

Detecting offsets in GPS time series: first results from the Detection of Offsets in GPS Experiment

*Original*

Detecting offsets in GPS time series: first results from the Detection of Offsets in GPS Experiment / Gazeaux, J.; Williams, S.; King, M.; Bos, M.; Dach, R.; Deo, M.; Moore, A. W.; Ostini, L.; Petrie, E.; Roggero, Marco; Teferle, F. N.. - In: JOURNAL OF GEOPHYSICAL RESEARCH. SOLID EARTH. - ISSN 2169-9356. - ELETTRONICO. - 118:5(2013), pp. 2397-2407. [10.1002/jgrb.50152]

*Availability:*

This version is available at: 11583/2507420 since:

*Publisher:*

John Wiley & Sons

*Published*

DOI:10.1002/jgrb.50152

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*  
AGU

Da definire

(Article begins on next page)

## Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment

Julien Gazeaux,<sup>1</sup> Simon Williams,<sup>2</sup> Matt King,<sup>1,3</sup> Machiel Bos,<sup>4</sup> Rolf Dach,<sup>5</sup> Manoj Deo,<sup>6</sup> Angelyn W Moore,<sup>7</sup> Luca Ostini,<sup>5</sup> Elizabeth Petrie,<sup>1</sup> Marco Roggero,<sup>8</sup> Felix Norman Teferle,<sup>9</sup> German Olivares,<sup>9</sup> and Frank H. Webb<sup>7</sup>

Received 4 December 2012; revised 7 March 2013; accepted 8 March 2013; published 8 May 2013.

[1] The accuracy of Global Positioning System (GPS) time series is degraded by the presence of offsets. To assess the effectiveness of methods that detect and remove these offsets, we designed and managed the Detection of Offsets in GPS Experiment. We simulated time series that mimicked realistic GPS data consisting of a velocity component, offsets, white and flicker noises (1/f spectrum noises) composed in an additive model. The data set was made available to the GPS analysis community without revealing the offsets, and several groups conducted blind tests with a range of detection approaches. The results show that, at present, manual methods (where offsets are hand picked) almost always give better results than automated or semi-automated methods (two automated methods give quite similar velocity bias as the best manual solutions). For instance, the fifth percentile range (5% to 95%) in velocity bias for automated approaches is equal to 4.2 mm/year (most commonly  $\pm 0.4$  mm/yr from the truth), whereas it is equal to 1.8 mm/yr for the manual solutions (most commonly 0.2 mm/yr from the truth). The magnitude of offsets detectable by manual solutions is smaller than for automated solutions, with the smallest detectable offset for the best manual and automatic solutions equal to 5 mm and 8 mm, respectively. Assuming the simulated time series noise levels are representative of real GPS time series, robust geophysical interpretation of individual site velocities lower than 0.2–0.4 mm/yr is therefore certainly not robust, although a limit of nearer 1 mm/yr would be a more conservative choice. Further work to improve offset detection in GPS coordinates time series is required before we can routinely interpret sub-mm/yr velocities for single GPS stations.

**Citation:** Gazeaux, J., et al. (2013), Detecting offsets in GPS time series: first results from the detection of offsets in GPS experiment, *J. Geophys. Res. Solid Earth*, 118, 2397–2407, doi:10.1002/jgrb.50152.

<sup>1</sup>School of Civil Engineering and Geosciences, Newcastle University, Newcastle, UK.

<sup>2</sup>National Oceanography Centre, Liverpool, UK.

<sup>3</sup>School of Geography and Environmental Studies, University of Tasmania, Hobart, Tas, Australia.

<sup>4</sup>CIIMAR/CIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal.

<sup>5</sup>Astronomical Institute, University of Bern, Bern, Switzerland.

<sup>6</sup>National Geospatial Reference Systems, Earth Monitoring Group, Canberra, ACT, Australia.

<sup>7</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.

<sup>8</sup>Politecnico di Torino, Torino, Italy.

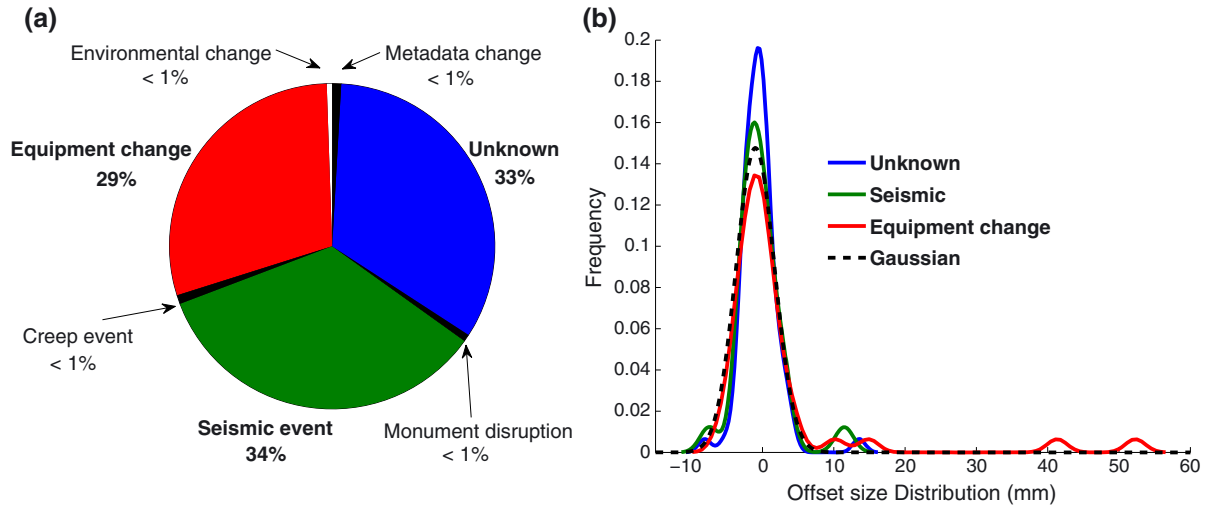
<sup>9</sup>Geophysics Laboratory, University of Luxembourg, Luxembourg.

Corresponding author: Julien Gazeaux, School of Civil Engineering and Geosciences, Newcastle University, Newcastle, UK. (julien.gazeaux@ncl.ac.uk)

## 1. Introduction

[2] Since the 1980s, GPS receivers have been established at a variety of geophysical sites to measure positions and velocities of Earth's surface. As data analysis approaches have improved, the time series have achieved increasingly higher precision [Santamaria-Gomez *et al.*, 2011]. However, further improvements are necessary in order to measure small geophysical signals or test competing models. For example, intra-plate deformations may be as small as a few tenths of a millimeter per year requiring precision below 0.1 mm/yr [e.g., Frankel *et al.*, 2011] and tide gauge vertical land movements need to be obtained with a precision and accuracy of around 0.1–0.2 mm/yr [Wöppelmann *et al.*, 2009] in order not to degrade measurements of sea level change. In such cases, even small errors in the GPS coordinate time series may be important.

[3] However, GPS coordinates time series remain disrupted by offsets occurring at times that are known (e.g., documented equipment changes) or unknown, and with



**Figure 1.** SOPAC offset (a) description and (b) magnitude distribution since 1995 over 340 sites (560 offsets).

magnitudes that are at best known imprecisely. Offsets in coordinates time series are defined as a sharp change of the mean resulting in a long-lasting effect on parameters, such as velocity estimation. Depending on their locations in the time series, undetected offsets may have a detrimental effect on velocity estimation. For example, when estimating uplift rates in East Antarctica, *Thomas et al.* [2011] reported velocities approximately 2.1 mm/yr lower than *Argus et al.* [2011], leading to very different interpretations of the data. *Thomas et al.* [2011] suggest that about 50% of the difference was due to differences in handling offsets. Both approaches are described in their respective auxiliary material. This velocity difference notes the difficulty in detecting offsets robustly.

[4] As the length of time series increases the number of offsets is likely to increase and the cumulative effect of even small offsets can significantly alter position and velocity estimates. The detection of offsets is therefore an important challenge when attempting to obtain an accurate understanding of Earth surface deformation.

[5] Offset detection, also known as data segmentation or homogenization, is a problem investigated in a large number of scientific studies. These include climate/meteorology [*Beaulieu et al.*, 2008; *Gazeaux et al.*, 2011], biostatistics research [*Olshen et al.*, 2004], image processing [*Pham et al.*, 2000], and quantitative marketing [*Fong and DeSarbo*, 2007]. Some studies have also been dedicated to the importance of offsets in GPS studies [*Williams*, 2003a] and their detection [*Khodabandeh et al.*, 2012; *Williams*, 2003a; *Borghi et al.*, 2012; *Vitti*, 2012]. *Williams* [2003b] notably highlights the role of offsets in velocity estimation of GPS time series as well as the impact of the position and magnitude of the offsets in the time series.

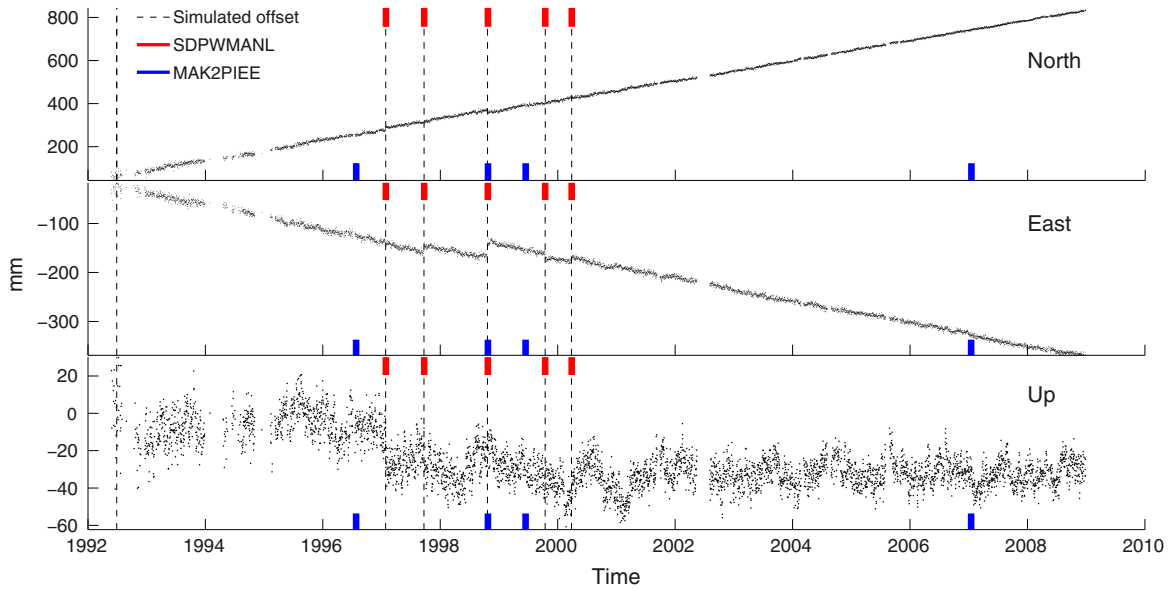
[6] As shown in Figure 1a, according to the SOPAC archive (the Scripps Orbit and Permanent Array Center, <http://sopac.ucsd.edu/>), around two thirds of the offsets have understood reasons (either equipment changes or seismic activity), with metadata available to determine their timings and hence allow estimation of their magnitudes. However, the last third of offsets are due to unknown reasons; these

are unknown to the analyst and have to be detected by some post-processing or possibly pre-processing approach. Figure 1b shows the size distribution of offsets from Figure 1a, taking all coordinate components together. From these data, all types of offsets have a symmetric distribution (i.e., symmetric around zero). Assuming that SOPAC analysts have correctly identified the majority of offsets, the main differences lay in the variance, which is higher for “equipment change” and “seismic event” offsets than for “unknown” offsets. As such, “unknown” type offsets tend to be smaller than those of other types making them especially difficult to detect. Offset detection methods are therefore essential to the removal of spurious effects of accumulated unknown events and to accurately estimate relevant site parameters such as position and velocity.

[7] In this paper, GPS-derived time series are simulated and are then examined by different offset detection methods. Realistic characteristics of the time series such as velocity, noise amplitude, timing, and magnitude of offsets are simulated with regards to state-of-the-art knowledge on GPS time series (based on articles such as *Williams* [2003a], *Langbein* [2008] or *Santamaria-Gomez et al.* [2011]). The time series are then blind-tested by analysts through a range of recently developed detection methods. The detected offset epochs and the consequent site velocities computed after considering them are compared to the actual simulated offset epochs and velocities. Finally, the methods are compared to each other in order to highlight the relative merits of each.

## 2. Methodology Description

[8] The Detection of Offsets in GPS Experiment (DOGEx) aims to consistently and objectively compare a range of offset detection methods applied to GPS time series analysis. To establish a known truth, we simulated three-dimensional coordinate time series containing known and realistic GPS signal, noise, offsets, and data gaps. We produce up to 18 years of simulated GPS daily coordinate time series for the three components (North, East, and Up) and for 50 idealized sites.



**Figure 2.** Example of simulated GPS daily coordinate time series. The vertical dashed black lines show dates of simulated offsets. The red and blue markers represent, respectively, the offsets detected by two different methods discussed in the text (SDPWMANL and MAK2PIEE). These methods, respectively, turn out to be the best and worst methods among those under investigation in this article. Note that, due to its location at the very beginning of the time series, the first offset occurring in 1992 was not detected by either method. The effect of offset location on velocity estimate is discussed in *Williams* [2003b].

[9] For each created site, we combine an intercept ( $a$ ), trend ( $f$ ), cycle ( $c$ ), noise ( $\epsilon$ ), gaps and offsets ( $\delta$ ) in an additive model (1) for each component (North, East, and Up) in daily time series ( $y$ ) given by

$$y(t) = a + f(t) + c(t) + \delta(t) + \epsilon(t), \quad (1)$$

and defined for each  $t$ , when there is no gap in data. The velocities were chosen randomly from a reasonable distribution of Earth surface velocities, and the annual and semi-annual were chosen randomly from past estimates of annual and semi-annual from real GPS data.

[10] Based on *Hosking* [1981], simulated noise characteristics ( $\epsilon$  in equation (1)) are based on those present in the state-of-the-art GPS reprocessing solutions using a white plus flicker noise model [see *Williams*, 2003a]. The noise is not necessarily time-constant at each site [cf. *Langbein*, 2008]. This time-dependent property of the noise allows the simulation of the decreasing GPS data uncertainty over the decades thanks to instrument and data analysis improvements [see *Santamaria-Gomez et al.*, 2011]. Note that the definition of  $\epsilon$  does not allow for simulation of outliers, and hence they are not taken into account in this initial DOGE.

[11] Offset time series ( $\delta$ ) are generated randomly over time and dates of the offsets follow a binomial distribution. The offset occurs in all three components and its magnitude varies according to component. Offsets are modeled as a stepwise signal with magnitudes changing for every offset, or remaining constant when no offset occurs. Amplitudes of the offsets are modeled as a symmetric Pareto distribution (see *DeGroot* [1970] for details on the Pareto distribution). That is a Pareto distribution multiplied by  $\pm 1$  with probability equal to  $\frac{1}{2}$ . While the difference from a Gaussian distribution is generally quite small, the adopted distribution

allows for a better representation of the smallest magnitude offsets that would not be well represented via a Gaussian distribution. Using a Pareto distribution results in the frequency of very large offsets being reduced (offsets as big as 42 mm or 52 mm shown in Figure 1b), but we do not consider this as a weakness as such offsets are easier to detect than very small ones.

[12] Finally, gaps are created at random times with random lengths following a Zeta distribution with parameter  $s = 2.0$  (see *Lin and Hu* [2001] for details on Zeta distribution). The periods between gaps follow a Zipf-Mandelbrot distribution [*Mouillot and Lepretre*, 2000]; during these periods, the data are continuous. Gaps being only 1 epoch long are given an increased probability to model short events such as instrument change.

[13] An example of a simulated time series for one site is represented in Figure 2. The three components (North, East, and Up) are displayed and simulated offsets are highlighted by dashed vertical black lines.

[14] Once the time series were simulated, the experiment was announced through an open call to the GPS community for analysts to submit solutions. The request to analysts was to estimate as accurately as possible the time and magnitude of offset occurrences and determine the three components of the velocities of each site from the simulated time series described above. Fifty simulated GPS site time series were tested through a range of commonly used detection methods often modified to suit GPS time series in some way. Both manual and automatic solutions were requested and no information about the sites, other than the three-component time series, was provided to the analysts. The exact velocity, type and temporal variation of noise, number of offsets, and offset timing are known but were not revealed to solution providers.

### 3. Tested Solutions

[15] A total of 25 solutions were submitted, some of which represent variants of the same solution strategy. Some solutions were provided after preliminary results. These results revealed how some solutions performed in terms of true or false offset detection but did not reveal offset epochs or the true velocity. For solutions provided later in the experiment, therefore, the experiment was no longer entirely blind as solutions providers could learn from these results how the methods were performing. For example, AIUBCOD2-3 and ULGLFD02-3 are adapted versions of initial versions AIUBCOD1 and ULGLFD01, respectively.

[16] Solutions can be split within two groups: a group of manual solutions and a group of automated or semi-automated methods. All solutions provided epochs, estimates of offset magnitudes, and three-component velocities for each site.

[17] In Figure 2, outputs of two methods are shown as an illustration of performance differences. We notice for example that the “SDPWMANL” (red dashes above) manual solution gives better results than the “MAK2PIEE” (blue dashes below) automated method in terms of offset timing detection.

[18] In the following, we describe the detection methods used in the article and give details of solution methodologies and any assumptions made. Note that the CPU-time required by each automated solution is not discussed in the article. Indeed, it has not been an issue because all solutions require less than a few tens of seconds to run for each time series. It is also worth specifying that all approaches work in the case of data gaps, although we do not compare offset detection correctness between time series where gaps do and do not occur.

#### 3.1. Manual Solutions

[19] The first group of solutions consists of individual GPS experts providing solutions they obtained manually on a site-by-site basis. No automated or semi-automated method were used, but experts were asked to graphically detect offsets. GPS-specialized software such as Tsview (see *Herring* [2003] for more details) was used, which allowed the removal of annual and semi-annual signals and, in the case of Tsview, data averaging. Velocity rates were estimated using common approaches such as Maximum Likelihood [*Le Cam*, 1990]. In the article, these methods are referenced as BOSM\_MLE, EJP\_MANL, NOCLMANL, SDPWMANL, and ULGLM001.

[20] Inspired by *Bos et al.* [2008], BOSM\_MLE uses the Maximum Likelihood Estimation method, with a standard power law plus a white noise model for estimating the velocities. Gaps in the time series were filled using simple linear interpolation and an annual signal was considered. Offsets were detected by visual inspection of the difference between the raw time series and the estimated signal. Offsets were added in an iterative process until the residual plots looked free from offset.

[21] ULGLM001 uses a “subjective and liberal approach” where anything that graphically could be identified as an offset was taken to be an offset. Offsets identified in one coordinate component were also included in other components. Finally, standard deviation ( $\sigma$ ) was calculated and offsets

were disregarded when their magnitudes were smaller than  $3\sigma$  of its uncertainty.

[22] The other handpicked solutions (EJP\_MANL, NOCLMANL, and SDPWMANL) use a simple linear regression (with or without annual or semi-annual component) with independent Gaussian noise assumption. These three solutions used the Tsview software allowing interactive picking of offsets after the removal of combination of a linear trend and annual and semi annual signals.

#### 3.2. Automated Solutions

##### 3.2.1. Picard and Lavielle Solutions

[23] A subset group of solutions called MAK1PIXX or MAK2PIXX uses a likelihood maximization approach with different penalty functions, either under constant variance assumption over time (homoscedasticity, denoted by an “O” as the second last letter) or under varying variance assumption over time (heteroscedasticity, denoted by an “E” as the second last letter). Penalized likelihood refers to the following equation,

$$\mathcal{L}_K = \mathcal{L} - \beta \text{pen}(K), \quad (2)$$

where  $\mathcal{L}$  represents the initial likelihood of the model, and the penalty function,  $\text{pen}$  is a function of the number of offsets,  $K$ . The penalty function allows the optimal estimation of the number of offsets. The penalty increases with the increasing number of offsets, and thereby prevents over segmentation of the time series during the likelihood maximization process. Based on *Picard et al.* [2005], the  $\text{pen}$  function of equation (2) is either based on *Lavielle* [2005] (in this case the last letter of the solution name is “A”) or on *Lebarbier* [2005] (“E” as last letter name). Both penalizing functions are based on *Birgé and Massart* [2001], but *Lavielle* [2005] uses the additional assumption that the number of change points is small compare to the length of the series. This assumption allows for using the asymptotic version of *Picard et al.* [2005] function. The heteroscedasticity assumption is relevant if one wants to model the decreasing uncertainty on actual GPS data over decades [c.f. *Santamaria-Gomez et al.*, 2011]. Dynamic programming is used to optimize the speed of the algorithm. For instance, MAK1PIAO method uses the Lavielle penalty function under homoscedastic assumption.

##### 3.2.2. GA Solution

[24] The GA (named after Geoscience Australia agency) offset detection algorithm is based on a moving filter. For each point in the time series,  $t$ , two sets of  $n$  data sets are selected prior to and after this epoch. A least squares linear trend is fitted to the two data sets. The residual is interpolated at  $t$  based on the linear trend from the first data set. The difference between the interpolated value and actual value is designated as  $d1$ . The same procedure is repeated at  $t$ , using the second data set to yield a value  $d2$ . The value  $d$  is adopted as the larger of  $d1$  and  $d2$  and is considered an outlier if (1) it is greater than 3 times the larger of the standard deviation of the two data sets, or (2) it is detected as an outlier based on students  $t$ -distribution, carried out at the 95% level of statistical significance. The values of  $d$ , which are above either of these thresholds, are tracked and the time when it achieves a maximum is identified as a candidate offset point. When the point,  $t$ , is close to the beginning or end of the data set,  $n$  is reduced to however many data points are available. As

the sample size gets reduced at the start and end of the full data set, the outlier test becomes less reliable as it becomes sensitive to noisy data. Thus, the point  $t$  in the moving filter begins and ends a few points before and after the final and initial epochs, respectively.

### 3.2.3. MAK2CS3D Solution

[25] This method consists of the use of a cumulative sum to find the offsets and remove them. This approach removes decreasing size offsets until a size threshold is reached, with each iteration composed of a series of steps:

[26] 1. Compute a 3-D displacement vector from the three components of the data at each site.

[27] 2. Compute the cumulative sum and detrend.

[28] 3. Identify peaks in first differences of the series  $\geq 10$  mm.

[29] 3. Remove the largest offsets (by estimating its magnitude in the original time series) and then iterate until all offsets are found.

[30] The simple 10 mm threshold is chosen arbitrarily and designed to capture large offsets only; this value would need to be changed for time series with very different noise characteristics. This method does not attempt to find small offsets ( $\leq 10$  mm).

### 3.2.4. MRPCV1 Solution

[31] In this solution, GPS time series are modeled as stochastic process plus a step function that represents the time series offsets. The offsets detection is based on a hypothesis test that assumes as null hypothesis  $H_0$  that the time series do not have any offset. This hypothesis is tested against a certain number of alternative hypotheses  $H_A$ , with a jump in a given epoch. An alternative hypothesis can be formulated for each observation epoch or for candidate epochs only. The adequacy of the model can be verified using the ratio test, which is known to have the  $\chi^2$  distribution. After detecting the offsets, they can be estimated and removed.

[32] The presented approach focuses on four steps and is intended to detect, estimate, and remove the level shifts, performing iteratively the so-called detection, identification, and adaptation procedure (DIA), presented in *Teunissen and Kleusberg* [1998], as applied in *Perfetti* [2006].

[33] 1. The horizontal coordinates E and N are transformed to radial and tangential coordinates through principal component analysis in order to highlight offsets in the horizontal components of time series. Height coordinates are not transformed in any way. Further information in the context of deformation monitoring may be found in *Teunissen* [2006], *Teunissen and Kleusberg* [1998] and *Perfetti* [2006].

[34] 2. Detect and remove the offsets through least squares estimation and testing the null hypothesis in the absence of discontinuities against a number of alternative hypotheses in the presence of discontinuities.

[35] 3. Detect velocity changes within the time series.

[36] 4. Fit of one or more linear models to remove the trend.

[37] However, instead of assuming an a priori functional model, the station motion is represented as a discrete-time Markov process. The state vector can be designed in 3-D and it is estimated by least squares, constraining the system dynamic by setting the system noise at a low value (the process is detailed in *Roggero* [2006]). Because offsets

do not necessarily affect horizontal and vertical components similarly, the vertical component is studied separately using the same approach. This approach also makes it possible to consider documented and undocumented offsets, to predict the station coordinates in data gaps, and to correctly represent pre-seismic and post-seismic deformations or other nonlinear behaviors.

### 3.2.5. Kehagias and Fortin Solution

[38] MAK1KF99 and MAK2KF99 solutions use, respectively, version-1 and version-2 of a shifting means Hidden Markov Model approach and the use of the Expectation-Maximization (EM) algorithm [e.g., *Kehagias and Fortin* 2006]. Both models describe a Gaussian process, the mean of which shifts at random epochs and with random amplitudes. Version-1 describes a zero-mean Gaussian process impacted by a random walk series, whereas version-2 describes a succession of piecewise Gaussian processes with shifting means. Hence, the two versions behave very similarly. The main difference is that, in version-2, values of the series after offsets do not depend on the pre-offset value of the series. The use of an EM algorithm makes the method fast and particularly easy to use.

[39] This approach is based on the assumption that the time series is driven by independent Gaussian noise which is randomly disrupted every epoch with a certain probability by another Gaussian distribution. However, this method assumes that the signal is piecewise stationary which does not allow a linear trend in the time series to be taken into account. To overcome this, the first step was applied to remove the nonzero velocity from each coordinate component. Each component was treated independently.

### 3.2.6. FODITS Solutions

[40] AIUBCOD1, AIUBCOD2, and AIUBCOD3 refer to solutions based on versions 1, 2, and 3 of FODITS (Find Outliers and Discontinuities in Time Series), which is a detection tool integrated within the Bernese Global Navigation Satellite Systems software [see *Ostini et al.*, 2008].

[41] The FODITS algorithm iteratively adapts a functional model to the time series of station coordinates where all three coordinate components (North, East, and Up) are treated at the same time. The principle is to reduce, step-by-step, the discrepancy between the functional model and the time series through using a statistical test to identify the next element to be added to the functional model. This process is based on the DIA procedure. The identification step was reformulated as an absolute value of the sum of the residuals. The parameters are added to the functional model starting with the previous estimated components (velocity and cycle) compensating the largest discrepancy between the current status of the model and the time series. New parameters are added (and insignificant parameters are removed) until a certain level of agreement between the model and the time series is achieved. FODITS assumes a functional model containing offsets, velocity changes, outliers, and annual and semi-annual periodic functions.

[42] The AIUBCOD1 solution was obtained by the algorithm presented in *Ostini et al.* [2008], where the parameters were removed step-by-step from the normal equation representing the model, and where the most probable offset was sought in the whole interval of the post-fit residual time series. In order to achieve more reliable results, for AIUBCOD2, the algorithm was modified to add elements

step-by-step to the functional model, which allows the introduction of a new offset, velocity change, or periodic function at each iteration (outliers may be proposed in groups to speed up the processing, see *Ostini* [2012] for complete details). As FODITS assumes a statistical model that is free of errors (i.e., post-fit residuals are normally distributed), searching in sub-intervals allows it to overcome the weakness of searching for offsets in the presence of colored noise. The third contribution AIUBCOD3 was provided with the same algorithm used for AIUBCOD2, except that the overall level of significance which controls the sensitivity of the detection was lowered to identify more offsets.

### 3.2.7. Neyman-Pearson Solutions

[43] These methods are based on the first difference (FD) time series in order to transform offsets into outliers and thus to apply the Neyman-Pearson test at every point of the FD time series. For each  $t$ , the first difference of  $y$  is defined by  $y'(t) = y(t+1) - y(t)$ . The test assumes that all the points of  $y'$  are independently and normally distributed as the FD whitens the raw position time series (i.e., it is assumed that  $y$  has independent increments). The null hypothesis  $H_0$  is defined such that there is no outlier in the FD time series, which corresponds to no offset in the raw position time series at the current epoch; the alternative hypothesis  $H_1$  is that there is an outlier at the current epoch, indicating a potential offset (see *Teunissen* [2006] for more details on testing theory). The unique feature of this method is that its efficiency is independent of the offset position, a current issue in automatic offset detection with some algorithms [*Williams*, 2003a].

[44] The first solution (ULGLFD01) runs the FD over the time series producing the signal  $y'$ , then computes the variance  $\sigma$  within fixed yearly intervals (aka local intervals). Thus, the Gaussian likelihood of  $H_1$ , i.e.,  $\mathcal{L}(\mu, \sigma|y')$ , is computed, with  $\mu \in [-2y'_{\max}, 2y'_{\max}]$  the estimated size of the prospective offset, and  $y'_{\max}$  the maximum value of the FD time series. Then, the ratio  $\mathcal{R}$  between  $\mathcal{L}(\mu, \sigma|y')$  and the likelihood of  $H_0$ , i.e.,  $\mathcal{L}(0, \sigma|y')$ , is obtained for each point  $t$  of the FD time series. If at  $t$ ,  $\mathcal{R} > 1$  and  $\mu > 3\sigma$ , that point is considered an offset of estimated size  $\mu$  in the time series at the  $3\sigma$  confidence level. Gaps and outliers within the raw position time series were not considered, and this is likely to give rise to a large number of false positives.

[45] In ULGLFD02 and 03, any outlier stemming from a gap in the raw position time series was considered a wrong detection and the threshold was set up at  $\mu > 4\sigma$ . Then, in order to reduce the computational time, the estimated offset size for every point was calculated using the maximum value within its local interval instead of the absolute maximum value  $y'_{\max}$  of the time series. Finally, the local standard deviation  $\sigma$ , necessary to implement the Neyman Pearson test at every point, was computed within fixed yearly intervals (ULGLFD02) with running yearly intervals centered on each point to be tested (ULGLFD03). In the latter case, the central point was dropped to calculate the local standard deviation, thus its influence, should it be an offset, would not affect the detection.

[46] Any outliers in the raw position time series yields pairs of outliers in the FD time series, thereby increasing the number of wrong detection. A method to detect and remove those outlier pairs, and thus to decrease the number of wrong detection, is still work in progress.

### 3.2.8. JPL Solution

[47] The first step of the JPL solution was to remove a bias, trend, and annual and semi-annual terms using *analyze\_tseri*, a time series analysis program in the QOCA package (<http://qoca.jpl.nasa.gov>, see *Dong et al.* [1998] for details). All further operations are performed on the resulting residual series.

[48] JPL\_STP1 algorithm works separately on each component, with each point in the time series tested by the step algorithm for the existence of a jump in a mini-series of specified width centered upon the candidate point. Within this window, a line is fit with an offset at the midpoint  $t_{\text{midpt}}$ :

$$\begin{aligned} y_i &= a(t_i - t_{\text{midpt}}) + b_1 \text{ for } t_i < t_{\text{midpt}} \\ y_i &= a(t_i - t_{\text{midpt}}) + b_2 \text{ for } t_i \geq t_{\text{midpt}}, \end{aligned} \quad (3)$$

where  $y_i$  is the coordinate value in the time series at time  $t_i$ ,  $a$  is a linear velocity that is constant across the offset,  $b_1$  and  $b_2$  are the coordinate biases. The time series offset is then  $b = b_2 - b_1$ .

[49] For each candidate point, this fit is repeated for windows of varying width from  $N=10$  points to a few hundred points, under the hypothesis that a true offset's existence is independent of the window size. Note that the length of the window, in units of time, included in the estimate may not be the same on either side of  $t_{\text{midpt}}$  since the window width is defined by the number of points. The window width is not permitted to become arbitrarily large, because the probability increases that the window contains more than one offset.

[50] Each offset estimate for a given  $N$  is then tested against a separate null hypothesis estimate (i.e., no offset)

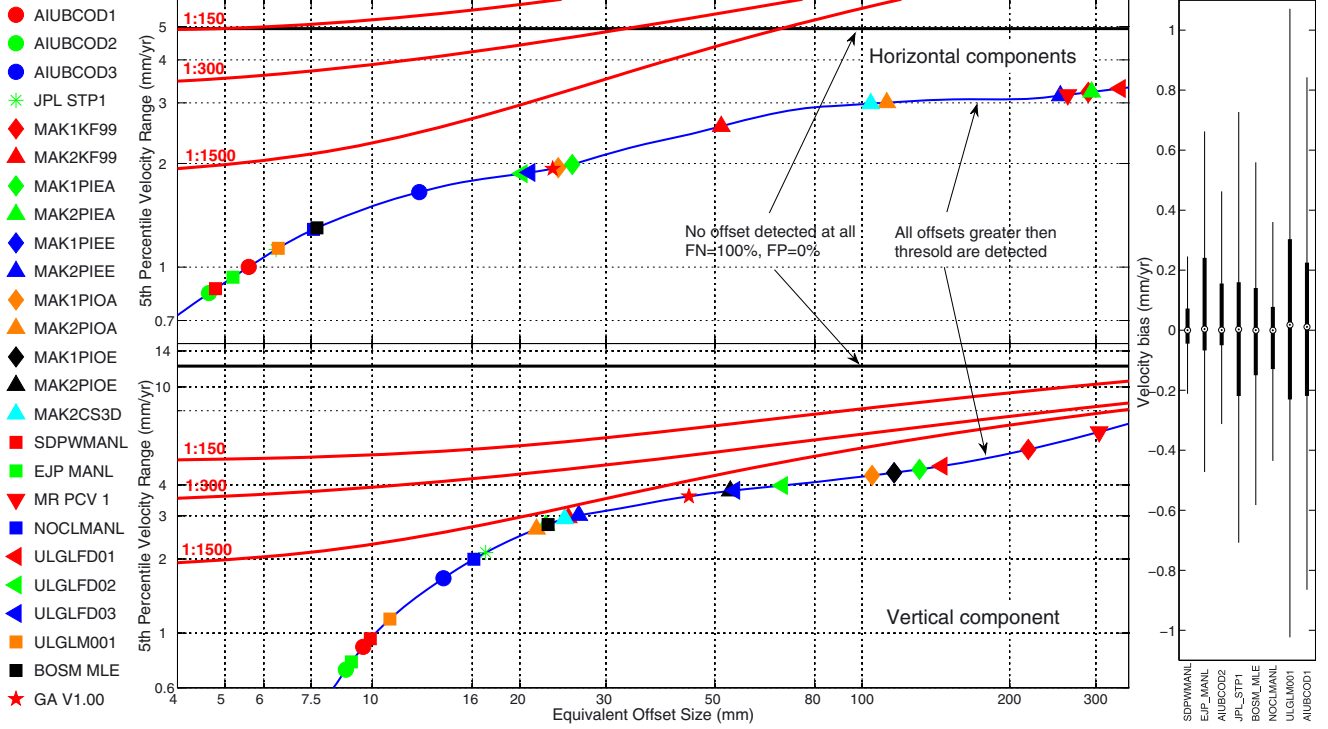
$$y_i = a(t_i - t_{\text{midpt}}) + b_1 \text{ for all } t_i \quad (4)$$

and the significance of the offset determined. The significance of the weighted RMS (WRMS) of the set of  $b$ 's is then tested using an F-test [see *Webb*, 2010]. An additional test is performed to identify outliers and reject them as offsets by comparing the WRMS fit with and without the point being tested. Candidate offsets with a minimum strength metric (a function of the estimated magnitude of the jump, the WRMS, and the probability that the variances in the with or without jump fits are not significantly different) are retained in the candidate list. The estimated jump size is further refined by another *analyze\_tseri* run, estimating an offset in both horizontal coordinate components where the step algorithm found a likely candidate in one or the other. Vertical offsets are treated separately. When multiple jumps are found at neighboring points, only the one with the most favorable statistics is retained. Finally, one last minimum magnitude (0.75 mm) criterion for a jump is applied.

## 4. Results

[51] All solutions are compared to the true offset epochs for each component and the true velocity. In order to correctly compare the solutions, we define three types of detection. A "True Positive" (TP) defines an offset that was originally simulated and also detected by a solution. A "False Positive" (FP) defines an offset that was not simulated but which has however been reported by the solution. Finally, a "False Negative" (FN) defines an offset that was





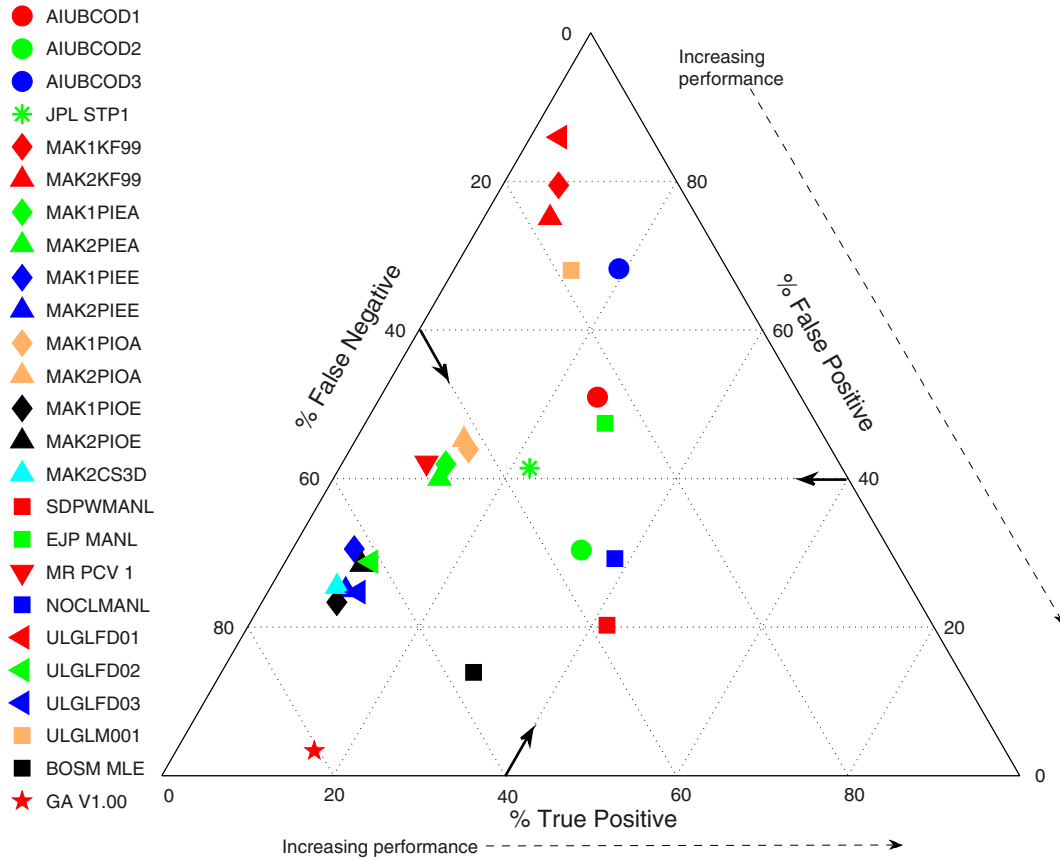
**Figure 3.** Top and bottom panels show the fifth percentile ranges (5% to 95%) in velocity absolute biases (from truth) as a metric for the performance of the different solutions. The top panel depicts the horizontal components whereas the bottom panel shows the vertical component. The red lines represent the effect of FPs, showing that over segmentation (increasing the ratio of FPs, given next to the lines) is detrimental to velocity estimation. The blue line has no FP but only offsets greater than the detection threshold are set as TPs. Manual solutions are represented by square markers. The right panel shows boxplots for the best solutions with regard to this metric.

simulated but which has not been detected. More precisely, a TP is defined as a detected offset, the date of which is 2 days either side of an actual offset. If there is more than one detected offset within this 5-day window, the extra offsets are considered as FP. A FP is also counted when an offset is detected and there is no actual offset within the 5-day window. Similarly, an FN is counted as the case when there is no offset detected within the 5-day window of an actual offset.

[52] We choose to present the fifth percentile ranges (5% to 95%) in velocity differences (from the truth) as one metric for the performance of the different solutions. That means 90% of the estimated velocities were within the specified range around the true velocity. Figure 3 ranks the solutions in order of their performance in this metric. The top panel gives the performance on the horizontal components (North and East), whereas the bottom panel describes the vertical component (Up). The lower the fifth percentile range, the more the solution is consistently closer to the truth. The blue line has no FP but only offsets greater than the detection threshold are set as TPs. The fifth percentile range estimated indicates the magnitude of TP offsets that are being detected by the different solutions. For instance, for the horizontal component, NOCLMANL solution (blue square) has an equivalent offset detection threshold of around 7.5 mm and 90% of the velocity estimates are 1.2 mm/yr distant to the simulated velocity. For the vertical component, the

equivalent offset detection threshold is equal to 16 mm and 90% of the estimates are 2.0 mm/yr distant to the simulated velocity. The effect of adding FPs to solutions (red lines) is quantified through the ratio of total offsets (TP + FP + FN) to time series length, as shown near the left axis. The greater the ratio of FPs, the worse the velocity estimation. One important consequence is that over identification of offsets can lead to velocity biases that are larger on average than ignoring all offsets (red lines go above the black line on the top panel of the figure). Note that velocities of the horizontal components (North and East) are generally better estimated than those of the vertical component (Up). This is because the horizontal components are less noisy than the vertical, which makes the offset detection easier (for those solutions who consider the different coordinate components separately). Despite the increased noise in the vertical component, the fifth percentile ranges in the velocity bias of some of the best solutions are slightly smaller on the vertical component than on the horizontal. The right panel of Figure 3 shows the boxplots of the velocity bias for the eight best solutions with respect to this metric. Horizontal and vertical solutions are considered together in this panel. On each box, the central mark is the median, the edges of the thick black lines are the 25th and 75th percentiles of velocity bias, the thin lines extend to the most extreme data point, defined as 1.5 times the 25th and 75th percentiles. For instance,





**Figure 4.** The ratio between the three performance measures of the solutions (TP, FP, and FN). Dotted arrows show the direction of a perfect solution. Manual solutions are represented by square markers.

the median of the velocity bias estimated by JPL\_STP1 is about 0.01 mm/yr, 50% of the velocity estimates are within  $-0.21$  and  $+0.18$  mm/yr of the truth and the two extreme points of the estimate distribution are  $-0.71$  mm/yr and  $0.73$  mm/yr.

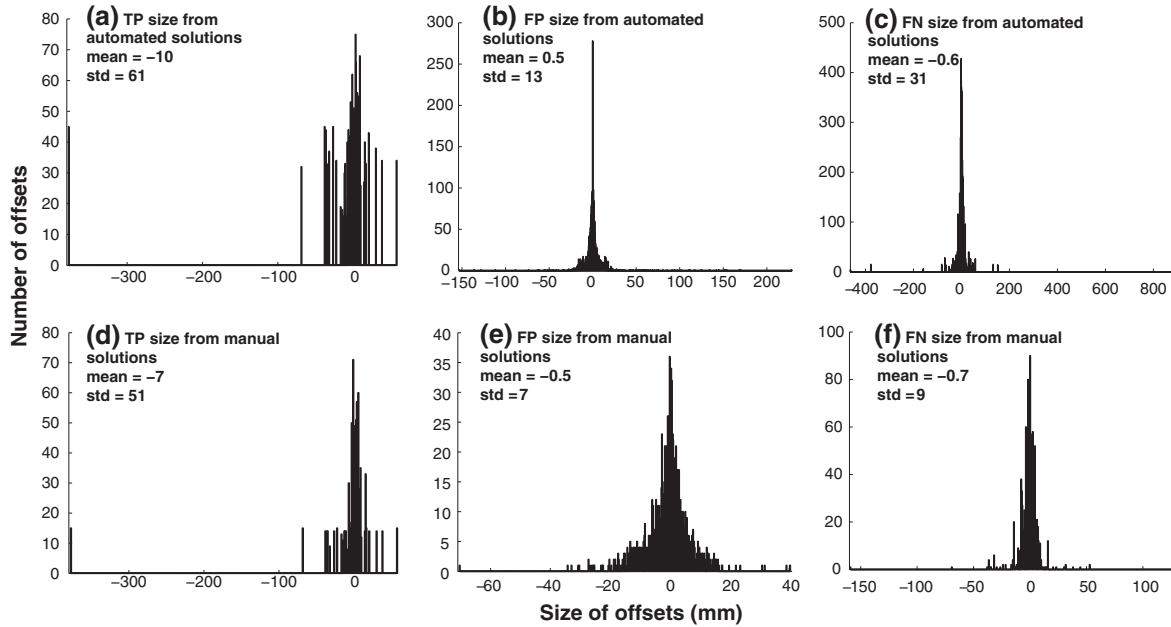
[53] Figure 4 depicts the ratios of the three variables, TP, FP, and FN by their position in an equilateral triangle. The plot highlights the trade off between the three. Over segmentation leads to a higher FP percentage, for example, the MAK[1,2]KF99 solution. The handpicked solutions have a low FN percentage but a higher FP rate compared to JPLSTP1 and AIUBCOD2. In Figure 3, a perfect solution would appear in the bottom left of the figure, while in Figure 4 a perfect solution would appear on the bottom right corner of the triangle (see dashed lines in the figure).

[54] The overall performance of the solutions was quantified by combining the detection results (FN, TP, and FP counts) and the estimated velocities. Let us consider, somewhat arbitrarily, that good solutions have fifth percentile ranges in velocity biases smaller than 3 mm/yr. We note that for many geophysical studies, a fifth percentile range of 3 mm/yr may be several times that desired. Sorting by increasing performance, the following methods meet this requirement: AIUBCOD1, ULGLM001, NOCLMANL, BOSM\_MLE, JPL\_STP1, AIUBCOD2, EJP\_MANL, and SDPWMANL. In terms of the detection process (TP, FP, and FN rates), let us consider methods with TP rate greater than 20%, FN rate less than 40%, and FP rate less than

40%. Methods that exceed these thresholds are AIUBCOD2, NOCLMANL, and SDPWMANL. JPL\_STP1 is the second best automated method and almost meets these requirements, but its FP rate is slightly too high. Note that all these methods already meet the 3 mm/yr velocity bias requirement. Of the three best solutions, two are manual (NOCLMANL and SDPWMANL) and the other automated (AIUBCOD2, based on the FODITS algorithm). Figures 3 and 4 show that the two metrics (the first one considering velocity estimates and the second considering TP-FP-FN ratio) are consistent, and together allow the assessment of the solution performances.

[55] Figure 5 shows the distribution of offset metrics for both automated and manual solutions, showing that there is no clear difference between manual and automated TP offsets size (Figures 5a and 5d).

[56] However, the difference between the maximum and minimum FPs and FNs size is larger for the automated than for the manual solutions (compare Figures 5b and 5e, and Figures 5c and 5f, as are the variances. This shows that automated solutions tend to detect larger magnitude FPs (Figures 5b and 5e) and false negatives (Figures 5c and 5f) compared to the manual solutions. The percentage of detected offsets by manual solution is higher and FN offsets are smaller than those from automated solutions. This explains why manual solutions give, in general, more accurate solutions than automated solutions in terms of the velocity estimates.



**Figure 5.** (a–f) Offsets size difference for manual and automated solutions. The  $X$  axis range of Figures 5a–5f reflects the minimum and maximum size of the offsets.

## 5. Discussion

[57] At this point, it is important to consider the DOGE<sub>x</sub> experiment design as the way in which the solutions were requested will affect the solution outcomes. The request to those who provided solutions was to detect possible offsets in the time series. The only information they had was that these time series were supposed to mimic realistic GPS data and that there was only one velocity in each time series. Consequently, the focus was made on offset estimation and not on estimation of velocities. We are aware that this question may have induced an overestimation of the number of offsets, which tends to be confirmed by the high rate of FPs in almost every solution. The alternative request could have been to focus on the estimate on the velocity rate rather than on offsets detection. However, we justify this choice first because we think it is the less influencing question. Second, in the case of overestimation of the number of detected offset, it can be legitimately assumed that the FPs offsets would also be the smallest and consequently would have the lowest impact on the rates estimation.

[58] Overall, most solutions work well in terms of detecting the largest offsets and also in terms of TP detection. The weak point of most of the presented solutions remains the high rate of FPs. As previously pointed out by *Thomas et al.* [2011], over segmentation has the effect of reducing velocity estimates towards zero. Another explanation could be that none of the methods simultaneously detects offsets and estimates the velocity. In most of the methods under study, the trend and annual effects are removed before the offset detection is applied, meaning that offset detections (both epoch and magnitude) depend on these first estimates. This implies that if the first estimates of velocity and annual cycle are not accurate enough, this might consequently impact the estimate of offsets detection and then also the number of FPs.

[59] Uncorrected offsets were shown by *Williams* [2003b] to mimic random walk behavior, and the GPS velocity error budget could be increased to account for this. However, since we do not, by definition, know how many undetected offsets we have in our series or their size, it is difficult to estimate how much to inflate our uncertainties by. An amplitude for the random walk noise of  $1 \text{ mm}/\sqrt{\text{yr}}$  would be a conservative choice.

[60] None of the solutions currently take into account all the peculiar properties of the GPS data. For example, the consideration of the white plus flicker noise instead of Gaussian noise would prevent FNs, as this assumption models time dependent noise; BOSM\_MLE manual solution did this, which probably results in its significantly reduced FP ratio, although the TP ratio of that solution remains low and FN ratio high. The consideration of information on the dependence between components of the GPS time series for instance (i.e., multivariate approach) would also likely reduce the number of FPs. (Note that AIUBCOD<sub>x</sub> solutions do this across all components and JPL do in the horizontal components.) Finally, solutions also do not make use of metadata. This information was not available in the present DOGE<sub>x</sub> data set but will, however, be included in the DOGE<sub>x</sub> v2 data set. Metadata gives relevant indications about when a subset of offsets likely to have occurred and the type of offset can also give a priori information about the likely magnitude of the offset (as shown in Figure 1b).

[61] This paper shows that, while geophysical studies require GPS velocities with an accuracy of up to  $0.1 \text{ mm/yr}$ , the detection of offsets remains a problem that prevents us from reaching such accuracy at the single site level. Spatial data averaging, if sufficient station density exists, may help reduce regional velocity biases toward zero (Figure 1). Nevertheless, further work is required in order to accurately determine offsets in GPS coordinate time series and

produce more sophisticated metrics for classifying the solutions. Aside from the applications of individual GPS time series, the International Terrestrial Reference Frame (ITRF) must reach an accuracy of 0.1 mm/yr to meet future science requirements [Altamimi *et al.*, 2011]. This target is approximately an order of magnitude greater than is achievable today and significant advances are required. Uncorrected offsets in GPS time series incorporated in ITRF realizations will slow progress toward this goal. Techniques developed in other fields will not provide an optimal solution for GPS, at least without modification.

## 6. Conclusion

[62] We simulated GPS time series for 50 GPS sites with realistic time series characteristics, including trends, periodic signals, noise types, offsets, and data gaps, with time series duration of 18 years. A range of detection approaches were then used to detect offsets in the simulations and the relative merits of each were discussed. The key findings of this work highlight the statistically significant high performance of handpicked solutions compared to automated solutions in terms of epoch detection and a posteriori velocity estimation.

[63] However, even the best solutions that show good performance (i.e., fifth percentile velocity range smaller than 1 mm/yr, and 25th percentile velocity range smaller than 0.2 mm/yr) do not reach the accuracy of 0.1 mm/yr required by science. Therefore, geophysical interpretation of single site surface displacements smaller than 1 mm/yr must be subject to particular caution depending on the offset detection method employed.

[64] The number of continuous GPS sites is now in the many thousands globally and this expansion is expected to continue for the foreseeable future. As this occurs, and as existing GPS time series length increases further, analysts need automated approaches to interrogate the time series and aid geophysical interpretation. Given the importance of time series offsets in this context, the purpose of DOGE<sub>x</sub> is to objectively test and evaluate different offset detection approaches. A particular emphasis is on automatic solutions but manual solutions are also considered; our results show that a repeatable approach to offset detections with a similar skill to expert analysts has not yet been realized.

[65] Further work needs to be done to further reduce offset-related velocity biases. For automated solutions, a way to achieve this task is to take into account all GPS data properties, especially in terms of noise estimates and sites interdependence. Further improvements will almost certainly be found by considering metadata that provide relevant information on the epoch and the magnitude of offsets (see Figure 1b) [e.g., Ostini *et al.*, 2008]. These metadata should benefit both the manual and automated approaches. Most of these challenges will be considered in the second version of DOGE<sub>x</sub>.

[66] **Acknowledgments.** This research was supported by the Leverhulme Trust, NERC, and COST Action ES0701. We thank Olivier Mestre and Mark Tamisiea for helpful discussions and the solution providers who elected not to be named as co-authors on this paper. We acknowledge SOPAC for making their offset information available. The research of the JPL authors was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Matt King is a recipient of an Australian Research Council Future Fellowship (project number FT110100207).

## References

- Altamimi, Z., X. Collilieux, and L. Mtivier (2011), Itrf2008 : An improved solution of the International Terrestrial Reference Frame, *J. Geod.*, **85**, 457–473, doi:10.1007/s00190-011-0444-4.
- Argus, D. F., G. Blewitt, W. R. Peltier, and C. Kreemer (2011), Rise of the Ellsworth mountains and parts of the East Antarctic coast observed with GPS, *Geophys. Res. Lett.*, **38**, L16,303, doi:10.1029/2011GL048025.
- Beaulieu, C., O. Seidou, T. Ouarda, X. Zhang, G. Boulet, and A. Yagouti (2008), Intercomparison of homogenization techniques for precipitation data, *Water Resour. Res.*, **44**, W02,425, doi:10.1029/2006WR005615.
- Birgé, L., and P. Massart (2001), Gaussian model selection, *J. Eur. Math. Soc.*, **3**, 203–268, doi:10.1007/s100970100031.
- Borghini, A., L. Cannizzaro, and A. Vitti (2012), Advanced techniques for discontinuity detection in GNSS coordinate time-series. An Italian case study, in *Geodesy for Planet Earth, International Association of Geodesy Symposia*, edited by Kenyon, S., M. C. Pacino, and U. Marti, pp. 627–634, vol. 136, Springer, Berlin Heidelberg, doi:10.1007/978-3-642-20338-1\_77, (to appear in print).
- Bos, M., R. Fernandes, S. Williams, and L. Bastos (2008), Fast error analysis of continuous GPS observations, *J. Geod.*, **82**, 157–166, doi:10.1007/s00190-007-0165-X.
- DeGroot, M. H. (1970), *Optimal Statistical Decisions*, p. 41, McGraw-Hill, New York.
- Dong, D., T. Herring, and R. King (1998), Estimating regional deformation from a combination of space and terrestrial geodetic data, *J. Geod.*, **72**, 200–214.
- Fong, D., and W. DeSarbo (2007), A bayesian methodology for simultaneously detecting and estimating regime change points and variable selection in multiple regression models for marketing research, *Quant. Market. Econ.*, **5**(4), 427–453, doi:10.1007/s11129-007-9030-8.
- Frankel, K. L., J. F. Dolan, L. A. Owen, P. Ganey, and R. C. Finkel (2011), Spatial and temporal constancy of seismic strain release along an evolving segment of the Pacific-North America plate boundary, *Earth Planet. Sci. Lett.*, **304**, 565–576, doi:10.1016/j.epsl.2011.02.034.
- Gazeaux, J., E. Flaounas, P. Naveau, and A. Hannart (2011), Inferring change points and nonlinear trends in multivariate time series: Application to West African monsoon onset timings estimation, *J. Geophys. Res.*, **116**, D05,101, doi:10.1029/2010JD014723.
- Herring, T. (2003), MATLAB tools for viewing GPS velocities and time series, *GPS Solutions*, **7**(3), 194–199, doi:10.1007/s10291-003-0068-0.
- Hosking, J. R. M. (1981), Fractional differencing, *Biometrika*, **68** (1), 165–176, doi:10.1093/biomet/68.1.165.
- Kehagias, A., and V. Fortin (2006), Time series segmentation with shifting means hidden Markov models, *Nonlinear Processes Geophys.*, **13**, 339–352, doi:10.5194/npg-13-339-2006.
- Khodabandeh, A., A. Amiri-Simkooei, and M. Sharifi (2012), GPS position time-series analysis based on asymptotic normality of m-estimation, *J. Geod.*, **86**, 15–33, doi:10.1007/s00190-011-0489-4.
- Langbein, J. (2008), Noise in GPS displacement measurements from Southern California and Southern Nevada, *J. Geophys. Res.*, **113**, B05,405, doi:10.1029/2007JB005247.
- Lavielle, M. (2005), Using penalized contrasts for the change-point problem, *Signal Process.*, **85**, 1501–1510, doi:10.1016/j.sigpro.2005.01.012.
- Le Cam, L. (1990), Maximum likelihood: An introduction, *Int. Stat. Rev.*, **58**, 153–171.
- Lebarbier, E. (2005), Detecting multiple change-points in the mean of Gaussian process by model selection, *Signal Process.*, **85**, 717–736, doi:10.1016/j.sigpro.2004.11.012.
- Lin, G. D., and C.-Y. Hu (2001), The Riemann Zeta Distribution, *Bernoulli*, **7**(5), 817–828.
- Mouillot, D., and A. Lepretre (2000), Introduction of Relative Abundance Distribution (RAD) indices, estimated from the Rank-Frequency Diagrams (RFD), to assess changes in community diversity, *Environ. Monit. Assess.*, **63**, 279–295, doi:10.1023/A:1006297211561.
- Olshen, A. B., E. S. Venkatraman, R. Lucito, and M. Wigler (2004), Circular binary segmentation for the analysis of array based DNA copy number data, *Biostatistics*, **5**(4), 557–572, doi:10.1093/biostatistics/kxh008.
- Ostini, L. (2012), Analysis and quality assessment of GNSS-derived paramter time series, Ph.D. thesis, Philosophisch-naturwissenschaftlichen Fakultät of the University of Bern.
- Ostini, L., R. Dach, M. Meindl, S. Schaer, and U. Hugentobler (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, *J. Geod.*, **80**, 381–396, doi:10.1007/s00190-006-0080-6.

- Pham, D., C. Xu, and J. Prince (2000), A survey of current methods in medical image segmentation, *Annu. Rev. Biomed. Eng.*, 2, 315.
- Picard, F., M. Robin, S. Lavielle, C. Vaisse, and J. Daudin (2005), A statistical approach for array CGH data analysis, *BMC Bioinf.*, 6, 27, doi: 10.1186/1471-2105-6-27.
- Roggero, M. (2006), Kinematic GPS batch processing, a source for large sparse problems, *Hotine-Marussi Symp. Theor. Comput. Geod.*, 6, 160–165., doi:10.1007/978-3-540-74584-6\_25.
- Santamaria-Gomez, A., M. Bouin, X. Collilieux, and G. Wöppelmann (2011), Correlated errors in GPS position time series: Implications for velocity estimates, *J. Geophys. Res.*, 116, B01,405, doi: 10.1029/2010JB007701.
- Teunissen, P. (2006), *Testing Theory, an Introduction*, Series on Mathematical Geodesy and Positioning, 2000–Mathematics, 147 pp., Delft University Press, The Netherlands.
- Teunissen, P., and A. Kleusberg (1998), *Quality Control and GPS. GPS for Geodesy*, pp. 271–318, Springer-Verlag, Berlin.
- Thomas, I., et al. (2011), Widespread low rates of Antarctic glacial isostatic adjustment revealed by GPS observations, *Geophys. Res. Lett.*, 38, L22,302, doi:10.1029/2011GL049277.
- Vitti, A. (2012), SIGSEG: A tool for the detection of position and velocity discontinuities in geodetic time-series, *GPS Solutions*, 16, 405–410, doi: 10.1007/s10291-012-0257-9.
- Webb, F., (2010), A software tool for detecting offsets in geodetic time series, *Tech. Rep. NTR-47707*, NASA New Technology Report.
- Williams, S. D. P. (2003a), The effect of coloured noise on the uncertainties of rates estimated from geodetic time series, *J. Geod.*, 76, 483–494, doi: 10.1007/s00190-002-0283-4.
- Williams, S. D. P. (2003b), Offsets in global positioning system time series, *J. Geophys. Res.*, 108, 2310, doi:10.1029/2002JB002156.
- Wöppelmann, G., C. Letetrel, A. Santamaria, M.-N. Bouin, X. Collilieux, Z. Altamimi, S. D. P. Williams, and B. M. Miguez (2009), Rates of sea-level change over the past century in a geocentric reference frame, *Geophys. Res. Lett.*, 36, L12,607, doi:10.1029/2009GL038720.