Joint segmentation of multiple GPS coordinate series

Title: Segmentation multiple de séries de coordonnées GPS

Julien Gazeaux, Emilie Lebarbier, Xavier Collilieux et Laurent Métivier

Résumé : Pour la première fois, une procédure de segmentation multiple de séries de coordonnées est proposée pour des stations GPS géographiquement proches. Elle permet d'estimer simultanément des vitesses de déplacements et des signaux saisonniers spécifiques à chaque série tout en déterminant un signal de déplacement commun à toutes les stations. Une extension du modèle proposé par Picard et al. (2011) et Bertin et al. (2014) est considérée afin de prendre en compte les différentes caractéristiques liées aux données GPS ainsi que la procédure d'estimation, procédure itérative. Les résultats obtenus sur quatre ensembles de séries réelles GPS sont très pertinents d'autant plus que la méthode permet de ne pas segmenter le signal physique en identifiant des ruptures liées au mouvement réel du sol.

Abstract: For the first time, a joint general segmentation procedure of multiple GPS coordinate series is proposed for nearby stations. It allows simultaneously estimating station-specific trends, seasonal signals and a common ground motion signal between series from adjacent stations. An extension of the model and the estimation procedure, which is an iterative procedure, proposed following Picard et al. (2011) and Bertin et al. (2014) is considered in order to take into account for the specificities of the GPS data. The tested approach has been shown to be efficient in providing meaningful offsets and is found to be a relevant method for avoiding segmenting the true physical signal.

Mots-clés : Segmentation, Homogénisation, GPS, Séries temporelles
Keywords: Segmentation, Homogenization, GPS, time series
Classification AMS 2000 : 35L05, 35L70

1. Introduction

Thanks to the various Global Navigation Satellite System (GNSS) and in particular the Global Positioning System (GPS), a wide set of very precise coordinate series of some materialized points at the Earth's Surface is available. Those data supplies snapshots of the Earth's shape for the last twenty years and are thus a precious source of information for geophysicists. The number of permanent stations equipped with GNSS receiver and antenna is constantly increasing due to the rather low cost of equipment and the increasing use of GPS products. Examples of applications are Earth's deformation monitoring - the one discussed in this paper - or measurement of water vapor in the atmosphere and of total electronic content in the ionosphere. As station coordinates are usually estimated on a daily basis using dedicated software packages, a huge amount of data

[1] Observatoire de Paris, Systèmes de Référence Temps-Espace, Paris, France.
E-mail : julien.gazeaux@obspm.fr
[2] IGN LAREG, Université Paris Diderot Sorbonne Paris Cité, Paris, France.
E-mail : xavier.collilieux@ign.fr and E-mail : laurent.metivier@ign.fr
[3] AgroParisTech UMR518, Paris 5e and INRA UMR518, Paris 5e, FRANCE.
E-mail : emilie.lebarbier@agroparistech.fr

becomes available and urgent need for automatic methods appears to control and assess estimated coordinates.

It has been found that obtained GPS station coordinate series show a significant number of change-points (Williams, 2003), generally called offsets or discontinuities in the geophysics and geodesy literature. If they are not identified or corrected, those lead to misinterpretation of long-term ground motion and affect coordinate forecasts. Indeed, a simple cinematic model is generally fitted to the series to derive station velocities. More particularly, those are used to study geodynamic large scale processes such as post-glacial rebound (King et al., 2010) or tectonic plate motion (Kreemer et al., 2003; Altamimi et al., 2012) or to study current ice melting indirectly by investigating Earth crust uplift (Wu et al., 2012, 2011; Wahr et al., 2013). In addition to the physical interpretation of the coordinate variations, a specific fitted coordinate model defines the Terrestrial Reference frame (Altamimi et al., 2002; Altamimi et al., 2011; Petit and Luzum, 2010) which is used for mapping purpose, to reference any geophysical measurement on the Earth or to estimate artificial Earth's satellite orbits.

As a priori knowledge on the location of the offsets is incomplete, automatic algorithms to detect offsets in GPS coordinate series have been proposed. Those were based on sequential testing (Perfetti, 2006; Ostini et al., 2008) or variational approaches (Vitti, 2012). However, none of the current tested methods have been found to be accurate enough for current coordinate accuracy requirement (Gazeaux et al., 2013). Visual inspection and offset amplitude significance test based on some simple noise model (Herring, 2003; Williams, 2008) is the currently used approach. In addition to the limitations of the current tested methods, all proposed algorithms are based on the analysis of a single time series, one coordinate component at a time whereas taking into account simultaneously the three components may decrease false detections.

In other fields, such as climatology, the simultaneous analysis of series from closed stations is typically used to detect artificial changes, for example temperature (Caussinus and Mestre, 2004). Their detection and correction is called homogenization. Such methods have the advantage of limiting assumptions on the form of the physical signal that is recorded. However, the climate signal at close stations needs to be considered the same so that the series of difference (of pairs of series) approximate a white noise series contaminated by offsets. Thus, such series can be analyzed with standard segmentation model such as Picard et al. (2005). While it is well performing, looking at difference of series requires the sequential detection and correction of offsets, since there is an ambiguity about to which of the two series offsets belong.

A joint segmentation model that simultaneously studies set of series has been developed by Picard et al. (2011b) in the context of genomics. They proposed a segmentation model in which a functional part is added in order to model common technical errors that affect all the series at the same position. The method they developed has been found appropriate by Mestre et al. (2014) to homogenize climate records. In this specific application, the common errors to all series is the climate signal which is aimed to be determined. To estimate the functional part, Picard et al. (2011b) proposed to use spline functions or a fixed effect. More recently, Bertin et al. (2014) have considered the same model but they proposed to estimate this part as a linear combination of a collection of known functions and select the relevant one using a lasso-type approach. This model has been applied to Height time series from four GPS stations located in the same observatory. However, having several GPS stations at the same location is rare. Stations separated by a few kilometers to several hundred of kilometers are common. However, their coordinate series can

show significantly different slopes and station specific seasonal biases.

In this paper, the model of Picard et al. (2011b); Bertin et al. (2014) was modified in order to allow multiple series to have fixed individual series-specific trends and individual seasonal periodic variations in addition to the common physical signal. Following the same estimation procedure, the parameters are estimated according to an iterative procedure. This method was applied to several groups of GPS series. The results were evaluated based on a priori knowledge on the location of the offsets and authors expertise.

Section 2 reports some background about GPS series and describe the data set used in this study. While section 3 describes the model and the proposed estimation procedure, section 4 describes the obtained segmentation based on real data.

2. Data

2.1. Background on GPS series

Highly precise coordinates of GPS permanent stations need to be post-processed in order to accumulate sufficient data to accurately determine GPS satellite positions in the sky. In addition to satellite positions, a large set of parameters as the result of some least-squares adjustments of one day of observations requires to be estimated to model any phenomenon that affect GPS code and phase measurements. Cartesian coordinates in a global Earth's centered frame are estimated for each stations but they are usually rotated and translated to a local frame along the East, North and vertical direction, which is the best frame to interpret measurements [1]. In that frame, the noise affecting height coordinate is about three times larger than the horizontal.

An example of GPS coordinate series is shown on Figure 1. Just before year 2012, an offset is visible in the North and Height component but it is not clear that other apparent variations should be considered as change-points or not. Earthquakes and equipment changes (station antenna, receiver, cable) are the two main reasons explaining the change-points. The reason why antenna changes cause offsets is not fully understood all the more since antennas are most of the time installed at exactly the same location. At least a portion can be explained by mis-calibration of antennas. More marginally, changes in the environment of the stations, bump on station, change in raw observation computation strategy or equipment progressive malfunction can explain offsets. They may affect one component only or several simultaneously as shown in Figure 1.

Gazeaux et al. (2013) gave statistics on the magnitude and frequency of offsets from a large database of GPS series. They showed that 33% of all detected offsets remain unexplained (i.e. neither by equipment change nor by seismic event). This number can decrease to less than 5% when metadata are collected and analyzed with care, which is only possible with smaller network (Santamaría-Gómez et al., 2012). A significant work has been recently carried out to provide a priori information on the offsets related to Earthquakes (Tregoning et al., 2013; Métivier et al., 2014). While such work provides the right magnitude of the effect in average, it can not be fully trusted. One has to verify whether the offset is detected in the observations, see Figure 1 for example of modeled induced earthquake displacements.

---

[1] Strictly speaking, the vertical is described here as the normal at a reference ellipsoid at the station location.
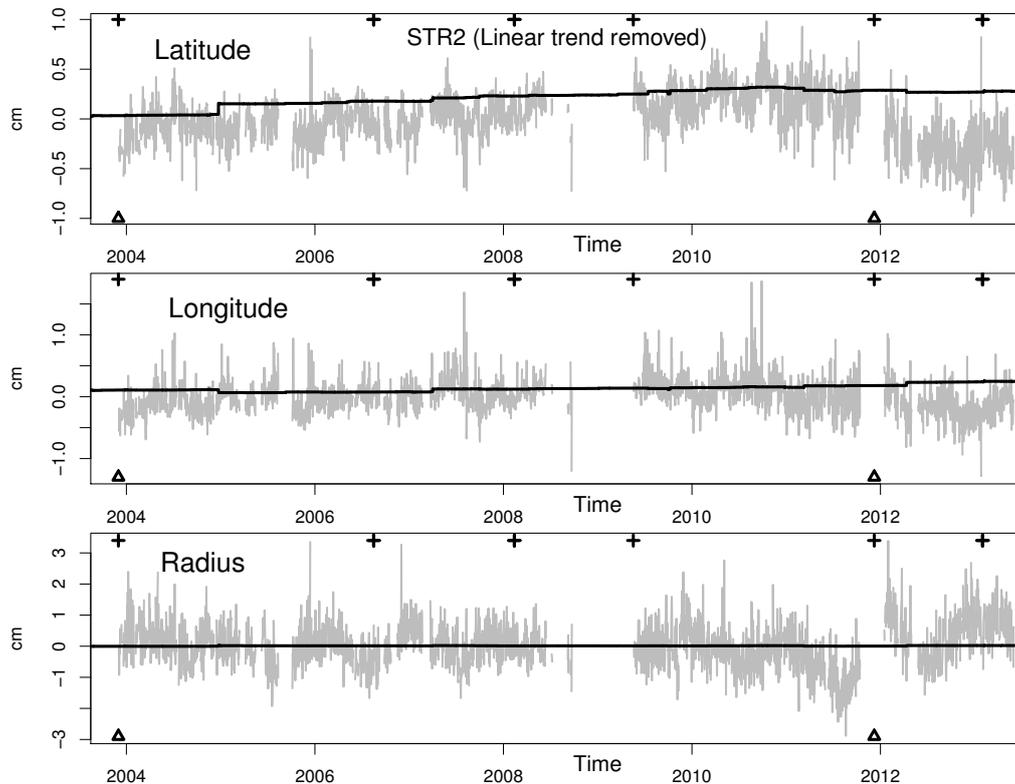
FIGURE 1. *Example of the coordinate series of station STR2 (subset STR1), see section 2.2 for the data source. A trend has been removed for each series to highlight time series variations. Antenna and receiver changes are represented as triangles and crosses. Theoretical co-seismic motions is super-imposed.*

The detection of offsets in GPS series is made difficult by the various type of signals contained in a time series. The dominant measured signal has a tectonic origin. About 70% of the Earth's surface is located on stable tectonic plate. In this case, station displacement can be generally approximated by a linear function. In the remaining part of the world, the signal is more complex to describe with simple mathematical functions. The observational noise, generally considered to be additive, is time-correlated (Wdowinski et al., 1997; Williams et al., 2004). Its amplitude has been shown to slightly decrease in the recent year (Santamaría-Gómez et al., 2011), which makes it difficult to model with a simple covariance model. Furthermore, this noise is spatially-correlated (Wdowinski et al., 1997; Dong et al., 2006; Amiri-Simkooei, 2009). Finally, periodic errors are superimposed on background noise at well known periods (Ray et al., 2008, 2013).

In addition to this spatially correlated noise, close series exhibit close physical signal related to the Earth's crust motions. Large-scale feature are notably related to fluid mass transfer at the Earth's surface, in the ocean, atmosphere or in the upper layer of the crust (Farrell, 1972). As an example, when atmospheric pressure increases, the load at the Earth's surface increases which induces a subsidence. This effect has been shown to be detectable in GPS results for more than two decades (Van Dam et al., 1994). Thus, GPS series from adjacent stations show strong correlations. However, they may show different slopes (velocities) - for example in case of local

motion of one station such as ground water pumping. In addition, different seasonal signal can be observed, even for close records (Collilieux et al., 2007; King and Williams, 2009). This means that if one wants to use this spatial redundancy to better detect offsets in the series, one should allow close series to show such different signals, see section 3.

2.2. Used series

Daily series computed by the Jet Propulsion Laboratory (JPL) are considered in this study [2]. They have been averaged at weekly interval to reduce the number of data points. The advantage of this dataset is that it has been homogeneously reprocessed which insure a minimum number of offsets related to the GPS raw observation computation strategy. In addition, the network computed by JPL is composed by 2389 stations [3], which makes it an extraordinary dataset for this work.

In order to apply the method of section 3, clusters of series have been generated. Within a cluster, all should show the same signal of deformation due to Earth's crust motion, but different seasonal periodic signals and long-term slopes. Long-term slopes are caused by tectonic activity and can significantly vary depending on site location. Therefore, we generated clusters that contain stations that are distant from less than 100 km so that respective crust motion signal can be considered similar to all coordinate series. However, unstable areas such as volcanic or seismic regions usually show poor correlated time series due to highly non linear signal. For this reason, GPS time series from unstable regions are not considered in this study. In Figure 2, each dot corresponds to a GPS station. Colors depend on Earth's crust deformation. Gray dots correspond to region with high deformation, such as volcanic or seismic regions. Black dots correspond to sites with low deformation and stars correspond to the subsets of sites that are close to less than 100 km and therefore have been selected for this study. Four representative sets of three-components series (latitude, longitude and radius) have been selected, see Figure 2. For example, time series in Figure 3 shows latitude coordinate time series of a north America subset of stations named after LAMT station. Each panel corresponds to the time series of a specific nearby sites. Finally, initial time series are weekly average in order to improve computing performance.

A priori information on the equipment changes at all these stations have been taken from the Scripps Orbit and Permanent Array Center (SOPAC) station logsheet database. In these so-called logsheets, any event that affects a station is normally recorded such as antenna, receiver or clock change, the epoch of the event being supplied. Those information are hereafter called metadata.

3. Statistical model and estimation procedure

This section recalls the model proposed by (Picard et al., 2011b; Bertin et al., 2014) and describes modifications as well as the estimation procedure.

M denotes the number of series with length n_m for series m, y_{mt} the observed signal of the series m at time t, K_m the number of segments of the mth series and $K = \sum_{m=1}^{M} K_m$ the total number of segments. Note that the segmentation is specific to each series. The model proposed by both (Picard et al., 2011b; Bertin et al., 2014) is the following :

$$y_{mt} = \mu_{mk} + f(t) + e_{mt}, \quad \text{for } t \in I_k^m = [\tau_{k-1}^m, \tau_k^m], \tag{1}$$

---

[2] They can be downloaded at ftp ://sideshow.jpl.nasa.gov/pub/JPL/GPS_Timeseries/repro2011b/post/point/.

[3] series downloaded in mid 2013. Indeed, Series available at JPL are updated everyday.
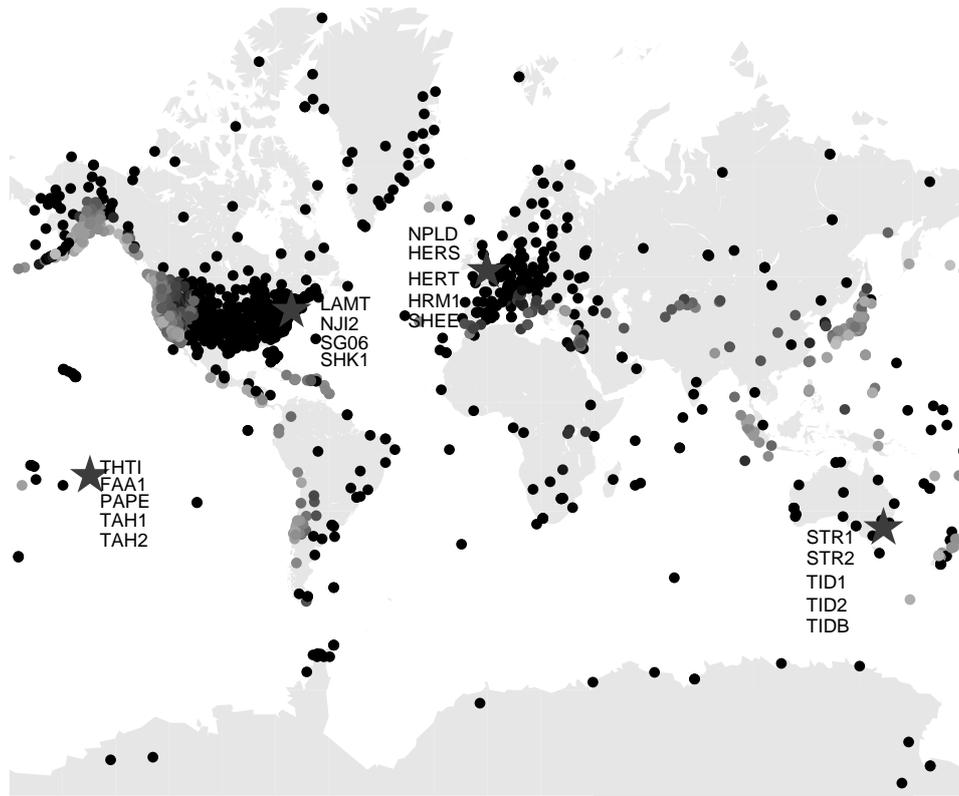
FIGURE 2. *Network of GPS stations whose data are processed at JPL. Stations in deformation zone are shown in gray. Selected set of stations for this study are shown in star symbol.*

where τ_k^m is the kth change-points of the series m, μ_k^m is the mean of the series m on the segment I_k^m and the $e_m(t)$ are i.i.d centered Gaussian with variance σ^2.

To estimate all these parameters (the function f, the means and the positions of the breakpoints), they proposed an iterative algorithm where f is estimated using spline functions or a fixed effect in (Picard et al., 2011b) and a lasso-type approach in (Bertin et al., 2014).

We propose here an extension of model (1) in order to take into account for the characteristics of the GPS data described previously :

$$y_{mt} = c_m t + \mu_{mk} + \sum_{i=1}^{p} (a_{mi}cos(2\pi w_i t) + b_{mi}sin(2\pi w_i t)) + f(t) + e_{mt}, \text{ for } t \in I_k^m = [\tau_{k-1}^m, \tau_k^m] \quad (2)$$

where the $e_m(t)$ are i.i.d centered Gaussian with variance σ^2, (a_{mi}, b_{mi}) model site-specific periodic variations at periods w_i, c_m are site-specific slopes that depend on tectonic activity and function f is the residual common signal found in all series that includes, among others, Earth's crust deformation. The parameters of the model to be estimated are the trends c_m, the means μ_{mk}, the change-points τ_k^m, the function f, the coefficients a_{mi} and b_{mi}, the variance σ^2 and the number of segments K.

As usual in the segmentation estimation framework, a two-step procedure is used : the parameters are estimated for a fixed number of segments K, then K is chosen using a model selection strategy. Following (Picard et al., 2011b; Bertin et al., 2014), an iterative procedure is proposed that alternates the estimation of each parameter the others being fixed : the power (h) of a parameter denotes its value at the iteration (h), then at the next iteration $(h+1)$ the procedure is summarized as follows :

- At iteration 0, a first estimation of the trend component is carried out and the difference between initial series and the later trend now defines y_m.
- c_m, a_{mi} and b_{mi} are estimated through classical least-squares on the vector of observations $\tilde{y}_{mt} = y_{mt} - \mu_{mk}^{(h)} - f^{(h)}(t)$.
- f is estimated using fixed effect or the lasso approach on $\tilde{y}_{mt} = y_{mt} - c_m^{(h+1)} t - \mu_{mk}^{(h)} - \sum_{i=1}^{p}(a_{mi}^{(h+1)}cos(2\pi w_i t) + b_{mi}^{(h+1)}sin(2\pi w_i t))$ (see (Picard et al., 2011b; Bertin et al., 2014) for details), denote by $f^{(h+1)}$ its new value.
- The means and the change-points are estimated simultaneously by :

$$(\mu_{mk}, \tau_{k-1}^m)^{(h+1)} = \underset{\mu,\tau}{\operatorname{argmin}} \sum_{m=1}^{M} \sum_{k=1}^{K_m} \sum_{t \in I_k^m} (\tilde{y}_{mt} - \mu_{mk})^2, \qquad (3)$$

where $\tilde{y}_{mt} = y_{mt} - c_m^{(h+1)} t - \sum_{i=1}^{p}(a_{mi}^{(h+1)}cos(2\pi w_i) + b_{mi}^{(h+1)}sin(2\pi w_i)) - f^{(h+1)}(t)$. The problem is then reduced to a classical segmentation problem : segment $\tilde{Y}$ into K segments. It is now well known that the exact solution of the problem (3) can only be obtained by the Dynamic Programming (DP) algorithm. This algorithm can here be applied since the quantity to be optimized is additive according to the segments (that explained the iterative procedure in order to use it). In particular here, we use the double-stage of DP proposed by Picard et al. (2011a) which is adapted (faster) to the segmentation of multiple series.

This procedure allows to estimate the parameters for a fixed number of segments. To estimate this number, the model selection criterion proposed by Picard et al. (2011a) is applied.

The overall model (2) has been verified for theoretical identifiability issues, especially the regressing functions (periodic signals, trend, the functions in the dictionary for lasso approach) have been selected so that they are not correlated (see Arribas-Gil et al. (2014) for details on this condition for the functions of the dictionary). However, in a practical point of view, some confusions can be observed. For example, when considering fixed effect to estimate the functional part (f), as mentioned in (Picard et al., 2011b,a), simultaneous change-points in a large fraction of series can be confounded. Nevertheless we have observed that if initial step consists in segmenting the signal, the trend or/and the periodic signals can be captured by it. The choice for the initialization of the iterative procedure is then clearly important. We have observed on simulated data that it is preferable to begin with the estimation and the correction of the trend.

4. Results

Our proposed method of section 3 has been applied to the four sets of series discussed in section 2.2, the three components East, North and Height were handled independently. Site-specific periodic variations are modeled with four periods $1/w_i$: 365.25, 350.5, 182.625 and 175.25 days. For the estimation of the functional part of model (2), we consider both the estimation by fixed effect ($f(t) = f_t$) and the estimation using Lasso approach. The results are given and discussed in section 4.1 and 4.2 respectively. The 350.5-days period, also called draconitic signal is not yet well explained but was shown to be part of GPS signal by Ray et al. (2013). These signals could originate from either local multipath effects at stations, mismodelling effect in satellite orbits or errors in a priori IERS (International Earth Rotation and Reference Systems Service) model. Results are assessed by looking at the offsets locations and by checking initial hypothesis on noise characteristics in residuals $\hat{e}_m$.

Independence of residuals has been checked by calculating partial autocorrelations. Those show no significant autocorrelation with an order higher than 1. Nevertheless, first order regressive parameter of residual is in most cases lower than 0.2 and therefore is neglected in our model. Detected offsets have been validated or not according to three criteria : correspondence with recorded station equipment changes, simultaneity for more than one coordinate component and visual inspection of time series.

Epoch of all detected offsets have been compared using a tolerance of 8 or 15 days (corresponding to 1 and 2 data points). Detected offsets are validated if they match the above criteria within the selected time window. It is worth mentioning that missing value are taken into account in those comparisons since data generally miss before an antenna or receiver being replaced. In the following, 'multivariate offset' stands for an offset occurring simultaneously (or almost simultaneously) on more than one component at the same epoch. Those have found to be common in case of equipment change or when stations actually move. The visual inspection of detected offsets is carried out on the time series of differences between two different station records. Those differences show an offset when one occurs on either one of the two station series. If the model of section 3 is correct, such series of differences should exhibit only a segmentation part, seasonal changes and additive white noise. Any deviation to this model should cause an over estimation of the number of change-points. Thus, visual inspection mainly aims to verify that detected change-point can be detected visually and to validate if the data verify the assumption of the tested model. Finally, outliers are defined as two consecutive detected offsets. Table 1 which supplies the total number of detected offsets and the validation criteria is discussed in the next two sections.

4.1. Fixed effect

As illustrated on Figure 3a), the estimated common signal, represented in black over each detrended series is rather close to the original series, which shows that series show similar variations. Detected change-points are plotted over by vertical black lines. Antenna and receiver changes are represented by triangles and crosses. Here, only three detected offsets can be explained by metadata. Figure 4a) shows time series corrected for the segmentation part for stations subset named LAMT.

According to Table 1, more than 50 change-points are detected in total for each subset of

TABLEAU 1. *Statistics on the detected change-points. Number of series, total number of offsets, number of offsets corresponding to receiver or antenna change according to a tolerance of 8 or 15 days, number of offsets that are visible on several components (counted twice in detected on two components), offsets validated by a visual inspection of the time series of differences, number of confusion i.e. offsets that are wrongly interpreted as change-points instead of being absorbed by function f and number of offsets related to outliers. These figures have been obtained by stacking the results from the three coordinates components for each subset of series.*

|	nb	Total	Receiver	Antenna	Multivariate	Visual	Confusion	Outlier
f modeled as a fixed effect								
LAMT 8 days	4	50	1	6	8	25	5	12
LAMT 15 days			4	6	10			
STR1 8 days	5	84	14	11	28	44	18	4
STR1 15 days			28	19	40			
THT1 8 days	5	46	5	3	18	22	5	12
THT1 15 days			7	3	24			
NPLD 8 days	5	77	5	8	36	50	3	10
NPLD 15 days			8	8	48			
NPLD run 2 8 days		60	6	8	20	35	3	10
NPLD run 2 15 days			8	8	26			
f modeled as periodic function								
LAMT 8days	4	20	1	5	4	15	3	0
LAMT 15days			4	5	4			
STR1 8days	5	28	6	6	14	16	10	2
STR1 15days			13	10	16			
THT1 8days	5	13	2	2	0	10	2	0
THT1 15days			3	2	2			
NPLD 8 days	5	74	8	6	20	15	3	4
NPLD 15 days			11	6	40			
NPLD run 2 8 days		56	5	8	13	20	4	8
NPLD run 2 15 days			7	8	17			
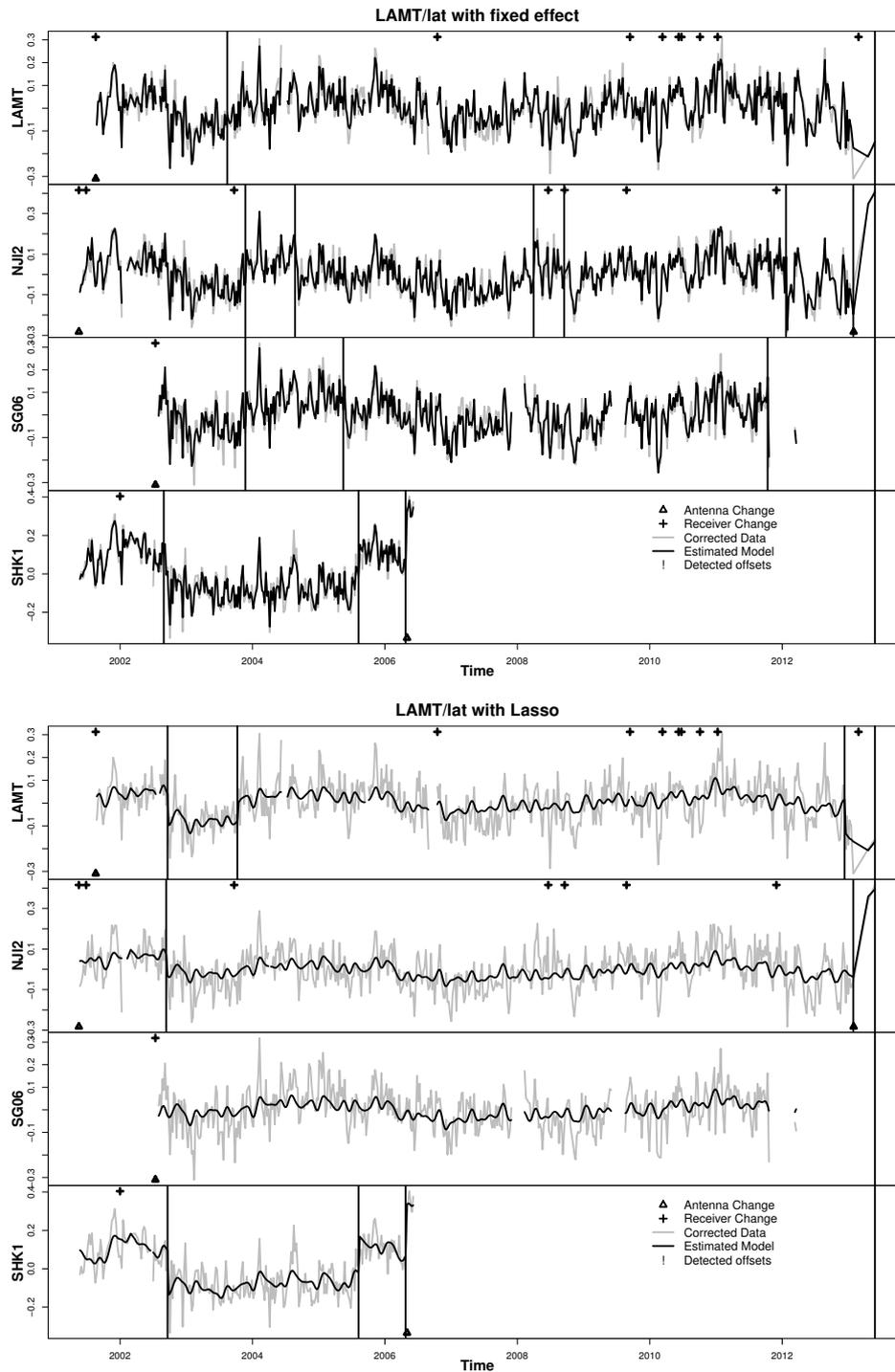
FIGURE 3. *Time series and method results for subset LAMT when a) f is modeled as fixed effect or b) f is modeled with Lasso. Each panel corresponds to a time series of latitude GPS coordinates. A trend has been removed for clarity. Black lines correspond to the total estimated model, including f, station specific periodic signal and the segmentation. Metadata from SOPAC archive are shown with black markers.*
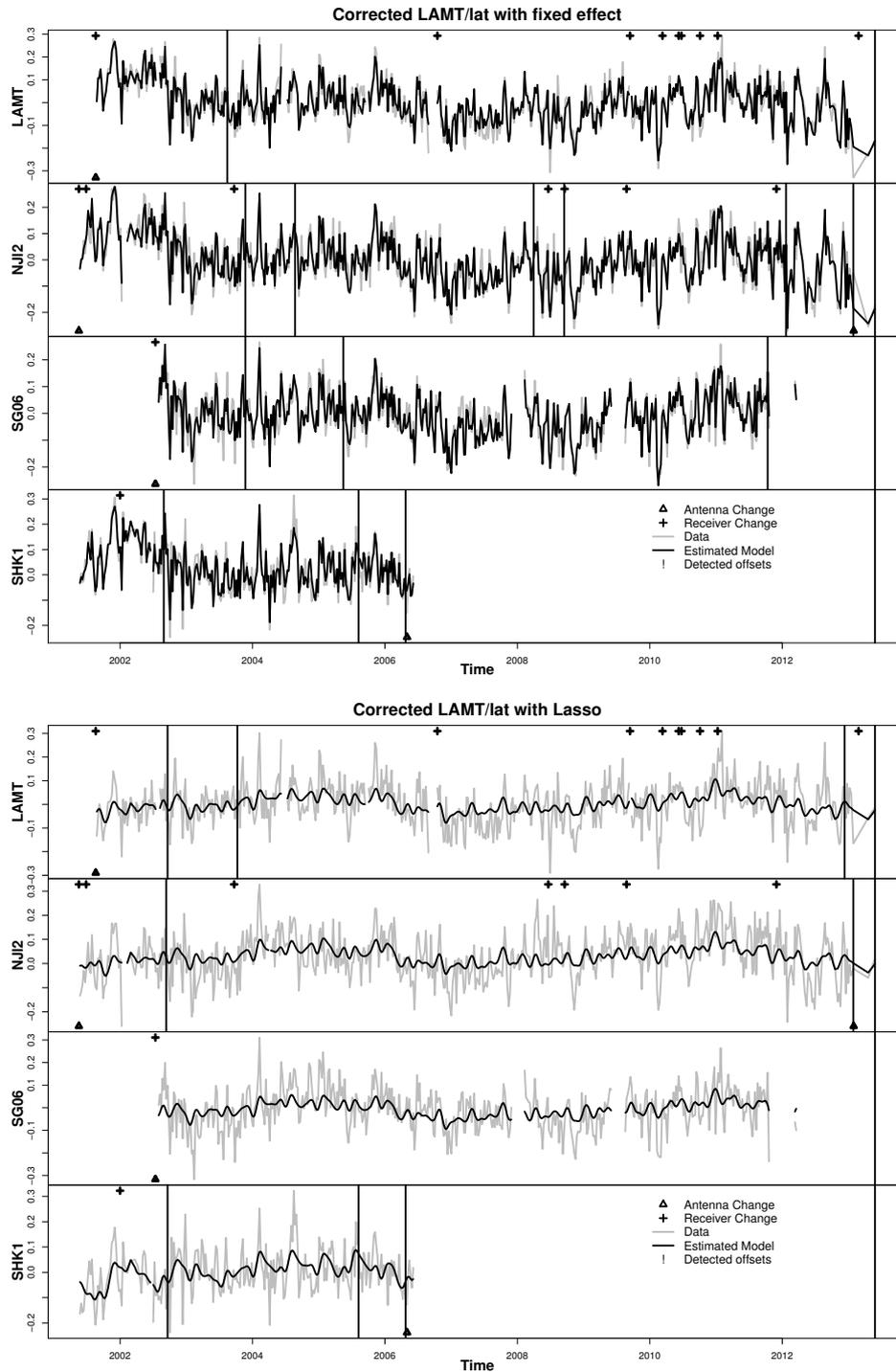
FIGURE 4. *Time series and method results for subset LAMT corrected for offsets effect when a) f is modeled as fixed effect or b) f is modeled with Lasso. Black lines correspond to the total estimated model, including f, station specific periodic signal but without segmentation. Metadata from SOPAC archive are shown with black markers.*

stations along three components. Only a few are related to metadata according to column 4 and 5. Among the detected one, only half can be confirmed by a visual inspection. As illustration, Figure 5a) shows the series of coordinate differences for some series of the same subset of stations in latitude component. First notification is that SHK1 seems not to behave as the others. This would imply that the detected offsets of series SHK1 are not real and can be caused by the peculiar motion of this station. In this case, the three discontinuities estimated in the other series around 2004.0 would be also fake : they would be related to a confusion between the signal f and the segmentation. Indeed, one would expect to have an offset in function f and one offset at the same time in SHK1. This confusion arises since the algorithm is iterative and the number of segments is re-estimated using the same function f and site-specific biases whatever the number of segment in the DP within each iteration. A specific column for those confusions between individual series segmentation and f are given in Table 1. Those are not considered to be a problem since they does not prevent for a real offset to be detected although it is not provided in the right series. The others discontinuities in series NJI2 and SG06 all have a very small magnitude, except the last one of NJI2, yielding a very high sensitivity to time series small apparent variations.

In addition to SHK1, two other series show different behaviors than the others of their groups, which generates several artificial offsets in Table 1 : Station STR1 in latitude and longitude and station FAA1 (subset THTI) in vertical. This partly explains why the subset STR1 show such a large number of change-points. The large number of change-points at NPLD cluster is related to a confusion between the segmentation of one series and its trends, see section 4.3. The high sensitivity of the algorithm has been shown to allow detecting outliers in the series : about 15% of the offsets. Among all the detected offsets, a significant part is multi-variate (between 35 and 47%), which tends to confirm them. However, outliers are often visible on several components and some small-scale features can appear simultaneously on two coordinate components of the same station, which supply at least 4 offsets (ex : TID1 in latitude and longitude - subset STR1). The large number of detected offsets also sometimes causes detected offsets to be simultaneous.

4.2. Lasso

Modeling f with a set of functions allows adding constraints on the time variations of f. Dictionary is filled with periodic functions with frequencies $\frac{2\pi \cdot i}{T}$ with T the length of the series. Well known periods $350.5/j$ days with j from 3 to 8 are also added. Indeed, not having the exact frequency could lead to a wrong segmentation even if close frequencies are selected in the Dictionary. Annual and semi-annual periods and periods $350.5/j$ days with j from 1 to 2 have been already removed from the data with station-specific amplitudes in the iterative process. Together with annual and semi-annual signals, this makes a Dictionary of 10 functions. Note that, for each subset of station, between 1 and 3 periodic functions from Dictionary were finally selected by the Lasso procedure.

Figure 3b) and 5b) show the obtained results for the subset of stations LAMT. Compared to Figure 3a), function f is much smoother and a fewer number of change-points have been estimated. An offset has been detected at the same time in 2002 for the three stations having data, pointing a confusion. This shows that function f was unable to model the abrupt variation of the common signal and that those multiple offsets are not real. The second offset of LAMT is clearly a false detection. It can be related to station SHK1 that do not show the same motion than the
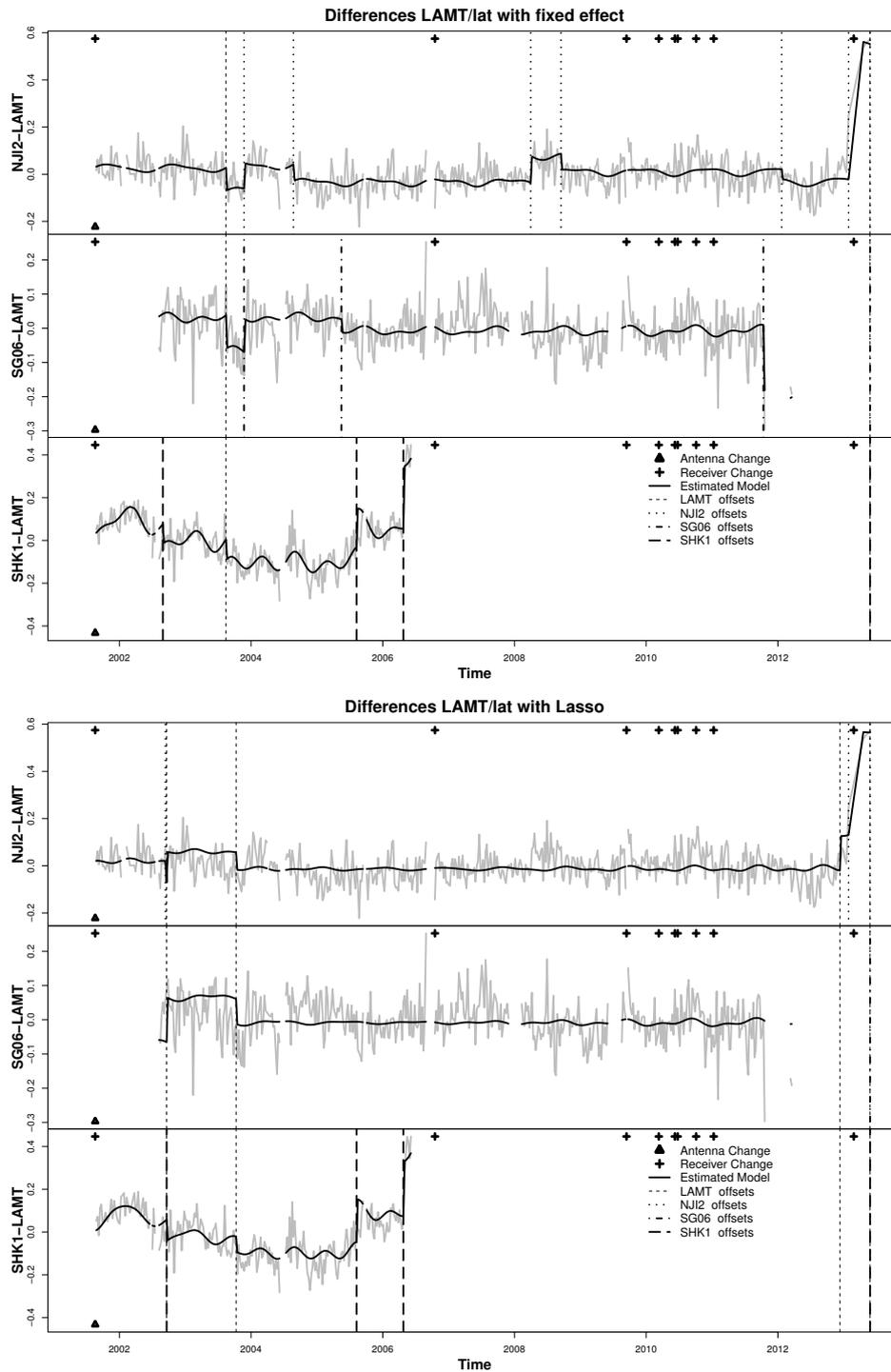
FIGURE 5. *Time series of differences between station pairs for subset LAMT. Each panel corresponds to the difference between LAMT time series and an other series of the same subset. Vertical lines represent detected offsets on each time series (LAMT, NJI2, SG06 and SHK1) when a) f is modeled as a fixed effect or b) f is modeled with periodic functions. Black lines correspond to the differences between station specific periodic signals and segmentation.*

other stations, which is also noted by the algorithm due to the detection of two additional offsets.

Except for subset NPLD, results for the Lasso strategy lead to segmentation that can be validated by the analysis as far as the few stations that show different motions are excluded. Most are clearly visible in the time series of differences or can be explained by outliers, see Table 1. For example for subset LAMT, 15 of the 20 change-point are visible in the time series of differences and 3 comes from a confusion. However, Lasso has been shown to be too permissive in a few cases and a few real offsets that have been detected in the fixed effect strategy have been missed. Most correspond to small segments (one for TIDB in longitude and one in latitude, one in TID2 in longitude and one in STR2 in longitude and latitude, all in the subset STR1). An offset of very small amplitude could also have been added for THTI in latitude around 2008 that corresponds to a receiver change.

4.3. Discussion and conclusions

Whatever the modeling of function f, the over-segmentation detected in NPLD subset is related to a confusion between individual series slopes and individual segmentation. In case of large offset in one series, the removal of the estimated trend at iteration 0 generates piece-wise triangle functions in the series of residuals $\tilde{y}_{mt}$ i.e. each segment between change-points in $\tilde{y}_{mt}$ shows a trend. This trends is erroneously segmented since it is interpreted as mean changes during the segmentation procedure. To illustrate this, algorithm on NPLD dataset was rerun by modeling two obvious large offsets (larger than 1 cm) at iteration 0. The results are supplied in Table 1 with name "run 2". The number of detected offsets is significantly smaller. The number of matching equipment changes remains similar and so remains the proportion of multivariate offsets. This means that offsets with large magnitude can affect detection and estimation procedure. Thus, it is mandatory to identify at first the largest offsets in the series for a more robust estimation of the trend at iteration 0. This could be easily achieved by sequential testing using time-running windows and a minimum offset amplitude of 1 cm.

The reason why modeling f by a fixed effect leads to a larger number of change-points can be understood. As fixed effect tends to fit series variations, the variance of the noise term is about four times smaller than in Lasso case. The under-estimation of the noise term variance likely implies an overestimated number of detected change-points. The small number of series coupled with the presence of data gap may also prevent for a robust estimation of the fixed effect. Indeed, more confusions have been reported compared to the Lasso case. However, the fixed effect strategy was able to find all real offsets compared to the Lasso case and was able to detect small-scale features in the series. This explains why many have been validated by visual inspection. Thus, the two strategies are complementary.

While some detected change-points are multivariate, which is likely to imply an event at the station, a few are reported by metadata, probably partly because metadata are sometimes not documented enough. It was also confirmed that not all equipment changes occurring at the stations network cause an offset, which was a well known results. An other cause for offset in GPS coordinates are Earthquakes. In order to test for a possible seismic origin, theoretical synthetic series of offsets related to Earthquake co-seismic motions have been derived following Métivier et al. (2014). It is found that no Earthquake is able to generate differential motions within a subset of tested coordinate series. Co-seismic common motions with cumulative effect up to 2 mm in

horizontal are expected. However, no obvious correlation has been found between theoretical motions and estimated f functions for THTI and STR1 clusters where the expected co-seismic motion is the largest.

5. Conclusions

In this paper, a modified version of the multivariate segmentation model by Picard et al. (2011b); Bertin et al. (2014) adapted to the joint analysis of multi-site GPS coordinates time series has been proposed. Following Mestre et al. (2014), offsets in coordinate series of nearby stations are simultaneously searched, so that the ground motion signal can be considered identical or sufficiently close. It was mandatory, due to strong trends in the series and in order to avoid confusion with the segmentation to explicitly add a trend term in the model for each series. Because GPS series from close stations can show different seasonal signal, station-specific periodic terms at well-known periods were added. These modifications have been introduced within the iteration procedure of Picard et al. (2011b); Bertin et al. (2014) in order to keep using dynamic programming.

The results that have been obtained on four different subsets of GPS station coordinates series have been shown to be relevant. Reliable segmentation has been obtained especially when the Lasso method is used to estimate the common signal as a set of periodic functions although a few change-points related to small segments have been missed. The estimation of that signal as a fixed effect shows a higher detection power but tends to over-segment GPS series in averaged according to authors experience. However, it is mandatory to identify beforehand the largest offsets ($> 1cm$) in the raw individual series to avoid over-segmentation, which could be easily implemented.

This joint analysis of GPS series is very powerful since it does only assume that close GPS stations should show close coordinate variations in addition to the individual segmentation and the trends. As can be observed by looking at Figure 3, a standard univariate analysis of the GPS series may likely have wrongly lead to a segmentation of the real physical signal which show large unpredictable time-scale features. Besides, this algorithm has been also found to be an interesting tool to identify stations which show inconsistent displacement compared to its neighbors. Ideally, one would need to form clusters of series that show similar ground motion before running the segmentation algorithm. The design of this cluster is a practical issue that will need to be solved since several thousand of station series are currently available. Preliminary analysis show that fixing a maximum distance between stations was not sufficient and that further investigations are necessary. Although computational cost of the method is quite high, work by Hannart et al. (2014) may give valuable hints about how to group stations together based on similarity metrics and in order to automatize station clustering.

6. Acknownledgment

Références

Altamimi, Z., Collilieux, X., and Métivier, L. (2011). ITRF2008 : an improved solution of the International Terrestrial Reference Frame. *Journal of Geodesy*, 85(8) :457–473.

Altamimi, Z., Métivier, L., and Collilieux, X. (2012). ITRF2008 plate motion model. *Journal of Geophysical Research (Solid Earth)*, 117 :7402.

Altamimi, Z., Sillard, P., and Boucher, C. (2002). ITRF2000 : A new release of the International Terrestrial Reference Frame for earth science applications. *Journal of Geophysical Research (Solid Earth)*, 107 :2214.

Amiri-Simkooei, A. R. (2009). Noise in multivariate GPS position time-series. *Journal of Geodesy*, 83 :175–187.

Arribas-Gil, A., Bertin, K., Meza, C., and Rivoirard, V. (2014). Lasso-type estimators for semiparametric nonlinear mixed-effects models estimation. *Statistics and Computing*, 24(3) :443–460.

Bertin, K., Collilieux, X., Lebarbier, E., and Meza, C. (2014). Segmentation of multiple series using a lasso strategy. *Submitted*. http ://arxiv.org/abs/1406.6627.

Caussinus, H. and Mestre, O. (2004). Detection and correction of artificial shifts in climate series. *Applied Statistics*, 53 :405–425.

Collilieux, X., Altamimi, Z., Coulot, D., Ray, J., and Sillard, P. (2007). Comparison of very long baseline interferometry, GPS, and satellite laser ranging height residuals from ITRF2005 using spectral and correlation methods. *Journal of Geophysical Research : Solid Earth*, 112(B12).

Dong, D., Fang, P., Bock, Y., Webb, F., Prawirodirdjo, L., Kedar, S., and Jamason, P. (2006). Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis. *Journal of Geophysical Research (Solid Earth)*, 111 :3405.

Farrell, W. E. (1972). Deformation of the Earth by Surface Loads. *Reviews of Geophysics and Space Physics*, 10 :761.

Gazeaux, J., Williams, S., King, M., Bos, M., Dach, R., Deo, M., Moore, A. W., Ostini, L., Petrie, E., Roggero, M., Teferle, F. N., Olivares, G., and Webb, F. H. (2013). Detecting offsets in GPS time series : First results from the detection of offsets in GPS experiment. *Journal of Geophysical Research : Solid Earth*, 118(5) :2397–2407.

Hannart, A., Mestre, O., and Naveau, P. (2014). An automatized homogenization procedure via pairwise comparisons with application to Argentinean temperature series. *International Journal of Climatology*, 34 :3528–3545.

Herring, T. (2003). MATLAB Tools for viewing GPS velocities and time series. *GPS Solutions*, 7 :194–199.

King, M., Altamimi, Z., Boehm, J., Bos, M., Dach, R., Elosegui, P., Fund, F., Hernandez-Pajares, M., Lavallée, D., Cerveira, P., Riva, R., Steigenberger, P., van Dam, T., Vittuari, L., Williams, S., and Willis, P. (2010). Improved constraints on models of glacial isostatic adjustment. *A review of the contribution of ground-based geodetic observations, Surveys in Geophysics*, 31(5).

King, M. A. and Williams, S. D. P. (2009). Apparent stability of GPS monumentation from short-baseline time series. *Journal of Geophysical Research : Solid Earth*, 114(B10).

Kreemer, C., Holt, W. E., and Haines, A. J. (2003). An integrated global model of present-day plate motions and plate boundary deformation. *Geophysical Journal International*, 154 :8–34.

Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E., Böhm, R., Aguilar, E., Guijarro, J., Vertachnik, G., Klancar, M., Dubuisson, B., and Stepanek, P. (2014). HOMER : a homogenization software - methods and applications. *Quarterly Journal of the Hungarian Meteorological Service*, 117(1) :47–67.

Métivier, L., Collilieux, X., Lercier, D., Altamimi, Z., and Beauducel, F. (2014). Global co-seismic deformations, GNSS time series analysis and earthquake scaling laws. *Journal of Geophysical Research, under revision*.

Ostini, L., Dach, R., Meindl, M., Schaer, S., and Hugentobler, U. (2008). Fodits : A new tool of the bernese gps software. *in Proceedings of EUREF 2008 Symposium, edited by Torres, J. A., and H. Hornik, Brussels, Belgium*.

Perfetti, N. (2006). Detection of station coordinate discontinuities within the Italian GPS Fiducial Network. *Journal of Geodesy*, 80(7) :381–396.

Petit, G. and Luzum, B. (2010). IERS Conventions (2010). (IERS Technical Note ; 36). Technical report, Frankfurt am Main : Verlag des Bundesamts für Kartographie und Geodäsie. inprint.

Picard, F., Lebarbier, E., Budinskà, E., and Robin, S. (2011a). Joint segmentation of multivariate gaussian processes using mixed linear models. *Computational Statistics & Data Analysis*, 55(2) :1160 – 1170.

Picard, F., Lebarbier, E., Hoebeke, M., Rigaill, G., Thiam, B., and Robin, S. (2011b). Joint segmentation, calling and normalization of multiple cgh profiles. *Biostatistics*, 12(3) :413–428.

Picard, F., Robin, S., Lavielle, M., Vaisse, C., and Daudin, J.-J. (2005). A statistical approach for CGH microarray data analysis. *BMC Bioinformatics*, 6 :27.

Ray, J., Altamimi, Z., Collilieux, X., and van Dam, T. (2008). Anomalous harmonics in the spectra of GPS position

estimates. *GPS Solutions*, 12(1).

Ray, J., Griffiths, J., Collilieux, X., and Rebischung, P. (2013). Subseasonal GNSS positioning errors. *Geophysical Research Letters*, 40(22) :5854–5860.

Santamaría-Gómez, A., Bouin, M.-N., Collilieux, X., and Wöppelmann, G. (2011). Correlated errors in GPS position time series : Implications for velocity estimates. *Journal of Geophysical Research*, 116 :1405.

Santamaría-Gómez, A., Gravelle, M., Collilieux, X., Guichard, M., Míguez, B. M., Tiphaneau, P., and Wöppelmann, G. (2012). Mitigating the effects of vertical land motion in tide gauge records using a state-of-the-art GPS velocity field. *Global and Planetary Change*, 98 :6–17.

Tregoning, P., Burgette, R., McClusky, S. C., Lejeune, S., Watson, C. S., and McQueen, H. (2013). A decade of horizontal deformation from great earthquakes. *Journal of Geophysical Research (Solid Earth)*, 118 :2371–2381.

Van Dam, T., Blewitt, G., and Heflin, M. (1994). Detection of atmospheric pressure loading using the global positioning system. *J. of Geophys. Res.*, 99(23) :939–23.

Vitti, A. (2012). Sigseg : a tool for the detection of position and velocity discontinuities in geodetic time-series. *GPS Solutions*, 16(3) :405–410.

Wahr, J., Khan, S. A., van Dam, T., Liu, L., van Angelen, J. H., van den Broeke, M. R., and Meertens, C. M. (2013). The use of GPS horizontals for loading studies, with applications to northern California and southeast Greenland. *Journal of Geophysical Research : Solid Earth*, 118(4) :1795–1806.

Wdowinski, S., Bock, Y., Zhang, J., Fang, P., and Genrich, J. (1997). Southern California permanent GPS geodetic array : Spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake. *Journal of Geophysical Research*, 102 :18057.

Williams, S. (2003). Offsets in Global Positioning System time series. *Journal of Geophysical Research (Solid Earth)*, 108(19) :2310.

Williams, S. (2008). CATS : GPS coordinate time series analysis software. *GPS Solutions*, 12(2) :147–153.

Williams, S., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R., Prawirodirdjo, L., M., M., and Johnson, D. (2004). Error analysis of continuous GPS position time series. *Journal of Geophysical Research*, 109(B18) :B03412.

Wu, X., Collilieux, X., Altamimi, Z., Vermeersen, B., Gross, R., and Fukumori, I. (2012). Accuracy of the international terrestrial reference frame origin and earth expansion. *Geophysical Research Letters*, 38(L13304).

Wu, X., Heflin, M. B., Schotman, H., Vermeersen, B. L. A., Dong, D., Gross, R. S., Ivins, E. R., Moore, A. W., and Owen, S. E. (2011). Simultaneous estimation of global present-day water transport and glacial isostatic adjustment. *Nature Geoscience*, 3(9) :642–646.