



REVISTA UD Y LA GEOMÁTICA

<http://revistas.udistrital.edu.co/ojs/index.php/UDGeo/index>

DOI:<http://dx.doi.org/10.14483/udistrital.jour.udgeo.2013.7.a05>

INVESTIGACIÓN

Análisis de tendencias de movimiento de estaciones GNSS/GPS mediante series de tiempo geodésicas en Colombia, 2008-2011

Movement trend analysis of GNSS/GPS stations using geodetic time series in Colombia, 2008-2011

Sindy Carolina Lizarazo^a, Héctor Mora Páez^b, Luis Fernando Santa^c

Para Citar este artículo: Lizarazo S.C., Mora Páez H. & Santa LF,. (2013). Análisis de tendencias de movimiento de estaciones GNSS/GPS mediante series de tiempo geodésicas en Colombia 2008-2011. UD y la Geomática, (7), pp. 41 – 51.

Fecha de recepción: 20 de septiembre de 2013 / **Fecha de aceptación:** 15 de diciembre de 2013

RESUMEN

Mediante el empleo de la metodología ARIMA se realiza el análisis de series geodésicas de tiempo de nueve estaciones de la Red Nacional de Estaciones Geodésicas Espaciales GNSS para investigaciones geodinámicas del Proyecto GeoRED del Servicio Geológico Colombiano, para el periodo 2008-2011. Los resultados obtenidos permiten identificar las tendencias de movimiento 3D en dichas estaciones geodésicas. Se estableció que los modelos univariados presentan poder de predicción limitado, aunque la tendencia del movimiento se conserva, siguiendo el comportamiento habitual enmarcado por los datos objeto de análisis, mientras que los modelos multivariados representan en forma más real el pronóstico de movimiento de las estaciones GPS, considerando además la tendencia característica y el elemento de variabilidad inherente a los procesos estudiados. Se concluye, además, la necesidad de mayor tiempo de observación para este tipo de análisis, en especial para la componente vertical.

Palabras clave: GNSS/GPS, estacionariedad, serie de tiempo, dependencia temporal, ARIMA, ruido blanco.

ABSTRACT

Through the use of ARIMA methodology, it is carried out a geodetic time series analysis of nine stations that are part of the National Network of Space Geodetic GNSS stations for geodynamics research, GeoRED Project, performed by the Colombian Geological Survey, for the period 2008-2011. The results obtained allow identifying trends of 3D motion in those geodetic stations. It was established that, in this case, the univariate models have limited forecasting power, although the trend of the movement is kept, following the usual behavior framed by the data under analysis, while the multivariate models represent more real forecast of GPS stations movement, considering also the trend characteristics and the variability inherent in the processes studied. It is concluded also the need for longer observation time for this type of analysis, especially for the vertical component.

Keywords: GNSS/GPS, stationarity, time series, temporal dependence, ARIMA, white noise.

^aServicio Geológico Colombiano (SGC) – Proyecto GeoRED, Bogotá, Colombia, slizarazo@sgc.gov.co.

^bUniversidad Distrital Francisco José de Caldas, Bogotá, Colombia.

^cMSc. Universidad Nacional de Colombia- Bogotá, Colombia, lfasantag@unal.edu.co.

1. Introducción

El proyecto GeoRED del Servicio Geológico Colombiano ha venido implementando en los últimos años la Red Nacional de Estaciones Geodésicas Espaciales GNSS con propósitos geodinámicos, más conocida como GeoRED, la cual tiene dos componentes fundamentales: una red permanente de estaciones activas de operación continua y una red pasiva de estaciones de campo de ocupación episódica.

Uno de los productos obtenidos con la red pasiva corresponde a las series de tiempo de las posiciones diarias obtenidas mediante el uso del software GIPSY-OASIS II, desarrollado por JPL-NASA. El análisis de dichas series permite la identificación de posibles tendencias asociadas a los movimientos, tanto horizontales como verticales, que experimentan las estaciones geodésicas espaciales GNSS que constituyen la red activa, de vital importancia en geodinámica. Las series geodésicas de tiempo son en la actualidad expresadas en función del Marco Internacional Terrestre de Referencia (*International Terrestrial Reference Frame*, ITRF), y su análisis permitirá la definición de modelos basados en el principio de correlación temporal, los cuales son empleados para la generación de pronósticos sobre las variables tridimensionales analizadas. Así, el análisis de observaciones geodésicas con series de tiempo puede considerarse como punto de partida para estudiar con mayor detalle procesos transitorios, acumulación de deformación en el límite de placas y rotaciones de bloques de corteza donde se encuentran localizadas las estaciones GNSS permanentes de operación continua.

2. Marco teórico

El análisis de series temporales es una técnica estadística de amplia aplicación en diferentes campos de las ciencias económicas, ingenieriles, biológicas y ambientales (Box et al., 2008), cuyo uso, en virtud de la tecnología asociada, ha podido extenderse a diferentes ramas de las geociencias.

En la actualidad, el gran volumen de información espacio-temporal obtenida a través de diferentes tecnologías, en este caso, información geodésica derivada del sistema de navegación global por satélite (*Global Navigation Satellite System*, GNSS), requiere el uso de esta técnica estadística con el fin, entre otros, de determinar tendencias asociadas a los movimientos observados en las estaciones geodésicas permanentes de operación continua y, también, la estimación de modelos basados en la estructura de correlación temporal empleada para la elaboración de pronósticos sobre el comportamiento de la variable espacial en un periodo determinado.

Para ello, es preciso considerar algunos conceptos fundamentales de geodinámica y de series temporales que proporcionarán mayor comprensión del fenómeno estudiado, su análisis y la presentación de resultados.

2.1. Movimiento de estaciones geodésicas GNSS

Una estación geodésica GNSS, debido a la acción de fuerzas de origen tectónico, volcánico, deslizamientos, entre otros, puede experimentar movimientos en sus tres componentes durante un periodo determinado, los cuales son expresados con respecto a una referencia específica en un determinado marco terrestre.

Dicho movimiento corresponde al cambio de posición que presenta la antena GNSS entre dos instantes diferentes, cuya magnitud dependerá de la acción de las fuerzas mencionadas. El comportamiento de dicho movimiento puede ser observado en una ventana temporal más amplia a partir de los gráficos de series temporales, los cuales proporcionan indicios sobre las tendencias, determinísticas o estocásticas, del movimiento asociado, que permiten obtener una estimación de la velocidad a la que la antena se mueve, expresada en milímetros o centímetros por año.

2.2. Series de tiempo

Una serie temporal se define como el conjunto de observaciones z_1, z_2, \dots, z_k generadas secuencialmente en el tiempo. Dichas observaciones son consideradas como la realización de un proceso estocástico, que es la sucesión de variables aleatorias Z_t observadas en diferentes instantes en que se captura la información de la variable de interés.

El análisis de series temporales tiene como objetivo primordial el desarrollo de modelos estadísticos para describir el comportamiento de la(s) variable(s) bajo estudio y, así mismo, determinar la tendencia futura de los datos a través del uso de modelos de pronóstico, bajo el supuesto de que las condiciones futuras serán análogas a las actuales o pasadas, lo que se conoce como enfoque de dominio temporal. Se distinguen dos clases de análisis de series de tiempo: por una parte, el análisis univariado, que estudia una característica en particular durante un periodo, basándose en la historia de la propia serie; y, por otra parte, el análisis multivariado, que busca estudiar la relación dinámica entre dos o más series que analizan de forma independiente la misma variable en condiciones heterogéneas, o simplemente la asociación existente entre variables totalmente diferenciadas.

2.3. Series de tiempo en geociencias

Según Gilgen (2006), las variables estudiadas en geociencias dependen tanto del tiempo como del espacio, y el comportamiento aleatorio puede implicar que:

- i. Un modelo de probabilidad (función de densidad probabilística) puede atribuirse a los datos observados y,
- ii. Los modelos no determinísticos generan un mejor ajuste a las observaciones, siendo más adecuados para pronósticos.

El análisis de series de tiempo orientado a aplicaciones en geociencias tiene un amplio uso práctico, gracias a que existen variables espaciales que han sido medidas por décadas y las cuales generan un registro de información geocientífica robusto para el estudio de diferentes fenómenos del sistema terrestre. Sin embargo, la adquisición de datos es un proceso desafiante por limitaciones asociadas a los instrumentos, sistemas de medición, regiones y/o periodos de observación, lo que causa inconvenientes en cuanto a la calidad de la información. En este sentido, el análisis de series de tiempo es de gran utilidad para separar la señal de interés del ruido de fondo (Gilgen, 2006), elemento significativo en el tratamiento de datos geodésicos, los cuales pueden ser afectados por factores asociados, tales como medio de propagación y otros de orden local como multitrayectoria, obstáculos, aspectos climáticos, por citar algunos.

2.4. Metodología ARIMA

La metodología más difundida dentro del enfoque de dominio temporal se conoce como metodología ARIMA (Autoregressive Integrated Moving Average), desarrollada por Box y Jenkins. Esta metodología se basa en el cumplimiento del supuesto de estacionariedad y en la dependencia existente entre los valores actuales y pasados de una serie de tiempo. Estos modelos son ajustados a datos de series de tiempo para entender en mejor forma los datos o para estimar futuros puntos en las series (pronóstico). En el modelo conocido como ARIMA (p, d, q), los parámetros p, d, q son enteros positivos referidos al orden de las partes autorregresivas, integradas y móviles respectivamente.

Se basa en un proceso iterativo compuesto por los pasos de:

- i. Identificación: a partir de la estructura de correlación temporal, se determina un conjunto de posibles modelos para los datos bajo estudio.
- ii. Estimación: cada uno de los modelos es estimado mediante máxima verosimilitud (MLE) o mínimos cuadrados ordinarios (OLS).
- iii. Selección: con base en los criterios de selección de modelos, a saber AIC (Akaike) y BIC (Bayesiano), se escoge aquel que representa de mejor manera la serie observada.
- iv. Diagnóstico: consiste en la comprobación de diferentes hipótesis sobre los residuos.
- v. Uso: aplicación del modelo seleccionado en pronósticos.

2.5. Estacionariedad

Este supuesto establece que las leyes probabilísticas que gobiernan un proceso estocástico permanecen invariantes con respecto al tiempo (Gubbins, 2004). Lo anterior implica que:

$$P(z_{t_1}, z_{t_2}, \dots, z_{t_k}) = P(z_{t_1+h}, z_{t_2+h}, \dots, z_{t_k+h}) \quad (1)$$

Donde t_k es el tiempo de cada observación y h representa el rezago temporal o el número de periodos que se desplaza la serie temporal.

La estacionariedad de un proceso puede clasificarse como fuerte o débil. En el primer caso, las propiedades del proceso no se ven afectadas por un cambio en el origen temporal, lo que conlleva a que todas las distribuciones marginales sean idénticas, manteniéndose la distribución de probabilidad. Por su parte, la estacionariedad débil o de segundo orden se cumple si la función de medias es constante en el tiempo $E(z_t) = \mu$, la función de varianzas también lo es $E(z_t^2) = \sigma^2$ y la covarianza depende únicamente del rezago h , es decir, $\gamma_z(t, t+h) = \gamma(h)$.

Para efectos prácticos, se consideran procesos estacionarios aquellos que presentan estacionariedad en sentido débil.

2.6. Dependencia temporal

La dependencia temporal entre las observaciones de una variable aleatoria implica que el proceso estudiado tiende a evolucionar con cierta inercia, lo que a su vez implica que las propiedades del conjunto de datos se mantienen a lo largo de la serie y a futuro con los pronósticos. La dependencia temporal se mide a partir de las funciones de autocovarianza y autocorrelación.

2.7. Función de autocovarianza

La autocovarianza mide la dependencia lineal entre dos puntos de la misma serie de tiempo observados en diferentes instantes. Es definida según la siguiente ecuación:

$$\gamma_z(t, t+h) = E[(z_t - \mu_t)(\mu_{t+h} - \mu_{t+h})] \quad (2)$$

Si el proceso estudiado es estacionario, la función de autocovarianza se define como:

$$\gamma_z(t, t+h) = \gamma_z(h) \quad (3)$$

2.8. Función de autocorrelación (ACF)

La autocorrelación mide la previsibilidad lineal de las series en el tiempo $t+h$, es decir, z_{t+h} , usando únicamente información hasta z_t (Honaker et al., 2011). Se define mediante

la siguiente ecuación:

$$\rho_z(t, t+h) = \frac{\gamma_z(t, t+h)}{\sqrt{Var(z_t) Var(z_{t+h})}} \quad (4)$$

Donde $\gamma_z(t, t+h)$ es la función de autocovarianzas y $Var(z_t) Var(z_{t+h})$ es el producto de las varianzas para los instantes t y $t+h$ respectivamente.

Si el proceso es estacionario, la función de autocorrelación se define como:

$$\gamma_z(h) = \frac{\gamma_z(h)}{Var(z_t)} \quad (5)$$

2.9. Función de covarianza cruzada

Se emplea cuando se quiere dar una medida de la predictibilidad de una serie Z_t a partir de otras series Y_t .

$$\gamma_{zy}(t+h, t) = E[(z_{t+h} - \mu_{z_{t+h}})(y_t - \mu_{y_t})] \quad (6)$$

Si las series Z_t y Y_t provienen de un proceso estacionario, la covarianza cruzada solo dependerá del retardo entre ellas y se define como:

$$\gamma_{zy}(t+h, t) = \gamma_{zy}(h) \quad (7)$$

con $h = 0, \pm 1, \pm 2, \dots$

Peña (2005) define las siguientes propiedades:

- Para $h > 0$ los coeficientes γ_{zy} , representan la influencia o relación causal de la variable z_t en la variable y_{t+h} .
- Para $h < 0$ los coeficientes γ_{zy} , representan la influencia o relación causal de la variable y_t en la variable z_{t+h} .

2.10. Función de correlación cruzada (CCF)

Se define como:

$$\rho_{zy}((t+h), t) = \frac{\gamma_{zy}((t+h), t)}{\gamma_z(t+h, t+h) \gamma_z(t, t)} \quad (8)$$

Donde $\gamma_{zy}((t+h), t)$ mide la dependencia lineal entre dos puntos de series diferentes observados en el mismo instante y $\gamma_z(t+h, t+h) \gamma_z(t, t)$ es el producto de las varianzas de las series Z e Y respectivamente (Honaker et al., 2011)

2.11. Matriz de covarianzas

La matriz de covarianzas de dimensiones $h \times h$ (donde h es el rezago temporal) está compuesta por los coeficientes de covarianza cruzada y mide la dependencia temporal para

un vector de k series $Z_t = (Z_{1t}, Z_{2t}, \dots, Z_{kt})$. La matriz queda definida por:

$$\Gamma(h) = \gamma(Z_t, Z_{t+h}) = E[(Z_t - \mu)(Z_{t+h} - \mu)] \quad (9)$$

$$\Gamma(h) = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1k} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k1} & \gamma_{k2} & \dots & \gamma_{kk} \end{pmatrix} \quad (10)$$

$$\Gamma(h) = \gamma(Z_{t-h}, Z_t) \quad (11)$$

Para $h = 0, \pm 1, \pm 2, \dots$, $i = 1, 2, \dots, k$ y $j = 1, 2, \dots, k$.

2.12. Matriz de correlaciones

Indica la asociación temporal entre varias series de tiempo. Está definida por:

$$\rho(h) = D^{-1/2} \Gamma(h) D^{-1/2} = [\rho_{ij}(h)] \quad (12)$$

Para $i = 1, 2, \dots, k$ y $j = 1, 2, \dots, k$, y donde D es la diagonal de la matriz de covarianzas; es decir, las varianzas del proceso.

La matriz de correlaciones queda entonces representada por:

$$R(h) = \begin{pmatrix} \rho_{11} & \rho_{12} & \dots & \rho_{1k} \\ \rho_{21} & \rho_{22} & \dots & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \dots & \rho_{kk} \end{pmatrix} \quad (13)$$

2.13. Procesos de series de tiempo

Se distinguen los siguientes procesos según la forma de dependencia temporal existente entre los datos de una serie de tiempo.

Proceso de ruido blanco (WN): Es un proceso estacionario caracterizado por presentar un vector de medias constante e igual a cero, $E(Z_t) = 0$, varianza constante y finita para todo t , $Var(Z_t) = \sigma^2$ y variables no correlacionadas para todos los retardos, $Cov(Z_t, Z_{t+h}) = 0$.

Procesos autorregresivos (AR-VAR*): En estos procesos, el valor actual de la serie puede explicarse como una función de p valores pasados, $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$, donde p es el orden del proceso y determina el número de pasos hacia el pasado para pronosticar el valor actual. Se define mediante la siguiente fórmula:

$$\tilde{Z}_t = \Phi_1 \tilde{Z}_{t-1} + \dots + \Phi_p \tilde{Z}_{t-p} + a_t \quad (14)$$

Donde $\tilde{Z}_t = Z_t - \mu$, siendo μ la media del proceso, a_t es un proceso de ruido blanco y Φ_i con $i = 1, 2, \dots, p$ son los parámetros autorregresivos del modelo.

La representación, usando la notación de operadores, B , es:

$$(1 - \Phi_1 B - \dots - \Phi_p B^p) \tilde{Z}_t = a_t \quad (15)$$

$$\Phi_p(B) \tilde{Z}_t = a_t \quad (16)$$

Donde B^p es el operador lineal que aplicado a una función temporal proporciona esa misma función retrasada p periodos y $\Phi_p(B)$ es una función lineal autorregresiva aplicada sobre la variable de interés.

Procesos de media móvil (MA-VMA*): se consideran procesos de memoria corta porque son función de un número finito (q) de innovaciones pasadas a_t , quedando definido de la siguiente manera:

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (17)$$

Donde a_t son las innovaciones del proceso y θ_i con $i = 1, 2, \dots, q$ son los parámetros de media móvil del modelo.

En notación del operador retardo, el proceso de media móvil queda expresado así:

$$\tilde{Z}_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad (18)$$

$$\tilde{Z}_t = \theta_q(B) a_t \quad (19)$$

Donde $\theta_q(B)$ es una función lineal de media móvil aplicada sobre las innovaciones del proceso.

Procesos integrados (I): son procesos no estacionarios que tienen la propiedad de que al diferenciarlos d veces se obtienen procesos estacionarios (Peña, 2005).

Los procesos integrados son muy importantes, dado que la mayoría de series prácticas provienen de procesos no estacionarios.

Procesos mixtos ARIMA-VARMA* Son una combinación de procesos autorregresivos y de media móvil. Su estructura queda definida en la siguiente ecuación:

$$\Phi_p(B) \tilde{Z}_t = \theta_q(B) a_t \quad (20)$$

Donde p y q representan el orden del proceso mixto para la parte autorregresiva y de media móvil, respectivamente.

Procesos mixtos ARIMA: Constituyen procesos integrados (d) que combinan componentes autorregresivos (p) y de media móvil (q) Son del tipo:

$$\Phi_p(B) \nabla^d \tilde{Z}_t = \theta_q(B) a_t + c \quad (21)$$

Donde $\nabla \tilde{Z}_t = Z_t - Z_{t-1}$ es el operador diferencial que se aplica a la serie no estacionaria de t observaciones para obtener una nueva serie estacionaria de $t - 1$ observaciones, y d indica el grado de diferenciación del proceso.¹

¹La notación VAR, VMA y VARMA corresponden a los procesos autorregresivos, de media móvil y mixto para el caso vectorial o multivariado

3. Metodología

El análisis se basa en las observaciones de nueve estaciones GNSS de operación continua del proyecto GeoRED, a saber: Bogotá - BOGT, Buenaventura - BUGT, Palmira - CIA1, Corozal - CORO, Manizales - MZAL, Popayán - POVA, San Andrés y Providencia - SAN0, Santa Elena - SEL1 y Tumaco - TUCO (figura 1). Las observaciones corresponden únicamente a la señal GPS.

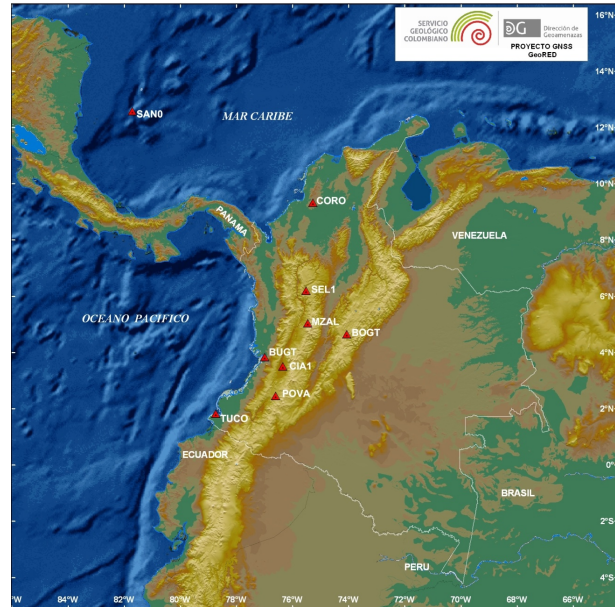


Figura1. Localización de estaciones empleadas en el estudio.

Los resultados del procesamiento de las estaciones citadas, realizado con el software científico GIPSY-OASIS II en el Centro de Procesamiento Científico de Datos GNSS del Proyecto GeoRED del Servicio Geológico Colombiano fueron utilizados para el análisis y corresponden a soluciones diarias para el periodo diciembre de 2008 - mayo de 2011.

Para el análisis se empleó el software estadístico R, que corresponde a un conjunto integrado de programas para manipulación de datos, cálculo y gráficos de código abierto. La versión utilizada fue R 2.13.1 (The Comprehensive R Archive Network, 2012).

4. Métodos

Depuración de la muestra: debido a la alta frecuencia de los datos, 908 por estación, se procedió a convertirlos a una frecuencia semanal, tomando como representativo de su conjunto la mediana de los datos que la componen. De esta manera, el conjunto de datos a analizar se redujo de 908 a 130, que abarcan las semanas GPS conocidas como G1509 hasta G1638. También se filtraron los datos de las

estaciones teniendo en cuenta como límite máximo de datos no disponibles (*missing values*), el 30% de la muestra. Así, de 11 estaciones inicialmente consideradas, correspondientes al mismo intervalo temporal, se redujo a un total de 9, quedando por fuera del análisis las estaciones MECE y PAL1.

4.1. Procesamiento estadístico

La imputación de datos ausentes consistió en el cálculo de datos no disponibles (*missing values*) dentro de las series de tiempo. Previamente, se verificaron los supuestos de normalidad y de MAR (*missing at random*) del conjunto de datos.

El análisis exploratorio de las series de tiempo correspondió al análisis preliminar de las características de los datos expresados tanto gráfica como numéricamente (medidas de tendencia central y de dispersión). Para ello, se elaboraron los gráficos de series de tiempo para cada componente geodésica (latitud, longitud y vertical) con el objetivo de evaluar la tendencia de movimiento de la estación GPS durante el periodo de estudio.

4.2. Metodología ARIMA univariada y multivariada

Para la identificación, se realizaron chequeos de estacionariedad mediante varias pruebas estadísticas. De la misma manera, se analizaron los correlogramas simples y parciales de cada serie para observar tendencias que indicaran procesos integrados. Lo anterior, con el objetivo de definir el orden de diferenciación de las series temporales para cumplir con el supuesto de estacionariedad. Para el análisis multivariado, adicional a la verificación de la estacionariedad conjunta de las series, se comprobó el supuesto de causalidad, definido como la posibilidad de que los p valores pasados de una serie Y originen el valor actual de una serie Z .

En cuanto a estimación y selección para el caso univariado, se hizo uso de las funciones implementadas en el software R, se realizó la identificación, estimación por máxima verosimilitud (MLE) y selección automática de modelos con base en los criterios AIC (Akaike), BIC (Bayesiano) y AICc (Akaike corregido), teniendo en cuenta el orden de diferenciación y la bondad de la misma, evaluada a partir la desviación estándar de las series originales versus las series diferenciadas.

Para el establecimiento del modelo vectorial, se realizó una identificación y selección automática con base en el criterio AIC. El método de estimación empleado fue mínimos cuadrados ordinarios (OLS), el cual proporciona los parámetros del modelo identificado por cada componente y las estadísticas básicas de los estimadores, como error estándar, el valor del estadístico de prueba, el p-valor y su significancia estadística en el modelo.

El diagnóstico de los modelos seleccionados, tanto univariados como multivariados, se hizo teniendo en cuenta que los $r = \max(p, q)$ residuos son iguales a cero por los limitantes asociados a las condiciones iniciales sobre los parámetros (Peña, 2005). Una vez aplicada la restricción mencionada, se verificaron los supuestos de la normalidad, blanquitud (entendida como la carencia de autocorrelación entre las observaciones de la serie), homocedasticidad (igual varianza) y estabilidad general de los residuos.

Con los modelos finales para cada estación, se realizó la estimación de pronósticos a un periodo de 52 semanas (un año) con niveles de confianza del 80% y 95% para el caso univariado, y de 95% para el caso multivariado.

5. Resultados y análisis

Los gráficos de series de tiempo en las diferentes componentes (figuras 2, 3 y 4) muestran la existencia de tendencias más o menos permanentes en algunas de las estaciones GPS del estudio.

Es importante destacar las tendencias determinísticas casi lineales presentes en todas las estaciones en la componente de latitud, señalando el sentido Norte del movimiento. La componente de longitud presenta un comportamiento ruidoso dado por la alta variabilidad de la serie. Finalmente, la componente vertical presenta tendencias de comportamiento menos claras, lo cual puede deberse a:

- i) menor precisión del sistema GPS en la componente vertical,
- ii) necesidad de mayor tiempo de observación para estos análisis,
- iii) influencia de algunos fenómenos locales, dependiendo del análisis del comportamiento con más tiempo de observación.

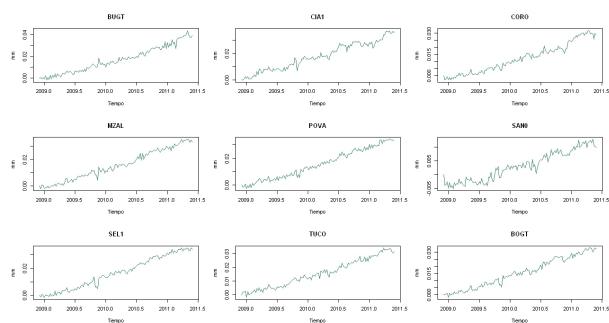


Figura2. Gráfico de series de tiempo de la componente latitud.

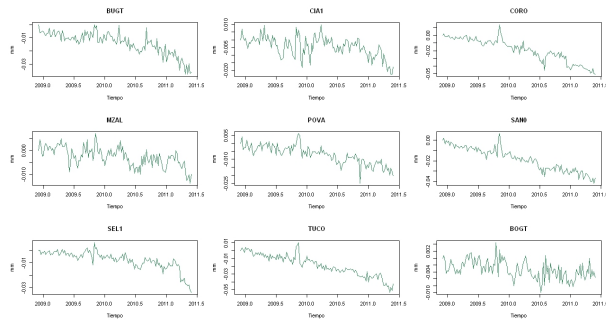


Figura3. Gráfico de series de tiempo de la componente longitud.

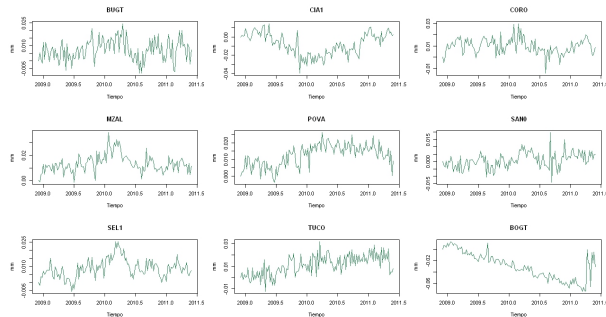


Figura4. Gráfico de series de tiempo de la componente vertical.

Como resultado del análisis univariado, se obtuvo un total de 27 modelos. La búsqueda del conjunto de modelos a estimar se basó en un proceso iterativo, teniendo en cuenta ciertas restricciones impuestas a través de algunos parámetros de la función, donde las más destacadas estuvieron asociadas con la omisión de los órdenes estacionales. Esto se hizo considerando la dificultad de encontrar procesos estacionales o SARIMA en series de frecuencia semanal o mayor, como las utilizadas en el desarrollo del proyecto. Estas limitaciones de periodicidad en las series temporales de frecuencias altas están en función de los 52 subprocesos semanales que, en teoría, deberían obedecer a la misma estructura de autocorrelación temporal, o lo que es equivalente, que su mecanismo de generación natural debería depender del mismo proceso estocástico $ARMA(p, q)$. Los aspectos anteriores resaltan los impedimentos estadísticos para encontrar estacionalidad en series con alta frecuencia de observaciones por unidad de tiempo (Lütkepohl, 2005).

La mayoría de los modelos univariados presenta un comportamiento integrado debido a que las series provienen de procesos no estacionarios. Lo anterior se confirma con las pruebas de estacionariedad y raíces unitarias aplicadas sobre el conjunto de datos original, que proporcionan suficiente evidencia estadística para aceptar la hipótesis de no estacionariedad. Al interior de las componentes espaciales analizadas, las únicas estaciones que presentan estacionariedad débil o de segundo orden son las estaciones de BUGT y CORO para la componente vertical; el resto de ellas pre-

senta raíces unitarias que conducen a la obtención de modelos integrados o diferenciados de primer orden.

El análisis visual de los correlogramas ratifica este hecho, ya que los coeficientes de autocorrelación simples disminuyen linealmente con el retardo, lo que, según Peña (2005), es indicación de la existencia de procesos integrados. Los modelos univariados definitivos se encuentran resumidos en las tablas 1, 2 y 3 para las componentes geodésicas de latitud, longitud y vertical, respectivamente. En ellos se denomina a Z_t como la serie de tiempo analizada; es decir, corresponde con la variable de interés medida. Los modelos de series temporales resultantes del análisis univariado son procesos que además de ser integrados presentan la combinación de características de procesos autorregresivos y de media móvil $ARMA(p, q)$. Dicha estructura de dependencia temporal fusiona las propiedades esenciales de los procesos básicos y tiene ventajas en cuanto a la representación de diferentes procesos, pues permite simbolizar el mecanismo generador de los datos con el uso de pocos parámetros.

Tabla 1. Modelos univariados componente latitud

MODELOS DE SERIES TEMPORALES LATITUD	
ID	MODELO ESTIMADO
BOGT	$(1 - B) Z_t = (1 + 0.608 B) a_t$
BUGT	$(1 - B) Z_t = (1 + 0.716 B) a_t$
CIA1	$(1 - B) Z_t = (1 + 0.403 B) a_t$
CORO	$(1 - B) Z_t = (1 + 0.624 B) a_t$
MZAL	$(1 - B) Z_t = (1 + 0.631 B + 0.0119 B^2) a_t$
POVA	$(1 - B) Z_t = (1 + 0.767 B) a_t$
SAN0	$(1 - 0.067 B) (1 - B) Z_t = (1 + 0.658 B) a_t$
SEL1	$(1 - B) Z_t = (1 + 0.545 B - 0.013 B^2 + 0.25 B^3) a_t$
TUCO	$(1 - B) Z_t = (1 + 0.608 B) a_t$

Tabla 2. Modelos univariados componente longitud

MODELOS DE SERIES TEMPORALES LATITUD	
ID	MODELO ESTIMADO
BOGT	$(1 - B) Z_t = (1 + 0.632 B) a_t$
BUGT	$(1 - B) Z_t = (1 + 0.693 B) a_t$
CIA1	$(1 - B) Z_t = (1 + 0.532 B + 0.131 B^2) a_t$
CORO	$(1 - B) Z_t = (1 + 0.476 B) a_t$
MZAL	$(1 - B) Z_t = (1 + 0.503) a_t$
POVA	$(1 - B) Z_t = (1 + 0.704 B) a_t$
SAN0	$(1 - B) Z_t = (1 - 0.151 B) a_t$
SEL1	$(1 + 0.58 B - 0.33 B^2) (1 - B) Z_t =$ $(1 - 0.03 B + 0.65 B^2) a_t$
TUCO	$(1 + 0.96 B) (1 - B) Z_t = (1 - 0.34 B + 0.40 B^2) a_t$

Tabla 3. Modelos univariados componente vertical

MODELOS DE SERIES TEMPORALES LATITUD	
ID	MODELO ESTIMADO
BOGT	$(1 - B) Z_t = (1 + 0.575 B) a_t$
BUGT	$(1 - 0.839 B) Z_t - 47.9 = (1 + 0.625 B) a_t$
CIA1	$(1 - B) Z_t = (1 + 0.5267 B) a_t$
CORO	$(1 - 0.407 B - 0.266 B^2) Z_t - 155.79 = a_t$
MZAL	$(1 - B) Z_t = (1 + 0.602 B) a_t$
POVA	$(1 + 0.94 B) (1 - B) Z_t = (1 - 0.15 B + 0.58 B^2) a_t$
SANO	$(1 - B) Z_t = (1 + 0.684 B) a_t$
SEL1	$(1 - B) Z_t = (1 + 0.429 B) a_t$
TUCO	$(1 + 1.64 B + 0.31 B^2 - 0.34 B^3) (1 - B) Z_t = (1 - 0.90 B + 0.86 B^2 + 0.79 B^3) a_t$

En general, se obtuvieron cinco procesos $IMA(d, q)$ para las series de CIA1, MZAL, SANO, SEL1 y BOGT en la componente vertical; ocho para BUGT, CIA1, CORO, MZAL, POVA, SEL1, TUCO y BOGT en la componente latitud; y siete para BUGT, CIA1, CORO, MZAL, POVA, SANO y BOGT en la componente longitud. Lo anterior constituye un total de veinte procesos integrados de media móvil en las series univariadas. Los procesos elaborados para las series de BUGT y CORO en la componente vertical fueron $ARMA(1, 1)$ y $AR(2)$ respectivamente. Finalmente, se obtuvieron cinco procesos $ARIMA(p, d, q)$ para las estaciones de POVA y TUCO en la componente vertical, SANO en componente latitud y SEL1 y TUCO en componente longitud. Los procesos que presentan parámetro $d = 1$, es decir, los procesos integrados, no varían de manera estable alrededor de una media fija, lo cual se relaciona fuertemente con las tendencias que sigue el movimiento tridimensional de las estaciones GPS debido a la dinámica terrestre.

Los modelos $ARMA(p, q)$ e $IMA(d, q)$ identificados son procesos invertibles, caracterizados porque sus coeficientes de media móvil son menores a 1 en valor absoluto, lo que indica que el efecto de los valores de observación previos decrece con el tiempo. Según Peña (2005), los procesos de promedio móvil son procesos de memoria muy corta, ya que asocian el valor actual con un número finito de observaciones precedentes. En vista de ello, se puede decir que los modelos identificados, es decir, los modelos $IMA(p, q)$, se generaron para estaciones cuya dependencia temporal es más o menos débil en comparación con aquellos procesos con parámetros autorregresivos que señalan una dependencia temporal más persistente.

Las series cuyos modelos iniciales presentaron diagnósticos deficientes fueron identificadas y estimadas nuevamente, considerando la serie GPS suavizada por atípicos. Este proceso se realizó para disminuir el sesgo introducido por valores atípicos, lo que, según Tsay (1988), puede conducir a conclusiones equívocas sobre el mecanismo generador de los datos medidos en el tiempo.

El diagnóstico de los modelos estimados, cuya bondad se basa en la comparación de los valores observados con relación a los valores predichos con el modelo ajustado, debe corresponder, según Brockwell y Davis (1987), a comportamiento consistente de los residuos evaluado a través del cumplimiento de diferentes supuestos. En cuanto a la normalidad de los residuos, se obtuvieron resultados que ofrecen evidencia estadística para aceptar la hipótesis nula de normalidad a un nivel de significancia del 5% y del 1%.

Los resultados de correlación serial entre residuos, evaluados mediante la aplicación de la prueba global de blancitud de Ljung-Box, permitieron definir la pertinencia de los modelos univariados identificados y estimados para cada estación GPS. Estos resultados indican que a un nivel de significancia del 5, las estaciones muestran un proceso de ruido blanco o falta de autocorrelación para los primeros 20 rezagos (Box et al., 2008), lo que representa la bondad del modelo ajustado a las observaciones GPS en las diferentes componentes geodésicas.

En los gráficos de diagnóstico adicionales se puede observar que los coeficientes de autocorrelación residual están dentro de los límites definidos por $\pm 1,96$, que constituyen un intervalo aproximado del 95% para la hipótesis nula de ruido blanco. Según lo anterior, es de esperarse que, en promedio, el 5% de los coeficientes de autocorrelación caiga fuera de dichas fronteras, lo que no significa que el modelo esté errado y requiera modificación (Brockwell y Davis, 1987). Los resultados analíticos para calificar el supuesto de varianza constante de los residuales muestran evidencia estadística para rechazar la hipótesis alternativa de heterocedasticidad (diferente varianza) a un nivel de significancia del 5% para la mayoría de las estaciones analizadas. Las variables que cumplen con el supuesto a un nivel $\alpha = 1\%$ son CORO en la componente latitud, y BOGT, BUGT, CORO y SANO en la componente longitud. Lo anterior indica que no existe dependencia temporal de la varianza.

Las cartas de control CUSUM permiten visualizar una conducta óptima de los modelos seleccionados para cada estación GPS en las diferentes componentes espaciales; así mismo, resaltan una estabilidad adecuada de los modelamientos ARIMA generados de forma automática por el software R, que los hace convenientes para su utilización en pronósticos.

Los pronósticos generados se elaboraron imponiendo intervalos de predicción del 80% y 95%, los cuales representan las regiones en que se espera estén los valores pronosticados; al mismo tiempo, dan una medida de la incertidumbre de las predicciones obtenidas. En general, se encontró que el comportamiento de los valores predichos varía significativamente y que en las series con una marcada tendencia lineal, las predicciones son muy pobres, tanto que las asemeja a una función determinística del tiempo con varianza prácticamente nula (cambios casi imperceptibles). Esta situación se presenta para la mayoría de las estaciones

GPS; sin embargo, las series de POVA y TUCO en la componente vertical y de TUCO en la componente de longitud presentan una pauta diferente, porque simulan de una manera más detallada la tendencia real de movimiento de la estación GPS. Los valores de pronóstico obtenidos van de la semana G1638 a la G1690. Los pronósticos se pueden visualizar en las figuras 5, 6 y 7 para las componente latitud, longitud y vertical, respectivamente.

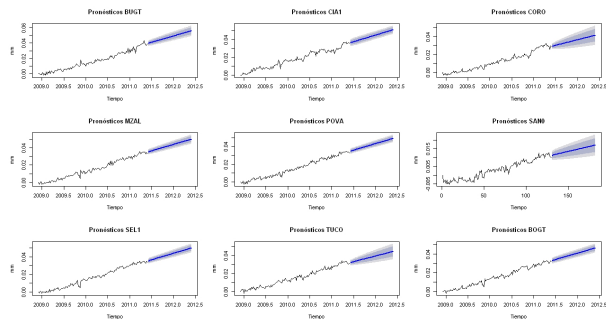


Figura 5. Pronósticos componente *latitud*.

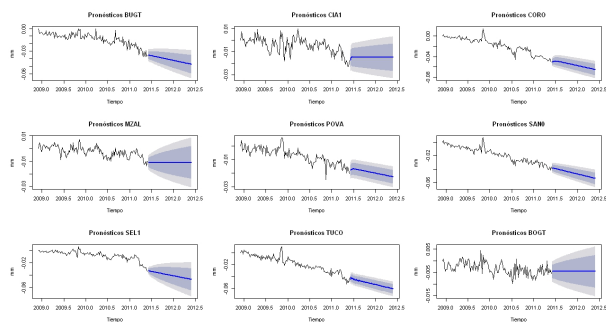


Figura 6. Pronósticos componente *longitud*.

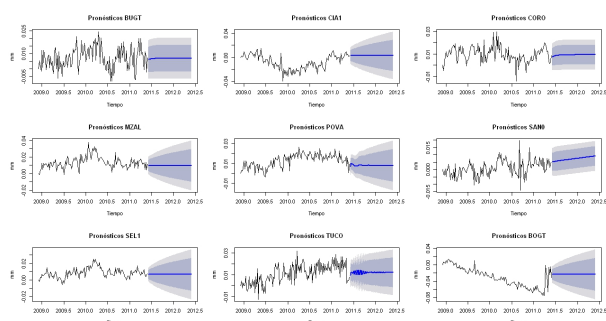


Figura 7. Pronósticos componente *vertical*.

En el análisis multivariado, las pruebas de raíces unitarias aplicadas para determinar estacionariedad en el conjunto de series indican claramente el incumplimiento del supuesto para los datos originales. Debido a esto, se diferenciaron las series individualmente, con lo que se obtiene estacionariedad de segundo orden para el vector de series. Por otra parte, la prueba general de Granger para determinar causalidad acepta la hipótesis nula de que no existe

influencia de los valores pasados de una serie en los valores actuales de las demás series a un nivel del 1%.

La metodología de Box-Jenkins aplicada al vector de series estableció modelos vectoriales autorregresivos de segundo orden $VAR(2)$ para las nueve estaciones GPS bajo análisis. En el caso de la componente vertical, los coeficientes estadísticamente relevantes constituyen un total de 47: 20 de ellos de primer orden y 27 de segundo orden, indicando que el valor actual de una serie es dependiente en una forma importante de los dos valores pasados de las demás series. Se resalta la existencia de relaciones directas significativas de primer orden entre la estación BUGT con las estaciones de CIA1, POVA y TUCO, y de segundo orden con las estaciones SAN0 y TUCO.

Para la componente latitud se tienen 12 coeficientes autorregresivos de primer orden y 15 de segundo orden, estadísticamente significativos. En esta componente, la asociación directa parece ser muy fuerte entre la serie bajo análisis consigo misma, lo que podría significar una dependencia temporal del valor actual de una serie respecto a sus propios valores pasados, y en menor medida de los valores pasados de las series de otras estaciones.

El modelo $VAR(2)$ estimado para la longitud contiene 37 coeficientes relevantes, de los cuales 24 son de primer orden y 13 de segundo orden, lo que permite inferir que el valor presente de la componente longitud para una serie cualquiera está en función de su valor inmediatamente anterior. Se reflejan asociaciones importantes entre las estaciones BUGT con SEL1 y BOGT, CIA1 con BUGT y TUCO, MZAL con SEL1, TUCO con POVA, SEL1 y SAN0, y BOGT con CORO, POVA y SAN0. De esta manera, se puede establecer como hipótesis preliminar para futuros trabajos en la materia que el movimiento de las estaciones geodésicas GPS está asociado con sus respectivas localizaciones geográficas, además de factores geológicos, tectónicos, de presión litosférica y otros, que pueden enriquecer los análisis aquí presentados.

Los modelos estimados fueron verificados a partir de la comprobación de los supuestos que, en general, arrojan resultados estadísticamente aceptables para optar por emplearlos en pronósticos a corto plazo. Los resultados de la prueba de correlación serial entre los errores fueron el criterio de mayor peso para la utilización del modelo. Las pruebas de estabilidad o ajuste general mostraron un buen comportamiento de los modelos para cada una de las variables en consideración, con lo que se asume que son correctos para su utilización en pronósticos. Los pronósticos multivariados muestran una pauta mucho más realista que en el caso univariado, ya que permiten observar unas oscilaciones del valor de la componente medida mucho más significativas desde el punto de vista gráfico; sin embargo, también se percibe un poder predictivo de muy corto plazo que se pierde gradualmente en función de los intervalos de confianza, que —como es de esperar— son más amplios a

medida que se aborda el horizonte de predicción establecido. En las figuras 8, 9 y 10 se pueden apreciar gráficamente los resultados de pronóstico.

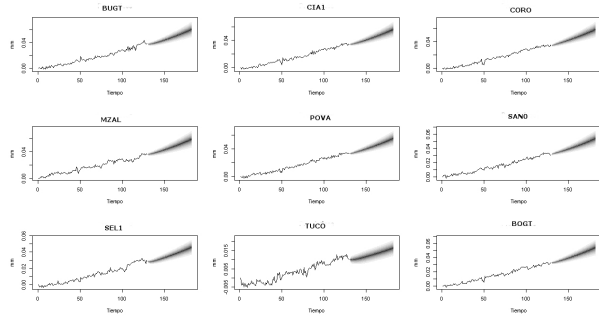


Figura 8. Pronósticos modelos multivariados para la componente de latitud.

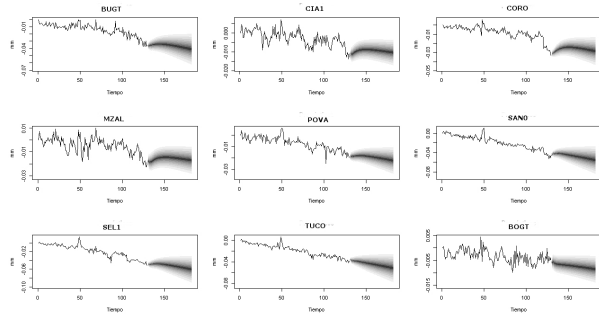


Figura 9. Pronósticos modelos multivariados para la componente de longitud.

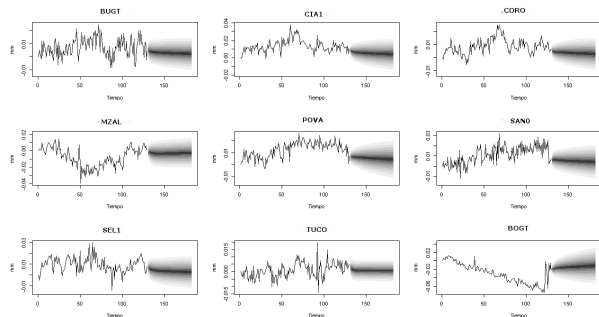


Figura 10. Pronósticos modelos multivariados para la componente de vertical.

6. Conclusiones

Se analizaron nueve estaciones GPS de la red activa del proyecto GeoRED en Colombia, con el propósito de identificar y desarrollar modelos de series temporales para los datos de posición en las componentes geodésicas de latitud, longitud y vertical.

De acuerdo con la metodología aplicada, se desarrollaron modelos univariados que representan de forma general el mecanismo generador de los datos bajo estudio.

Para las componentes de latitud y longitud, los modelos de series univariadas definitivos fueron integrados de media móvil o $IMA(d, q)$. Para la componente vertical, los modelos obtenidos son diversos, presentándose así procesos $ARIMA(p, d, q)$, $ARMA(p, q)$ e $IMA(d, q)$ que se asocian con eventos o fenómenos de carácter regional y/o local y los cuales, no son generalizados para las estaciones geodésicas trabajadas.

Por otra parte, los modelos multivariados desarrollados fueron procesos vectoriales autorregresivos de segundo orden $VAR(2)$. Los resultados de su estimación y diagnóstico permitieron configurar de forma conjunta la información derivada de las nueve estaciones GPS y, con ello, la determinación de relaciones dinámicas significativas entre las diferentes series, las cuales parecen estar asociadas con la ubicación geográfica y geológica de las estaciones permanentes en el territorio colombiano, aspecto que requiere de mayor estudio.

Se realizó la estimación de los parámetros para los modelos de series temporales univariados y multivariados con base en los métodos de máxima verosimilitud condicional (caso univariado) y mínimos cuadrados ordinarios por ecuación (caso multivariado). Dichas estimaciones se diagnosticaron con base en el análisis de los residuos, los cuales proporcionaron las herramientas necesarias para implementar los modelos en pronósticos, buscando determinar el comportamiento a futuro de la variable posición en las componentes de latitud, longitud y altura para un periodo de 52 semanas.

Se identificaron las tendencias de movimiento existentes en las estaciones GPS en cada componente geodésica con base en los pronósticos de corto plazo elaborados. Para las componentes de latitud y longitud, el movimiento recurrente de las estaciones se presentó en sentido Norte (N) y Este (E) respectivamente, mientras que para la componente vertical, la tendencia observada es particular a cada estación y no muestra el mismo grado de generalización que en las componentes horizontales. Por lo anterior, se requiere mayor investigación en el modelamiento de la componente vertical que contribuya a la definición precisa de la tendencia del movimiento y de las posibles causas que lo originan.

A pesar de que los modelos univariados presentan un poder de predicción limitado, la tendencia del movimiento se conserva, siguiendo el comportamiento habitual enmarcado por los datos objeto de análisis. La limitación de los pronósticos está vinculada con la clase de modelos identificados y estimados, que, en general, presentan componentes de media móvil o de memoria corta caracterizados por restringir el valor actual a la influencia de pocos valores precedentes. Por otra parte, los modelos multivariados representan de forma más realista el movimiento espacial de las estaciones GPS a futuro, porque además de considerar la tendencia característica, introducen el elemento de variabilidad inherente a los procesos estudiados. En general, estos modelos permitieron confirmar el comportamiento en la

tendencia de los datos GPS en las diferentes componentes geodésicas, lo cual resalta la existencia de estructuras de dependencia temporal permanentes y la evolución habitual que sigue el movimiento tridimensional de los puntos examinados.

7. Reconocimientos

Al Proyecto GeoRED del Servicio Geológico Colombiano al proporcionar la información básica para el estudio preliminar del análisis del movimiento de las estaciones geodésicas mediante el análisis de series de tiempo, lo cual permitió la realización del trabajo de grado en Ingeniería del primer autor de este artículo.

El Proyecto GeoRED es un proyecto de investigación y desarrollo basado en instrumentación, ejecutado por la Dirección Técnica de Amenazas Geológicas del Servicio Geológico Colombiano. Es resultado de consideraciones de pertinencia técnica, científica, social, económica y política, orientado a incrementar el grado de conocimiento de la geodinámica de la esquina noroccidental de Suramérica, el cual es a su vez insumo esencial en la gestión del riesgo por fenómenos naturales.

Referencias

- Box, G., Jenkins, G. & Reinsel, G. (2008). Time series analysis: Forecasting and control (4th ed.). New Jersey: John Wiley. ISBN 978-0-470-27284-8.
- Brockwell, P y Davis, R. (1987). Time series: theory and methods. New York: Springer. ISBN 3-540-96406-1.
- Gilgen, H. (2006). Univariate Time Series in Geosciences - Theory and Examples. Springer. ISBN: 978-3-540-23810-2
- Gubbins, D. (2004). Time series analysis and inverse theory for geophysicists. Cambridge. ISBN 978-0-521-52569-5.
- Honaker, J., King, G. y Blackwell, M. (2011). Package Amelia II: A program for missing data. The R Project for Statistical Computing, Institute for Statistics and Mathematics, Vienna University of Economics and Business. [Versión electrónica]. Recuperado de: <http://cran.r-project.org/web/packages/Amelia/vignettes/amelia.pdf>
- Lütkepohl, H. (2005). New introduction to multiple time series analysis. Berlín: Springer. ISBN 3-540-26239-3.
- Peña, D. (2005). Análisis de series temporales. Madrid: Alianza. ISBN 84-206-9128-3. The Comprehensive R Archive Network. (2012). Disponible en: cran.r-project.org.
- Tsay, R. (1988). Outliers, level shifts and variance changes in time series. Journal of Forecasting. Recuperado de: <http://www.unc.edu/jbhill/tsay.pdf>



