Online detection of offsets in GPS time series

A case study of eruptive fissure opening at Mount Etna

Giuseppe Nunnari¹ · Flavio Cannavó²

Received: 18 November 2019 / Accepted: 25 August 2020 / Published online: 2 October 2020
© The Author(s) 2020

Abstract

This paper deals with the online offset detection in GPS time series recorded in volcanic areas. The interest for this problem lies in the fact that an offset can indicate the opening of eruptive fissures. A Change Point Detection algorithm is applied to carry out, in an *online* framework, the offset detection. Experimental results show that the algorithm is able to recognize the offset generated by the Mount Etna eruption, occurred on December 24, 2018, with a delay of about 4 samples, corresponding to 40 min, compared to the best offline detection. Furthermore, analysis of the trade-off between success and false alarms is carried out and discussed.

Keywords Volcano monitoring · Eruptive fissures · Offsets in GPS time-series

Introduction

The Global Positioning System (GPS) has become an essential tool for ground deformation monitoring in environments subject to the risks of natural disasters, such as active volcanic and tectonic areas. Fast detection of potentially hazardous events (such as earthquakes and volcanic eruptions), can be decisive for safeguarding human lives and infrastructures. However, data available in real time by GPS monitoring networks are not themselves enough for a reliable evaluation of the phenomena in progress, unless appropriate analysis tools be available too. One of the problems that prevent an effective use of GPS sub-daily solutions for real time applications, is that they are usually affected by a significant amount of noise due to different sources of uncertainty (e.g.

non-precise ephemerides). In particular, the shorter the processed period, the higher the level of noise affecting the GPS solutions. Thus, the algorithms for detecting true displacement transients must be as robust as the sampling time is lower. Due to the large number of noise sources, often not well known, GPS noise is generically modeled as a mixture of white noise, flicker noise and random walk noise (Mao et al. 1999).

Offsets are one of the components of GPS time series, sometimes considered as a source of noise, which could contribute to degrade the accuracy of GPS time series. They can be due to the equipment problems, such as antenna or receiver changes, but also to natural phenomena, such as post seismic effects of earthquakes and, in volcanic areas, to the opening of eruptive fissures. Offsets can have various sizes, from very small, which are very difficult to detect in presence of noise, to quite large, which, on the contrary, can be easily detected. In Gazeaux et al. (2013) results of a competition, launched to various research teams, with the purpose of assessing the effectiveness of methods to detect and remove offsets, are reported. In the competition, the data set, consisting of simulated GPS time series, 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 of this experiment showed that almost always manual methods, where offsets are hand-picked, gave better results than automated or semi-automated methods. However, while hand-picked methods can be considered for offline

Communicated by: H. Babaie

✉ Giuseppe Nunnari
giuseppe.nunnari@dieei.unict.it

Flavio Cannavó
flavio.cannavo@ingv.it

¹ Dipartimento di Ingegneria Elettrica Elettronica e Informatica, Università degli Studi di Catania, Viale A. Doria, 6, 95125 Catania, Italy

² Istituto Nazionale di Geofisica e Vulcanologia, Piazza Roma, 2, Catania, Italy

applications, they have no utility for monitoring purposes, where automatic approaches are mandatory.

This paper deals with the problem of online detection of offset in high rate (sub-daily) GPS time series, reporting a case study concerning a data set recorded in the Mount Etna volcanic area. The task was tackled by using the Change Point Detection (CPD) approach, already widely described in literature. A recent survey of methods for CPD can be found in Aminikhanghahi and Cook (2017). This paper is organized as follows: in “[A general model of GPS time series](#)” a general model of GPS time series is described, while in “[Change point detection](#)” the mathematical background about change point detection algorithms is reported. In “[Online detection of offsets: a case study](#)”, the results of a case study concerning online transient detection in sub-daily GPS time series at Mount Etna is reported. Finally, the conclusions are drawn in “[Conclusions](#)”.

A general model of GPS time series

In order to describe more precisely what is referred to as offset in GPS time series, it can be useful to know that GPS time series can be modeled as described in Eq. 1

$$x(t_i) = x_0 + bt_i + \sum_{j=1}^l a_j H(t_i - T_j) + \sum_{n=1}^m A_n \sin(\omega_n t_i + \varphi_n) + \eta(t_i) \quad (1)$$

where, for the epoch t_i :

- the first term x_0 is the site coordinate,
- the second term bt_i is the linear rate,
- the third term, consists of a sum of l Heaviside step functions $H(t_i - T_j)$, each having amplitude a_j .
- the fourth term consists of m periodic components, mainly the annual and semiannual signals.
- the final term, $\eta(t_i)$, is the noise component of the GPS times series.

The first four terms of this model, which represent the deterministic part, are easy to understand. As an example, consider the GPS daily solutions reported in Fig. 1, recorded on Mount Etna by the station referred to as ECPN, from January 2011 to November 2018.

Such a station belongs to the GPS monitoring network installed on Mount Etna, whose current configuration is shown in Fig. 2.

In particular, the Heaviside step functions $a_j H(t_i - T_j)$ attempt to model what are referred in this paper as offsets. A clear example of Heaviside step can be appreciated looking at the horizontal component in Fig. 2, which was due to the opening of the eruptive fissure occurred at Mount Etna on 28-Dec-2014 (Gambino et al. 2016).

While offsets are usually evident in the horizontal GPS components, periodic fluctuations, mainly due to mass loading, are particularly evident on the Up component, i.e. in direction of the gravity field (Liu et al. 2017). Thus, it is a common practice to analyze separately the horizontal and the vertical components. The noise term, which is the stochastic term, attempts to model the effects

Fig. 1 Horizontal and Up daily GPS components recorded at station ECPN

Fig. 2 The GPS measuring network managed by INGV. The eruptive fissure and the lava flow refers to the 24-Dec-2018 eruptive event, considered in this paper as case study

of several phenomena, such as propagation effects of GPS signals, multi-path etc. Several papers have been devoted in literature to characterize the noise term, based on the approaches described in Mao et al. (1999) and Williams et al. (2004). These methods, applied also recently (Birhanu et al. 2018), characterize the term $\eta(t_i)$ as a combination of white noise, flicker noise, with a spectral density proportional to $1/f$, and random walk noise, with a spectral density proportional to $1/f^2$, being f the frequency.

Change point detection

A change point represents a transition between different states in a process that generates the time series. Change point detection can be defined as the problem of choosing between two alternatives: no change occurred and the alternative hypothesis, a change occurred. CPD algorithms are traditionally classified as online or offline. Offline algorithms consider the whole data set at once and try to recognize where the change occurred. Thus, the objective, in this case is to identify all the sequence change points in batch mode. In contrast, online, or real time, algorithms run concurrently with the process they are monitoring, processing each data point as it becomes available, with the goal of detecting a change point as soon as possible, after it

occurs, ideally before the next data point arrives. In practice, no CPD algorithm operates in perfect real time because it must wait for new data before determining if a change point occurred. However, different online algorithms require different amounts of new data before a change point can be detected. Based on this observation an online algorithm, which needs at least ε samples in the new batch of data to be able to find a change, is usually denoted as ε -real time. Therefore, offline algorithms can then be viewed as ∞ -real time while the best online algorithm is 1-real time, because for every data point, it can predict whether or not a change point occurs before the new data point. Smaller ε values may lead to stronger, more prompt change point detection algorithms.

To find a change point in a time series, a global optimization approach can be employed with the following basic algorithm:

1. Choose a point and divide the signal into two sections.
2. Compute an empirical estimate of the desired statistical property for each section.
3. At each point within a section, measure how much the property deviates from the empirical estimate and at the end, add the deviation for all points.
4. Add the deviations section-to-section to find the total residual error.

5. Vary the location of the division point until the total residual error attains a minimum.

As mentioned above, the search for a change point k can be formulated as an optimization problem where the cost function $J(k)$ to minimize can be written, in the general case as

$$J(k) = \sum_{i=1}^{k-1} \Delta(x_i; \chi([x_1, \dots, x_{k-1}])) + \sum_{i=k}^N \Delta(x_i; \chi([x_k, \dots, x_N])) \quad (2)$$

where $\{x_1, x_2, \dots, x_N\}$ is the time series, χ is the chosen statistic and Δ is the deviation measurement. In particular, when χ is the mean, the cost function assumes the following form:

$$J(k) = \sum_{i=1}^{k-1} (x_i - \langle x \rangle_1^{k-1})^2 + \sum_{i=k}^N (x_i - \langle x \rangle_k^N)^2 \quad (3)$$

where the symbol $\langle \cdot \rangle$ indicates the mean operator.

Another aspect to be considered, when formulating the optimization problem, is that signals of practical interest have more than one change point. Furthermore, the number of change points K is often not known a priori. To handle these features, the cost function can be generalized as

$$J(K) = \sum_{r=0}^{K-1} \sum_{i=k_r}^{k_{r+1}-1} \Delta(x_i; \chi([x_{k_r}, \dots, x_{k_{r+1}-1}])) + \beta K \quad (4)$$

where k_0 and k_K are respectively the indexes of the first and the last sample of the signal. In expression (4) the term βK is a penalty term, linearly increasing with the number of change points K , which avoid the problem of overfitting. Indeed, in the extreme case (i.e. $\beta = 0$), $J(K)$ reaches the minimum value (i.e. 0) when every point becomes a change point (i.e. $K = N$).

The algorithm described above for a univariate time-series, can be easily extended to the case of multivariate time series, which is the case, for instance, of a data set recorded

by a GPS network. In this case the cost function is evaluated, of course, over the whole set of available time series. Recent developments concerning algorithms for CPD can be found in Soh and Chandrasekaran (2017) and Celisse et al. (2018).

The software considered in this work was implemented in MATLAB based on the CPD algorithms described in Killick et al. (2012) and Lavielle (2015). It can operate both by fixing the maximum number of change points to be detected (K), or the maximum acceptable residual error (i.e. fixing a threshold on $J(K)$). The first choice is appropriate when operating offline, since one can estimate the number of offsets, for instance by visual inspection of the time series. If this choice is performed, the indicated number of offsets will be detected and ordered by amplitude from larger to smallest. The choice of operating by threshold is instead more appropriate when an estimation of the number of offsets in the GPS time series cannot be easily performed.

As an example, we have performed the offline detection of the offset shown in Fig. 1, by considering the daily horizontal component recorded at five stations from January 2011 to November 2018. To this purpose, we have set to 1 the maximum number of change points, thus obtaining the result shown in Fig. 3. It can be interesting to observe that the time stamp associated with this offset was 29-Dec-2014, which corresponds to one sample after the starting of the eruptive event, occurred at Mount Etna on 28-Dec-2014 (Gambino et al. 2016).

Therefore, the offline CPD algorithm was able to correctly detect the offset due to the eruption event.

Online detection of offsets: a case study

To implement a warning system based on offset detection, the daily GPS time series, considered in the previous section, are not appropriate since 1 day is quite a long time for detecting very fast phenomena such as the opening of an eruptive fissure, which may onset in times of the order of a few tens of minutes or some hours. For this purpose, we have taken into account the time series obtained from

Fig. 3 Offline detection of the largest offset in the GPS displacement strength daily component

10-min kinematic GPS solutions to detect possible offsets, recorded in a time interval which include the eruption occurred on Mount Etna in December, 2018. Such event took place about 10 years after the last flank eruption, from 13 May 2008 to 6 July 2009. Between these two flank eruptions, only summit eruptions occurred, some of which were extremely explosive (see the Etna Eruption that occurred on 28 December 2014, mentioned in “[Change point detection](#)”) and leading to the birth of a new summit cone, the New South-East Crater (NSEC).

This eruptive event has been studied in Bonforte et al. (2019) and Novellis et al. (2019) by using DinSAR (Differential Interferometric Synthetic Aperture Radar) data, and in Cannavó et al. (2019) from an integrate geophysical perspective. A description of such events can be summarized as follows (AA.VV 2018). Shortly after 11:00 (hereinafter all the times are expressed in GMT), an eruptive fissure opened up at the southeastern base of the New Southeast Crater, from which a violent Strombolian activity emerged, which rapidly formed a gaseous plume, rich in dark ash. A second small eruptive fissure simultaneously opened slightly more to the North, between the New Southeast Crater and the Northeast Crater, which produced only a weak Strombolian activity, lasting a few tens of minutes. At the same time, also the Northeast Crater and the Bocca Nuova produced a continuous Strombolian activity of variable intensity. In the following two hours the eruptive fissure spread south-eastwards, exceeded the edge of the western wall of the Valle del Bove, until it reached a minimum altitude of about 2400 m above sea level (see Fig. 2).

Fig. 4 10 min sampling GPS time series recorded at the ECPN station with indication of the offset due to the 24 December 2018 Etna Eruption

Dealing with outliers

One of the main problems dealing with sub-daily GPS time series is the pervasive presence of outliers due to the quite poor geodetic quality of real time GPS positioning. As an example, the presence of outliers can be appreciated in Fig. 4, which shows the horizontal GPS components measured at one of the summit stations referred to as ECPN.

In our application, the risk, when finding outliers, is that offsets produced by the opening of eruptive fissures could be erroneously confused with outliers, thus losing useful information.

Outliers detection can be performed by using various approaches, but the most popular are based on evaluating the distance between measured samples and some statistical feature of the time series. The most popular features are:

- The median (Hampel filter (Hancong et al. 2004)). Outliers are defined as elements more than a prefixed number of times (usually three) the scaled Mean Absolute Deviation (MAD) from the median. In statistics, the MAD is a robust measure of the variability of a univariate sample of quantitative data.
- The mean. Outliers are defined as elements more than a prefixed number of times (usually 1.5) the standard deviations from the mean.
- The quartiles. Outliers are defined as elements more than 1.5 interquartile ranges above the upper quartile (75%) or below the lower quartile (25%). This method is useful when the data is not normally distributed.

Fig. 5 Finding outliers with various approaches (median, mean, quartiles, grubbs) and filling by clipping

Furthermore, outlier detection can be performed by appropriate test, such as:

- The Grubbs test, which removes one outlier per iteration based on hypothesis testing. This method assumes that the data is normally distributed (Grubbs 1950).
- The extremes Student test, an iterative method similar to the Grubbs test, but can perform better when there are multiple outliers masking each other (Rosner 1983).

As concerning the criterion for filling outliers, we have chosen to use the clip method, which replaces the higher outliers with the upper threshold value. Indeed, this criterion is easy to implement for online applications and reduces the risk of losing useful information, i.e confusing offsets related to the eruptive activity with outliers. By using this approach the lower outliers are filled with the lower threshold value.

Results obtained in detecting and filling outliers from the horizontal GPS time series, recorded at the ECPN station, by using four of the above mentioned approaches, namely computing the distance from the median, from the mean, from the quartiles and performing the Grubbs test, are shown in Fig. 5.

It can be seen that among the inter-compared approaches, the only method that preserves the offsets recorded during the 24-Dec-2018 eruption, is computing the distance from the mean and was thus adopted in this paper.

A comparison between the true time series and the one obtained by using the adopted approaches is shown in Fig. 6.

Offline detection of the offset produced by the 24 Dec, 2018 eruption

The effects of the 24 December 2018 Etna Eruption on the GPS signal, after processing for outliers, can be recognized looking at the horizontal GPS component, mainly at seven out the nineteen available recording stations, as shown in Fig. 7.

The figure shows that among the nineteen available GPS recording stations, the first and most intense offset is

Fig. 6 True and filled horizontal GPS component at the ECPN station. The clipping method was considered for filling outliers. For clarity reasons the displacement was reported log scale

Fig. 7 Horizontal displacement strength and up components at seven stations, immediately before and after the eruptive event occurred on December 24, 2018

recorded by the ECPN station. Of the remaining stations, only six others recorded the ground breaking phenomenon in various ways, and in any case with some delay, compared to the ECPN station. It could be interesting to know that there are no appreciable evidences of the occurring of the eruptive event in the Up components (see Fig. 7, at the bottom).

In order to compare the performance of the *online* CPD algorithm, which will be discussed later, we carried out the detection of the offset produced by the December 24, 2018 Etna eruption on the horizontal component, by using data recorded from August 1 up to December 31, 2018, thus operating *offline*. Results obtained after this trial are shown in Fig. 8.

It is possible to see that the CPD algorithm gives the time of detection at 24-Dec-2018 10:59, which also agrees with the data tips reported in Fig. 9.

Indeed, the figure shows that the visible displacement phenomenon started at about 10:39 of December 24 and has its maximum extension at about 11:19.

Online detection of the offset produced by the 24 Dec, 2018 eruption

For *online* detection it is important to evaluate the promptness of the CPD algorithm, compared to the *offline* detection.

To this end, we have set up the following procedure, which is based on the use of the *offline* algorithm, but in an *online* framework.

1. Consider the largest batch of data ending with the sample before the hand-picking and verify that the CPD algorithm does not detect any offset.

Fig. 8 Offline detection of the offset produced by the 24 December 2018 Etna eruption

Fig. 9 Horizontal displacement strengths at 17 stations, with data tips

2. Add to the batch of data one sample at a time until the offset is detected.
3. The number of samples supplied is assumed as the ε index, here also referred to as the delay, and associated with the batch length.
4. Repeat the procedure from step 2, after reducing the length w of the data batch by 1000 samples, subtracted from the left side (i.e. the oldest recorded data), until $w < 1000$.

Results obtained by this algorithm for the batch of data from 1000 to 16000 are shown in Fig. 10.

It is possible to see that for $w \leq 11000$ samples, we obtain $\varepsilon \leq 4$, while for larger w the delay significantly increases.

Fig. 10 Detection delay versus the batch data length

Fig. 11 Comparison between the delay of the online CPD algorithm when searching on no-filled or filled outliers data-set

In order to evaluate the benefit of removing the outliers before searching for the change points, we applied the algorithm described above, directly to the original time series, i.e. without replacing the outliers. Results are shown in Fig. 11, in comparison with the delay computed after filling outliers.

It is possible to see that even by using batch of data with length $w = 1000$ samples, we get $\varepsilon = 12$ samples. Moreover, Fig. 11 shows that the ε index sharply increases for batch length $w > 3000$. We have also experimentally observed that for $w > 10000$ the algorithm which operates on true data, i.e., without replacing outliers, is not able to report for any change point, when required to search for a maximum number of change points set to 1.

Fig. 12 Number of false alarms and delay versus the batch length

Robustness analysis

From results described above, one might be led to think that for the online detection of change points the advantage is to work with relatively small batch of data, for instance for $w \leq 2000$. Unfortunately, this is not really true. Indeed, as it will be shown below, small data batch is not a good representative of the process, as it may lead to an increase of false alarms (FAs). Here, with the term FA we mean the detection of any change point that is not related to the opening of the 24-Dec-2018 eruptive fissure, which is the only significant ground displacement event in the considered time period (Aug-Dec 2018).

The number of false alarms and the detection delay, versus the batch length are shown in Fig. 12.

It is possible to see that, while FA decreases for increasing w , the detection delay ε increases. Thus, the best w can be obtained as the trade-off between these two quantities. The figure suggests choosing w between 7000 and 8000 samples, which lead $\varepsilon \leq 4$ samples, with a rate of $FA = 4$. Since the explored data set covers a time interval of about 153 days, a level of FA equal to 4, correspond to a frequency of $4/153 = 0.026 \text{ days}^{-1}$, which could be considered acceptable for practical applications.

Conclusions

In this paper, we have explored, on an experimental basis only, given a GPS data set made available by the INGV sez. di Catania, the possibility of performing the *online* detection of the opening of eruptive fissures on Mount Etna. To this end, after several attempts, we decided to perform this task by using a Change Point Detection (CPD) algorithm. Starting from a CPD algorithm implemented in the Matlab library for *offline* detection, we have implemented an *online* CPD algorithm, through which we obtained the results reported in this paper. We have also evaluated the frequency of false alarms that can be occur, using the described algorithm for practical applications. One of the weak points of the work carried out up to now, is that the performance was evaluated based on an individual case study, as high frequency GPS data sets for similar kinds of events are not available. Therefore, one of the future developments of this work could be to apply the framework to other similar case studies, in order to improve the confidence of results and generalize the approach.

Further, since false alarms are heavily affected by the presence of outliers, margins for improvements may derive from the use of more appropriate online data filtering

algorithms, that take into account the spatial correlation of GPS time series recorded in the same area. To this end work is in progress.

Acknowledgements This work was supported by the University of Catania under the project Piano della Ricerca 2016-2018, Linea Intervento 2.

Funding Open access funding provided by Università degli Studi di Catania within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- AA.VV (2018) Etna weekly bulletin n. 52/2018. Istituto Nazionale di Geofisica e Vulcanologia 52:1–11
- Aminikhanghahi S, Cook DJ (2017) A survey of methods for time series change point detection. *Knowl Inf Syst* 51:339–367
- Birhanu Y, Williams S, Bendick R, Fisseh S (2018) Time dependence of noise characteristics in continuous gps observations from East Africa. *J Afr Earth Sci* 144:83–89
- Bonforte A, Guglielmino F, Puglisi G (2019) Large dyke intrusion and small eruption: the December 24 2018 Mt. Etna eruption imaged by Sentinel data, Terra Nova. <https://doi.org/10.1111/ter.12403>
- Cannavò F, Sciotto M, Di Grazia G (2019) An integrated geophysical approach to track magma intrusion: the 2018 Christmas eve eruption at Mount Etna. *Geophys Res Lett* 46(14):8009–8017. <https://doi.org/10.1029/2019GL083120>
- Celisse A, Marot G, Pierre-Jean M, Rigai G (2018) New efficient algorithms for multiple change-point detection with reproducing kernels. *Comput Stat Data Anal* 128:200–220
- Gambino S, Cannata A, Cannavò F, La Spina A, Palano M, Sciotto M, Spampinato L, Barberi G (2016) The unusual 28 December 2014 dike-fed paroxysm at Mount Etna: timing and mechanism from a multidisciplinary perspective. *J Geophys Res: Solid Earth* 121(3):2037–2053
- Gazeaux J, Williams S, King M, Bos M, Dach R, Deo M, Moore AW, Ostini L, Petrie E, Roggero M, Teferle F, Norman O, German W, Frank H (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
- Grubbs FE (1950) Sample criteria for testing outlying observations. *Ann Math Stat* 21(1):27–58. <https://doi.org/10.1214/aoms/1177729885>

- Hancong L, Shah S, Jiang W (2004) On-line outlier detection and data cleaning. *Comput Chem Eng* 28(9):1635–1647. <https://doi.org/10.1016/j.compchemeng.2004.01.009>
- Killick R, Fearnhead P, Eckley I (2012) Optimal detection of changepoints with linear computational cost. *J Am Stat Assoc* 107:1590–1598
- Lavielle M (2015) Using penalized contrasts for the change-point problem. *Signal Process* 85:1501–1510
- Liu B, Dai W, Liu N (2017) Extracting seasonal deformations of the Nepal himalaya region from vertical gps position time series using independent component analysis. *Adv Space Res* 60:2910–2917
- Mao A, Harrison CA, Dixon T (1999) Daily clearness index profiles cluster analysis for photovoltaic system. *J Geophys Res* 104:2797–2816
- Novellis VD, Atzori S, Luca CD, Manzo M, Valerio E, Bonano M et al (2019) Dinsar analysis and analytical modeling of mount etna displacements—the december 2018 volcanotectonic crisis. *Geophys Res Lett* 46. <https://doi.org/10.1029/2019GL082467>
- Rosner M (1983) Percentage points for a generalized ESD many-outlier procedure. *Technometrics* 25(2):165–172
- Soh YS, Chandrasekaran V (2017) High-dimensional change-point estimation—combining filtering with convex optimization. *Appl Comput Harmon Anal* 43:122–147
- Williams SP, Bock Y, Fang P, Jamason P, Nikolaidis RM, Prawirodirdjo L, Miller M, Johnson DJ (2004) Error analysis of continuous gps position time series. *J Geophys Res* 109:1–19

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.