INVESTIGATION OF TIME-CHANGEABLE SEASONAL COMPONENTS IN GPS HEIGHT TIME SERIES: A CASE STUDY FOR CENTRAL EUROPE

Marta GRUSZCZYNSKA *, Anna KLOS, Maciej GRUSZCZYNSKI and Janusz BOGUSZ

Military University of Technology, Faculty of Civil Engineering and Geodesy, Kaliskiego 2, 00-908 Warsaw, Poland

*Corresponding author’s e-mail: marta.gruszczynska@wat.edu.pl

ABSTRACT

Nowadays, beyond the dispute we should take into account the time-varying parameters of seasonal signals in the GPS-derived position time series. Either real geophysical effects or system-specified artefacts can introduce non-sinusoidal changes. For this study, we used 18 daily position time series from Central European stations provided by the Jet Propulsion Laboratory (JPL) processed in the GIPSY-OASIS software in a Precise Point Positioning (PPP) mode. We tested two different approaches to subtract the seasonal signals: Least-Squares Estimation (LSE) and Singular Spectrum Analysis (SSA). The SSA approach is suitable for all (stationary and non-stationary) time series, without prior knowledge about the data characteristics, which is an undisputable advantage. We extracted periodicities from GPS position time series, and demonstrated the usefulness of the SSA approach on the example of the vertical component. We showed, that the reassembled signal, containing only the first four Reconstructed Components, has a larger correlation coefficient than LSE-extracted signals, with respect to the original time series. Moreover, using the Akaike Information Criterion and Fisher-Snedecor test we tested optimum length of the sliding window and significance of the obtained RCs, respectively.

INTRODUCTION

The seasonal signals in geodetic time series are well described in scientific literature (e.g. van Dam et al., 2007; Ray et al., 2008; Davis et al., 2012 or Bogusz and Klos, 2015). The origin of signals, should be sought both in the real geophysical effects, and within the artefacts of the navigation system (Dong et al., 2002). Environmental loadings can cause substantial variation in the position, predominately in the annual period (Jiang et al., 2013). Moreover, aliasing in short periods may cause artificial annual and semi-annual oscillations, which can be seen in the coordinates (Penna and Stewart, 2003), while repeating time of the GPS satellite constellation with respect to the Sun (Agnew and Larsson, 2007) appears as draconitic period of about 351 days (Amiri-Simkooei, 2013). These variations, if being present in the time series, can also bias the reference frame realization, when seasonal signals are incorrectly defined (Freymueller, 2009; Collilieux et al., 2012). Seasonal variations in GPS time series have a significant impact on determining a station’s velocity in a reliable manner. Blewitt and Lavallée (2002) defined the Dilution of Precision of GPS-derived velocity, when not considering seasonals, but their research was based on the white noise assumption, which is not so frequent in the time series. Wherefore, Bos et al. (2010) extended their consideration for power-law dependencies, and concluded that under equal noise conditions, the seasonal signals have to be included in the estimation of the linear trend, due to significant differences in trend errors when seasonals are not considered.

Initially, the seasonal signals were typically presented as the sum of sine waves (e.g. Amiri-Simkooei et al., 2007; Kenyeres and Bruyninx, 2009 or Bogusz and Figurski, 2014). However, there is no reason to expect that seasonal signals in the GPS-derived time series are time-invariant. Following the growing demands for precise geodynamic analysis it needs to be taken into consideration in the coordinate time series analysis. Freymueller (2009), Tesmer et al. (2009) and Bogusz et al. (2015) used non-parametric methods, to estimate the annual signal from GPS time series, for which this term is not a sine function. Freymueller (2009) and Tesmer et al. (2009) used a window weighted mean, while Bogusz et al. (2015) used the wavelet decomposition, to estimate annual variations. The above described methods, demonstrated an opportunity to obtain a better approximation of time-changeable annual oscillations. Following that, the aim of this paper, is to use the non-parametric Singular Spectrum Analysis (SSA) method, when extracting seasonal signals with time-variable amplitudes and phases.

DATA

For this study, we used 18 daily position time series of Central European stations with the minimum data length of 12 years. The longest time series have the time span of 22 years (1992-2014). This time series was provided by the Jet Propulsion Laboratory (JPL), which uses GIPSY-OASIS software in a Precise Point Positioning (PPP) mode. More information regarding the analysis and software used in the collection of GPS data can be found at https://gipsy-oasis.jpl.nasa.gov. Figure 1 shows the layout of the permanent GPS stations included in our research. We chose the longest available time series, and discarded ones with significant gaps. For time series where small gaps (which did not exceed the length of 30 days) were present, we interpolated missing epochs in three steps. Firstly, we estimated the linear trend and seasonal signals with Least-Squares Estimation (LSE) using all effective observations. Secondly, when deterministic part were modelled, we used residuals to make linear interpolation of missing coordinates adding white noise with equal variance to the all effective residual variance. In the last step, we reassembled time series using model determined previously and completed residuals.

Initially, we pre-processed the data in two stages. In the first stage, we used the STARS algorithm (Sequential t-test Analysis of Regime Shifts) (Rodionov, 2004) to detect epochs of offsets. Then, we eliminated outliers with the median absolute deviation criterion (for details see Klos et al., 2015). Before analyzing the time varying terms from the chosen set of 18 stations, we estimated the linear trend with LSE, and then removed it from each time series. This gave more accurate insight and allowed us to better demonstrate annual and semi-annual oscillations’ change over time.

SINGULAR SPECTRUM ANALYSIS

Here, we propose to use Singular Spectrum Analysis (SSA) for daily GPS solutions, to model the non-stationary seasonal changes. This approach is used for the time series analysis and as a tool for signal extraction. Moreover, the SSA is a technique which integrates many fields of study, such as signal processing, geometry, elements of classical time series, probabilistic, and statistical elements. It is a method for investigation of nonlinear dynamics, which is based on spectral eigenvalue decomposition. This approach is suitable to all stationary and non-stationary time series without prior knowledge about the characteristics of the data. The aim of SSA, is to mathematically decompose the original time series into a sum of components, such as a slowly varying trend, oscillatory components, and noise (Elsner and Tsonis, 1996; Golyandina and Zhigljavsky, 2013).

The basic principles of the singular spectrum approach have been already described by Broomhead and King (1986), Vautard et al. (1989) or Ghil and Taricco (1997). This method has become a widely used tool in various fields, such as climatology, oceanology, financial mathematics, and geophysics (e.g. Ghil et al., 2002; Korotchenko et al., 2014; Hassani and Zhigljavsky, 2009; Cholyi, 2014 or Kumar et al., 2015). In geodesy, Rangelova et al. (2010) used the Multi-channel Singular Spectrum Analysis (MSSA) method, to study the water mass changes in GRACE monthly data, and modelled the glacial isostatic adjustment (GIA) signal. Rangelova et al. (2012) analyzed a 6-year, weekly GRACE time series and hydrology models, using MSSA to extract non-periodic mass variations. Khelifa et al. (2012) analyzed a 6-year, weekly GRACE time series and hydrology models, using MSSA to extract non-periodic mass variations. Khelifa et al. (2012) used Singular Spectrum Analysis and the Wavelet Transform, to assess noise levels in DORIS (Doppler Orbitography and Radiopositioning Integrated by Satellite) time series, to better determine the non-linear station motion. Zerbini et al. (2013) applied the Empirical Orthogonal Functions (EOF) technique, as well as Singular Value Decomposition (SVD) to analyze GPS positions of stations located in Europe with time series of atmospheric pressure, terrestrial water storage, and surface mass anomalies provided by the GRACE gravity mission. Chen et al. (2013) applied the SSA method on GPS time series, to model seasonal variation within weekly coordinates. They also applied Least-Squares Estimation, as well as Kalman Filtering, and compared results from both approaches concluding that SSA is a feasible tool for the exploration of modulated oscillations. Xu and Yue (2015) used Monte Carlo SSA (MCSSA), to extract the time-variable seasonal signals from daily GPS time series, and conducted statistical analysis on the colored noise.

Singular Spectrum Analysis consists of two stages, which are: decomposition and reconstruction. The first step in the decomposition process, is the embedding of a time series \( \{ X(t): t = 1, \ldots, N \} \) in a vector space to the dimension \( M \) (Ghil et al., 2002).
In this process, the window of length $M$ is sliding over a time series (here: day-by-day), thus giving the trajectory vector:

$$\tilde{X}(t) = [X(t), X(t+1), ..., X(t+M-1)]'$$

(1)

The vectors $\tilde{X}(t)$ are indexed by $t = 1, ..., N'$, where $N' = N - M + 1$, with $N$ being the length of the original data.

The search for an optimal window size, was the subject of previous research studies (e.g. Vautard et al. 1992; Ghil et al., 2002; Rengelova et al., 2012; Chen et al., 2013). It was demonstrated that the optimum window size has to be greater than the periodicity being extracted, but lower than half of the length of the entire time series. Based on the simulated time series, it was confirmed that for the extraction of annual and semi-annual oscillations, a two or three year-long lagged window is suitable (Chen et al., 2013). In our research, we estimated Akaike Information Criterion (Akaike, 1974) values, to identify the optimal dimension of a window size $M$. Since we deal with the variability of 1 year, a window with an optimal size was selected from a pool of potential candidates, between 1 to 5 years, in length, which thus minimized the value of the $AIC$ index. This can be computed using residual sums of squares, from the regression statistic as follows (Burhman and Anderson, 2002):

$$AIC = N \cdot \ln \left( \hat{\sigma}^2 \right) + 2 \cdot K$$

(2)

where

$$\hat{\sigma}^2 = \frac{\sum_i \hat{\varepsilon}_i^2}{N}$$

with:

$N$ – the number of observations;

$\hat{\varepsilon}$ – the estimated residuals for a particular candidate model;

$K$ – the number of parameters in the model. This parameter is significant if we used the $AIC$ index to compare models with a different amount of parameters. Here, we have the same amount of parameters: $K = 2$. The first parameter is $M$ (window length) and the second parameter is residual variance parameter ($\hat{\sigma}^2$) (Burhman and Anderson, 2002).

The lowest value of the $AIC$ index, was obtained here for a 3-years window, so we adopted this approach to further process GPS time series with SSA. Figure 2 shows the first forty eigenvectors for SVTL (Svetloe, Russian Federation) station that explain the percentage of total variance, respectively for each eigenvector of a particular window. In this paper, the time series from SVTL was used as a leading example. We tested data from all stations in the Central European region, but the results were very similar to SVTL.

In the next step, we computed the lag-covariance matrix $C_y$ with the Toeplitz structure of constant diagonal for the observed time series $X(t)$ (Vautard and Ghil, 1989):

$$C_y = \frac{1}{N^2 - 1} \sum_{i=1}^{N^2} X(t_i)X(t_i + 1)$$

(3)

We have used the algorithm known as VG to calculate the covariance matrix $C_y$ of $X(t)$. Broomhead and King (1986) proposed the covariance matrix $C_y$ to be computed by calculating the product of the trajectory matrix $D$ and that the singular value decomposition be applied to $C_y$, using the following formula:

$$C_y = \frac{1}{N'} D' D$$

(4)

This formula is known as the Broomhead and King (BK) algorithm. We also explored the Vautard and Ghil (VG) approach, which was also studied by a number of authors (e.g. Allen and Smith, 1996; Ghil and Taricco, 1997), who compared both BK and VG approaches. They presented the differences and similarities of both methods on how to compute eigenelements of $C_y$. The VG method is better than BK at reducing noise for short and stationary time series. However, if the original time series is non-stationary, it is better to use SVD than the Toeplitz matrix (Golyandina and Zhigljavsky, 2013).

Once we computed the lag-covariance matrix $C_y$, we then decomposed it into eigenvectors (also called Empirical Orthogonal Functions - EOF) and eigenvalues. The eigenvalues are sorted in a decreasing order. The sum of all eigenvalues (the trace of covariance matrix), returns us to the total variance of the original time series (Kondrashov and Ghil, 2006). Figure 3 presents the first two leading pairs of EOFs, each of which is in quadrature.

The eigenvalues obtained in the previous step were sorted in decreasing order, and their corresponding eigenvectors were used to calculate the $k$-th Principal Component (PC) in the form of:

$$A_k(t) = \sum_{j=1}^{N} X(t + j - i)E_k(j) \text{ for } 0 \leq i \leq N - M$$

(5)

The first Principal Component contains the maximum amount of variance, while each subsequent component contains a declining amount of residual variance. Each of the pairs is characterized by either periodicity or amplitude variability. Figure 4 presents the first four PCs for the SVTL station, with all of them being in quadrature. The first two PCs have the same period of about one year (annual). The third and the fourth PC have a period of about six months (semi-annual), with clearly visible variability over time.
We then produced each Reconstructed Component (RC) for the original time series $X(t)$. The RC can be created, using previously calculated Principal Components and Empirical Orthogonal Functions (Vautard and Ghil, 1989; Vautard et al., 1992):

$$R_k(t) = \begin{cases} \frac{1}{t-j} \sum_{j=1}^{t} A_k(t-j+1) E_k(t) & \text{for } 1 \leq t \leq M-1 \\ \frac{1}{M} \sum_{j=M}^{N} A_k(t-j+1) E_k(t) & \text{for } M \leq t < N' \\ \frac{1}{N-t+1} \sum_{j=N-S+M}^{N} A_k(t-j+1) E_k(t) & \text{for } N' \leq t \leq N \end{cases} \quad (7)$$

There are two main advantages of SSA: there is no loss of signal in the reconstruction process and it lacks of edge effect. If we sum all of the individual RCs, we will receive the original time series (Ghil et al., 2002). Nevertheless, some kind of filtration effect can be obtained, when not all, but only selected components are included in the reconstruction process. This way, we can choose specified periodic signals of interest (e.g. annual and semi-annual) or values that are statistically significant for the time series. To do so, in our research, we used the Welch method (Welch, 1967) to determine the frequency of particular RCs. Using Welch’s periodogram, we obtained the exact periods that can characterize each station’s Reconstructed Component. For example, the SVTL station’s annual signal occurs in the first and second Reconstructed Components, while the semi-annual signal is seen in the third and fourth RCs (Fig. 5). Each of the RCs, which are sorted by decreasing variance, corresponds to a specified period. If we would like to investigate annual and semi-annual oscillations, we need to identify which Reconstructed Components represent these periods. For most of the researched stations, the first two eigenvectors (the first and the second) explain the maximum amount of variance, corresponding to the annual signal, while the third and the fourth components represent the semi-annual signal.

To determine the statistical significance of Reconstructed Components we used the Fisher-Snedecor test (Johnson et al., 1995). To test the null hypothesis, whether the first $j$ modes are significant, the $F$-statistic was computed using (Tianpo et al., 2012):

$$F(1, N-j) = \frac{(N-j) \hat{\lambda}_j}{\sum_{i=j}^{N} \hat{\lambda}_i} \quad (8)$$

where $\hat{\lambda}_j$ are the eigenvalues of each Reconstructed Component.

This procedure allowed us to separate variance corresponding to each of the significant components that differs from the variance of statistically insignificant components. The purpose of using the Fisher-Snedecor distribution, is to determine the threshold of significant values. From Figure 6 we can notice that the first four components are significant at a 99.7 % level equivalent to the 3-sigma criterion. The first two components correspond to the annual signal, while the third and the fourth components to the semi-annual.
In our numerical example, we estimated annual and semi-annual signals varying over time, for a set of 18 stations located in Central Europe. Figure 7 presents the variance contribution of annual and semi-annual signals in the Up component of the time series, as a percentage of the total variance of data. For most of the stations, the annual signal has the greatest influence on data variability in the time series, while the semi-annual signal in the Up component has much smaller contribution in the total variance. For three of the stations (WTZR, WTZZ, ZIMJ), annual and semi-annual signals have a similar percentage of explained variance, in relation to the total variance. Stations located in the South-Eastern part of Europe are characterized by greater variance in the annual signal compared to other stations. For these stations, more than 30% of the total variance is explained by the annual signal. Possible explanations of origin of these variances are beyond the scope of this paper.

Finally, we compared the non-parametric SSA approach implemented here with a parametric Least Squares Estimation. When using LSE, we assume constant amplitudes and phase shifts over time. Figure 8 presents a comparison between SSA and LSE for Up component for the SVTL station. The maximum difference in seasonal oscillations is equal to 1.4 mm. This difference, if unnoticed is likely to cause the velocity of a permanent station to change of c.a. 0.3 mm/yr (Bogusz and Figurski, 2014).

The seasonal signal identified for the height time series with the SSA, using the first four Reconstructed Components, has larger correlation than an LSE-extracted signal, in respect to the original time series. Figure 9 presents the Root-Mean-Square Deviation (RMSD) for differently determined seasonal signals and the original time series, as well as correlation coefficients between the seasonal signal and the original time series. Seasonal signals extracted by SSA approach are highly correlated to time series of about 4 % on average and are characterized by lower RMSD of about 0.4 mm on average in comparison with LSE signal.
Fig. 7  Variance contribution of annual and semi-annual signals in the height time series, as a percentage of total variance. Bars colored in black relate to the annual signal, while ones in gray represent semi-annual variations. Symbol “1” (1 year) refers to annual signal (black) and “½” (half of the year) refers to semi-annual signal (gray). “10 %” refers to height of the referenced rectangle meaning a percentage of total variance.

SUMMARY

Singular Spectrum Analysis is a non-parametric method, which can be applied to analyze the geodetic time series, with the ability to retrieve information on time-variability of seasonal signals. However, no error propagation is carried out during the analysis, which means there is no possibility for a quantitative evaluation of the analysis, but lack of edge effect is an undeniable advantage. Moreover, the SSA-based results are very responsive to the size of the sliding window and this research should be conducted prior to the major study to assure proper determination of changes of our interest. Using AIC values, we proved that a 3-years window is optimal for investigation of the annual and semi-annual changes. On the other hand, there is no reason to claim that subsequent annual overtones have such large variability and can be determined with Least Squares Estimation without any loss of information. For the set of stations with high-quality time series located in Central Europe which was our case, the changes of annual signal in height were proved to be significant at the level of 2 % with respect to LSE, while semi-annual signal contributes in a small percentage to the total variance. However, these numbers are to change when considering some other parts of the world due to a different origin of the residual annual oscillations in GPS-derived position time series.
ACKNOWLEDGMENTS


Maps were drawn in the Generic Mapping Tool (GMT) (Wessel et al., 2013).


This research was supported by the Polish National Science Centre, grant No. UMO-2014/15/B/ST10/03850.

REFERENCES


Cholyi, V.Ya.: 2014, Precision estimation and forecasting of the Earth’s Orientation Parameters with the singular spectrum analysis method. Earth’s Rotation and Geodynamics, 31, No. 4, 205–212. DOI: 10.3103/S0884591315040030


Golyandina, N. and Zhigljavsky, A.: 2013, Singular spectrum analysis for time series. Springer-Verlag Berlin Heidelberg. DOI: 10.1007/978-3-642-34913-3


Klos, A., Bogusz, J., Figurski, M. and Kosek, W.: 2015, On the handling of outliers in the GNSS time series by the
noise and probability analysis. IAG Symposium, 143. DOI: 10.1007/1345_2015_78


Xu, Ch. and Yue, D.: 2015, Monte Carlo SSA to detect time-variable seasonal oscillations from GPS-derived site position time series. Tectonophysics, 665, 118–126. DOI: 10.1016/j.tecto.2015.09.029

Fig. 2  Tests on the optimal window size using SVTL height time series. On the basis of $AIC$ values, a 3-years window was adopted.

Fig. 8  Comparison of LSE (blue) and SSA (red) of the Up topocentric changes for SVTL station. With gray the original time series is presented.