Time reversibility

Natural time analysis

August 26, 2023

Dataset

Catania daily precipitation

- 2002/2022
- 7669 datapoints
- minimum not null value: 0.2 mm
- maximum value: 161.6 mm
- mean value (over all days ≥ 0 - 7669 datapoints): 1.71 mm
- mean value (over wet days > 0 - 1686 datapoints): 7.79 mm

Questions

1. Why is there a plateau on the PDF's plot?
2. What about applying time reversal over intertimes above std dev threshold and over daily volumes above volume threshold or intensity classes?
3. Can we compare the PDF across different decades of daily volumes without threshold?
4. Can we compare the two rain gauge SIAS and RBI data from 2002 to 2018?
5. Can we test the predictive power of time reversal?

1. Daily volumes and Histogram Representations

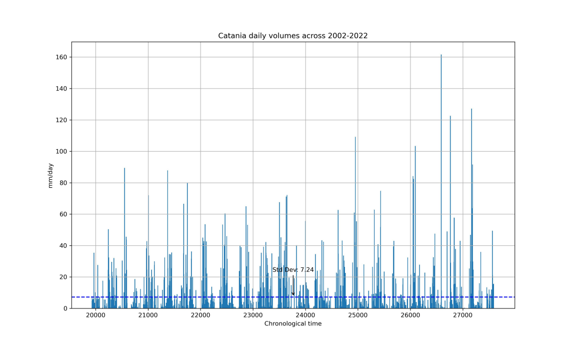
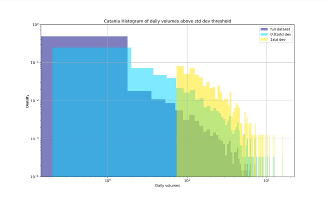


Figure 1: Blue bars: [mm] of rain per day. Blue dashed line: standard deviation



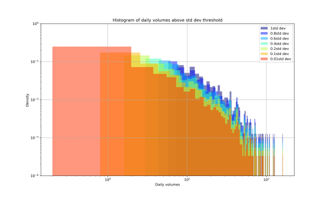


Figure 2: Top: comparison between the histogram of the full dataset without threshold and two filtered dataset (above 100% and 1%). Bottom: comparison between histograms over filtered dataset (above 100, 80, 60, 40, 20, 10 and 1 %)

1. PDF of daily volumes norm_pdf

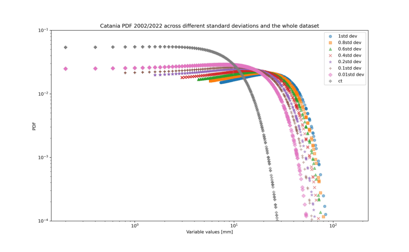


Figure 3: Comparison between PDF of full dataset and data filtered by standard deviation

$$PDF = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{1}$$

x = [mm]

σ = standard deviation

μ = mean value

Assumption = continuous data normally distributed

Risk = Overfitting

1. PDF of daily volumes kde_pdf

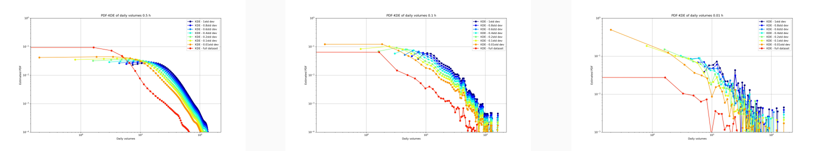


Figure 4: Comparison between Probability Density Function of full dataset and data filtered by standard deviation - from the left h=0.5,0.1,0.01

$$PDF = \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-x_i}{h})^2} \qquad (2)$$

n = # of datapoints

h = bandwidth

σ = standard deviation

x = value at which PDF is computed

x_i = datapoints

Assumption = continuous data

Risk = Overfitting for too large bandwidth

1. PMF of daily volumes

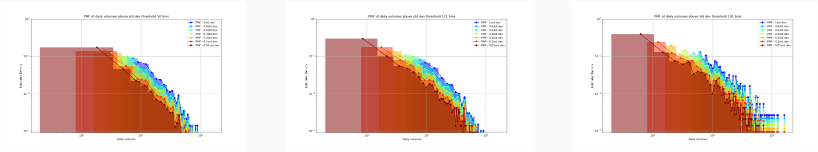


Figure 5: Comparison between Probability Mass Function of data filtered by standard deviation depending on bins - from the left bins=50,121,191

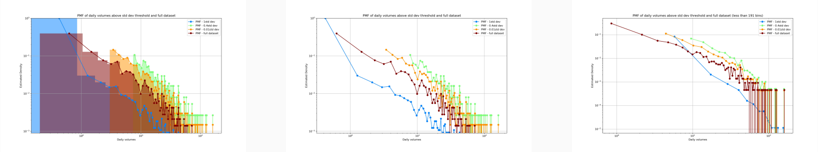


Figure 6: Comparison between Probability Mass Function of data filtered by standard deviation depending on bins- from left bins=191,191,121

$$PMF = HIST = \frac{n_x}{n} \tag{3}$$

n = # of datapoints

n_x = # of occurrences

Assumption = discrete data

NB = the sequence of datapoints is divided into BINS

1. WARNING

- Daily precipitations are NOT normally distributed
- The full dataset (7668 datapoints) of daily precipitation is CONTINUOUS
- The data filtered based on the std dev threshold (1686 $\div$464 datapoints) are DISCRETE
- Would it be more appropriate to employ

  $\longmapsto$ PDF-KDE for full dataset;

  $\longmapsto$ PMF for filtered dataset

  $\longmapsto$???

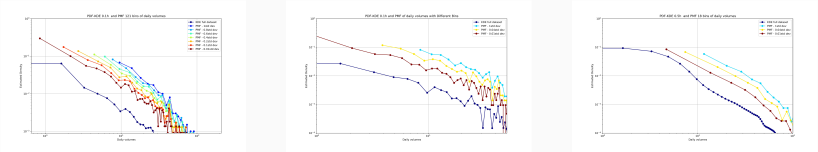


Figure 7: Comparison between Probability Mass Function and Probability Density Funtion. Left: 0.1h for full dataset and 121 bins for filtered dataset. Center: 0.1h for full dataset and customized bins for filtered dataset. Right: 0.5h for full dataset and 18 bins for filtered dataset.

1. And so?

By adjusting the parameters of the choosen Probability Function, such as bandwidth h and bin count BINS, we can enhance the similarity between the probability function representations of filtered and unfiltered data (as shown in center panel Fig.7).

2. Intertimes time reversal

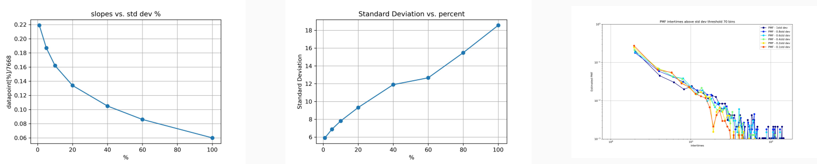


Figure 8: Left: slopes of filtered datapoints/full dataset vs std dev% . Center: Standard deviation of the antisymmetric sum vs std dev %. Right: Probability Mass Funtion for intertimes (based on std deviation filtering)

2. Time reversal

Here follows some tables about the standard deviation value of the antisymmetrical sum: each table has been obtained applying the time reversal

- over intertimes between events above different standard deviation thresholds
- directly over above-threshold volumes (min threshold)
- directly over precipitation's intensity classes (min/max threshold)

2. Tables

mm	%	std dev [*days*]	#
≥ 0.2	1	5.91	1686
≥ 0.4	5	6.88	1436
≥ 0.8	10	7.81	1245
≥ 1.6	20	9.32	1028
≥ 3.0	40	11.89	808
≥ 4.4	60	12.66	605
≥ 5.8	80	15.48	560
≥ 7.4	100	18.56	464

Time reversal applied over intertimes (days) between events above std deviation threshold (mm,%). "std dev" is referred to the antisymmetrical sum; "#" rapresents the n.o of datapoints.

mm	std dev [*mm*]	#
≥ 0	5.13	7669
> 0	9.69	1686
≥ 15	15.30	248
≥ 45	18.54	44

Time reversal applied over daily volumes above [*mm*] threshold.

int class	std dev [*mm*]	#
$[0, 15]$	2.70	1438
$[15, 45]$	5.51	205
$[45, 90]$	7.94	36

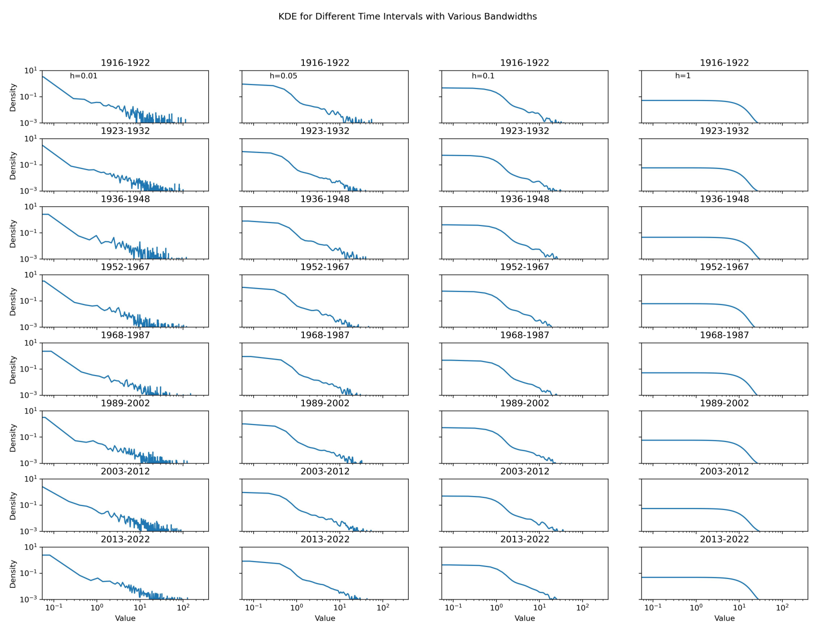
Time reversal applied over daily volumes (mm) depending on intensity classes.

2. And so?

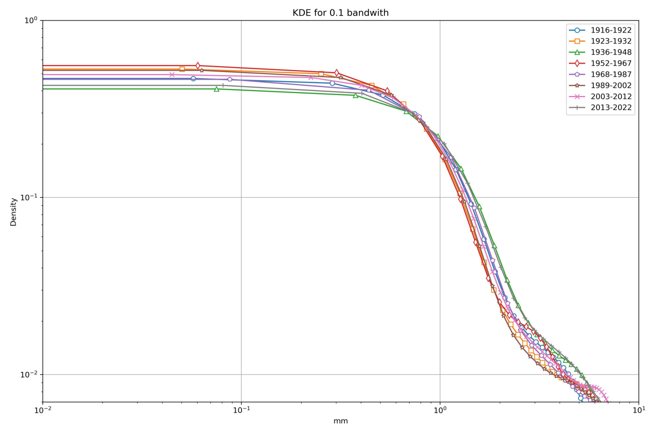
Except the full dataset with 7669 datapoints, the datasets with both a statistical relevance and meaningful daily volumes threshold are made up of 1668 and ca. 1438. The lowest value for the antisymmetric sum's standard deviation:

- 7669: full dataset$\geq$0
  std dev = **daily volumes** time reversal 5.13 mm
- 1686 : wet days > 0
  std dev = **intertimes** time reversal 5.91 days
- 1438 : soft precipitation intensity class ,i.e $[0, 15]$
  std dev = **daily volumes** time reversal 2.70 mm

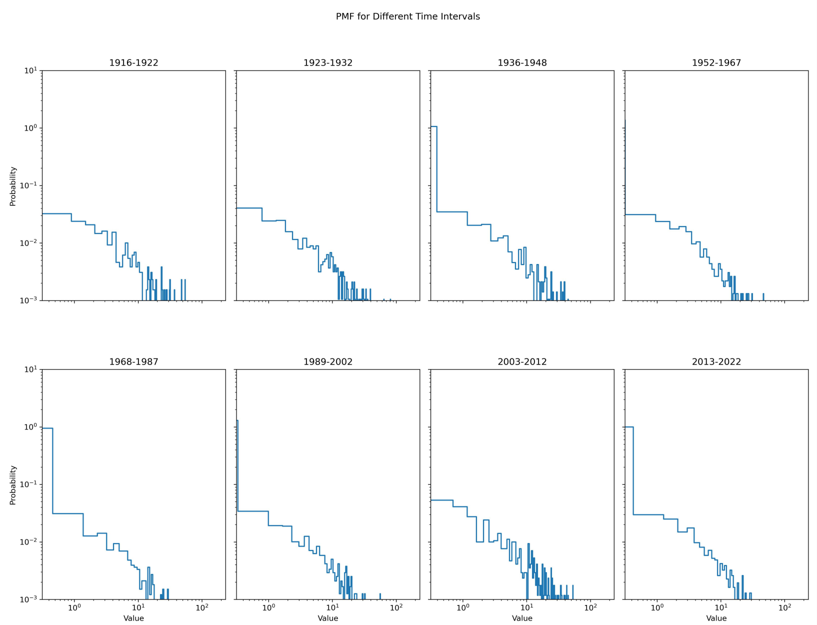
3. Decades comparison - kde_ pdf with different bandwidths



3.Decades comparison - kde_ pdf with 0.1 bandwidth



3. Decades comparison PMF



4. Servizio Agrometeorologico Regione Sicilia & Rete Bacino Idrografico

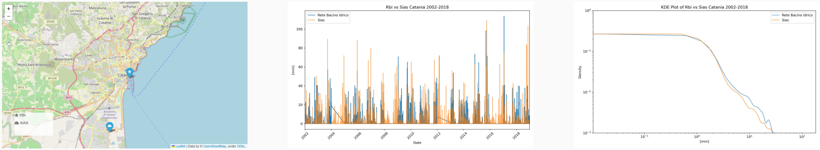


Figure 11: Comparison between SIAS and RBI rain gauges (distance ca. 7.3 km) from 2002 to 2018

5.Testing the predictive power of time reversal

Algorithm

1. Select a dataset to predict
2. Choose suitable historical dataset over which apply the time reversal procedure
3. Overlap present and past natural time series

5. Selection of the dataset

The last 21-years dataset consists of 7668 datapoints: the filtering process affects heavily the no of datapoints. As we can see on Tables (pag 12) if we exclude dry days (i.e days with 0 precipitation) the whole dataset is reduced to almost 1686 datapoints. So when we decide to choose a dataset to predict the last 21 years, it must consist of the same no of datapoints of the choosen last 21-years dataset, which we will refere to as PRESENT DATASET.

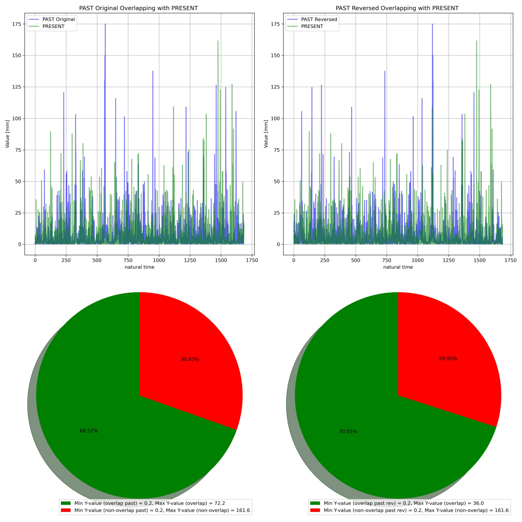
NB For any filtered present, 21 past years are never enough to satisfy the equality about the no of datapoints: we can deduce that the no of dry days increased during the last 21 years

5. Dataset

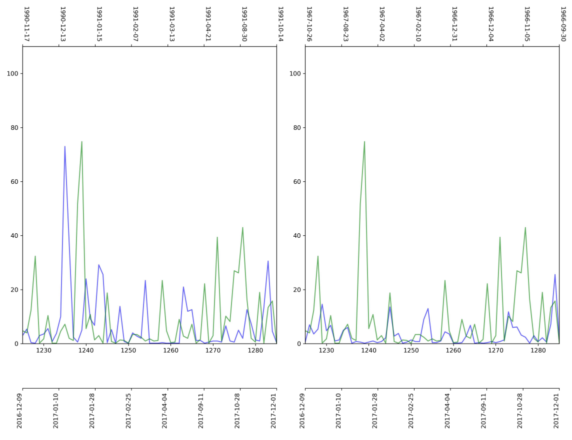
- Filtered data > 0
- Datapoints 1686
- PRESENT 21/01/2002 - 31/12/2022 21y
  antisymm std dev 9.69 [mm]
- PAST 03/03/1961 - 31/12/2001 26y
  antisymm std dev 10.38 [mm]
- PAST PAST 03/03/1921 - 02/03/1961 26y
  antisymm std dev 10.13 [mm]

5 PAST → PRESENT prediction

- Overlap PAST-PRESENT
- Overlap PAST REVERSE-PRESENT
- tolerance 10.38 $[mm]$
- Mean common datapoints 1158 over 1686



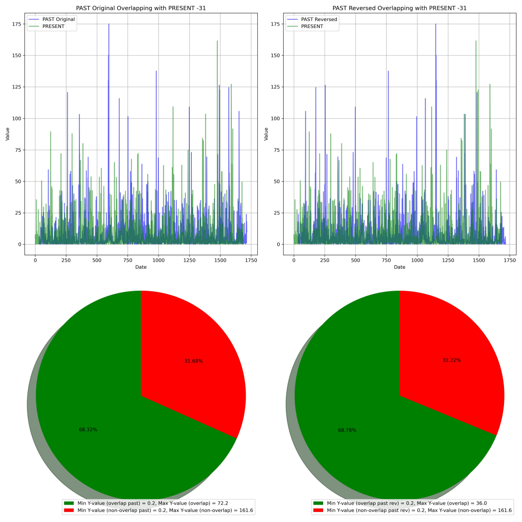
5 Zooming the overlap



- Let's shift and overlap (blue-past ; green-present)

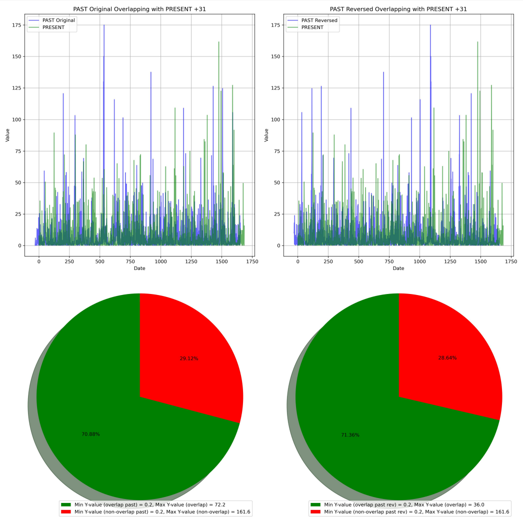
5 PAST $\rightarrow$ PRESENT prediction with a negative shift [*days*]

- Overlap PAST-PRESENT
- Overlap PAST REVERSE-PRESENT
- tolerance 10.38 [*mm*]
- shift of PAST to the left -31 days

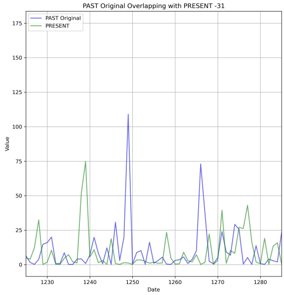


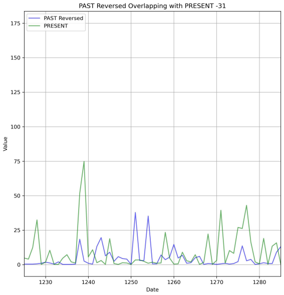
5 PAST $\rightarrow$ PRESENT prediction with a positive shift [*days*]

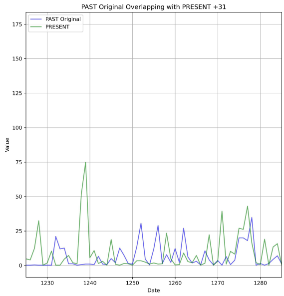
- Overlap PAST-PRESENT
- Overlap PAST REVERSE-PRESENT
- tolerance 10.38 [*mm*]
- shift of the PAST to the right +31 days

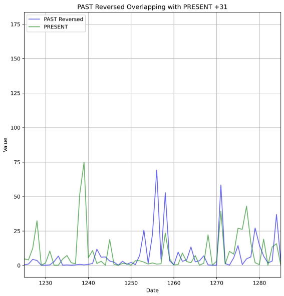


5. Zooming shifted data







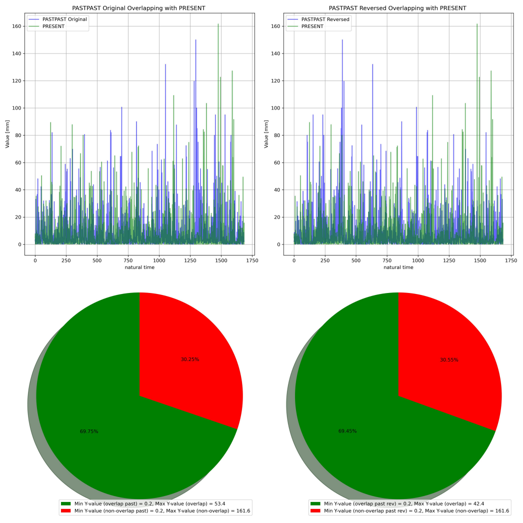


5. Remarks about tolerance and shifting

- The standard deviation of the volumes antisymmetric sums [*mm*] represents the tolerance within the y coordinate (i.e precipitatiom volumes)
- The standard deviation of the intertimes antisymmetric sums [*days*] represents the tolerance within the x coordinate (i.e datapoints-days shifting)

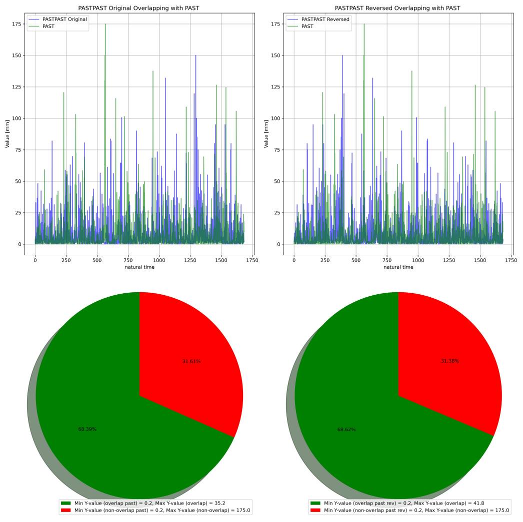
5. PASTPAST → PRESENT prediction

- Overlap PASTPAST-PRESENT
- Overlap PASTPAST REVERSE-PRESENT
- tolerance 10.38 $[mm]$
- Mean common datapoints 1141 over 1686



5. PASTPAST → PAST prediction

- Overlap PASTPAST-PAST
- Overlap PASTPAST REVERSE-PAST
- tolerance 10.38 $[mm]$
- Mean common datapoints 1123 over 1686



Comments

- REV timeseries have a slightly higher overlapping percentage (more common datapoints)
- OR timeseries overlap for higher values of [mm] (common datapoints with higher [mm])
- The overlap between PRESENT and PAST ($\simeq$ 70%) is better than PRESENT-PASTPAST ($\simeq$ 69%) and PAST-PASTPAST ($\simeq$ 68%)
- A positive shift increases the overlapping percentage

Coming soon

- Reverting PRESENT time series and overlapping with the first semester of 2023 (which we will refer to as CURRENT)