

La Geoinformación al Servicio de la Sociedad

# Memorias



Sociedad Latinoamericana en Percepción Remota y Sistemas de Información Espacial Capítulo Colombia



Medellín, Colombia 29 de Septiembre al 3 de Octubre de 2014

### DESARROLLO DE UNA RED NEURONAL MULTICAPA PARA LA PREDICCION DE ALTURAS NIVELADAS A PARTIR DE MEDICIONES GPS Y GRAVIMETRIA.

#### Miguel Fernando Arias Patiño<sup>1</sup>, Luis Hernán Ochoa Gutiérrez<sup>2</sup>

Palabras Claves: Altura elipsoidal, Aprendizaje de Máquina, Geodesia, Gravimetría, Nivelación Geométrica, Redes Neurales Artificiales, Sistemas de referencia.

El modelamiento de la relación entre los valores de Gravedad Observada y las Alturas Geométrica y Elipsoidal resulta ser un problema complejo que relaciona variables altamente correlacionadas pero que en la práctica no resulta ser fácilmente resuelto. Es por esto que se presenta una alternativa basada en algoritmos de aprendizaje de máquina que permiten encontrar relaciones para estimar una de las variables con base en las otras, mediante el desarrollo de una red neuronal multicapa.

Cómo insumos se tienen los datos de gravimetría, nivelación geométrica y posicionamiento GPS, del Instituto Geográfico Agustín Codazzi – IGAC. Inicialmente, se realizó el procesamiento de la información cruda, aplicando las correcciones y ajustes pertinentes; con el fin de consolidar el conjunto de datos de aprendizaje y evaluación.

Para el entrenamiento se tuvieron en cuenta dos conjuntos de datos: Uno compuesto con valores reales para las variables de estudio y otro con valores interpolados de gravedad por medio de un modelo geoestadístico cuya fuente son los modelos globales geopotenciales.

<sup>&</sup>lt;sup>1</sup> Ingeniero Catastral y Geodesta, Candidato a Magister en Geomática mifariaspa@unal.edu.co

<sup>&</sup>lt;sup>2</sup> Ingeniero Civil, MsC. Geofísica, MsC. Geomática, Candidato a Doctorado en Sistemas y Computación lhochoag@unal.edu.co

La predicción obtenida basada en los datos, arrojó valores de correlación bastante altos, permitiendo su aplicación dentro del territorio Colombiano.

#### ABSTRACT

Key Words: Ellipsoidal height, Machine Learning, Geodesy, Gravimetry, Spirit Levelling, Artificial Neural Network, Reference Systems.

Modeling the relationship between Observed Gravity, ellipsoidal and leveled heights, is a very complex problem that is not easy to solve because of the high correlation between the variables. The main of this article is to propose a new alternative of Machine Learning algorithm, that makes possible to find data based on the relationship between the variables, implementing a Multilayer Neural Network.

The training data used, is a set of Observed Gravity Spirit Levelling and GPS positioning, obtained for each surveyed station by the Instituto Geografico Agustin Codazzi – IGAC The raw data was previously calculated and adjusted according each case. Then, it was consolidated in a database to be used as knowledge base for the model training and evaluation.

There were considered two sets of data. The first one consists in stations where the variables were determinated directly in the field and the second one, where gravity data was not available; this was determined by interpolation from a geoestatistical model with the source data from global geopotential models. Correlations found were so high and it makes possible to be applied for the Colombian Territory.

#### **INTRODUCCIÓN**

En los últimos años, el avance de las tecnologías geoespaciales, ha cambiado de manera radical la percepción del mundo. La posibilidad de obtener grandes precisiones a la hora de medir las variables geométricas y físicas de la Tierra, ha permitido determinar modelos fieles a las dinámicas propias del planeta. Una de las ramas en la que se evidencia un adelanto significativo, es el posicionamiento con el Sistema global de Navegación por Satélite (GNSS), permitiendo que un usuario no experto, a través de un dispositivo, capture las coordenadas latitud, longitud y altura elipsoidal, de cualquier sitio sobre la superficie terrestre.

Estas mediciones están referidas a un elipsoide, elemento matemático y geométrico, que es una buena representación de la forma y figura de la tierra (Heiskanen & Moritz, 1967). Las mediciones GPS, tienen altas posibilidades de operabilidad, en cualquier hora y condiciones propias del entorno, además de ser una técnica económica que ofrece precisiones de pocos centímetros, por medio de técnicas diferenciales (Hofmann-Wellenhof *et al.* 2003). Sin embargo, la precisión no es tan alta en medidas verticales (alturas elipsoidales).

Al comparar mediciones de alturas referidas al nivel medio del mar, capturadas a través de técnicas clásicas de la geodesia y alturas medidas con dispositivos GPS, se encuentran grandes diferencias, debido a que están medidas sobre dos superficies diferentes: El elipsoide y por otro lado una superficie equipotencial de referencia, que sufre alteraciones y deformaciones por fuerzas físicas (Figura 1). En virtud del auge de las tecnologías GNSS, ha surgido la necesidad de combinar los dos sistemas de referencia (físico y geométrico), con el fin de reducir esfuerzos y obtener mejores mediciones (Sánchez, 2003)



Figura 1. Relación de los sistemas de referencia geométrico y físico. Tomado de Martínez (2011).

La técnica que utiliza la relación entre altura elipsoidal, ondulación geoidal y altura nivelada es conocida cómo nivelación GPS. Para Colombia, la altura obtenida por este método es comparable con alturas obtenidas a partir de Nivelación Trigonométrica, con precisiones de 80 cm, siempre y cuando haya un ajuste con un punto de la red geodésica nacional, con una distancia menor a 20 kilómetros. (Sánchez & Martínez. 1997). La ondulación geoidal se obtiene de la diferencia del elipsoide y la superficie equipotencial, ortogonal al vector gravedad, conocida cómo Geoide, aproximada al nivel medio del mar (Heiskanen & Moritz, 1967)

Debido a la densificación de las redes de nivelación nacionales, hay zonas en las que encontrar una estación con cota conocida tiene una dificultad enorme siendo un obstáculo en el desarrollo de proyectos de ingeniería y estudios geofísicos Por ello es necesario implementar una técnica, con la que se pueda obtener datos de áreas poco muestreadas, teniendo en cuenta las variables, altura nivelada, gravedad observada y mediciones GPS.

De acuerdo a la relación entre las tres variables, es posible determinar un modelo basado en datos, dentro del contexto de la minería de datos espacial. Uno de los algoritmos más utilizados es la red Neuronal Artificial (ANN por sus siglas en inglés) que es un modelo de ordenador que iguala o adopta las funciones básicas del cerebro (Copín, 2004). Representa una serie de características propias del aprendizaje humano, cómo es el conocimiento a través de la experiencia, generalización de ejemplos previos a nuevos casos y abstracción de información a partir de una serie de datos (Basheer & Hajmeer, 2000).

Dentro de una red neuronal artificial se distinguen tres partes esenciales (Figura 2): la capa de entrada que contiene el conjunto de datos de entrenamiento, las capas ocultas, donde se asignan los diferentes pesos y se encuentran las funciones de activación y transferencia. Por último se encuentra la capa de salida que contiene la predicción de variables continuas o clasificación para variables discretas.



Figura 2. Diagrama de una red neuronal Adaptado de Pyle, (1999)

El aprendizaje en una red neuronal artificial se lleva a cabo en dos pasos: en el primero la red procesa el set de datos de entrada y su correspondiente valor de salida, multiplicando por un peso generalmente arbitrario. Los datos ponderados se funden y pasan a la neurona cómo respuesta a las particularidades de las entradas. La red busca las relaciones entre el vector de entrada y hace las estimaciones de los valores de salida, ignorando en primer instancia el valor real. En el segundo paso, la red compara el valor estimado con el valor de salida. Si se encuentran diferencias entre la predicción y el vector de salida, este error se regresa a través de la red, desde el argumento hacia los datos de entrada, reasignando los pesos, ajustando la red haciendo que el error sea menor. Este método se conoce cómo back propagation BP – ANN (Basheer & Hajmeer, 2000).

#### MATERIALES Y MÉTODOS.

Para la predicción de alturas niveladas a partir de información de gravimetría y GPS, se desarrollaron tres grandes pasos (Figura 3):

- Recopilación, unificación y procesamiento de la información gravimétrica, de nivelación y posicionamiento GPS.
- Procesamiento de la información de los modelos globales geopotenciales (GGP por sus siglas en inglés), por medio de un análisis geoestadístico.

✓ Desarrollo, implementación y evaluación del algoritmo de aprendizaje de máquina
Como recursos de software se tienen los siguientes:

 Procesamiento Gravimétrico: GRAVDATA y GRAVNET, desarrollados por el DGFI (Deutsches Geodätisches Forschungsinstitut, Drewes, 1978)

- 2. Minería de datos:
  - KNIME (Open Source, Disponible en http://www.knime.org/)
  - WEKA (Open Source, Disponible en http://www.cs.waikato.ac.nz/ml/weka/ Hall et al. 2009)
- 3. Análisis estadístico:
- R- Project Version 15. (Open Source, Disponible en http://www.r-project.org/ R Development Core Team, 2012)



Figura 3. Modelo conceptual para la predicción de alturas niveladas.

#### Unificación, evaluación y procesamiento de los datos

Los datos provienen del Instituto Geográfico Agustín Codazzi –IGAC. Para obtener el conjunto de datos de aprendizaje, se realizó la consulta de las bases de datos SIGNAR (gravimetría), SAUCE (Nivelación) y GEOCARTO (GPS), con el objetivo de encontrar estaciones que contaran con las tres variables de interés.

Para aumentar el conjunto de datos, se tuvo en cuenta la información cruda capturada por el IGAC a partir del año 2008. Para ello se realizó el procesamiento de 495 datos crudos de gravimetría aplicando las siguientes correcciones implementadas en el software GRAVDATA y GRAVNET (Drewes, 1978).

- Conversión de medidas instrumentales a miligales
- Corrección por Deriva
- Corrección por Mareas Lunisolares (Wenzel, 1994): para el cálculo de los radiovectores Tierra-Sol, Tierra-Luna, se utilizó la altura elipsoidal al igual que las coordenadas procesadas para cada punto.
- Ajuste de las observaciones a la red SIGNAR.

Después de todo el análisis y procesamiento de las diferentes bases de datos, se obtuvo un corpus definitivo de 383 estaciones.

## Modelo Geoestadístico para los Valores de los Modelos Geopotenciales Globales (GGP).

Para las zonas poco muestreadas se tiene cómo fuente alterna de datos los GGP, que permiten obtener un valor aproximado de gravedad y de anomalías, a través de la representación del potencial gravitacional a partir de armónicos esféricos (Balmino *et al.* 1996). De acuerdo a la fuente de datos y el grado de resolución de la función del potencial se seleccionaron cinco modelos para su evaluación a partir del análisis estadístico de los residuales con respecto a los valores de gravedad de la Red SIGNAR (Tabla 1).

Dada la distribución espacial de los valores calculados por los GGP, se desarrolló un modelo Geoestadístico para la generalización del algoritmo de aprendizaje de máquina y su aplicación en el territorio colombiano<sup>2</sup>. Dentro de los procedimientos ejecutados se tiene:

a) Análisis exploratorio de datos espaciales: La distribución de los datos está concentrada hacia los valores altos, siendo una muestra de la distribución de masas dentro del territorio colombiano. Al evaluar la estacionariedad del fenómeno se encuentra que el fenómeno no tiene media ni varianza constante, debido a la relación inversa entre gravedad y altura.

Tabla 1.	GGP	Evaluados
----------	-----	-----------

MODELO <sup>3</sup>	AÑO	GRADO	FUENTE DE DATOS
EIGEN-6C2	2012	1949	S(Goce,Grace,Lageos),G,A
EGM96	1996	360	EGM96S,G,A
GOCO02S	2011	250	S(Goce,Grace,)
dV_ELL_RET2012_plusGRS80	2014	2190	Topography
EGM2008	2008	2190	S(Grace),G,A

- b) Modelo de tendencia: Aplicación de un modelo de regresión de segundo grado, donde la variable explicada es la gravedad y las variables explicativas son las coordenadas Gauss-Krügger, removiendo las tendencias de los datos.
- c) Análisis Estructural: Para la estimación del semivariograma experimental omnidireccional se utilizó el estimador clásico o de Matheron, presentando en la Figura 4a, Se verificó la isotropía del fenómeno a través de los semivariogramas direccionales. Para el ajuste de del modelo teórico de semivarianza se realizó el ajuste a sentimiento, donde se buscan los posibles modelos al igual que los valores iniciales de los parámetros para luego ser usados en los métodos estadísticos de estimación. Para este ejercicio se evaluaron tres modelos: Exponencial, Esférico y Gausiano (Tabla 2)

<sup>&</sup>lt;sup>3</sup> Disponibles en http://icgem.gfz-potsdam.de/ICGEM



Figura 4. Semivariogramas experimentales a) omnidireccional b) direccionales.

Cómo métodos de ajuste de parámetros de los modelos teóricos de semivarianza se desarrollaron los mínimos cuadrados ordinarios (OLS, por sus siglas en inglés) y mínimos cuadrados ponderados (WLS, por sus siglas en inglés) con sus dos variaciones: Número de pares (n-pairs) y el método Cressie. De acuerdo a la Tabla 3 el modelo Gaussiano tiene mayor ajuste por mínimos cuadrados ponderados por número de pares. (Figura 5).

Tabla 2. Parámetros del semivariograma con ajuste a sentimiento

cov.model	sigmasq	phi	tausq	practicalRange
exponential	37903.32	429729.73	0	1287355.22
spherical	23489.38	462162.16	0	462162.16
gaussian	18150.89	178378.38	1601.55	308740.579

d) Predicción espacial Kriging: Se utilizó el método Kriging universal que permite la inclusión de un modelo de regresión en la estimación realizando una predicción lineal cuando se tiene una estructura de la media no estacionaria (Waller & Gotway, 2004).



Figura 5. Modelos Teóricos de semivarianza para los residuales.

MODELO	Método	sigmasq	phi	tausq	Practical	Sum	Error
	Ajuste				Range	Square	
	WLS-	139271	1845474	0	5528547	1.37E+12	3.249234
	npairs						
Exponential	OLS	206271.7	2810015.5	0	8418054	4.30E+06	3.36848
	WLS-	4055691	57518366	0	17230962	24960.17	3.547804
	cressie						
	WLS-	32461.42	656187.86	0	656187.9	1.17E+12	3.32108
	npairs						
Spherical	OLS	35162.9	721175.5	0	721175.5	3803409	3.384238
	WLS -	987489.7	21012915.4	0	21012915	24763.37	3.549064
	cressie						
	WLS-	19971.22	199289.34	2104.00	344933.7	7.83E+11	2.479771
Gaussian	npairs						
	OLS	19758.49	187975.43	1546.35	325351.3	3421219	3.493878
	WLS -	18847.67	171475.26	1237.24	296792.5	30578.96	4.192589
	cressie						

Tabla 3. Parámetros ajustados del modelo teórico de semivarianza.

#### Aprendizaje de máquina

La metodología para el diseño del algoritmo de predicción, consistió en la implementación de las regresiones disponibles dentro del software KNIME Y WEKA, configurando los

parámetros de cada uno de ellos de acuerdo al comportamiento de los datos. La técnica de evaluación para todos los casos fue la validación cruzada con 10 folds con muestreo aleatorio. Cómo variables explicativas se tienen las coordenadas Gauss-Krüger, la gravedad en mGal y la altura elipsoidal y la anomalía de Bouguer. Por último, cómo método de evaluación se realizó una prueba de hipótesis para muestras pareadas (t-paired test), y un análisis Getis Ord verificando la variable dependiente (altura nivelada) y su predicción.

#### **RESULTADOS Y DISCUSIÓN.**

Terminado el análisis, depuración y consolidación de las bases de datos del IGAC, se obtuvo 383 valores (329 entrenamiento, 54 Evaluación; Figura 6). Si bien el conjunto de datos de aprendizaje es pequeño, significativamente es representativo ya que recoge información de zona noroccidental del territorio colombiano, que contiene la zona Andina, la costa atlántica y parte del pacifico.

#### Evaluación de los GGP.

La comparación entre el valor real de la gravedad y el valor calculado por cada uno de los modelos permitió evaluar los GGP seleccionados. Para ello se analizaron los residuales por medio de un test pareado La hipótesis nula (H\_0) es que las diferencias entre el valor real y el calculado son iguales a cero (0). La hipótesis alternativa (H\_1) es que las diferencias son diferentes a cero (0) siendo una prueba de dos colas.



Figura 6. Distribución espacial del conjunto de datos de aprendizaje.

La Tabla 4 contiene los resultados para la evaluación de los diferentes modelos donde el pvalue, es menor que el nivel de significancia (<0.025) siendo evidencia estadística para rechazar la hipótesis nula. Adicionalmente se verifica que el modelo que presenta menor intervalo de confianza y menor media de los residuales es el modelo EGM\_2008.

Tabla 4. Evaluación de los GGP a través de test pareado
---

	Estadís	sticos	Intervalo de c	onfianza al 95%	Estimados	
Modelos	t	p-value	Intervalo	Intervalo Inferior	Media de las diferencias	
Evaluados			Superior			
REAL –	-4.5373	5.78E-06	-5.200898	-2.06277	-3.631834	
GOCO						
REAL -	34.242	2.20E-16	21.20428	18.90799	20.05613	
EIGEN						
REAL -	6.2747	3.67E-10	5.76265	3.019174	4.390912	
EGM2008						
REAL –	38.0309	2.20E-16	22.61438	20.3974	21.50589	
EGM96						
REAL - dv-	-631.6007	2.20E-16	-493.8978	-490.8416	-492.3697	
ELL						

### Modelo geoestadístico para los datos de gravedad relativa calculados a partir de los GGP.

De acuerdo al GGP seleccionado se construyó un modelo geoestadístico con un muestreo de 20000 puntos distribuidos cada 10 km. La predicción obtenida y su varianza se pueden apreciar en la Figura 7. Los valores altos se concentran en la costa Caribe, aproximándose a 978400 mGal, y los valores bajos en las zonas montañosas por el orden de 977400 mGal. Si bien la costa pacífica y la región caribe no presentan diferencias de alturas considerables, presentan variaciones del orden de 400 mGal respecto a los valores de gravedad relativa. Los llanos orientales tienen un comportamiento homogéneo a esta escala, fluctuando entre 977888 y 977700 mGal.



Figura 7. Modelo geoestadistico para la gravedad calculada a partir de GGP (en Miligales)a) Predicción b) Varianzas

#### Desarrollo de la Red Neuronal multicapa para la predicción de alturas niveladas.

Para la predicción del objetivo se desarrollaron pruebas con diferentes algoritmos de regresión basada en sistemas inteligentes y aprendizaje de máquina, siendo la Red Neuronal

multicapa la técnica de mejor resultados. Inicialmente se planteó la red sin incluir las coordenadas cómo variables independientes. Sin embargo los resultados no fueron los mejores, constituyendo un argumento para suponer que la red neuronal es capaz de estimar las relaciones espaciales entre los puntos de muestreo.

Entre las estrategias implementadas para la optimización del aprendizaje se encuentra la normalización de los datos de acuerdo con Basheer & Hajmeer (2000) quienes manifiestan que las funciones de transferencia mejoran su rendimiento con una serie de datos con valores entre 1 y 0. Los resultados evidenciaron que la mejor técnica de normalización es la transformación Z-SCORE. Todas las variables fueron normalizadas a excepción de las coordenadas, ya que de ellas dependen las relaciones espaciales de los diferentes elementos que componen el conjunto de aprendizaje.

El flujo de trabajo en KNIME se presenta en la Figura 8a. Adicionalmente se muestra el diseño de la red neuronal con 5 variables de entrada y 8 nodos ocultos. La validación cruzada se muestra en la Figura 8c en la que se divide el conjunto de aprendizaje en diez subconjuntos por un muestreo aleatorio.

Los parámetros de la red neuronal multicapa son los siguientes:

- Normalización: Método Z-Score
- LearningRate: 0.4
- Training Time: 50000
- Función de trasferencia: Sigmoide y lineal.
- Cross Validation: 10 folds

- HiddenLayers: 8
- Momentum: 0.3
- ValidationThereshold: 20



Figura 8. Desarrollo de la red neuronal multicapa: a) Flujo de Trabajo en KNIME, b) Diseño de la red c) Validación cruzada.

#### Predicción a partir de datos reales.

La Figura 9 muestra la distribución de las diferencias entre el valor real y el valor calculado para la altura nivelada. De acuerdo al gráfico de dispersión de los errores un 90% de las predicciones cumplen con la precisión requerida. Esta tendencia se ve reflejada en el error medio cuadrático que es de 0.82 para el conjunto de entrenamiento, siendo un indicador de la bondad del modelo planteado. Los residuales superiores al metro pueden corresponder a anomalías locales, que no son capturadas plenamente por el modelo de predicción, que corresponden a los puntos rojos cafés y amarillos (Figura 9).

Para verificar si se presentan patrones de concentración de los residuales altos, se desarrolló una prueba estadística Getis Ord.



Figura 9. Distribución Espacial y dispersión de los errores de la predicción con la red neuronal multicapa.

La Figura 10 refleja la distribución espacial de los valores del estadístico Getis Ord (z value), donde los valores altos (>2) muestran las agregaciones de los valores altos de la variable. De Acuerdo a esto hay una concentración significativa de residuales altos en la zona del Valle del Cauca al igual que en la cordillera oriental en el departamento del Huila.



Figura 10. Mapa de conglomerados y mapa de significancia

Por último se evaluó la generalización de la red neuronal multicapa, aplicando la predicción sobre el conjunto de datos test (54 datos), teniendo un error medio cuadrático de 2.40 m, que si bien puede ser muy alto, es consecuencia de la presencia de tres predicciones por encima del metro. En este sentido el 95% de los valores simulados mantienen la precisión submétrica (Figura 11).



Figura 11. Distribución espacial de los residuales de predicción (en metros) para el conjunto de datos test.

#### Predicción a partir datos interpolados.

Por medio de la interpolación de los valores de gravedad a partir de los modelos globales geopotenciales, se obtuvo un conjunto de 1351 valores. Para el diseño de la red se realizó un muestreo aleatorio simple en el que se subdividió el total de datos en dos subconjuntos con el 80% para aprendizaje (1068) y un 20% para pruebas (267). La distribución espacial y los errores de predicción de los dos subconjuntos se puede observar en la Figura 12.

El error medio cuadrático para el conjunto de datos de entrenamiento es de 0.445 metros y para el conjunto de datos test es de 0.526 m, siendo evidencia de la bondad de la predicción de la red neuronal, ya que los errores del modelo, en promedio, son submétricos, cumpliendo así los requerimientos de precisión.



Figura 12. Distribución los errores de predicción para los datos interpolados; conjunto de entrenamiento (a) y prueba (b).

#### CONCLUSIONES.

La implementación de algoritmos de inteligencia artificial y de aprendizaje de máquina en ciencias de la tierra, cada vez tiene mayor auge, dada la eficiencia en los resultados, al igual de la precisión alcanzada. No obstante, la calidad de los datos de entrenamiento interviene significativamente en los resultados obtenidos.

Los modelos Globales Geopotenciales constituyen una nueva fuente de datos para estudios de tipo geofísico y geodésico; La aplicación de los mismos ha sido reducida en el país. Su

disponibilidad para todo el territorio Colombiano representa una ventaja considerable ya que permiten obtener información de zonas sin captura de datos

Las relaciones entre los dos sistemas de referencia (físico y geométrico), fueron capturadas por una red neuronal multicapa, en las que se estableció la correspondencia entre las variables de entrada: Altura Nivelada, Altura Elipsoidal, Gravedad, Anomalía Bouguer y las coordenadas de las estaciones. Estas últimas permiten modelar las relaciones espaciales entre los datos. Si bien no es un desarrollo geoestadístico formal, la red neuronal desarrollada con las coordenadas obtuvo mejores resultados que una que no contenía este tipo de información espacial.

#### BIBLIOGRAFÍA

- Balmino, G., Sabadini, R., Tscherning, C., & Woodworth, P. (1996). Modern concepts, concerns and satellite projects in the determination and use of the Earth's Gravity Field. Recuperado el Agosto de 2013, de ESA Publications: http://cct.gfy.ku.dk/
- Basheer, I., & Hajmeer, M. (2000). Artificial neural networks; fundamentals, computing, desing and application. Journal of Microbiological Methods, 43, 3-31.
- Coppin, B. (2004). Artificial Intelligence Illuminated. Sadbury: Jones and Bartlett Publishers.
- Drewes, H. (1978). Zur Ausgleichung von Gravimeternetzen. ZfV, 485-496.
- Guo, D., & Mennis, J. (2009). Spatial data mining and geographic knowledge discovery— An introduction. Computers, Environment and Urban Systems, 403-408.
- Hall, M., Eibe, F., Holmes, G., Pfahringer, B., Reuteman, P., & Witten, I. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations, 11(1).
- Heiskanen, W., & Moritz, H. (1967). Physical Geodesy. San Francisco: W.H. Freeman.
- Hofmann-Wallenhof, B., Legat, K., & Wieser, M. (2003). Navigation, Principles of positioning and guidance. Wien Austria: Springer Verlag.

- Martínez, W. (2011). Introducción Sistemas de Referencia en Colombia. Instituto Geográfico Agustín Codazzi-IGAC.
- Pyle, D. (1999). Data Preparation for Data Mining. San Francisco: Morgan Kaufmann Publishers.
- Sánchez, L. (2003). Determinación de la superficie vertical de referencia para Colombia. Dresde: Technische Universität Dresden, Institut für Planetare Geodäsie.
- Sánchez, L., & Martínez, W. (1997). Guía Metodológica para la obtención de alturas sobre el nivel medio del mar utilizando el sistema GPS. Bogotá: Instituto Geográfico Agustín Codazzi.
- Waller, L., & Gotway, C. (2004). Spatial Exposure Data. Applied Statistics for Public Health Data, 272-306
- Wenzel, H. (1994). Earth tide analysis package ETERNA 3.0. Marees Terr. Bull. d'Inf. Bruxules, 8719-8727.