Homogenization of GNSS tropospheric products using subspace-based techniques

Mostafa Hoseini, Fadwa Alshawaf, Galina Dick and Jens Wickert
GFZ German Research Centre for Geosciences, Potsdam
Contact: hoseini@gfz-potsdam.de

Abstract

Heterogeneous tropospheric time series might lead to erroneous interpretation in climatic or meteorological applications. Inhomogeneities are present in GNSS-derived time series can be caused by various reasons. So, the homogenization procedure for such a data set is considerably more complicated than other meteorological data. An adaptive decomposition process has been used with the aim of developing a robust homogenization method based on subspace-based techniques. The method exploits data-driven basis functions to extract fundamental components of the time series and applies a classification method to explore the relation between the extracted components and the origins of heterogeneities.

Method

All of time series related to GNSS atmospheric products (e.g. Integrated Water Vapor (IWV), Zenith Total Delay(ZTD), Slant Total Delay(STD), Pressure and Temperature) can be generally considered as a combination of some additive components.

$$\mathrm{F} = (\mathrm{f}_1, \mathrm{f}_2, ..., f_N), f_i \in R,\ i = 1, 2, ..., N$$

$$F = F_{trend} + F_{periodicty} + F_{anomaly} + F_{heterogeneity} + F_{noise}$$

Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis has been used here as a general time series analysis tool for trend and constitutive components extraction, denoising, gap filling and change-point detection. The main steps of basic SSA are:

1- Embedding Original Time Series in L-dimensional Vector Space:

$$\overbrace{f_1, f_2, ..., f_L}^{window\ \rightarrow}, f_{L+1}, f_{L+2}, ..., f_N \Rightarrow X_1 = (f_1, f_2, ..., f_L)$$

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ f_3 & f_4 & f_5 & \cdots & f_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_L & f_{L+1} & f_{L+2} & \cdots & f_N \end{bmatrix}, \begin{cases} 1 < L < K \\ K = N - L + 1 \end{cases}$$

$$= \begin{bmatrix} X_1 & X_2 & X_3 & \ldots & X_K \end{bmatrix}$$

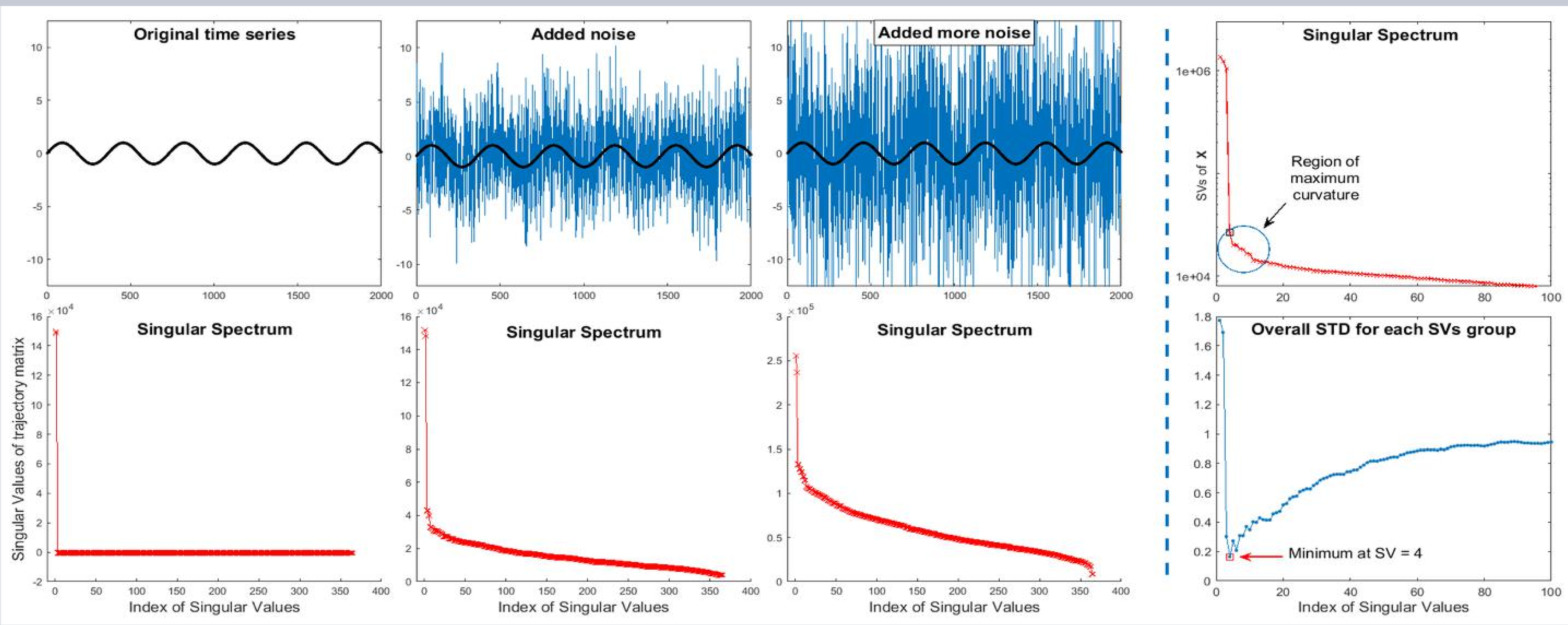
2- Decomposition:

Singular Value Decomposition (SVD) of the Trajectory Matrix (X):

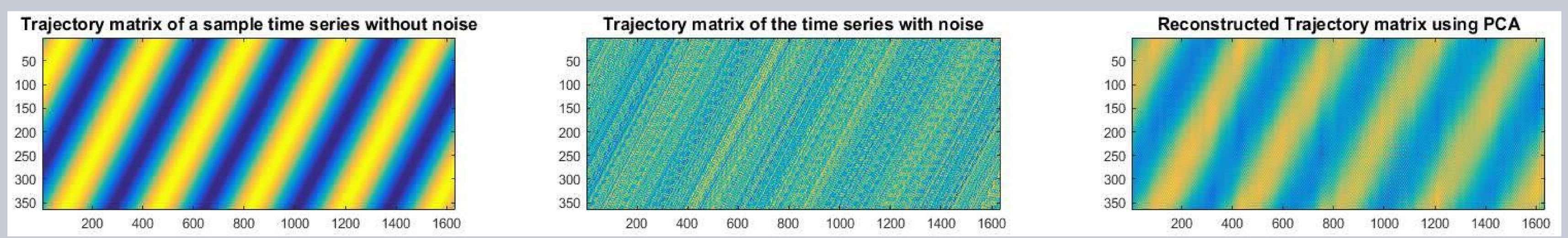
$$\mathrm{X} = \mathrm{U\Sigma V^T} \quad \mathbf{X} = \mathbf{X_1} + \mathbf{X_2} + ... + \mathbf{X_d}\ ,\ \mathbf{X_i} = \sqrt{\lambda_i} U_i V_i^T, d = max\{i \mid \lambda_i > 0\}$$

3- Grouping Singular Values and Corresponding Singular Vectors:

selecting proper number of singular vectors based on behavior of singular values:

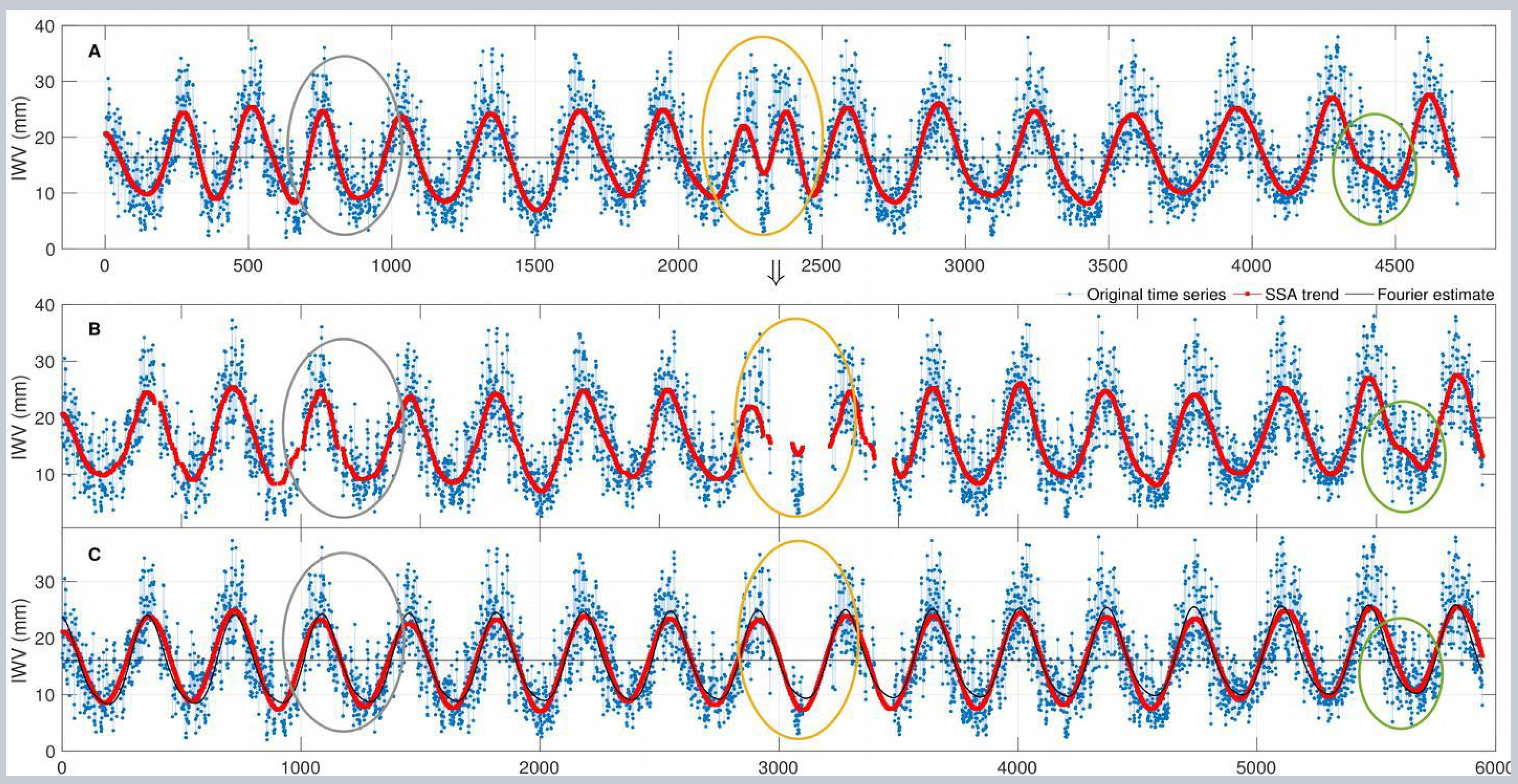


4- Reconstruction and cross-diagonal averaging:



Filling data gaps

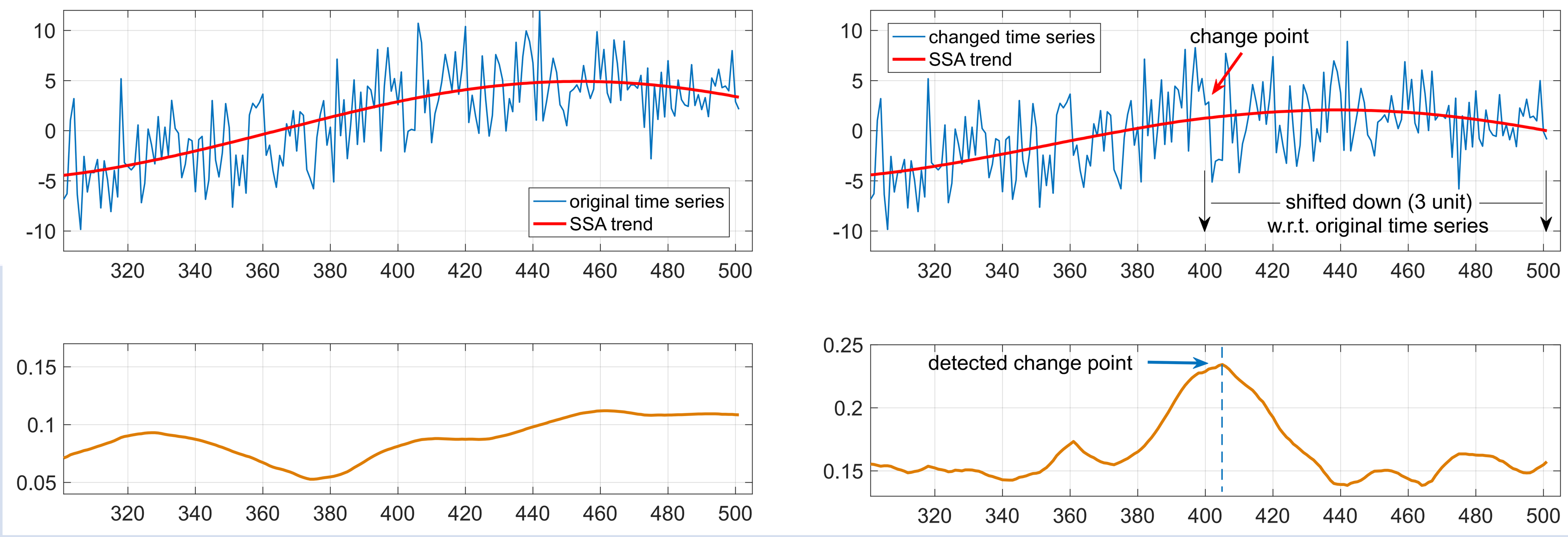
Removing effect of data gaps on SSA trend extraction:



www.gfz-potsdam.de

Homogeneity Check

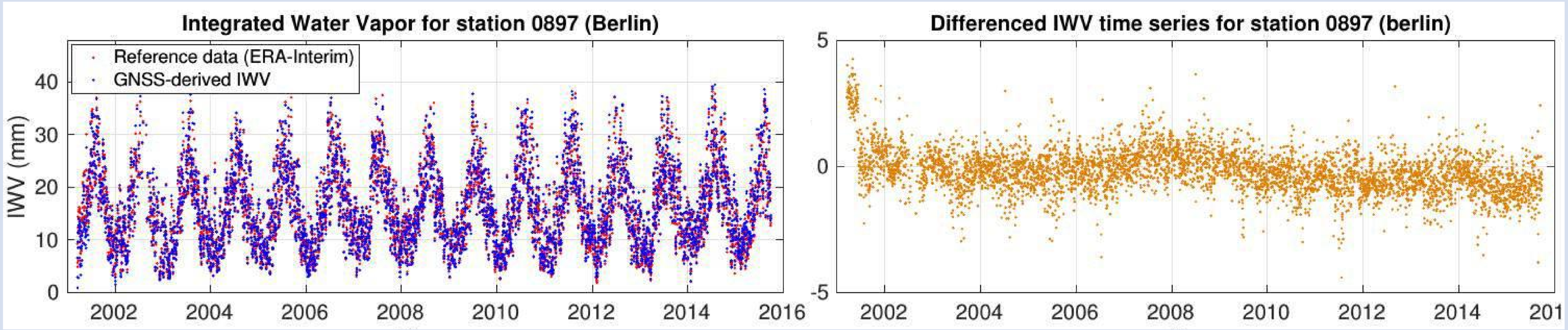
Change point detection by measuring variations with respect to SSA trend:



Data Sets

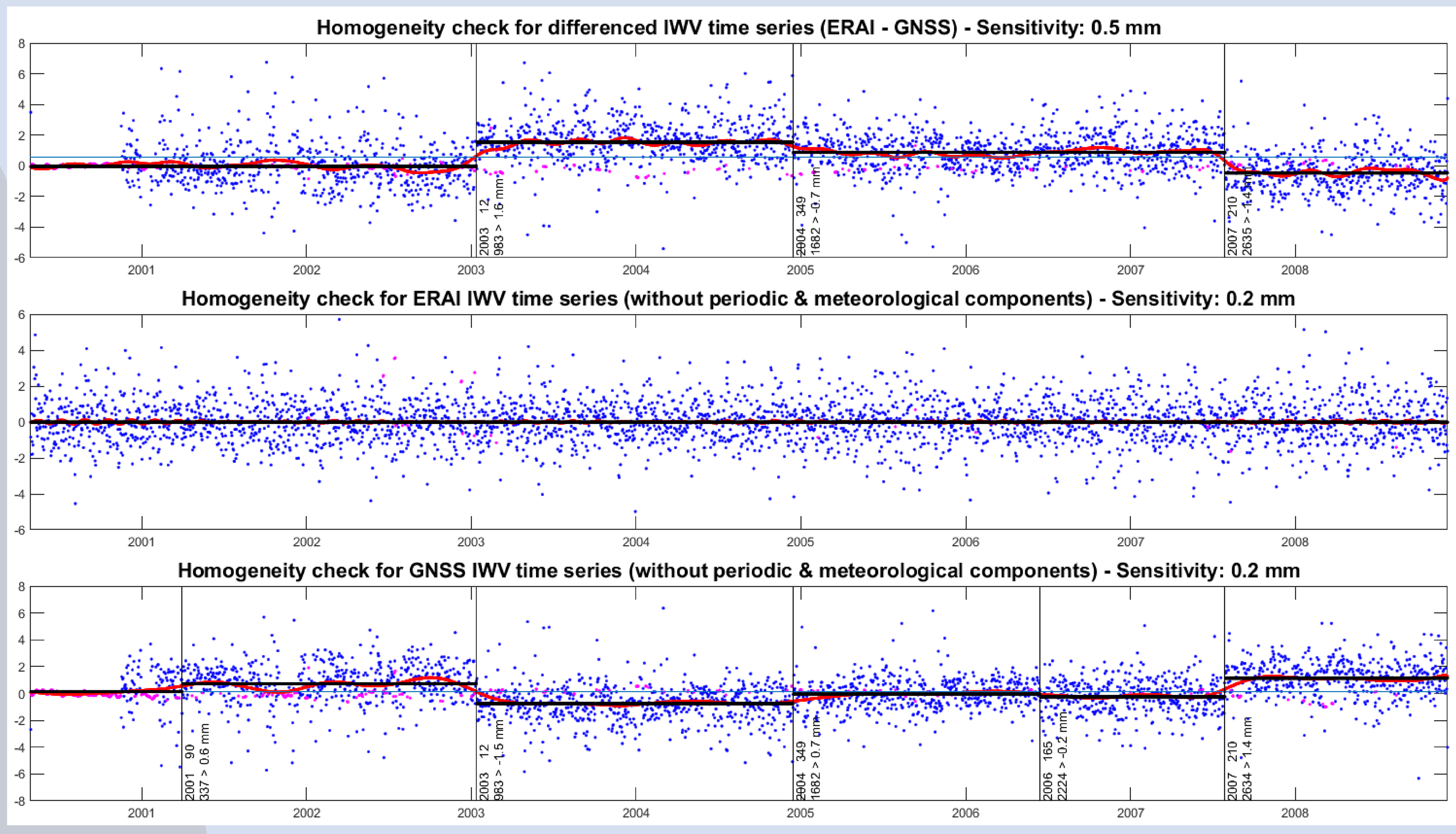
1- GFZ near real time GNSS-derived atmospheric products based on precise point positioning (PPP) technique (more than 300 stations in Germany)

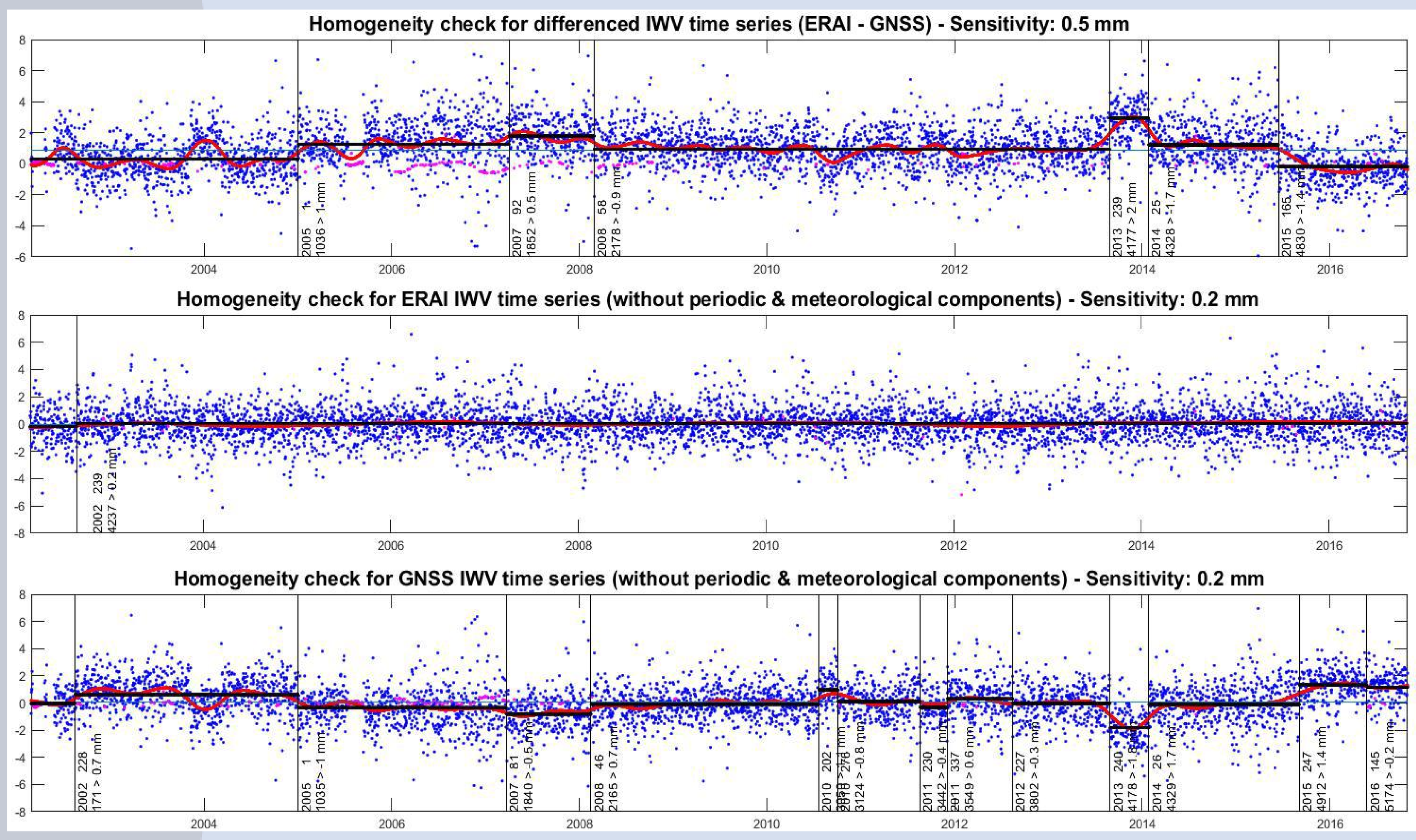
2- ERA-Interim, a reanalysis product by ECMWF (simulated at each station as the reference data set)



Results

Results of homogeneity check for two sample stations in Germany





Conclusion

◆ Subspace-based methods such as Principal Component Analysis and SSA which are used here, indicated powerful features in different stages of time series analysis including denoising, anomaly assessment, gap filling, periodicity reduction and change point detection.

◆ Although a rather homogeneous reference time series is usually used in data homogenization, existence of inhomogeneities in the reference time series has been given a serious consideration in this work. For this reason, a zero-difference homogenization process is being developed which its first results are shown in the result section of this poster.

◆ The proposed method can detect all of the possible change points simultaneously, so it can significantly enhance the process of homogenization and simplifies the automation procedure. The method generally is applicable on other time series.