

IMPLEMENTACIÓN Y EVALUACIÓN DE MÉTODOS NO PARAMÉTRICOS PARA DETECTAR VARIACIONES BRUSCAS EN SERIES DE TIEMPO GNSS

IMPLEMENTATION AND EVALUATION OF NON-PARAMETRIC METHODS TO DETECT ABRUPT VARIATIONS IN GNSS TIME SERIES

Micaela Alejandra Carbonetti^{1,2}, Mauricio Alfredo Gende^{1,2}

1 Facultad de Ciencias Astronómicas y Geofísicas, UNLP. Paseo del Bosque s/n. B1900FWA, La Plata, Buenos Aires, Argentina.

2 Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Calle 8 1467 B1904CMC, La Plata, Buenos Aires, Argentina
E-mail: carbonettimicaela@gmail.com

RESUMEN

Los Sistemas Globales de Navegación por Satélite (GNSS) proveen a sus usuarios de series de tiempo de coordenadas de alta precisión y permiten fácilmente vincular mediciones al Marco de Referencia Terrestre (ITRF). Asimismo dan una herramienta significativa a la comunidad científica para observar y modelar la dinámica del planeta Tierra. Para comprender mejor los fenómenos geodésicos a escala regional, existe una creciente demanda en la detección de desplazamientos milimétricos a sub-milimétricos, lo que requiere mejoras en la sensibilidad de las soluciones GNSS, y enfatizan la importancia de mantener la consistencia de los soluciones en el tiempo.

Se decidió hacer uso de métodos no paramétricos para contribuir a la localización de discontinuidades abruptas de manera automática. Estos tienen la ventaja frente a las aproximaciones funcionales, de no exigir conocer a priori el comportamiento de la señal. Se implementaron distintos algoritmos matemáticos de Cambio de Régimen (Regime Shift) y Análisis de Cambio Puntual (Change-point Analysis) para detectar offsets en las series de tiempo, llamados Promedio en Bloque, Promedio Secuencial y Suma Acumulativa. Posteriormente se hizo un análisis y comparación de los resultados obtenidos entre sí, y con un método clásico de maximización, denominado aquí estimador F, propuesto por Basseville y Nikiforov (1993).

Los distintos algoritmos fueron aplicados a series de tiempo de 21 estaciones pertenecientes a la red SIRGAS-CON, que presentaban discontinuidades conocidas de distintas magnitudes. Luego de aplicados los algoritmos se les impusieron condiciones adicionales a los resultados de manera de minimizar la cantidad de falsos positivos detectados, sin sacrificar la sensibilidad de la detección.

Todos los algoritmos propuestos fueron capaces de detectar los saltos en la gran mayoría de las estaciones analizadas. Su aplicación se vuelve más robusta al combinar las técnicas, y al cotejar para cada estación las ubicaciones de los offsets en las tres componentes. El método de Promedio Secuencial fue eficaz en un 87% de los casos analizados, mientras que el Promedio en Bloque fue exitoso en el 95% de ellos. El algoritmo de Promedio en Bloque fue el método más eficaz de los analizados para encontrar saltos en series de tiempo de coordenadas GNSS y cuantificarlos.

Palabras clave: GNSS, Detección de discontinuidades, Geodesia, Análisis de Series Temporales.

ABSTRACT

Global Navigation Satellite Systems (GNSS) provide users with high precision coordinate time series and enable measurements to be easily link to the International Terrestrial Reference Frame

(ITRF). They also give a significant tool to the scientific community to observe and model the dynamics of planet Earth. To better understand geodetic phenomena at the regional level, there is an increasing demand on detecting milimetric to sub-milimetric displacements. This situation requires improvements in the sensitivity of GNSS time series and emphasizes the importance of maintaining the solutions consistency over time.

In order to contribute to the localization of abrupt discontinuities automatically, it was decided to use non-parametric methods instead of functional approximations, since they do not require knowing the behaviour of the signal beforehand. Different mathematical algorithms of regime shift and change-point analysis were implemented to detect offset changes in the time series, called Block Average, Sequential Average and Cumulative Sum. Subsequently, an analysis of the obtained results was made, as well as a comparison with a classical method of maximization, referred here as estimator F, proposed by Basseville and Nikiforov.

The algorithms were applied to time series of 21 stations belonging to the SIRGAS-CON network. Each one of them presented known discontinuities of different magnitudes. After applying the algorithms, additional conditions were imposed on the results in order to minimize the amount of false positives detected, without sacrificing the detection sensitivity.

All the proposed algorithms were able to detect the jumps in the great majority of the analyzed stations. Its application becomes more robust when combining the techniques, and when comparing, for each station the locations of the offsets in the three components. The Sequential Average method was effective in 87% of the cases analyzed, while the Block Average was successful in 95% of them. The Average Block algorithm was the most efficient method for finding and quantifying jumps in GNSS coordinate time series.

Keywords: GNSS, Detection of discontinuities, Geodesy, Time Series Analysis.

INTRODUCCIÓN

Las series de tiempo GNSS han logrado incrementar su precisión durante la última década y son una de las herramientas fundamentales para materializar el Sistema de Referencia Terrestre a nivel global y regional, por su amplia distribución espacial y por su bajo costo de mantenimiento.

Las estaciones permanentes GNSS tienen en la geodesia un rol preponderante al monitorear la evolución de las coordenadas sobre la superficie terrestre (Bock and Melgar, 2016). Es por esto que las mismas son permanentemente evaluadas y mejoradas. A medida que la longitud de las series de tiempo aumenta, el efecto acumulativo que se produce debido a discontinuidades abruptas puede alterar significativamente las estimaciones de las coordenadas. Si no se corrigen, estas discontinuidades pueden dominar los errores en las estimaciones de la velocidad (Williams, 2003b). La precisión de las series de tiempo debe mantenerse estable para testear la competencia de los modelos geodinámicos de velocidades actuales, de forma de garantizar la consistencia del marco de referencia en el tiempo (Griffiths y Ray, 2016).

Por otra parte, la localización y caracterización de discontinuidades abruptas es fundamental para poder entender los procesos físicos que se derivan de las variaciones temporales de las coordenadas. Se han utilizado series temporales GNSS para estudiar numerosos efectos geofísicos como la deformación causada por terremoto pre, co- y post-sísmicos (Montillet et al., 2015, Bock et. Al, 1997), el ajuste isostático y el postglacial (Milne et al., 2001), el efecto de la carga hidrológica (Bevis et al., 2005, Tregoning et al., 2009), atmosférica (Petrov and Boy, 2004) y mareal (Bos et al., 2015). A modo de ejemplo, las deformaciones de la corteza debido a la carga atmosférica pueden alcanzar los 3 milímetros para la componente horizontal y son del orden de la decena de milímetros para la componente vertical (Petrov and Boy, 2004). Asegurar que las series de coordenadas GNSS se mantengan en precisiones milimétricas es una exigencia concreta de la comunidad científica para poder medir señales geofísicas pequeñas.

Hay numerosos trabajos dedicados a la localización de offsets en señales GNSS (Lavielle and Ludena, 2000; Rodionov, 2004; Perfetti, 2006; Montillet et al., 2015; Griffiths and Ray, 2016). Sin embargo, aún se utiliza con frecuencia la inspección directa de los datos de la estación para identificar discontinuidades en series de tiempo (Gazeaux et al, 2013), lo que se vuelve impráctico para redes de gran tamaño. En este trabajo buscamos analizar y comparar métodos que nos permitan encontrar automáticamente discontinuidades en las series de tiempo y analizar su sensibilidad, indicando el orden de magnitud de las discontinuidades que cada método propuesto es capaz de detectar.

Las discontinuidades en las series de coordenadas pueden atribuirse tanto a causas naturales como la ocurrencia de terremotos, así como a cambios en el instrumental o en el software utilizado para su procesamiento. Sin embargo hay alrededor de un tercio de las discontinuidades que tienen aún una causa desconocida (Gazeaux et al 2013). Detectar el salto es el primer paso para buscar su origen y eventualmente corregirlo si no es consecuencia de un desplazamiento real de la corteza.

A continuación, se expondrá la descripción de los algoritmos, el análisis y la cuantificación de los saltos encontrados en series de tiempo GNSS, que fueron abordados por un grupo de técnicas matemáticas que se denominan de Cambio de Régimen (Regime Shift) y Análisis de Cambio Puntual (Change-point Analysis).

DATOS UTILIZADOS

La fuente de información utilizada en el siguiente análisis corresponde a las series de coordenadas provistas por la red del Sistema de Referencia Geocéntrico para las Américas (SIRGAS). La definición de SIRGAS es idéntica a la del Sistema Internacional de Referencia Terrestre y su realización es una densificación regional del Marco Internacional de Referencia Terrestre, en América Latina y El Caribe (Sánchez et al., 2013).

En la actualidad SIRGAS es materializado por una red de cerca de 400 estaciones GNSS de funcionamiento continuo con coordenadas de alta precisión, y sus variaciones a través del tiempo están dadas por un modelo de velocidades lineales a trozos. Esta realización recibe el nombre SIRGAS de Operación Continua (SIRGAS-CON) y es calculada semanalmente por los centros de procesamiento y combinación de SIRGAS. Las coordenadas semanales de las estaciones SIRGAS-CON se refieren a la época de observación y al mismo marco de referencia utilizado por el IGS para el cálculo de las órbitas finales de los satélites GNSS.

En el presente trabajo se tomaron como series de tiempo a las soluciones semanales ajustadas de estaciones pertenecientes a la red SIRGAS-CON, en las coordenadas de latitud, longitud y altura. La extensión temporal de las mismas es variable, pero en todos los casos las mediciones están acotadas entre la semana GPS 1000 y la 1650. Las soluciones semanales semilibres calculadas hasta la semana GPS 1400 (noviembre de 2006) han sido reprocesadas utilizando el IGS05 como marco de referencia. Las soluciones semanales hasta la semana 1650 (comprendidas entre el 5 de noviembre de 2006 y el 16 de abril de 2011) se refieren directamente al IGS05.

Siendo el objetivo principal implementar y comparar distintas técnicas matemáticas para identificar discontinuidades en las series de datos GNSS, se les exigió a las mismas que presentaran saltos comprobados, aunque de magnitudes variables, en una o más de sus componentes. Asimismo, la longitud de las series debía superar los cuatro años de datos, y en el caso de presentar gaps, estos no debían superar el 20% de la cantidad de muestras de la serie. Este conjunto de criterios redujo el número de estaciones de la red SIRGAS-CON a 21, cuya distribución espacial está representada en la Figura 1.

DESCRIPCIÓN DE LOS ALGORITMOS IMPLEMENTADOS

Las posiciones de las estaciones GNSS no son constantes sino que se ven modificadas por una serie de fenómenos físicos de diferentes magnitudes, tanto estacionales como esporádicas.

Durante el procesamiento de los datos, los centros de cómputo SIRGAS corrigen a las mediciones por la respuesta elástica de la corteza terrestre a las mareas oceánicas y atmosféricas, y aplican correcciones de los centros de fase de las antenas publicados por el IGS, por lo que las series de tiempo se ven libres de estas perturbaciones (Sánchez et al., 2013). Sin embargo, el efecto sobre la corteza debido a la carga por redistribución de masas de agua, así como por variaciones atmosféricas no mareales provenientes por ejemplo de cambios de temperatura, condiciones meteorológicas, cambios bruscos del relieve subyacente, entre otros, están aún presentes en las series de tiempo. Asimismo, las series de coordenadas se ven modificadas por acontecimientos abruptos como terremotos, cambios en el instrumental, etc. La superposición de estos eventos hace que no sea posible conocer a-priori el comportamiento de la serie, tanto en las coordenadas horizontales como en la vertical, a pesar que esta última se caracteriza por una marcada estacionalidad en la gran mayoría de las estaciones analizadas.

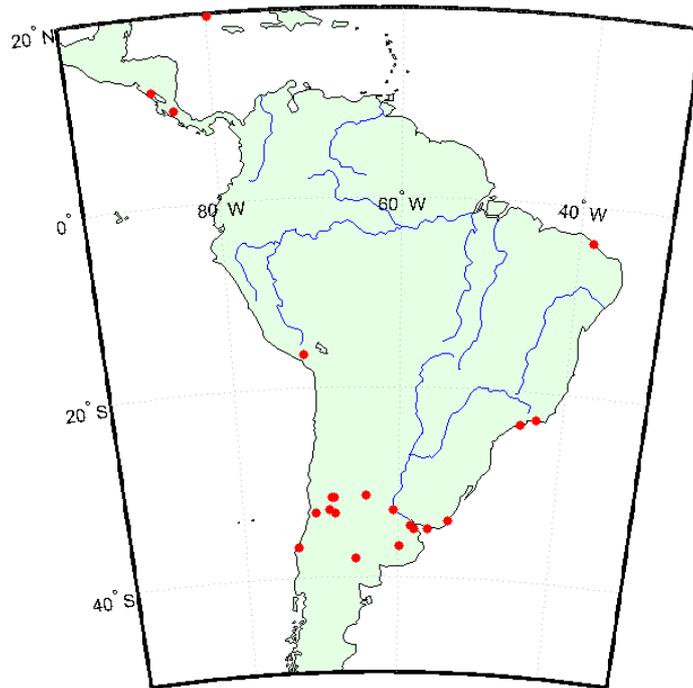


Figura 1 Distribución espacial de las 21 estaciones SIRGAS-CON utilizadas.
Figure 1 Spatial distribution of 21 SIRGAS-CON stations.

Es por esta razón que se propuso la búsqueda de discontinuidades abruptas mediante algoritmos que no requieran hacer un ajuste funcional o paramétrico de las series de tiempo GNSS. En este sentido, abordar esta problemática con métodos no paramétricos permite la detección de saltos a partir de estimadores estadísticos que se derivan de valores provistos por la misma serie, casi exclusivamente. El único parámetro externo que se debe incorporar es la longitud de las ventanas elegida para el desarrollo del algoritmo, cuyo valor se puede optimizar haciendo pruebas sucesivas. Por lo tanto, la principal ventaja sobre los métodos de ajuste funcional es que no se requiere hacer una suposición de la forma de la serie de coordenadas para detectar las discontinuidades.

En el análisis estadístico, la detección de puntos de cambio intenta identificar tiempos en los que cambia la distribución de probabilidad de una serie temporal (Taylor, 2000). En general, el problema se refiere tanto a ser capaces de detectar si ha ocurrido un cambio o no, como a identificar el tiempo de ocurrencia.

Se buscó detectar saltos aplicando algoritmos estadísticos de cambio de régimen, utilizando tanto ventanas fijas en un caso, como un análisis secuencial en otro. Dichos algoritmos involucran la comparación de valores medios de ventanas adyacentes que se desplazan sobre los datos. De acuerdo al método las mismas se solapan o no. La desventaja que presentan es que son susceptibles al tamaño de ventana elegida para efectuar los cálculos. Cualquier método que busque discontinuidades debe hacer una relación de compromiso entre la tasa de ocurrencias de falsas alarmas y asegurarse que toda discontinuidad sea encontrada. Es por esta razón que la búsqueda de la longitud de la ventana se vuelve un problema crítico.

Posteriormente se utilizó un análisis de cambio puntual para analizar si los resultados coincidían, eliminar los falsos positivos que se pudieran haber encontrado, y dar una medida de la probabilidad de que el cambio encontrado se haya producido.

Promedio en Bloque

En primer lugar se implementó el algoritmo de Promedio en Bloque. Este método encuentra saltos a partir de la comparación del valor promedio de dos ventanas adyacentes de longitud ajustable con un valor umbral elegido. En este caso, las ventanas no deben solaparse. El umbral fue determinado a partir del desvío estándar de cada una de las ventanas consideradas. Posteriormente se comparan los saltos encontrados en las tres componentes para hallar el resultado final.

Es importante notar que una vez definida la longitud de la ventana, solo podremos encontrar saltos entre los bloques, lo que conlleva a una estimación aproximada de la posición del salto, que tiene como error la cantidad de muestras de la ventana. Se puede repetir el procedimiento, cambiando el valor inicial en el que se toma la primera ventana, de manera de independizarse de ese problema. En esos casos el salto final quedará determinado por la variación más significativa dentro de cada ventana temporal.

Para hacer aún más robusto al método se recorren los datos en sentido ascendente y descendente. A los saltos encontrados se les exige como condiciones que el salto figure en el cálculo hacia adelante y hacia atrás, que en ambos cálculos la diferencia entre los valores medios sea del mismo orden y que el offset sea mayor que un 95% de las variaciones semanales. Esta serie de condiciones permitieron reducir el número de offsets que el programa registraba, de manera de minimizar la cantidad de falsos positivos.

Promedio Secuencial

En segundo lugar se implementó el algoritmo de Promedio Secuencial. Este método también compara el valor promedio de dos ventanas adyacentes, y nuevamente la elección de la longitud de la ventana tiene un rol significativo para encontrar los resultados óptimos. Sin embargo, a diferencia del algoritmo anterior, en este caso los datos se analizan secuencialmente, permitiendo que las ventanas se solapen entre sí.

El algoritmo comienza tomando una ventana inicial, a la que se le calcula su valor promedio y un estimador del desvío no sesgado. A partir de estos datos se calculan unos límites superior e inferior sumando o restándole a la media local un múltiplo del desvío. Luego, se analiza punto a punto si los elementos de la serie de tiempo exceden por exceso o por defecto los límites propuestos. (Riley, 2008)

Cuando se encuentra un posible offset, el siguiente bloque de datos es examinado para confirmar o rechazar el salto. Se toma una nueva ventana que parte desde el punto encontrado y que es de la misma longitud que la ventana inicial, y se le calcula el promedio local. Posteriormente se testea si el promedio local supera los límites impuestos. En este caso el salto es confirmado y se utiliza a este valor medio como referencia de los nuevos límites, desde esa posición en adelante.

En este algoritmo se analizan los datos en forma ascendente y descendente, con el fin de detectar posibles offsets que solo cumplen las condiciones especificadas en una sola dirección. Y a las discontinuidades encontradas también se les exige que superen al 95% de las variaciones semanales.

Suma Acumulativa

El método de Suma Acumulativa (CUSUM) es un algoritmo de Análisis de Cambio Puntual (Change-point Analysis, CPA) que se implementa en datos ordenados cronológicamente, para analizar los cambios en los valores medios, y localizar tanto visualmente como matemáticamente saltos en las series de tiempo (Basseville and Nikiforov, 1993). El método permite identificar si han ocurrido uno o más cambios abruptos, y para cada posible valor encontrado, proporcionan un nivel de confianza que indica la probabilidad de que se haya producido un salto y un intervalo de confianza, que indica cuándo puede haber ocurrido ese cambio. Si bien este método no es un sustituto de los otros algoritmos descritos anteriormente, puede proporcionar información adicional y utilizarse de manera complementaria.

Cuando se analizan conjuntos de datos especialmente grandes, este tipo de algoritmo es preferible para controlar la representación gráfica de los datos. Una desventaja del método CUSUM, es que no se puede implementar en series que aún estén recopilando información, sino que requiere que la serie esté completa. En cambio, los métodos secuenciales pueden ser utilizados a medida que se incorpora nueva información y la serie se va actualizando.

Dada una serie de datos ordenados cronológicamente x_1, x_2, \dots, x_n . En primer lugar se calcula el valor promedio de toda la serie. $(\bar{X} = 1/n \sum_{i=1}^n x_i)$. Se define a la serie Suma Acumulativa, de $n+1$ elementos, tomando como primer elemento al cero. Para cada uno de los elementos posteriores, se realiza la adición acumulativa entre el valor previo y la diferencia entre el dato i -ésimo y el valor promedio de toda la serie. La expresión general de esta función es, entonces:

$$\begin{cases} S_0 = 0 \\ S_i = S_{i-1} + (x_i - \bar{X}) \end{cases} , \text{ con } i = 1, \dots, n \quad \text{ec. (1)}$$

Al llegar al último elemento, estas diferencias se cancelan, de modo que el último elemento de la función siempre equivale a cero ($S_n = 0$). (Basseville and Nikiforov, 1993)

Al graficar la función S, se obtiene información del comportamiento de la serie original. Su pendiente punto a punto indica la importancia de cada valor respecto del promedio acumulado. Supongamos que durante un período de tiempo los valores tienden a estar por encima del promedio general. La mayoría de los valores agregados a la suma acumulada serán positivos y la suma aumentará. Por lo tanto, un segmento del gráfico con una pendiente positiva indica un período donde los valores tienden a estar por encima de la media general. Del mismo modo, un segmento con una pendiente hacia abajo indica un período de tiempo en el que los valores tienden a estar por debajo del promedio general. Un cambio abrupto en la pendiente indica la presencia del posible salto.

Para determinar un nivel de confianza del valor encontrado, se toma como estimador de la magnitud del cambio a la variable Sdiff que representa la diferencia entre el máximo y el mínimo valor que toma la función Suma Acumulativa:

$$S_{diff} = \max_{i=1, \dots, n} S_i - \min_{i=1, \dots, n} S_i \quad \text{ec. (2)}$$

Posteriormente se reordena aleatoriamente la serie de datos iniciales un número grande de veces ($N > 1000$), haciendo un muestreo sin repetición, estableciendo una base de comparación para la estimación del salto. Sobre estas nuevas series se calcula la suma acumulada, así como la diferencia entre valor máximo y mínimo para la posición de interés. Una vez hecho esto, se comparan dicha diferencia en las series reordenadas con la diferencia original Sdiff. El salto será considerado válido si la diferencia entre extremos es mayor en la serie original que en las re combinaciones, al menos en un 95% de las veces.

Estimador F

Se implementó otro método clásico, comúnmente implementado en la detección de discontinuidades, que se utiliza para series de tiempo off-line. El algoritmo consiste en el cálculo y la búsqueda de extremos de una función que depende para cada punto de la diferencia entre las medias hasta ese punto y posteriores al mismo (Williams, 2003b; Basseville and Nikiforov, 1993). Dicho estimador tiene la expresión:

$$F(p) = p \cdot (N - p) \cdot (\bar{X}_1 - \bar{X}_2)^2 \quad \text{ec. (3)}$$

, siendo

$$\bar{X}_1 = \frac{\sum_{i=1}^p X_i}{p} \quad \text{ec. (4)}$$

$$\bar{X}_2 = \frac{\sum_{i=p+1}^n X_i}{n - p} \quad \text{ec. (5)}$$

Este cálculo se basa en la idea de que una variación brusca de offset modifica el promedio antes y después del cambio. Al analizar cuán bien se ajustan los valores de la serie de tiempo a los dos promedios estimados para distintos valores de p, aquel valor que genere un mínimo o máximo de la función F (p) será candidato a ser el último punto antes del cambio.

Cotejar los valores encontrados con este procedimiento clásico sirvió como comparación externa de los otros algoritmos implementados.

RESULTADOS

En la Figura 2 se ilustra la aplicación del método Promedio en Bloque. La serie de tiempo en color azul representa a la coordenada vertical de la estación MZAC en función de la semana GPS, una vez removida su media aritmética. Se marcaron con funciones cajón los bloques utilizados en una de las iteraciones. Puede observarse que las ventanas no se solapan. Superpuesto en color verde se observa el promedio de cada bloque considerado. Cuando la diferencia entre los promedios excede el valor umbral calculado para la serie de tiempo se determina la estimación de la posición del salto, marcado en el gráfico con una línea vertical roja. Se puede apreciar claramente que para las

muestras de la ventana posterior al salto hallado, el valor promedio no pertenece a la ventana anterior.

Al aplicar el método de Promedio en Bloque para las series de tiempo analizadas se pudieron ubicar las ventanas que contenían los offsets para todas las estaciones. En la Figura 3 se muestran los resultados obtenidos para las estaciones AREQ, CONZ, MEXI, MZAC en las coordenadas Este, Norte y Altura, expresadas en milímetros. El programa determina aquellos bloques que cumplen las condiciones para que se encuentre una discontinuidad abrupta, y marca su primer elemento en el gráfico con una línea vertical verde. Luego de aplicarse las restricciones se determina la posición del salto más probable, exigiéndole que se encuentre en más de una componente. El salto determinado queda marcado con una línea punteada vertical roja.

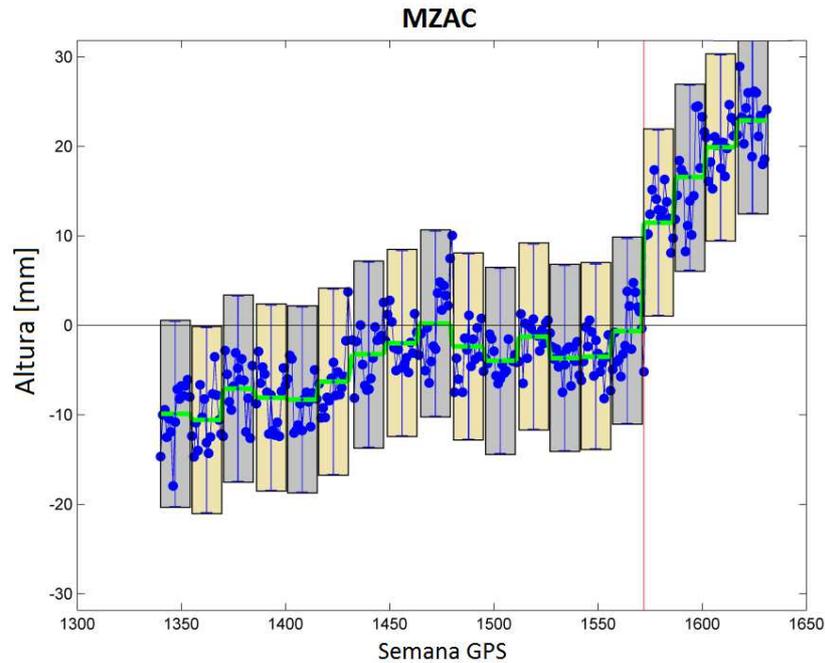


Figura 2 Aplicación del método Promedio en Bloque para la coordenada vertical sin media de la estación MZAC.
Figure 2 Implementation of the Block Average method for the vertical coordinate of station MZAC.

Posteriormente se implementó el método de Promedio Secuencial. En la Figura 4 se ilustra la aplicación del algoritmo para la coordenada Este de la estación IGM1, representado con una serie de puntos azules. En color gris se encuentra representada la primera ventana analizada, cuyo centro vertical es el promedio de la ventana, y su alto está acotado por los límites inferior y superior calculados, prolongados luego en líneas azules horizontales gruesas. Utilizando ventanas que se solapan secuencialmente, se estudian los puntos sucesivos. Cuando un punto excede dichos límites se analiza la media de la ventana que lo contiene, marcada en verde sobre la ventana local. Si la media está contenida por los límites azules, el salto es rechazado (caso marcado en el gráfico como "Posible Salto"). Si por el contrario la media excede los límites impuestos es confirmado (caso marcado en el gráfico como "Verdadero Salto"). Nuevos límites, ilustrados con líneas azules horizontales, son calculados desde ese punto en adelante.

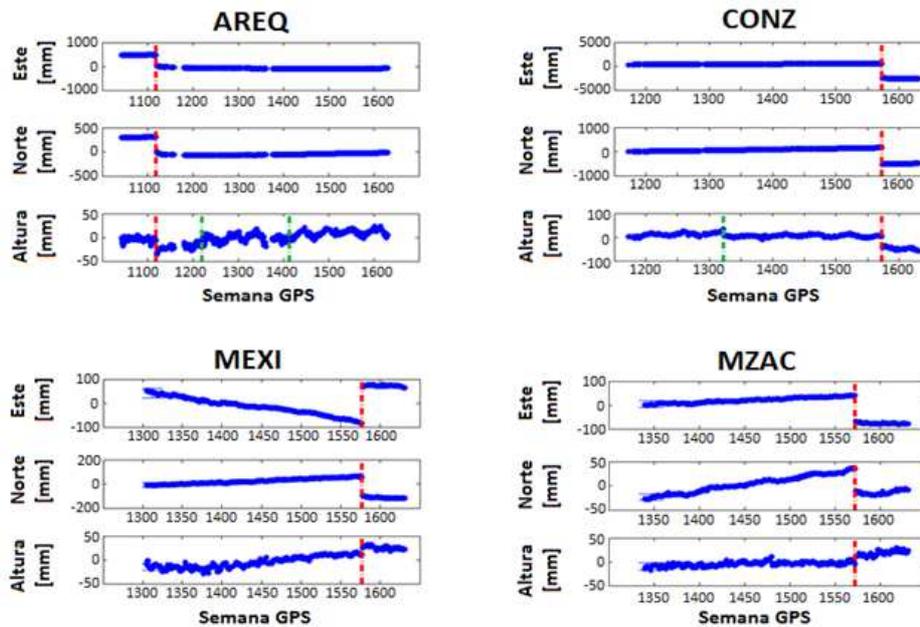


Figura 3 Resultados del método Promedio en Bloque en las coordenadas Este, Norte y Altura para las estaciones AREQ, CONZ, MEXI y MZAC.
Figure 3 Results of the Block Average method for the coordinates East, North and Height on AREQ, CONZ, MEXI and MZAC stations.

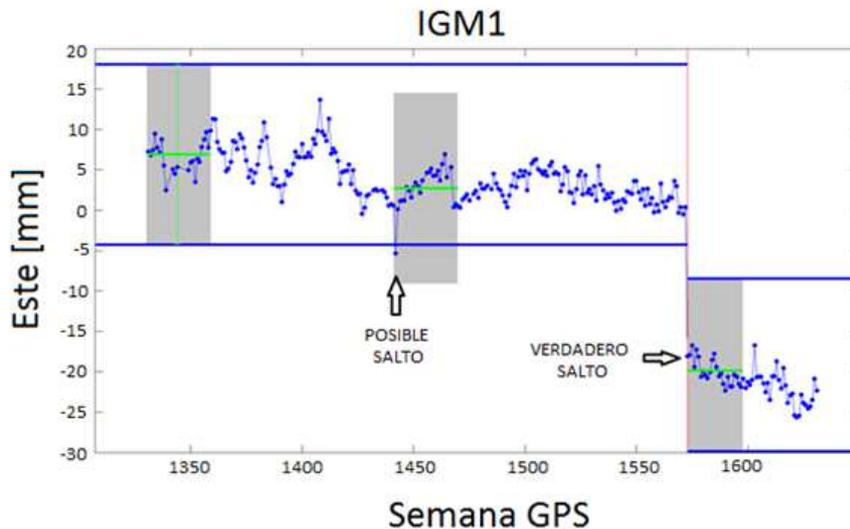


Figura 4 Aplicación del método Promedio Secuencial para la coordenada Este de la estación IGM1.
Figure 4 Implementation of the Sequential Average method for the East coordinate of the IGM1 station.

El análisis de las componentes Este, Norte y Altura se muestra en la Figura 5. Se eligieron las mismas estaciones para hacer luego un análisis comparativo entre los métodos. Las coordenadas están representadas en azul. Y se muestran en líneas verticales verdes los posibles saltos hallados, y en líneas verticales rojas los saltos verdaderos.

El método de Suma Acumulativa permite calcular y luego visualizar con facilidad donde pudo haber ocurrido un cambio abrupto de la serie de tiempo. En la Figura 6 se puede observar en color azul la función Suma Acumulativa en función de la Semana GPS, para las estaciones AREQ, CONZ, MEXI, MZAC. Se puede notar a simple vista un cambio abrupto en el comportamiento de la función, dado por un máximo o un mínimo absoluto. Superpuestas en color rojo están representadas, a modo de ejemplo, las funciones Suma Acumulada de 100 muestras reordenadas aleatoriamente de la serie

original. Esto nos permite ver a simple vista, que el comportamiento de cambio abrupto que estamos detectando en la serie Suma, es en efecto un cambio real de la serie de tiempo original.

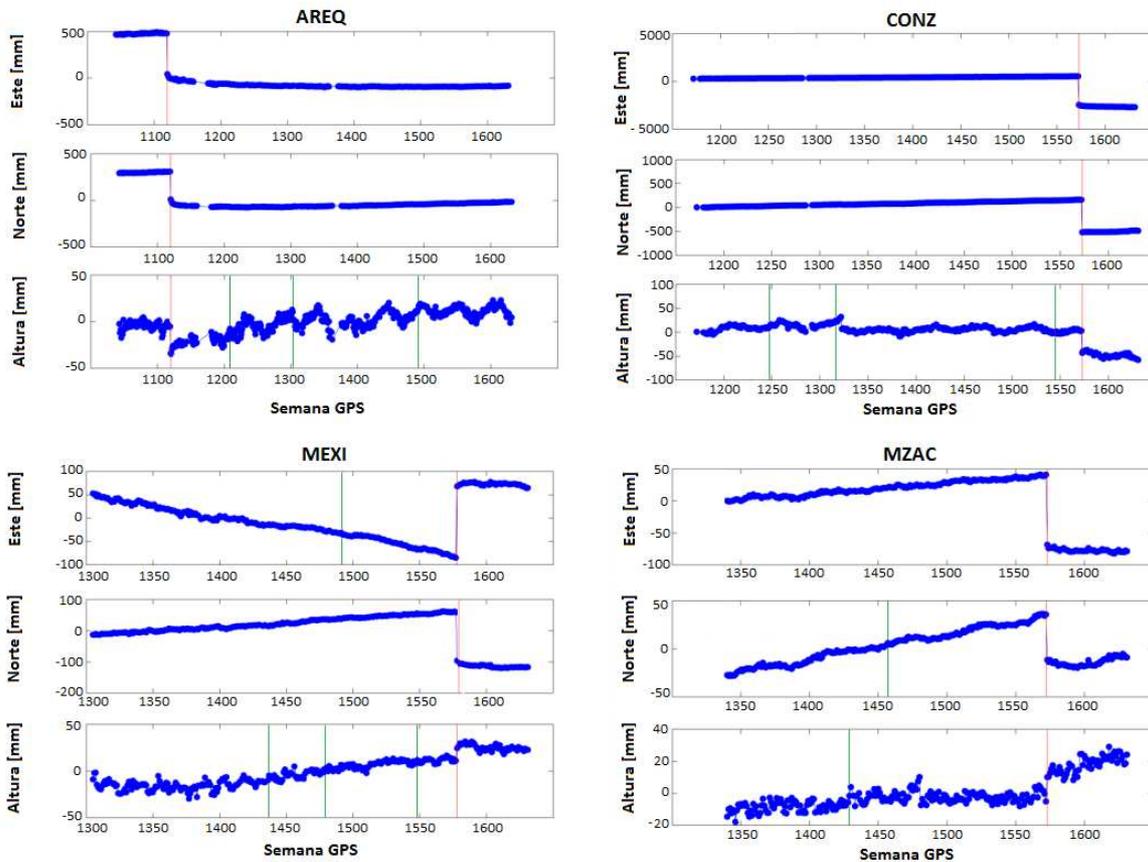


Figura 5 Resultados del método Promedio Secuencial en las coordenadas Este, Norte y Altura para las estaciones AREQ, CONZ, MEXI y MZAC.
Figure 5 Results of the Sequential Average method for the coordinates East, North and Height on AREQ, CONZ, MEXI and MZAC stations.

En la Tabla 1 se muestran para cada una de las estaciones los órdenes de magnitud de los saltos de mayor magnitud detectados en cada una de las componentes de las coordenadas con el método de Promedio en Bloque. En la última columna además se asocia a cada salto ubicado en el tiempo por la semana GPS, con su origen dado por un evento sísmico de magnitudes significativas.

- El 23 de junio de 2001 corresponde a un sismo de magnitud 8.4 en el sur de Perú.
- El 27 de febrero de 2010 corresponde a un terremoto con epicentro en el mar chileno que alcanzó una magnitud de 8,8 M_w .
- El 8 de enero de 2009 corresponde al sismo de Costa Rica que tuvo una magnitud de 6,1 M_w según el USGS.
- El 4 de abril de 2010, corresponde al Terremoto de Sierra El Mayor-Cucapah, que alcanzó una magnitud de 7,2 M_w .

Hay 4 estaciones en las que se evidencia un salto como resultado de los algoritmos, a la vez que es corroborado visualmente por los analistas, y sin embargo, estas discontinuidades no se deben a eventos sísmicos. Las mismas pueden ser atribuidas a cambios en el hardware del equipo de medición, así como a cambios en la estrategia de procesamiento de la información.

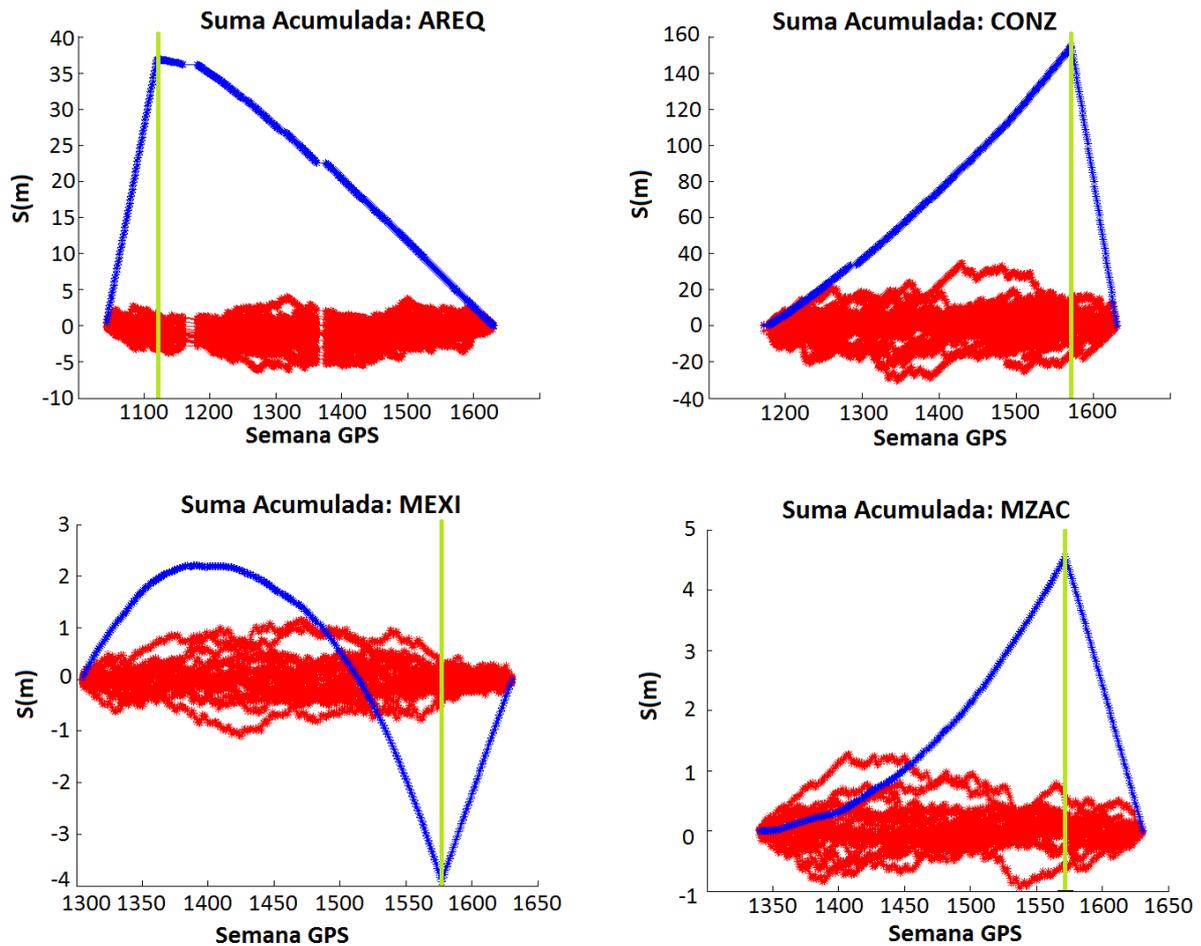


Figura 6 Gráficos de Suma Acumulativa, componente este, de las estaciones AREQ, CONZ, MEXI y MZAC.
Figure 6 Cumulative Sum graphs for the East component of AREQ, CONZ, MEXI and MZAC stations.

DISCUSIÓN

Haciendo una comparación entre los métodos mencionados podemos decir que el algoritmo de Promedio en Bloque es el más sencillo de implementar, tiene bajo costo computacional y es el que mejor maneja las pendientes pronunciadas (o drift) si las hubiera en los datos de entrada, ya que trabaja solo con los promedios de cada ventana.

La limitación más significativa del algoritmo de detección de salto es que utiliza ventanas de análisis de un tamaño fijo, aunque seleccionable. En consecuencia, la magnitud de los saltos detectados no son necesariamente las de los datos, sino que responden simplemente al cambio en el valor medio en el límite de las dos ventanas de análisis.

El parámetro crítico a considerar para que el método funcione de forma óptima es el tamaño de la ventana. Al modificar la cantidad de muestras en las ventanas a comparar, cambia el rango que se utiliza para confirmar o rechazar un salto. Si se aceptan ruidos más grandes se puede dificultar la detección de saltos más pequeños. Para elegirlo se evalúa la relación de compromiso entre resolver adecuadamente las discontinuidades y un mejor promedio del ruido. El valor óptimo encontrado fue, en la mayoría de los casos, el 8% de la cantidad de datos disponibles y, en todos los casos, superior al 5% de los valores de la muestra.

Como el programa detecta discontinuidades solo entre ventanas adyacentes, esto implica que el salto puede ser detectado sólo en ciertos lugares en el registro. Esta limitación fue salvada cambiando la posición inicial de las ventanas, de manera iterativa, de forma de pasar por todos los puntos del registro.

	Este (mm)	Norte (mm)	Altura (mm)	Evento Sísmico
AREQ	433,2	297,8	27,5	23/06/01
AZUL	23,8	1,66	13,4	27/02/10
BRFT	14,3	5,9	1,6	-
CBSB	6,7	1,2	0,4	-
CFAG	33,5	16,8	1,2	27/02/10
CONZ	2977,3	681,3	46,2	27/02/10
ETCG	5,8	15,3	5,4	08/01/09
IGM1	18,4	4,21	8,22	27/02/10
LHCL	66,4	6,2	3,79	27/02/10
LPGS	17,4	1,66	0,72	27/02/10
MEXI	152,5	155,2	13,3	04/04/10
MZAC	109,3	47,2	15,2	27/02/10
MZAE	106,5	37,9	17,5	27/02/10
ONRJ	8,0	5,6	9,5	-
SANT	257,8	144,5	20,5	27/02/10
UBAT	30,7	3,5	1,2	-
UCOR	23,0	10,6	0,7	27/02/10
UNRO	20,1	6,5	3,2	27/02/10
UNSJ	29,6	17,1	0,1	27/02/10
UYMO	12,1	1,3	9,5	27/02/10
UYRO	7,1	0,5	5,8	27/02/10

Tabla 1 Estimaciones de la magnitud de los saltos detectados con el método de Promedio Secuencial, en las 3 componentes de cada estación analizada, y la fecha del evento sísmico asociado si lo tienen.

Table 1 Estimates of the magnitude of the jumps detected with the Sequential Average method, in the 3 components of each station analysed, and the date of the associated seismic events.

Analizando la totalidad de las series de tiempo con este método se observa que el número de casos favorables que detecta es significativamente menor a los otros métodos. Por lo que se lo considera el más eficaz en reducir la cantidad de falsos positivos. La componente vertical, por su gran variabilidad, es aquella en la que se detectan más saltos posibles. Sin embargo es capaz de confirmarlos o rechazarlos exigiéndole condiciones adicionales, como estar presente en más de una coordenada.

Para una misma estación, las componentes horizontales tienen valores similares, tanto el orden de magnitud de sus variaciones semanales como en el de sus parámetros estadísticos, pero no ocurre lo mismo en la coordenada vertical. En la Figura 3 se puede ver el caso de la estación Arequipa (sup. derecha), en la que el salto en la componente vertical no es confirmado debido a que la variación de la altura presenta una gran variabilidad, por lo que al exigirles a los posibles candidatos que la discontinuidad supere cierto valor umbral, dicha condición no se verifica. En estos casos, la detección en las otras dos coordenadas y el análisis del observador son necesarios para confirmar el salto.

El método de Promedio Secuencial tiene la capacidad de detectar un salto en cualquier punto en el registro de datos, excepto en las primeras n muestras de la serie (siendo n el largo de la ventana elegida). Para salvar esta dificultad se puede realizar el análisis de los datos en ambas direcciones.

Al analizar los resultados que obtenemos con este método podemos apreciar que encuentra varios falsos positivos, especialmente en la componente vertical que tiene una variación semanal más marcada que el resto. Al igual que en el algoritmo anterior se observa que las condiciones externas que se les exigen a los posibles candidatos son fundamentales para minimizar la cantidad de discontinuidades encontradas. En especial el requisito de estar presente en más de una componente.

Una desventaja del método Promedio Secuencial es que su efectividad depende fuertemente del promedio calculado al principio de cada régimen, por lo que nuevamente es significativo el tamaño de

la ventana utilizada. En este trabajo se probaron con valores que variaban desde el 5 al 15% de la cantidad de muestras de la serie de tiempo. El valor óptimo que se encontró fue, en la mayoría de las estaciones cercano al 8%. Otra dificultad que se presenta es implementarlo en estaciones con pendientes pronunciadas, ya que a partir de un cierto valor, el mismo comportamiento de la serie genera que tanto un punto como la media de la ventana subsiguiente se excedan de los parámetros iniciales.

Analizando ahora el último método, la principal ventaja del método Suma Acumulativa es que nos permite identificar a simple vista, a partir de su gráfico, la posición de los saltos más significativos y validar el resultado obtenido por los otros métodos en caso de que haya encontrado más de un salto. Y dicha información se puede verificar mediante la comparación con la serie reordenada para obtener un nivel de confianza. El salto hallado será considerado válido si la diferencia entre extremos es mayor en la serie original que en las recombinaciones en al menos un 95% de las veces. La desventaja es que la confiabilidad se vuelve significativa si comparamos con un número grande de series (del orden de 1000) y eso genera un mayor costo computacional. Este método puede ser considerado complementario de los anteriores ya que no permite cuantificar la dimensión del salto por sí mismo.

Implementando este método a continuación de los dos anteriores, se puede resolver con mayor efectividad el problema de encontrar la ubicación de los saltos, ya que se excluyen aquellos valores hallados previamente que eran falsos positivos.

En la Figura 7 se observan dos gráficos para cada una de las estaciones AREQ, CONZ, MEXI, MZAC. En la imagen superior está representada la coordenada Este en función de la semana GPS, mientras que en la imagen inferior está graficado el estimador F para el mismo período de tiempo. Se puede apreciar que el extremo del estimador F coincide en el tiempo con la posición del salto en la coordenada.

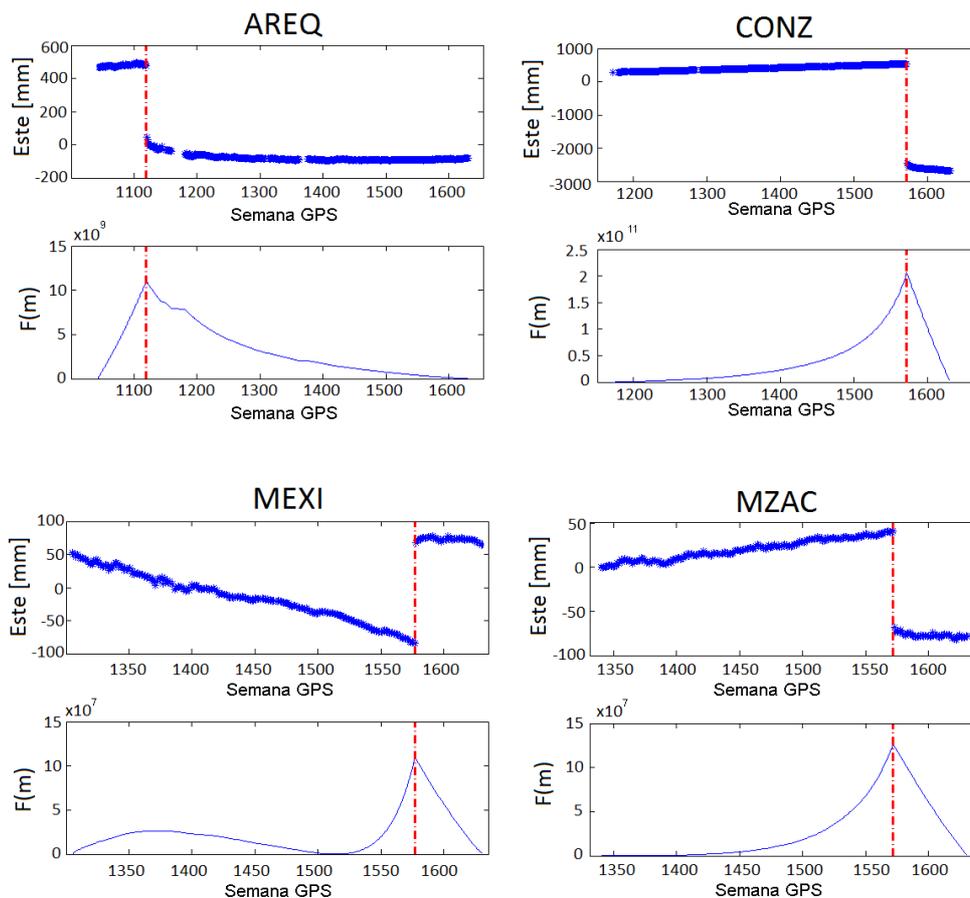


Figura 7 (Arriba) Gráficos de la componente este, de las estaciones AREQ, CONZ, MEXI y MZAC. (Abajo) Estimador F en función del tiempo.

Figure 7 (Above) East component graphs of AREQ, CONZ, MEXI and MZAC stations. (Below) Graph of F-estimator over time.

La principal ventaja de este método es su sencilla implementación automática y su bajo costo computacional. Sin embargo, no funciona adecuadamente en series de tiempo con una marcada tendencia lineal.

Otra ventaja de los métodos no paramétricos frente al estimador F, es su capacidad de encontrar más de una posible discontinuidad en cada serie de tiempo. Como se ilustra en la Figura 8, el método de Promedio en Bloque es capaz de resolver dos saltos en las coordenadas horizontales de la estación UCOR, mientras que el estimador F solo encuentra el más significativo de ellos para cada coordenada. En este caso, como las posiciones del salto encontradas con el estimador F son diferentes entre las dos coordenadas horizontales, en este caso el método no resolvería el problema.

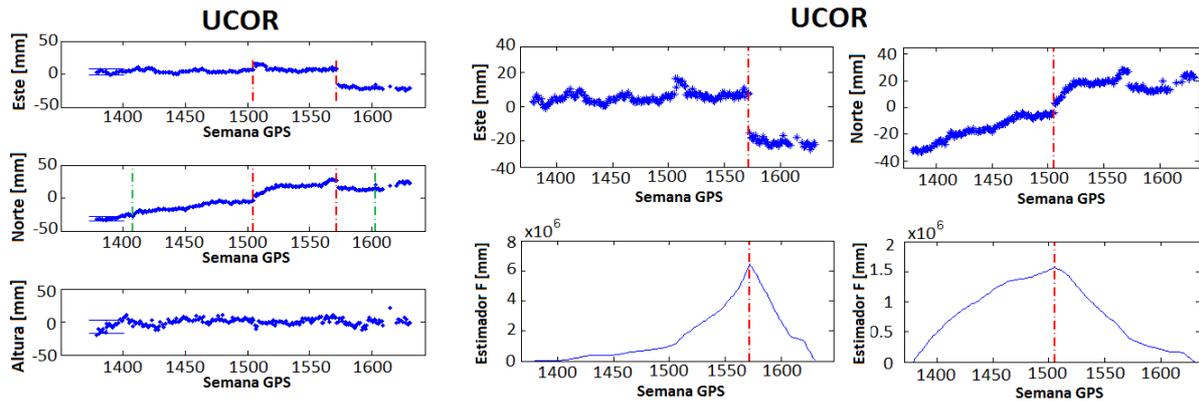


Figura 8 (Izquierda) Series de tiempo para la estación UCOR. Se puede observar que el algoritmo de Promedio en Bloque es capaz de resolver dos saltos en las coordenadas horizontales. (Derecha) Coordenada Este y Norte de la misma estación. Se marca el único salto hallado con el estimador F.

Figure 8 (Left) Time series for the UCOR station. It can be observed that the Average Block algorithm is able to solve two jumps in the horizontal coordinates. (Right) East and North coordinates of the same station. Only one jump is found with the estimator F.

CONCLUSIONES

Analizando la variabilidad de los datos de entrada, se observa que en las coordenadas horizontales Este y Norte la variación semanal es, en el 95% de los casos, menor a los 3mm. En la coordenada vertical, la variación semanal es, en el 95% de los casos, menor a los 8mm para la gran mayoría de las estaciones.

Los algoritmos Promedio en Bloque y Promedio Secuencial, junto con un gráfico de Suma Acumulativa, fueron razonablemente eficaces para encontrar saltos en series de tiempo de coordenadas GNSS. Su aplicación se vuelve más robusta al combinar las técnicas y al cotejar para cada estación las ubicaciones de los offsets en las diferentes componentes.

El parámetro crítico de los primeros métodos fue la longitud de la ventana y se buscó un compromiso entre aceptar señales con mayor cantidad de ruido y una buena resolución de las discontinuidades. Este se vuelve determinante para la detección del salto, necesitando en todos los casos una cantidad mayor a 12 muestras, valor que ronda el 8% de la longitud de la serie de datos.

Para optimizar el resultado y descartar falsos positivos, a los saltos encontrados se les exigieron distintas condiciones para aceptarlos como verdaderos. En particular se consideraron aquellos cuya amplitud superara un valor umbral. El mismo fue calculado, ordenando de forma creciente todas las variaciones semanales de cada componente, y estableciendo como valor de corte aquel que superara al 95% de las variaciones. Otra condición que se impuso a los resultados fue optar, en caso de existir saltos consecutivos, por aquellos que estuvieran presentes en ambas coordenadas horizontales y cuya amplitud fuera de mayor consideración. Estos requisitos razonables permitieron mejorar la calidad de las soluciones encontradas.

En la Figura 9 están representadas la cantidad de falsos positivos que obtienen cada uno de los métodos implementados para la determinación del salto en la coordenada Este. El Promedio Secuencial encuentra el salto con buena eficacia, pero en estaciones con gran deriva encuentra también otros saltos, en puntos que simplemente se salieron de los límites impuestos por la pendiente misma de la señal. El método de Suma Acumulativa encuentra la solución verdadera en la mayoría

de los casos, pero también presenta gran cantidad de falsos positivos. El método del Estimador F es eficaz para aquellas estaciones cuyas coordenadas se mantienen cercanas a un valor constante antes y después del salto, se vuelve ineficaz si las series temporales presentan una marcada pendiente agregando otro punto posible de discontinuidad. La principal desventaja de estos dos últimos métodos es que ambos deben ser apoyados por un análisis visual del gráfico, en detrimento de la automatización de la búsqueda de discontinuidades. A su vez no permiten cuantificar el salto encontrado.

Tanto el algoritmo de Promedio en Bloque como el Promedio Secuencial son capaces de encontrar las coordenadas de los saltos en las estaciones estudiadas para la mayoría de las estaciones. A su vez, ambos métodos son capaces de resolver más de un salto para cada serie de coordenadas. Teniendo en cuenta la totalidad de los casos analizados se puede concluir que el método de Promedio Secuencial fue eficaz en un 87%, siendo superado por el método Promedio en Bloque, que encontró con éxito las discontinuidades en el 95%.

La problemática que el algoritmo de Promedio Secuencial no es capaz de resolver adecuadamente, se da para las series de tiempo con pendientes pronunciadas. Este drift en los datos de entrada, provoca que la media aumente (o disminuya) paulatinamente, y genera que eventualmente se salga del intervalo definido a partir de la ventana de entrada.

Se puede observar que el Método de Promedio en Bloque encuentra la solución verdadera generando la menor cantidad de soluciones espurias, y por lo tanto es el más eficaz a la hora de detectar discontinuidades automáticamente.

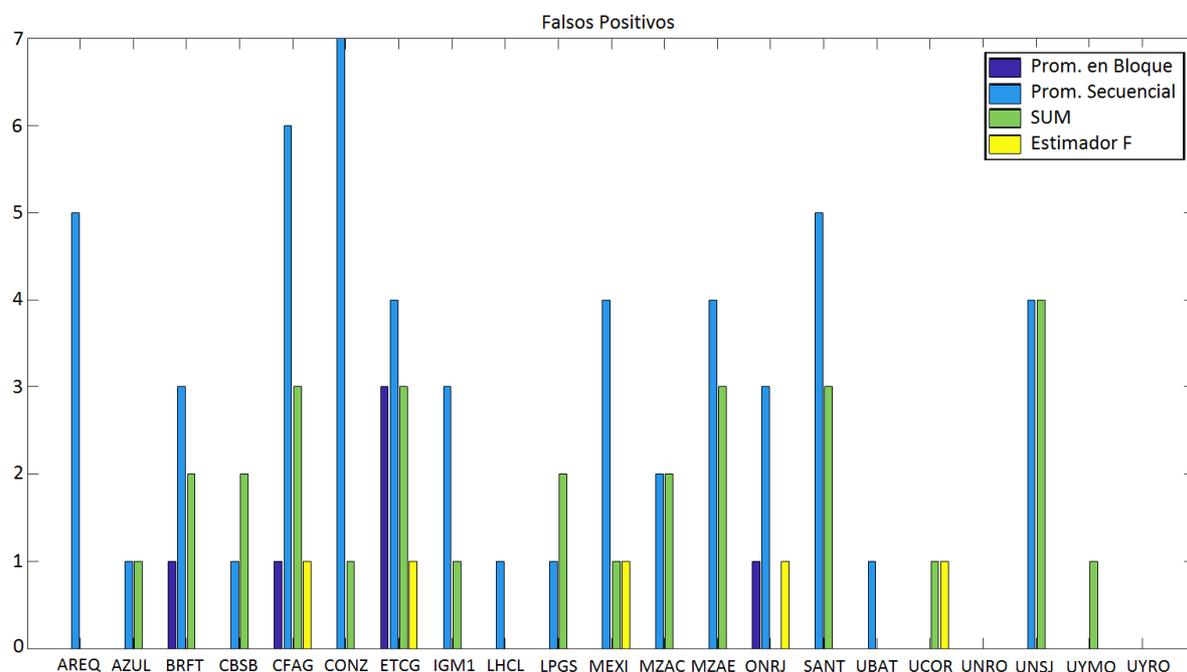


Figura 9 Gráfico de barras que muestra el número de falsos positivos hallados con cada método en la coordenada horizontal Este para cada estación analizada.

Figure 9 Bar graph showing the number of false positives found with each method in the East horizontal coordinate of every station analyzed.

Agradecimientos: Los autores agradecen a la red SIRGAS y a sus centros de procesamiento por la provisión de las soluciones semanales de coordenadas.

REFERENCIAS

- Basseville, M., y Nikiforov, I. V. (1993). *Detection of abrupt changes: Theory and application*. Englewood Cliffs, New Jersey: Prentice Hall.
- Bevis, M. (2005). Seasonal fluctuations in the mass of the Amazon River system and Earth's elastic response. *Geophysical Research Letters*, 32(16), L16308. doi: 10.1029/2005GL023491

- Bos, M. S., Penna, N. T., Baker, T. F., y Clarke, P. J. (2015). Ocean tide loading displacements in western Europe: 2. GPS - observed anelastic dispersion in the asthenosphere. *Journal of Geophysical Research: Solid Earth*, 120(9), 6540–6557. doi: 10.1002/2015JB011884
- Bock, Y., Wdowinski, S., Fang, P., Zhang, J., Williams, S., Johnson, H., ... Gurtner, W. (1997). Southern California Permanent GPS Geodetic Array: Continuous measurements of regional crustal deformation between the 1992 Landers and 1994 Northridge earthquakes. *Journal of Geophysical Research: Solid Earth*, 102(B8), 18013–18033. doi: 10.1029/97JB01379
- Bock, Y., y Melgar, D. (2016). Physical applications of GPS geodesy: A review. *Reports on Progress in Physics*, 79(10), 106801. doi: 10.1088/0034-4885/79/10/106801
- Gazeaux, J., Williams, S., King, M., Bos, M., Dach, R., Deo, M., ... Webb, F. H. (2013). Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment: DETECTING OFFSETS IN GPS TIME SERIE. *Journal of Geophysical Research: Solid Earth*, 118(5), 2397–2407. doi: 10.1002/jgrb.50152
- Griffiths, J., y Ray, J. (2016). Impacts of GNSS position offsets on global frame stability. *Geophysical Journal International*, 204(1), 480–487. doi: 10.1093/gji/ggv455
- Lavielle, M., Ludeña, C., y Ludena, C. (2000). The Multiple Change-Points Problem for the Spectral Distribution. *Bernoulli*, 6(5), 845. doi: 10.2307/3318759
- Milne, G. A. (2001). Space-Geodetic Constraints on Glacial Isostatic Adjustment in Fennoscandia. *Science*, 291(5512), 2381–2385. doi: 10.1126/science.1057022
- Montillet, J.-P., Williams, S. D. P., Koulali, A., y McClusky, S. C. (2015). Estimation of offsets in GPS time-series and application to the detection of earthquake deformation in the far-field. *Geophysical Journal International*, 200(2), 1207–1221. doi: 10.1093/gji/ggu473
- Perfetti, N. (2006). Detection of station coordinate discontinuities within the Italian GPS Fiducial Network. *Journal of Geodesy*, 80(7), 381–396. doi: 10.1007/s00190-006-0080-6
- Petrov, L. (2004). Study of the atmospheric pressure loading signal in very long baseline interferometry observations. *Journal of Geophysical Research*, 109(B3), B03405. doi: 10.1029/2003JB002500
- Riley, W. J. (2008). Algorithms for frequency jump detection. *Metrologia*, 45(6), S154–S161. doi: 10.1088/0026-1394/45/6/S21
- Rodionov, S. N. (2004). A sequential algorithm for testing climate regime shifts: ALGORITHM FOR TESTING REGIME SHIFTS. *Geophysical Research Letters*, 31(9), n/a-n/a. doi: 10.1029/2004GL019448
- Rodionov, S.N. (2005). A brief Overview of the regime shift detection methods. Joint Institute for the Study of the Atmosphere and Ocean. *Journal of Large-scale disturbances and recovery in aquatic ecosystems: challenges for management toward sustainability*, 17-24.
- Sánchez, L., Seemüller, W., Drewes, H., Mateo, L., González, G., da Silva, A., ... Cimbaro, S. (2013). Long-Term Stability of the SIRGAS Reference Frame and Episodic Station Movements Caused by the Seismic Activity in the SIRGAS Region. In Z. Altamimi y X. Collilieux (Eds.), *Reference Frames for Applications in Geosciences* (Vol. 138, pp. 153–161). doi: 10.1007/978-3-642-32998-2_24
- Taylor, W.A. (2000). Change point analysis: a powerful new tool for detecting changes. Deerfield, IL: Baxter Healthcare Corporation. Recuperado de <http://www.variation.com/cpa/tech/changepoint.html>
- Tregoning, P., Watson, C., Ramillien, G., McQueen, H., y Zhang, J. (2009). Detecting hydrologic deformation using GRACE and GPS: HYDROLOGIC DEFORMATION FROM SPACE. *Geophysical Research Letters*, 36(15), n/a-n/a. doi: 10.1029/2009GL038718
- Williams, S. D. P. (2003). Offsets in Global Positioning System time series: OFFSETS IN GPS TIME SERIES. *Journal of Geophysical Research: Solid Earth*, 108(B6). doi: 10.1029/2002JB002156

Recibido: Abril, 2018

Aceptado: Septiembre, 2019