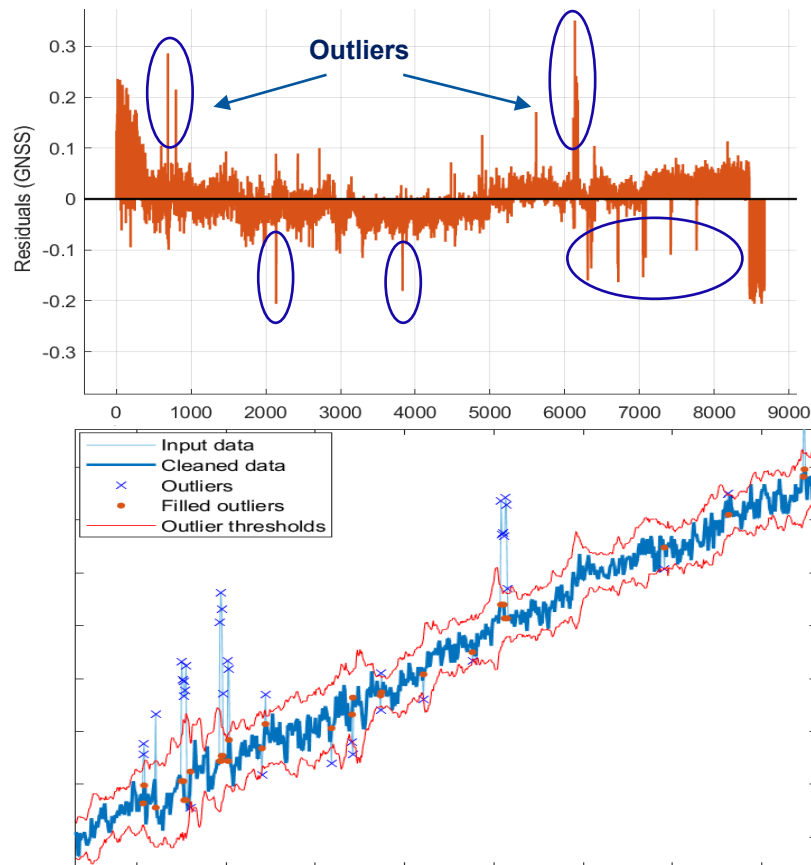


“Selection of an optimal algorithm for outlier detection in GNSS time series”

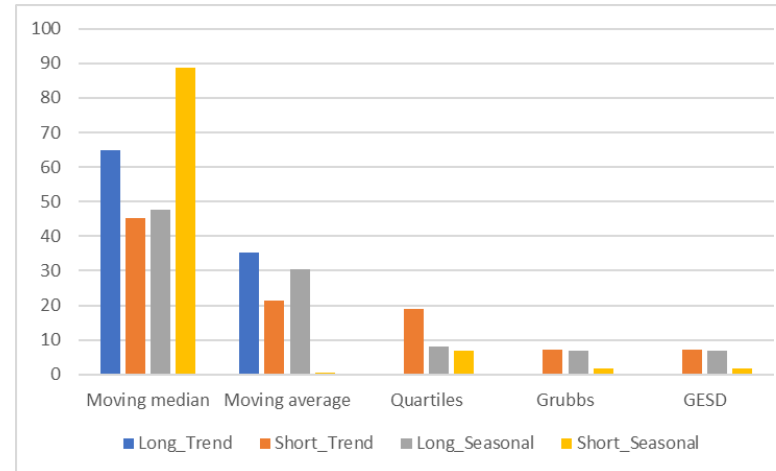
Nhung Le Thi, Benjamin Männel, Mihaela Jarema, Gopi Krishna Seemala, Kosuke Heki, Harald Schuh



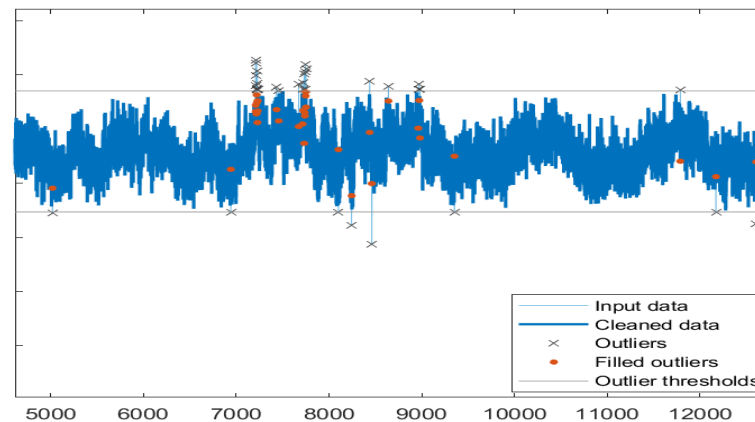
Filter outliers using **the moving median** for 26-year GNSS time series of ~ 9000 observations (Probability 95%, the sliding window at 30)

➡ **Aim:** Investigate five outlier filtering algorithms (Moving average, Moving median, Quartiles, Grubbs, and GESD) to select **optimal algorithms** to apply for GNSS time series data.

➡ **Study design:** Use two patterns (**trend and seasonal**) of **long and short term** time series for comparative investigations.

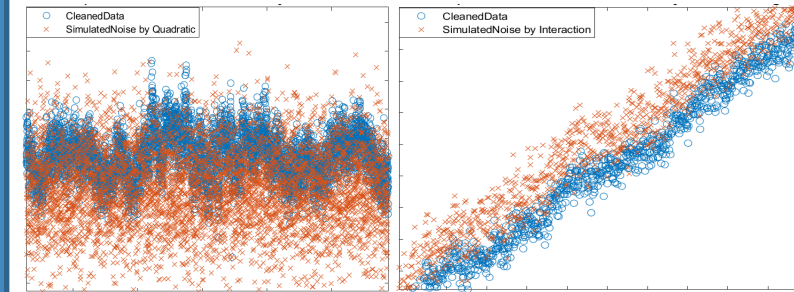


Outlier detection percentage of filtering algorithms for trend and seasonal patterns of long and short term time series

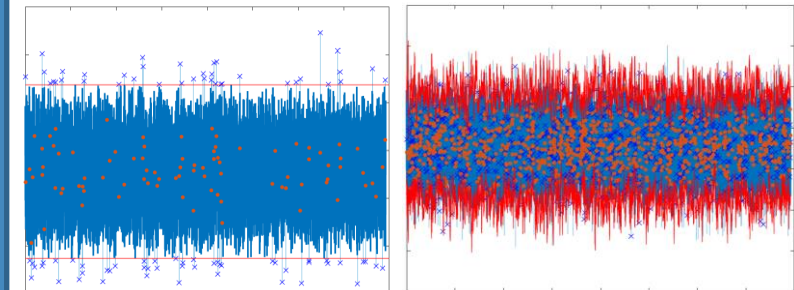


Filter outliers using **the quartiles** for 10-day TEC time series of ~ 16000 observations (Probability 95%)

➡ **Validate:** Simulate **Gaussian white noise** based on three algorithms (interaction, quadratic, linear) to create a new time series (the same normal distribution with original time series) for **testing the sensitivity** of filtering algorithms.



Simulate noise for seasonal and trend time series



Filter noise by the quartile and moving median algorithms

- ➡ **Sum up :**
- **The moving median algorithm** is a robust statistic and **the most sensitivity** for outlier filtering on both time series patterns.
 - **The quartile algorithm** is suited for seasonal pattern and **effective in deformation analysis**.



SELECTION OF AN OPTIMAL ALGORITHM FOR OUTLIER DETECTION IN GNSS TIME SERIES

Nhung Le Thi^{1,5}, Benjamin Männel¹, Mihaela Jarema², Gopi Krishna Seemala³, Kosuke Heki⁴, Harald Schuh¹

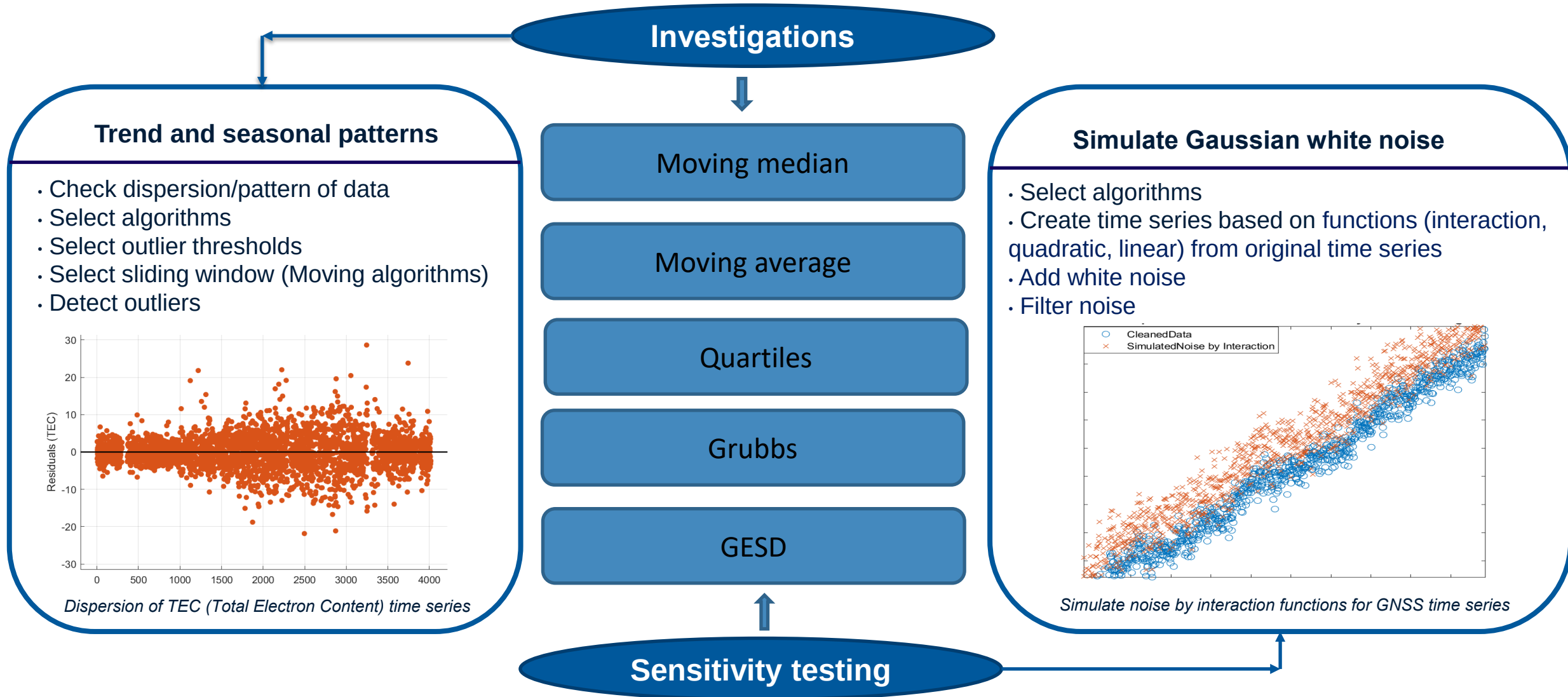
¹GFZ German Research Centre for Geosciences, Germany (nhung@gfz-potsdam.de)

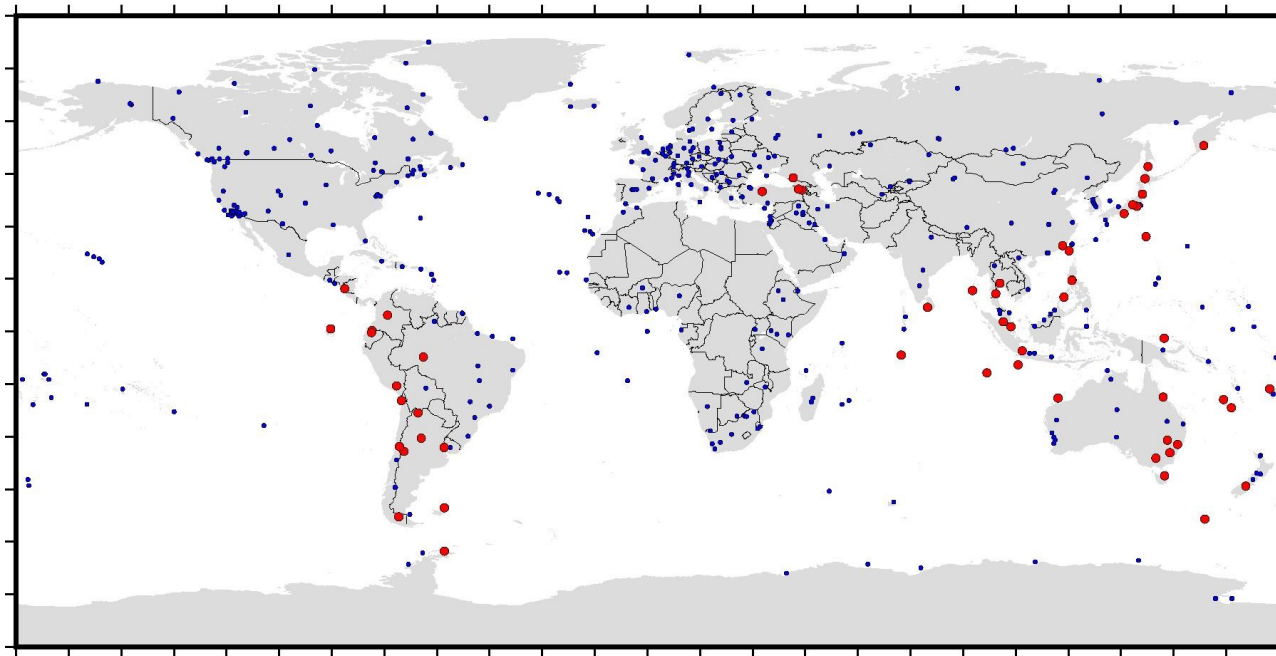
²MathWorks, Germany (mjarema@mathworks.com)

³Indian Institute of Geomagnetism, India (gopi.seemala@gmail.com)

⁴Hokkaido University, Japan (heki@sci.hokudai.ac.jp)

⁵Hanoi University of Natural Resources and Environment, Vietnam

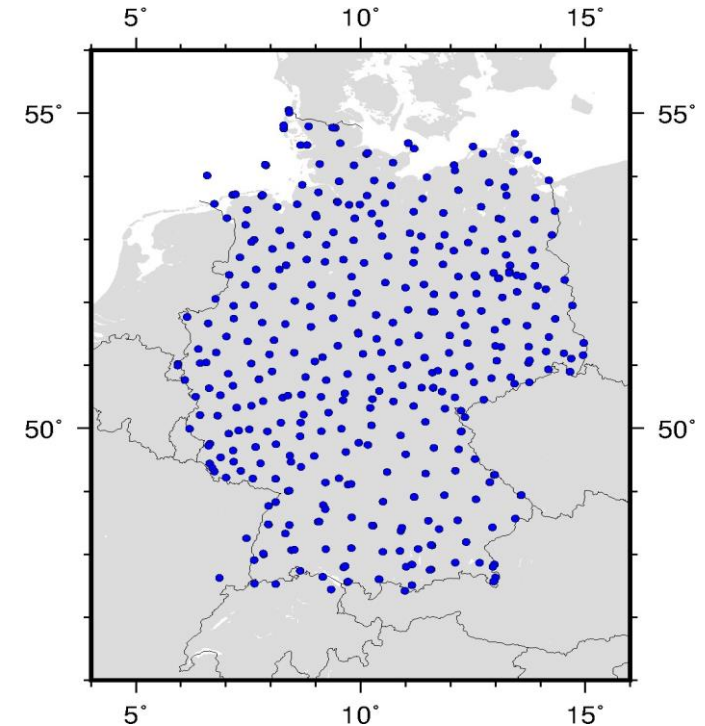




IGS stations

Note: Red plots present stations selected to compute TEC time series

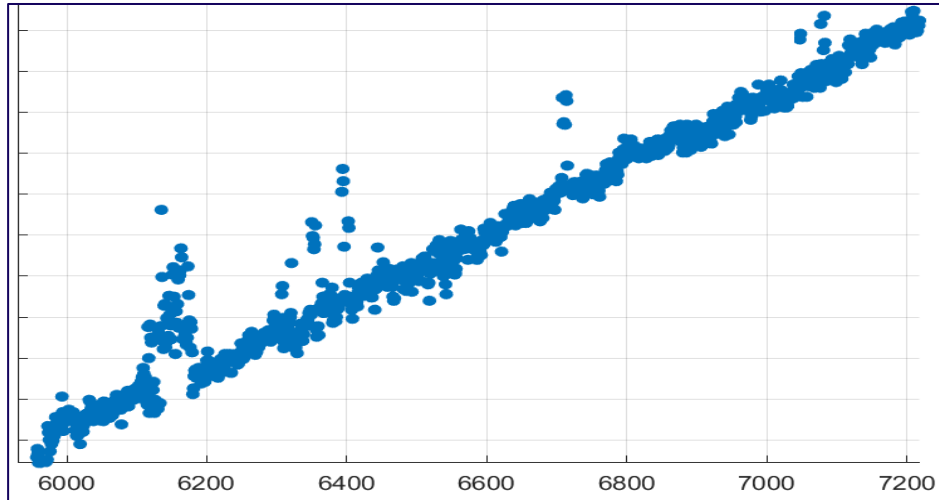
Cross-sectional investigations are performed on data observed from IGS stations (International GNSS Service) in the Ring of Fire. GPS-TEC software (Gopi K.Seemala) is used to get TEC time series.



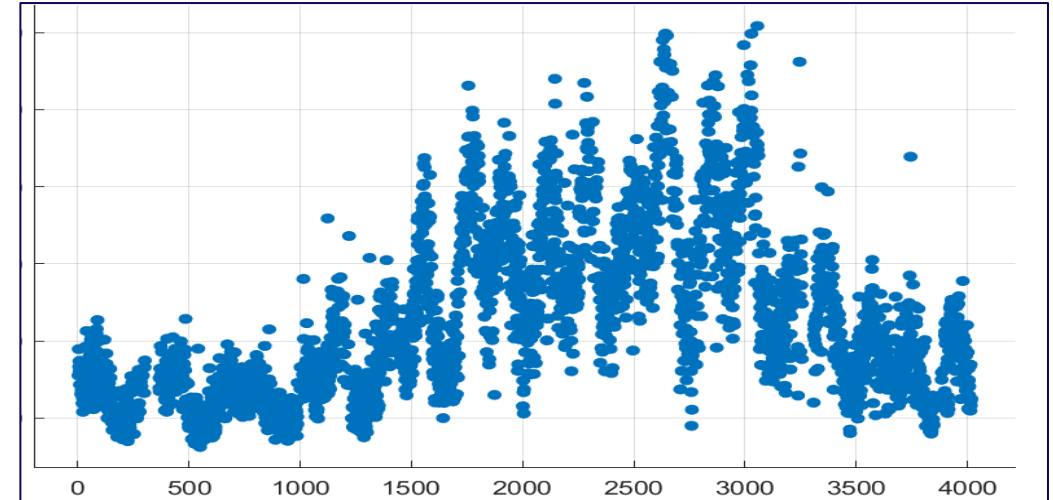
Permanent GNSS stations in Germany

EPOS.P8 software is used to get GNSS coordinate time series (GFZ repro3 network solution, and dedicated Precise Point Positioning) for permanent stations in Germany.

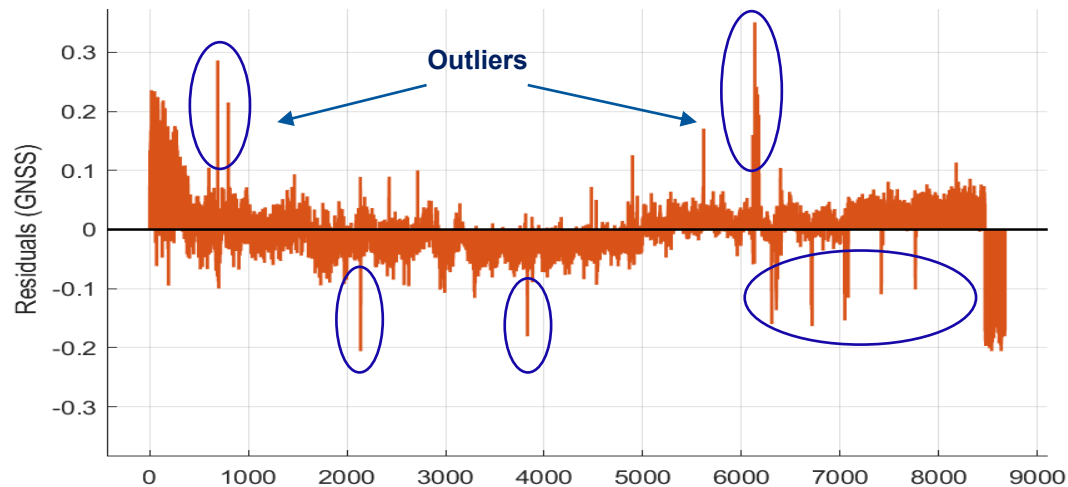
<https://meetingorganizer.copernicus.org/EGU21/EGU21-4148.html>



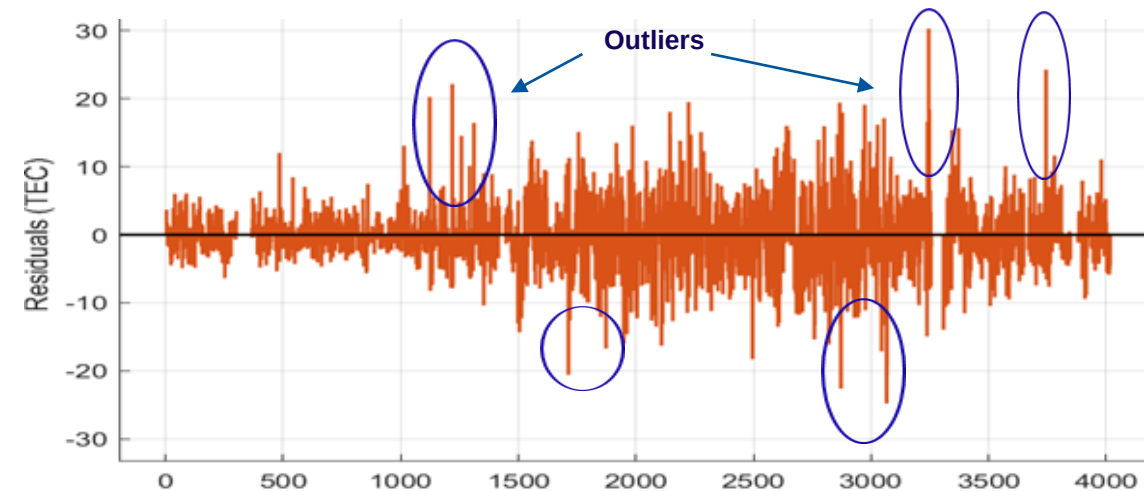
26-year GNSS time series (based on repro3 network solution) processed by EPOS software (GFZ center) at the IGS station POTS



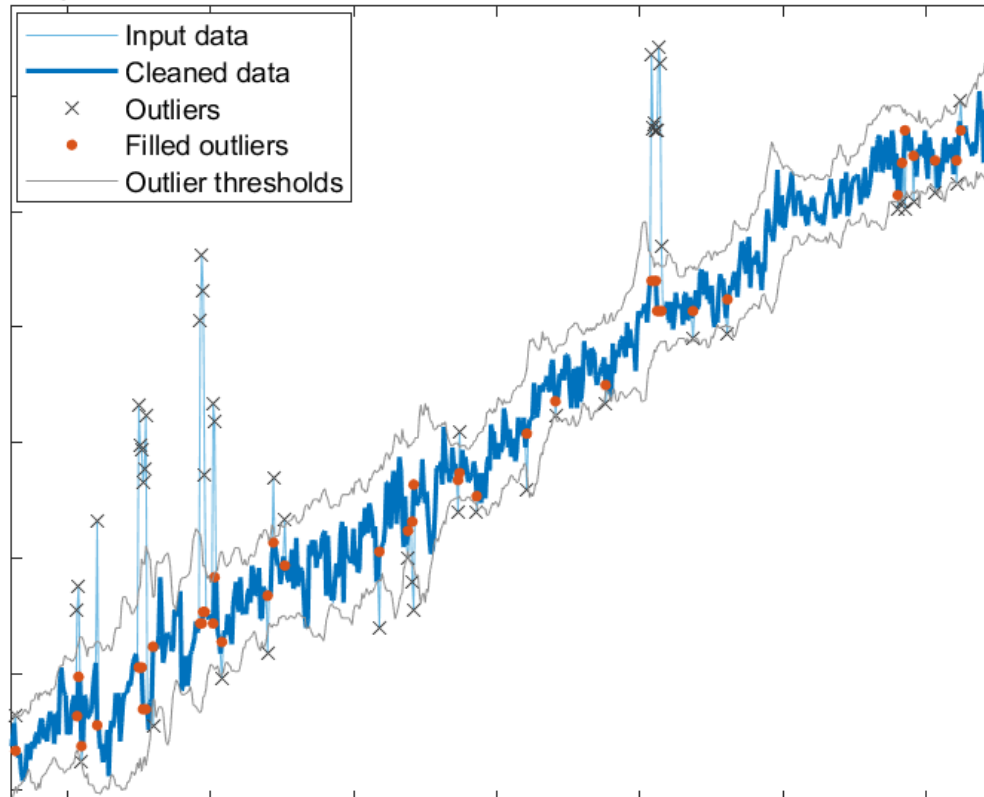
11-year TEC time series processed by GPS-TEC software (Gopi K.Seemala) at the IGS station BAKO



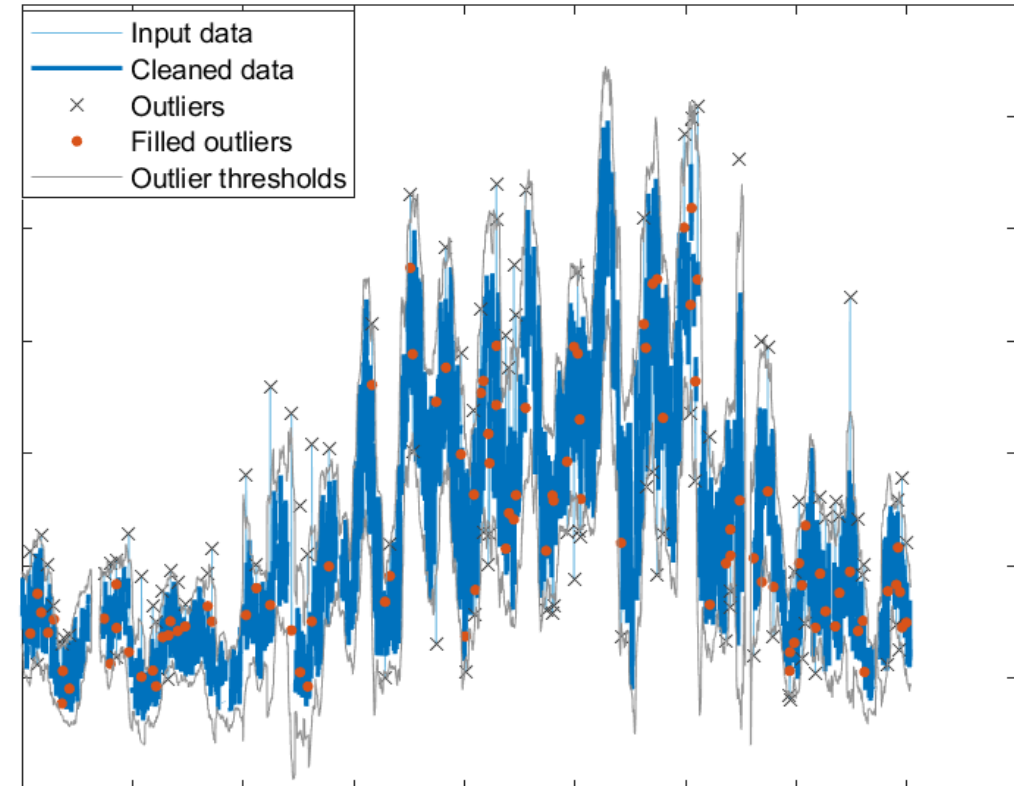
Dispersion of PPP time series at the IGS station POTS



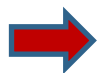
Dispersion of TEC time series at the IGS station BAKO



Filter outliers by the **moving median** algorithm for trend pattern of 26-year PPP (Precise Point Positioning) time series with sliding window 30, threshold 2.5 at the IGS station POTS



Filter outliers by the **moving average** algorithm for seasonal pattern of 11-year TEC time series with sliding window 30, threshold 2.5 at the IGS station BAKO



The moving median and average is effective on both patterns of time series.

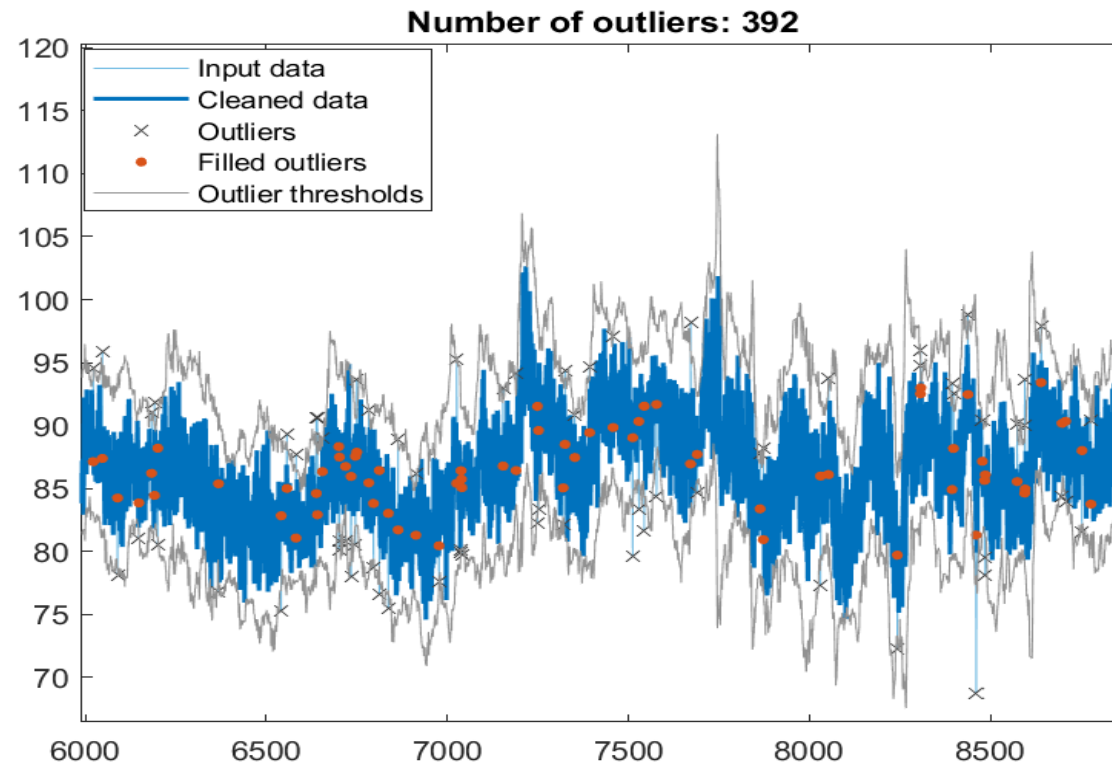
$$\bar{a}_{SM} = \frac{x_n + x_{n-1} + \dots + x_{M-(n-1)}}{M}$$

$$\bar{a}_{SM} = \frac{1}{M} \sum_{i=0}^{n-1} x_{M-i}$$

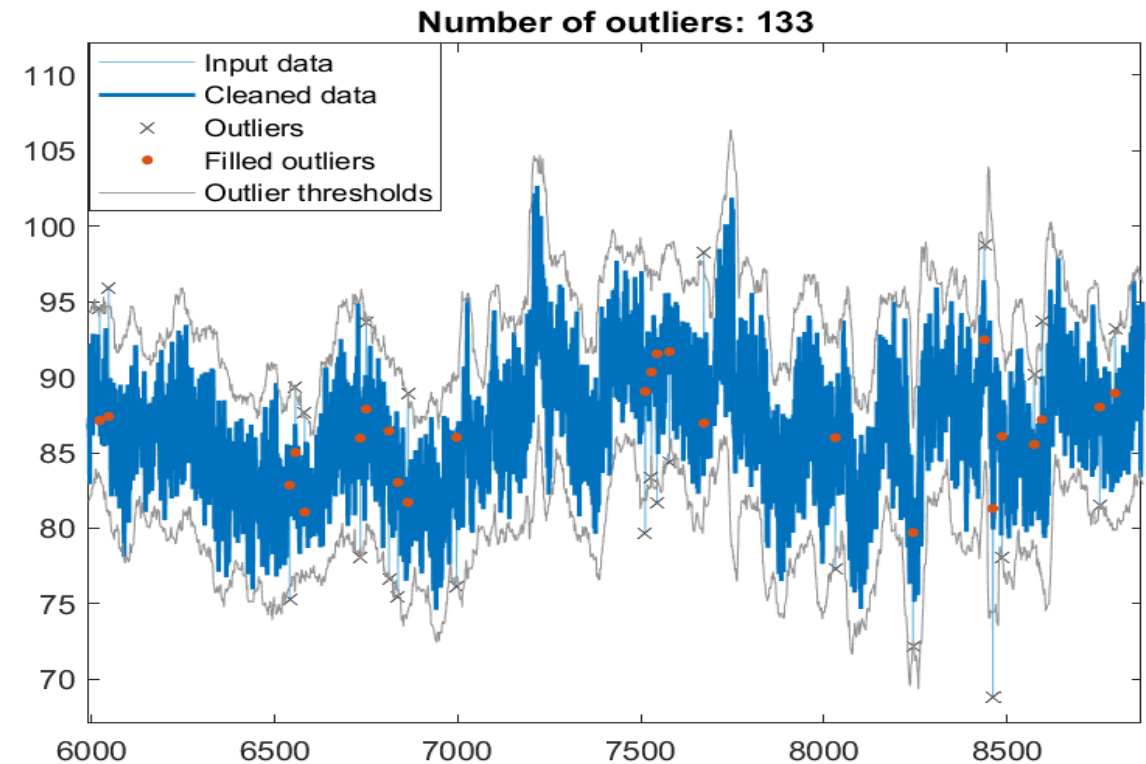
Where:

n – Number of data points

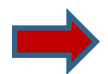
M – Size of sliding window



Filter outliers by the **moving median** algorithm on seasonal 10-day dense TEC time series (one-minute intervals) with sliding window 30, threshold 2.5 at the IGS station PARK



Filter outliers by the **moving average** algorithm on seasonal 10-day dense TEC time series (one-minute intervals) with sliding window 30, threshold 2.5 at the IGS station PARK



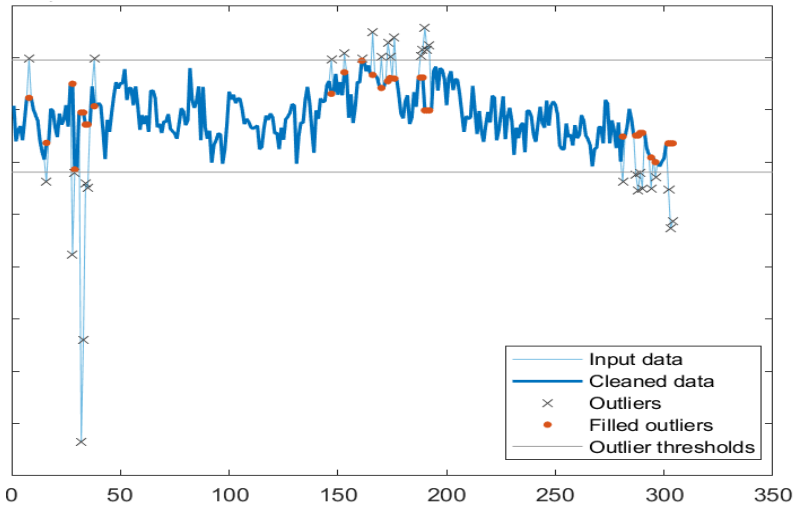
The moving median and average are sensitive for outlier detection and best suited to smoothen time series.

$$\bar{a}_{SM} = \frac{x_n + x_{n-1} + \dots + x_{M-(n-1)}}{M}$$

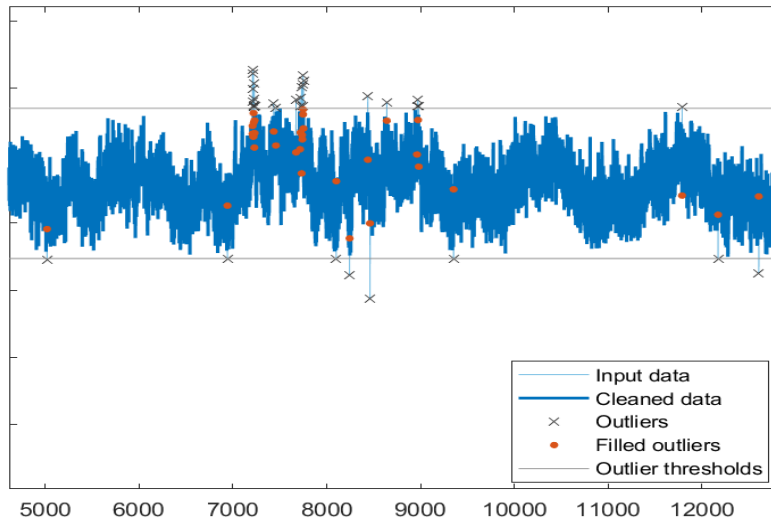
$$\bar{a}_{SM} = \frac{1}{M} \sum_{i=0}^{n-1} x_{M-i}$$

Where:

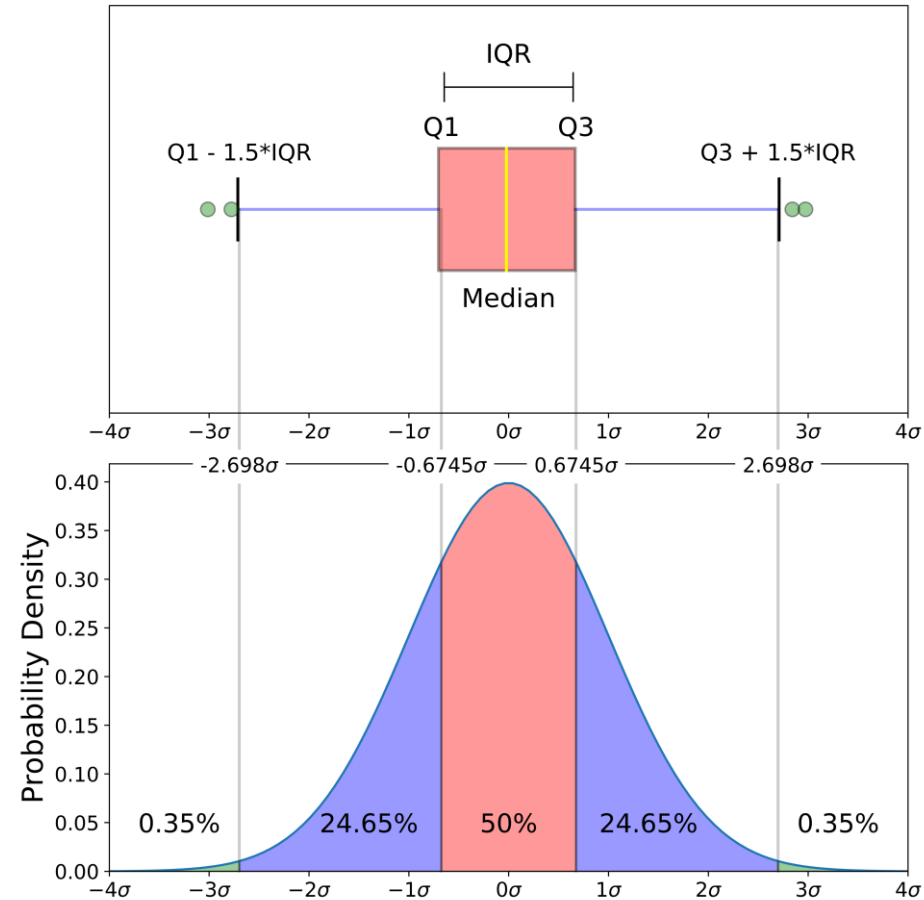
n – Number of data points
M – Size of sliding window



Filter outliers by the **quartile** algorithm for the seasonal pattern of PPP time series (daily intervals) at the SAPO station HOHE in 2019 (with threshold 2)



Filter outliers by the **quartile** algorithm for the seasonal pattern of TEC time series (one-minute intervals) at the IGS station PARK (with threshold 2)

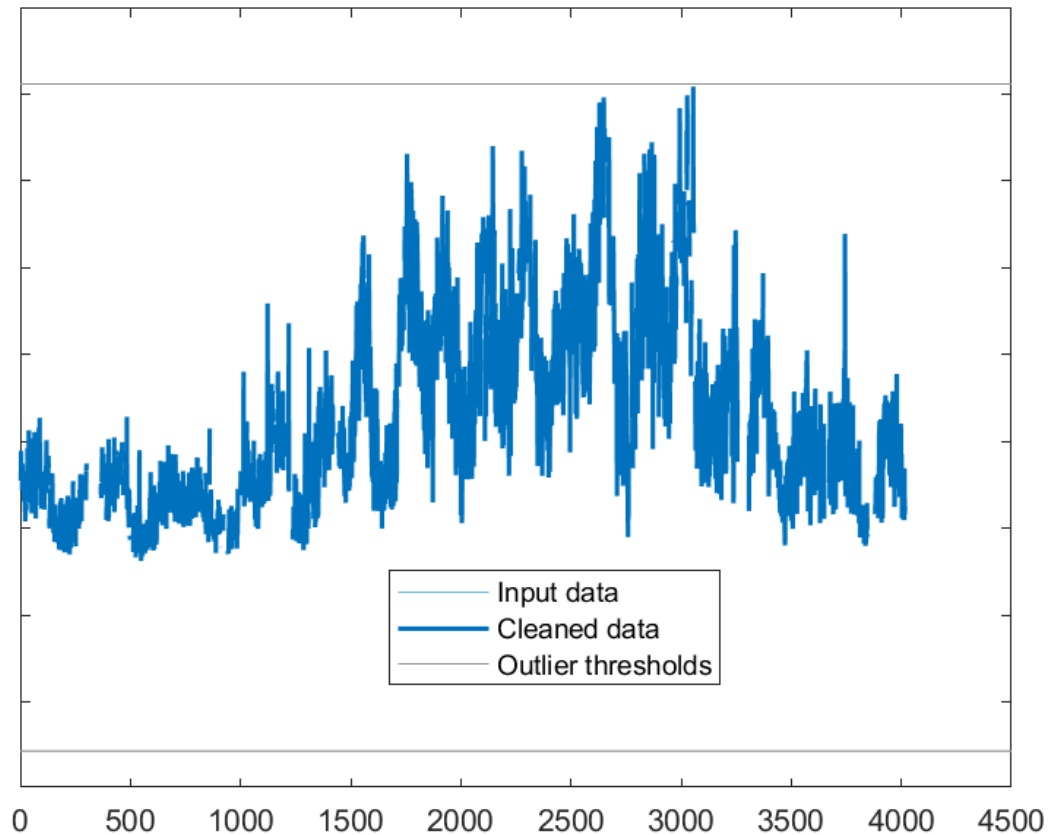


<https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>

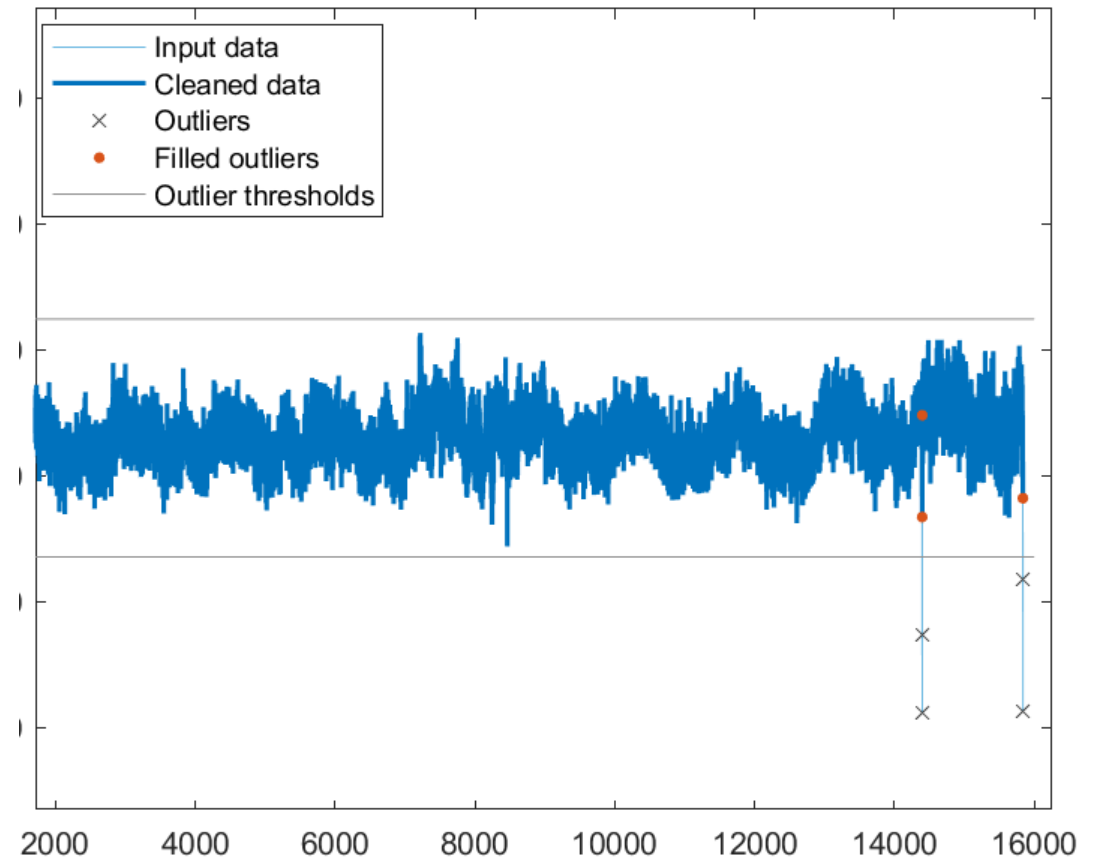
The quartile algorithm and the distribution of data



The quartiles is suited for seasonal pattern of sparse and dense time series.



Filter outliers by the **Grubbs** algorithm for the seasonal pattern of 11-year TEC time series (daily intervals) at the IGS station BAKO with threshold 0.5

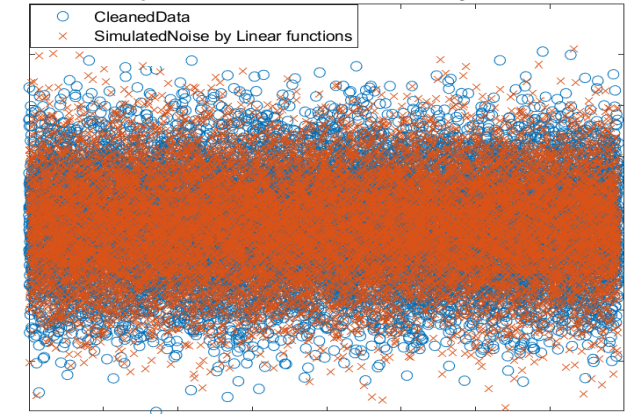
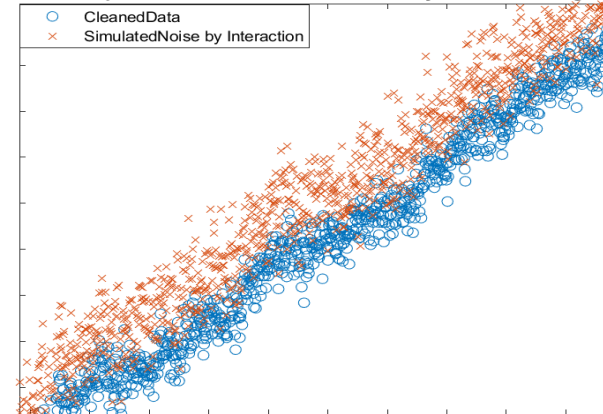
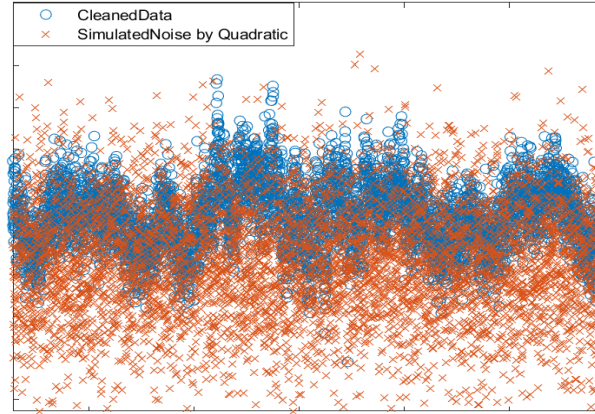


Filter outliers by the **GEDS** algorithm for the seasonal pattern of 10-day TEC time series (one-minute intervals) at the IGS station BAKO threshold 0.5

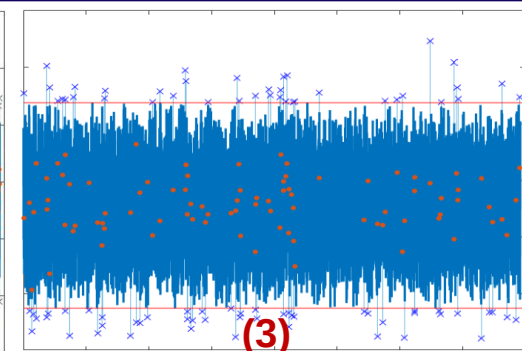
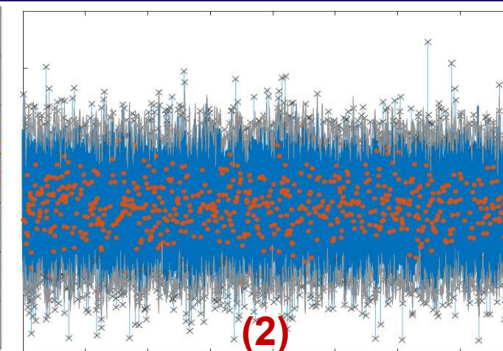
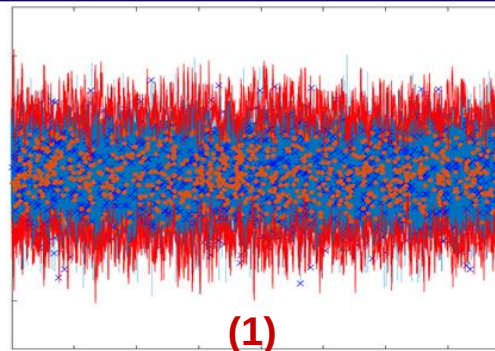


Grubbs and GESD are least sensitive for outlier detection on patterns of time series.

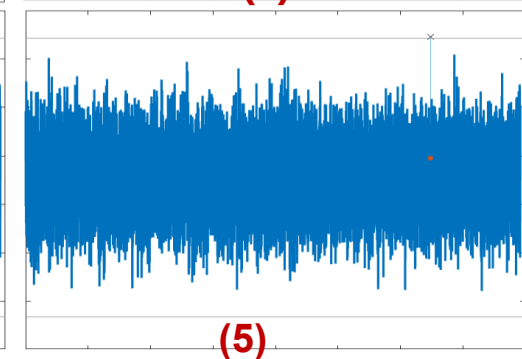
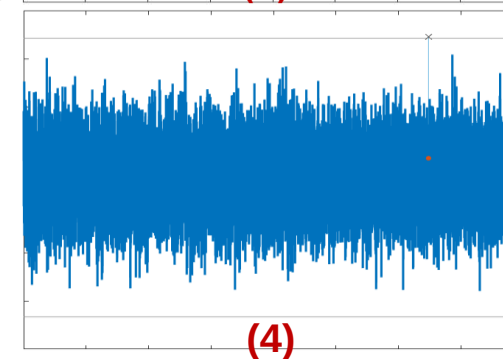
Simulate Gaussian white noise for seasonal and trend time series by three functions (quadratic, interaction, linear) to create a time series (the same normal distribution with the original time series)

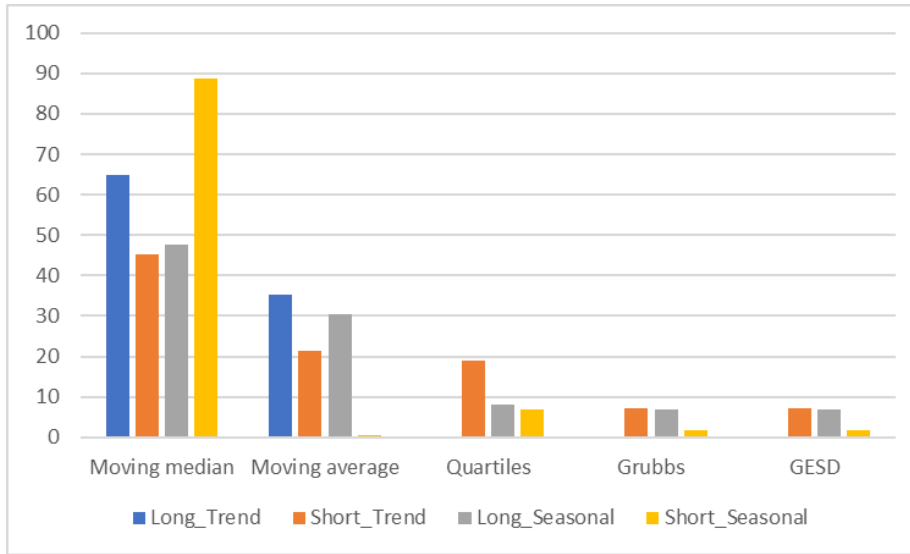


- The moving median is a robust statistic and most sensitive to outliers.
- Grubbs and GESD are least sensitive for outlier detection.



Filter noise by the moving median, moving average, quartiles, Grubbs and GESD, respectively





Outlier detection percentage of filtering algorithms

Outlier detection percentage of filtering algorithms for trend and seasonal patterns of long - short term time series

Algorithms	Trend pattern		Seasonal pattern	
	Long	Short	Long term	Short term
Moving median	64.80	45.24	47.58	88.82
Moving average	35.20	21.43	30.36	0.62
Quartiles	non_applied	19.05	8.09	6.83
Grubbs	non_applied	7.14	6.99	1.86
GESD	non_applied	7.14	6.99	1.86

Note: Under one-year data is considered as short term time series for investigations. The samples are selected by the cross-sectional investigation method from 19 TEC time series and 23 PPP time series.

- **The moving median is a robust statistic and most effective.**
- **The moving average is affected by noise in time series, that shows in investigations on short term time series.**
- **It is necessary to:**
 - + Have further investigations to the moving average on short term time series.
 - + Increase the number of investigation samples to Grubbs and GESD to search for more difference in outlier detection between two algorithms.



- The moving median and average are sensitive for outlier detection on both patterns of long-short term time series and best suited to smoothen time series applied in forecast, **in which the moving median shows a robust statistic and most effective.**
- The quartiles is suited for seasonal pattern of both long and short term time series.
- Grubbs and GESD are least sensitive for outlier detection.

The authors gratefully acknowledge:

- The International GNSS Service (IGS)
- The National Aeronautics and Space Administration (CDDIS)
- Satellitenpositionierungsdienst der Deutschen Landesvermessung (SAPOS®)
- MathWorks
for providing data, software, open-code sources.

The authors would like to thank colleagues at GFZ German Research Centre for Geosciences; Hanoi University of Natural Resources and Environment; and Institute of Geological Sciences, Vietnam Academy of Sciences and Technology for giving valuable suggestions.

The authors: Nhung Le Thi^{1,5}, Benjamin Männel¹, Mihaela Jarema², Gopi Krishna Seemala³, Kosuke Heki⁴, Harald Schuh¹

¹GFZ German Research Centre for Geosciences, Germany (nhung@gfz-potsdam.de)

²MathWorks, Germany (mjarema@mathworks.com)

³Indian Institute of Geomagnetism, India (gopi.seemala@gmail.com)

⁴Hokkaido University, Japan (heki@sci.hokudai.ac.jp)

⁵Hanoi University of Natural Resources and Environment, Vietnam

