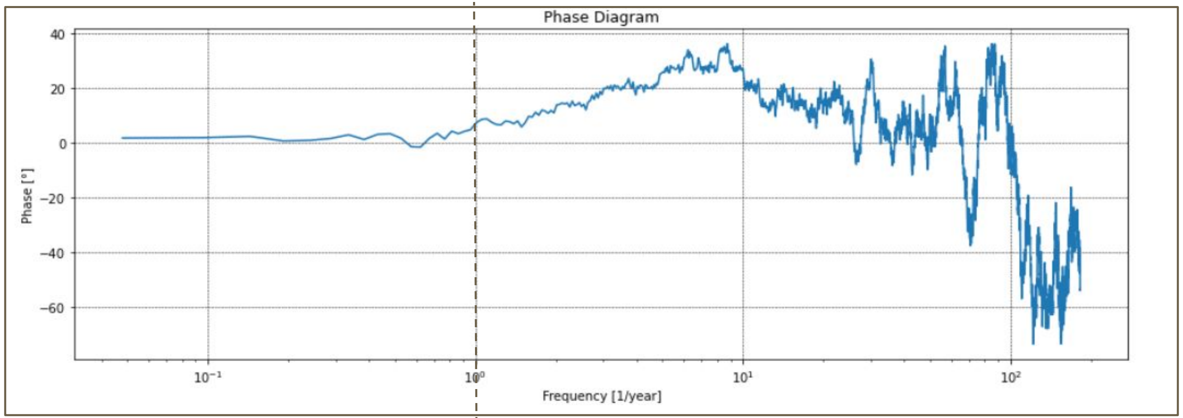
Seasonality and noise

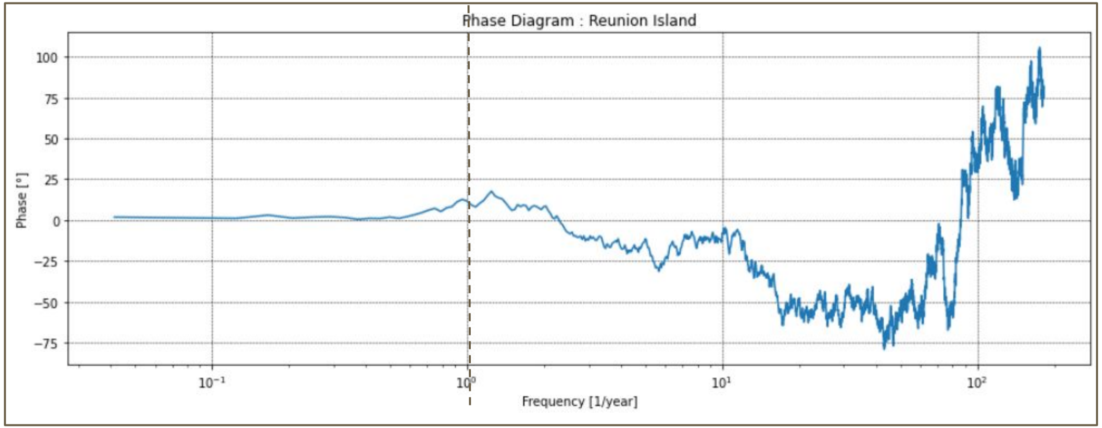
- Visible increase in temperature
- Pressure and humidity oscillate

Part 1 | Part 2 | **Part 3**

Phase Diagram

Guyane →





Phase switching after f > 1

← Reunion island

Conclusion

Findings ?

Residuals → white noise + signal left
Climate change and global warming
Year duration =
Mean temperature per year in Guyane =

Which tools ?

FFT Amplitude - Phase
Moving average
Autocorrelation
Residuals analysis (Normal distribution)
Forecasting with SARIMA model
Inverse FFT

Future Studies ?

Filtering

Are there more time series describing or observing the same phenomenon? **Or is there a model which describes the phenomenon in your time series**? **Compare your time series to at least one** other time series of observations or to a time series from a model.

A brief introduction/motivation to the problem at hand, relevant details about the data, **additional relevant scientific information from searching the web**, for example, and what is to be addressed.

A presentation of the **results of your analysis**, with **special focus on the interpretation of your results**. Are there limitations your study might suffer from? What could possible future studies contain?