

# DETERMINACIÓN DE LA EXACTITUD VERTICAL DEL MODELO GEOMÉTRICO DEL GEOIDE MGH44 MEDIANTE LA APLICACIÓN DEL ESTÁNDAR DE LA NSSDA

*Ing. Jorge Moya Zamora*

Escuela de Topografía, Catastro y Geodesia  
Universidad Nacional. Heredia, Costa Rica  
jmoya@una.ac.cr

## RESUMEN

El modelo geométrico del geoide MGH44 es el resultado de una comparación directa entre mediciones GPS y de nivelación convencional sobre puntos de una red geodésica ubicada en la zona urbana de 50 km<sup>2</sup> de la ciudad de Heredia, Costa Rica. Con la grilla del MGH44 se obtiene la ondulación del geoide para cualquier punto de esa zona, valor que se puede utilizar para estimar la altura sobre el nivel medio del mar a partir de mediciones de altura elipsoidal con GPS.

En este documento se describen los procedimientos y cálculos realizados para evaluar la calidad vertical del modelo MGH44 por medio de la aplicación del estándar de la National Standard for Spatial Data Accuracy (NSSDA). A través de la generación de una nueva grilla, con solo 36 datos denominada MGH36, se obtuvieron nuevos valores de la ondulación del geoide para los restantes 20 puntos escogidos como control. En el procesamiento de la información se aplicaron diferentes algoritmos para corroborar si los datos de los 20 puntos de control siguen una distribución normal y, además, verificar que en este conjunto no se tuvieran errores groseros. El valor promedio de la ondulación del geoide de los puntos de control es de 14,287 m y el cálculo según el estándar de la NSSDA brindó una exactitud vertical de los datos de  $\pm 0,045$  m. Posteriormente, por medio de la técnica de Bootstrap, se calcularon con un 95% de probabilidad los valores 14,233 m y 14,353 m como límites del intervalo de confianza del promedio.

**Palabras clave:** Estándar NSSDA, Modelo MGH44, Criterio de Chauvenet, Test de Grubbs, Bootstrap

## ABSTRACT

The geometric model of the geoid MGH44 is the result of a direct comparison between GPS measurements and leveling points on a conventional geodetic network located in the urban area of 50 km<sup>2</sup> of the city of Heredia, Costa Rica. With the grid MGH44 obtained from the undulation of the geoid to any point in that area; value that can be used to estimate the height above mean sea level measurements from GPS ellipsoid height.

This paper provides a description of the procedures and calculations to evaluate the quality of the model vertical MGH44 by applying the standard of the National Standard for Spatial Data Accuracy (NSSDA). Through the creation of a new grid with only 36 named MGH36 data were obtained new values of the undulation of the geoid for the remaining 20 chosen as a control. In the information processing algorithms were applied to verify whether the data of 20 monitoring points follow a normal distribution and also verify that this set errors were not rude. The average value of the undulation of the geoid of the control points is 14,287 m and the calculation according to the standard of NSSDA gave a vertical accuracy of the data of  $\pm 0045$  m. Later through the Bootstrapping technique, was estimated at 95% probability values of 14,233 m and 14,353 m limits of the confidence interval of the average.

**Keywords:** NSSDA standard, MGH44 model, Chauvenet criterion, Grubbs test, Bootstrap

## INTRODUCCIÓN

En este documento se describen las actividades y los procedimientos realizados en el

control de la exactitud posicional del modelo geométrico del geode MGH44, mediante la aplicación de test de la National Standard for Spatial Data Accuracy (NSSDA). De acuerdo con Ariza (2002), este procedimiento es de acatamiento obligatorio para las agencias federales de los Estados Unidos que se dediquen a labores de producción cartográfica, ya sea digital o analógica.

El conjunto puntual examinado en este trabajo consiste en datos obtenidos durante el proyecto de investigación “Determinación Empírica de la Ondulación del Geode en los Alrededores del Vértice ETCG”, desarrollado por (Moya y Dörries, 2005), para la Escuela de Topografía, Catastro y Geodesia de la Universidad Nacional, Costa Rica. La red de este proyecto está constituida por 56 puntos, sobre los que se realizaron mediciones GPS y mediciones de nivelación convencional, se obtuvo el valor de la altura ortométrica [H] (referida al nivel medio del mar) y el valor de la altura elipsoídica [h] (referida al elipsoide WGS84).

La magnitud que relaciona ambas coordenadas altimétricas es la ondulación del geode denominada como [N], y su determinación es uno de los temas más importantes dentro del campo de estudio de la geodesia y particularmente dentro de la geodesia física. Sin embargo, no se desarrollarán aquí los conceptos teóricos que dan fundamento al problema para la determinación de la ondulación del geode. Solo se mencionará que en la actualidad existen modelos matemáticos locales, regionales y globales que permiten obtener por interpolación y otras técnicas el valor de N como los modelos Carib97, Mexico96, EGM96.

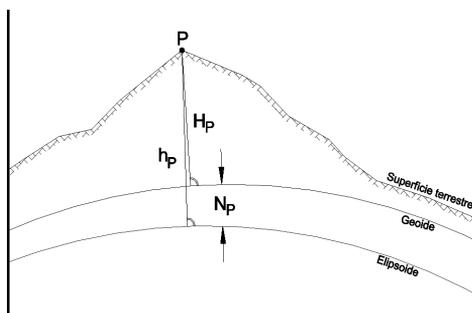
## EL ORIGEN DE LOS DATOS

La ondulación del geode [N] es la distancia lineal entre las superficies del elipsoide y del geode, medida a lo largo de [h], y brinda información sobre cómo se comporta el geode con respecto al elipsoide. Por su naturaleza, ambas superficies no son paralelas; el geode es una superficie de carácter físico que responde a la distribución heterogénea de las masas internas de la Tierra, por lo tanto su curvatura es irregular y resulta imposible representarla matemáticamente.

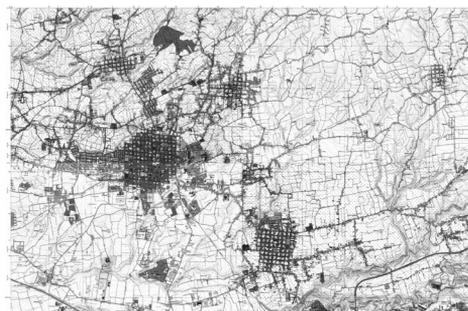
Por su parte, el elipsoide es una superficie de revolución, que se obtiene de hacer girar una elipse alrededor de su eje menor, y genera un sólido de curvatura regular y definida. Matemáticamente las tres magnitudes se relacionan de la siguiente manera:  $h = N + H$ .

En la figura 1 se muestran estos tres valores para un punto en la superficie terrestre.

El objetivo fundamental del proyecto académico “Determinación empírica de la ondulación del geode en los alrededores del vértice ETCG”, según Moya y Dörries (2005) consistió en comparar los valores de ondulación del geode dados por el modelo regional Carib97 y los valores obtenidos en forma “directa” por medio de las mediciones GPS y la nivelación convencional.



**Figura 1. Altura elipsoídica, altura ortométrica y ondulación del geode**



**Figura 2. Zona de estudio del proyecto geode**

La zona de trabajo, de unos 50 km<sup>2</sup>, cubre la parte central urbana de la provincia de Heredia, Costa Rica, que equivale a unas cuatro hojas cartográficas a escala 1:10000, ver figura 2.

Los resultados de esa investigación demostraron que el valor promedio de la ondulación del geode en la zona de trabajo, según el modelo Carib97, es de  $+13,498 \text{ m} \pm 14 \text{ mm}$ , mientras que la ondulación del geode a través de las mediciones del proyecto dio un valor promedio de  $+14,266 \text{ m} \pm 16 \text{ mm}$ . Para este trabajo final, se toma como conjunto de datos los valores de las ondulaciones obtenidos por el cálculo directo.

## EL MODELO MGH44

En el marco del proyecto de investigación, y luego de verificar que existe una diferencia sistemática de unos 76 cm entre los dos “geoides”, se decide crear un modelo geométrico del geode para la zona de estudio denominado como MGH44, siglas del Modelo del Geode para Heredia, zona de 4 X 4 minutos de arco.

**Cuadro 1. Nombre, coordenadas y ondulación del geode de los puntos del MGH36**

N°	Nombre	Norte [m]	Este [m]	N [m]	N°	Nombre	Norte [m]	Este [m]	N [m]
1	ALTO	221552.947	528074.511	14.284	29	LORE	222549.201	528657.913	14.149
2	ARMO	222887.364	522025.261	14.156	30	MABE	220464.406	525824.995	14.336
3	AUTO	221047.016	523514.585	14.238	31	MENO	221071.887	522105.720	14.129
4	BARV	222711.111	522754.929	14.203	32	MOMO	223747.566	528463.511	14.157
5	BENA	219556.482	522797.125	14.242	33	NEXO	220776.729	527352.612	14.352
6	BENE	222127.069	524463.079	14.256	34	PAST	219765.837	524662.597	14.326
7	BM04	220234.473	522973.651	14.196	35	PG01	222441.575	526133.688	14.061
8	BOMB	220316.262	522780.834	14.208	36	PG02	222741.446	527017.800	14.107
9	BRAC	217617.092	529093.693	14.571	37	PINO	220851.123	528411.152	14.313
10	CARA	223703.065	521386.370	14.037	38	PIRO	219798.064	523888.743	14.282
11	CARB	221119.346	522805.976	14.164	39	PISI	221472.734	525120.360	14.288
12	CEME	217095.401	526638.765	14.519	40	QUIZ	218544.776	526923.367	14.451
13	CHAN	218935.647	525707.845	14.422	41	RIBE	220515.135	526712.201	14.351
14	CIBE	220125.491	523720.016	14.244	42	ROK1	221943.000	521335.780	14.055
15	COMA	220517.909	523617.812	14.226	43	ROSA	219411.693	525581.866	14.375
16	COME	222648.960	524855.269	14.315	44	ROSI	217492.962	525657.754	14.441
17	CORA	220838.648	524009.111	14.220	45	SAPE	223607.380	521177.690	14.092
18	CUBU	220377.455	522260.391	14.133	46	SOCO	217789.164	528651.559	14.546
19	D012	218532.852	525985.445	14.443	47	SPIN	220020.037	525778.265	14.343
20	ETCG	220316.053	524723.072	14.341	48	SRC0	223667.469	527602.314	14.094
21	GUSA	217468.556	522393.804	14.286	49	TANQ	219110.830	524634.534	14.317
22	IRA1	220419.190	525326.035	14.303	50	TOMI	218882.599	528231.759	14.493
23	IRA2	220712.158	525288.412	14.306	51	TOP6	223057.242	527680.351	14.148
24	ISAJ	221762.269	524716.579	14.260	52	TRIA	220527.071	521230.586	14.106
25	ISLA	220905.375	524764.331	14.276	53	TRIV	221494.318	524595.745	14.239
26	KATU	219455.532	521931.079	14.168	54	UNAB	216911.028	522298.778	14.257
27	LAGU	216836.061	522822.440	14.254	55	VALE	216901.000	523881.270	14.341
28	LITE	220475.234	524081.531	14.262	56	VETE	217801.108	522448.080	14.355

El modelo MGH44 se generó haciendo uso del programa de interpolación propio del Carib97 denominado geoid.exe, escrito en ForTran y fundamentalmente consistió en crear una nueva grilla con los datos obtenidos de la investigación. Los datos de esta superficie son las ondulaciones del geoide de los 56 puntos. Para utilizar el programa de interpolación del Carib97 con esta nueva grilla, fue necesario desarrollar una serie de pasos previos, porque se debe generar un archivo en formato ASCII y/o binario, leerlo con el programa geogrd.exe para generar la grilla en el formato binario respectivo de Carib97 y luego editar el archivo de texto area.par para utilizar efectivamente la grilla MGH44.

### LOS DATOS DE PARTIDA

En el cuadro 1, se enlistan los 56 puntos originales que se usaron como insumo para la generación del MGH44. Las columnas indican número, nombre del punto, las coordenadas norte y este en la proyección Lambert para Costa Rica, así como el valor de la ondulación del geoide de acuerdo con los resultados de las mediciones directas.

### EL MODELO AUXILIAR MGH36

De acuerdo con Ariza (2002), la aplicación de cualquier estándar requiere la comparación de dos conjuntos de datos, uno de los cuales debe ser de mayor exactitud. El conjunto de puntos de control es un aspecto controvertido, sin embargo, las recomendaciones hablan de al menos 20 puntos independientemente del tamaño del área por controlar, y con una distribución lo más homogénea posible (MPLMIC, 1999) en Ariza (2002).

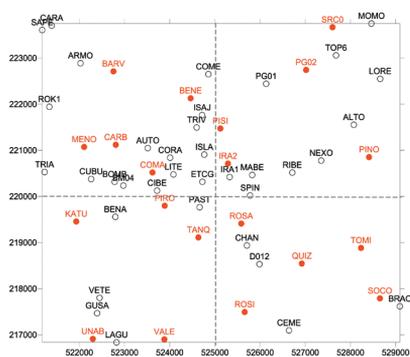
Para determinar la exactitud vertical del modelo MGH44 de acuerdo con las pautas del estándar de la NSSDA, se seleccionaron 20 puntos como control, y con los restantes 36 se generó un nuevo modelo denominado como MGH36. El procedimiento para la generación de esta superficie consistió en:

- Considerando las recomendaciones de ubicación y espaciamiento de los puntos de control, se escogieron los 20 puntos cuya ubicación (marcada con círculos

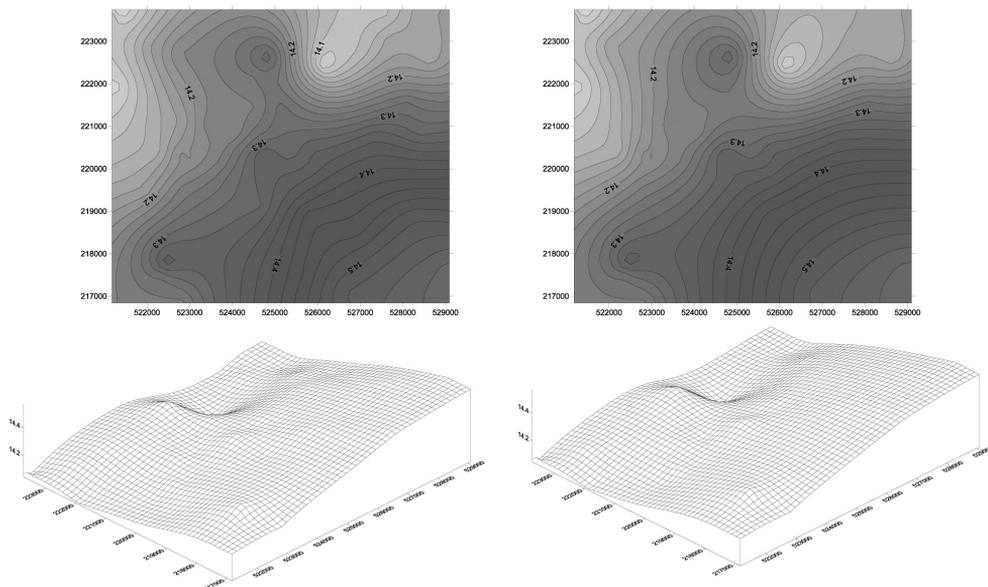
rellenos de color rojo) se presenta en la figura 2.

- Con los 36 puntos restantes se generó un nuevo modelo de ondulaciones denominado MGH36.
- Para su creación se tuvo que hacer una serie de pasos preliminares que coinciden con los descritos en el apartado 3. En primer lugar, se generó una nueva grilla en formato ASCII en el programa Surfer 8.0.<sup>®</sup> con 36 puntos denominada MGH36.grd.
- Este archivo se editó de acuerdo con el formato que pide el Carib97 en cuanto al encabezado y la disposición de los datos asignándole el nombre de MGH36.asc.
- Posteriormente, con el programa geogrd.exe, se convirtió el archivo MGH36.asc en el MGH36.geo, el cual contiene la grilla de 36 puntos en formato binario de Carib97.
- Luego, se editó el archivo de cabecera area.par con el nombre de la grilla MGH36.geo.
- Finalmente, con el programa geoid.exe, se calculó la ondulación del geoide para los 20 puntos de control. Para este paso se creó un archivo ASCII en el formato de entrada propio, con las coordenadas geodésicas y nombre de los puntos. Con el mismo programa, se generó un archivo ASCII de salida con el valor de la ondulación.

En la figura 3 de la siguiente página, se tiene la planta y perspectiva de los modelos MGH44 y MGH36 a izquierda y derecha, respectivamente.



**Figura 3. Distribución de los puntos del modelo MGH36. Los 20 de color rojo son los de control**



**Figura 3. Planta y perspectiva de las superficies de los modelos geométricos MGH44 (izquierda) y MGH36 (derecha)**

**ANÁLISIS PREVIO DE LOS DATOS**

Antes de iniciar efectivamente, con la aplicación del estándar de la NSSDA, se realizó una serie de procedimientos previos de cálculo con el objetivo de conocer fundamentalmente dos aspectos: la función de densidad de probabilidad de los datos (también llamada distribución) y determinar si existen errores groseros en el conjunto de los 20 puntos de control.

Para cada uno de los procesos se elabora una explicación general sobre la aplicación propia del algoritmo, sin embargo se invita al lector a revisar las referencias bibliográficas utilizadas.

**LA DISTRIBUCIÓN DE LOS DATOS**

De acuerdo con Kennedy y Neville (1982), los resultados de un experimento, a menudo, se consideran como elementos de una población, pero no se puede estar completamente seguro de que ese conjunto de datos sigan una cierta distribución estándar, por ejemplo normal. Sin embargo, de acuerdo con experiencias anteriores, cabe suponer que

ciertas observaciones se ajustan a una determinada distribución y la investigación referente a determinar si los datos efectivamente siguen la distribución supuesta, literalmente “...constituye una de las más importantes aplicaciones de la estadística a los problemas de medición.” (Kennedy y Neville, 1982).

Los diferentes cálculos realizados en este trabajo se hicieron considerando dos herramientas: hoja de cálculo de Microsoft Office Excel 2007® y el programa MatLab®.

Los 20 puntos de control se agrupan en el vector de observaciones  $X$  y para iniciar el procedimiento de distribución de los datos, se eligieron un total de  $m = 5$  clases, con un ancho de clase  $dx = 0,100$  m entre los cuales se agruparon los datos con subtotales de 6, 5, 4, 2 y 3, respectivamente. Las medidas de centralización media, mediana y moda se presentan el cuadro siguiente:

**Cuadro 3. Medidas de centralización en [m]**

Media	Desv. Est.	Mediana	Moda
14,287	0,129	14,259	14,258

Con esta información se graficó el correspondiente histograma como indicador primario de la distribución de los datos. En la figura 4 se presenta el gráfico respectivo.

En Gámez y Marín (2000) se describen la medida de simetría y la medida de curtosis como coeficientes para analizar la forma y aplastamiento de la distribución de una serie de datos. Contemplando estos cálculos y la forma del histograma de la figura 5, se puede concluir que el conjunto de datos tiene una distribución “acampanada” asimétrica. Los valores de estos dos parámetros son 0,473 y -0,650 respectivamente e indican que la curva asimétrica tendiente más a valores positivos de forma más aplastada que en relación a la curva normal denominada como platicúrtica.

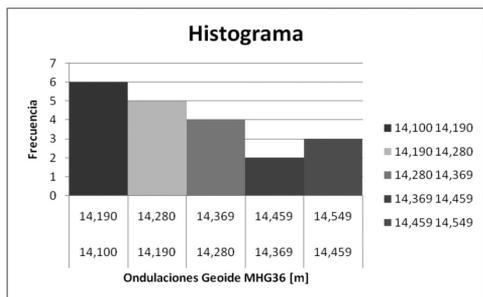


Figura 4. Histograma de los datos

Se ejecutaron también cuatro pruebas de normalidad adicionales al conjunto de datos:

- a. El test de Kolmogorov-Smirnov
- b. El criterio de Lilliefors
- c. El criterio de Jarque-Bera
- d. Descriptivo para determinar normalidad

En el test de Kolmogorov-Smirnov (KS), fundamentalmente, hace una prueba que compara la distribución de frecuencia acumulativa de la distribución teórica con la distribución de frecuencia acumulativa observada. Se determina el punto en el que estas dos distribuciones muestran la mayor divergencia.

En el criterio de Lilliefors (TL) básicamente se tiene una adaptación del test de (KS) y prueba si un

conjunto de datos proviene población normalmente distribuida. El criterio de Jarque-Bera (JB) es un test más sensible que el de Lilliefors y se basa en la medida de simetría y de curtosis siendo altamente sensible a la existencia de posibles errores groseros.

Tanto el test (KS), como el (TL) y el (JB) son rutinas disponibles en ambiente de MatLab® y en el cuadro 4 se muestran los resultados.

Cuadro 4. Resultados de los test KS, TL y JB aplicados al vector de datos

Parámetro/ Test	H	STAT	CV
KS	1	1	0,2941
TL	0	0,1366	0,1900
JB	0	1,3792	5,9915

Las columnas del cuadro anterior indican los parámetros H, STAT y CV, cuya nomenclatura es la usada por MatLab® para identificar con H el resultado de la hipótesis nula; 0 si esta se acepta y 1 en el caso contrario. En la columna STAT está el estadístico o valor de prueba y en la columna CV el valor crítico con el que se compara el valor de STAT.

Los resultados del cuadro 4 no rechazan la hipótesis de normalidad para el test Lilliefors y el test de Jarque-Bera, a diferencia del test de Kolmogorov-Smirnov que suele ser más exigente que los otros dos.

El Método Descriptivo, Mendenhall y Sincich (1997) se basa en el cálculo del cociente entre el intervalo intercuartiles (IQR) y la desviación estándar (s) de la muestra. Este cociente debe ser aproximadamente igual a 1,3 y en el cuadro 5 se muestran los resultados.

Cuadro 5. Resultados del método descriptivo de normalidad

Método descriptivo para normalidad				
Cuartil Superior	Cuartil Inferior	Desv. Est. (s)	IQR	(IQR)/s
14,359	14,184	0,129	0,184	1,426

A pesar de que el cociente obtenido no es exactamente 1,3, se continuó con el método y se graficó la curva de probabilidad normal

(ver figura 5). En Mendenhall y Sincich (1997) se dice que “...si los datos vienen de una distribución normal, una observación será aproximadamente igual a su valor esperado,

por lo tanto una tendencia lineal en la gráfica de probabilidad normal sugiere que los datos provienen de una distribución aproximadamente normal...”.

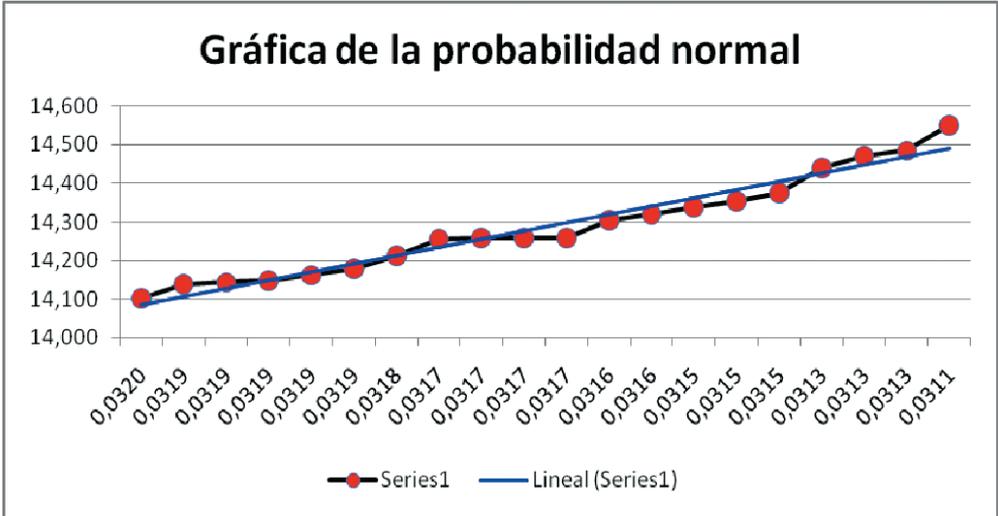


Figura 5. Gráfica de la curva de probabilidad normal

En la figura anterior se muestra dicha curva y se marca la tendencia de los datos con una línea de color azul. La discusión sobre estos resultados se hará al final de este apartado.

Adicionalmente, se realizó un estudio de ajuste de la curva normal de la distribución de frecuencias tal como se describe en Kennedy y Neville (1982). El procedimiento debe satisfacer la condición de que el área bajo la curva es igual al número de observaciones realizadas, que coincide con el perfil de la curva del histograma. Sin embargo, esto se hace en el caso de que se conozca la desviación estándar de la población, pero para muchas aplicaciones es mejor trabajar con probabilidad en vez de la distribución de frecuencias, pues la suma de las probabilidades debe ser uno, y el área bajo la curva de distribución probabilística deber ser también unitaria. La curva “normalizada” está dada por:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Luego de varios supuestos para *m* y de otros cambios de variables, se considera a *s* como la unidad de desviación a partir de la media, por lo que se dice que:

$$z = \frac{x - \mu}{\sigma} = \frac{X}{\sigma}$$

El cambio de variable de *x* a *z* no altera la probabilidad buscada (Kennedy y Neville, 2000), por lo que la probabilidad de que *x*<sub>1</sub> se encuentre entre (*x* y *x*+Δ*x*) es la misma para un valor *z*<sub>1</sub> se encuentre en el intervalos (*z* y *z*+Δ*z*). Por consiguiente se tiene que:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

La ecuación de la curva normal de frecuencias para el conjunto de datos analizados es:

$$y = 62,056e^{-30,246X^2}$$

Con respecto a los cuadros 4 y 5, se puede decir, en primer lugar, que de cuatro pruebas aplicadas para la verificación de la normalidad, tres dieron resultados positivos en el sentido de que el conjunto de datos se puede considerar como una muestra de una población que aproximadamente se distribuye normalmente. Además, Mendenhall y Sincich en (1997) afirman que existen pruebas que proporcionan cierta confiabilidad a la inferencia, sin embargo son muy sensibles a pequeñas desviaciones respecto a la normalidad, y tienen a rechazar la hipótesis de normalidad para cualquier distribución que no sea perfectamente simétrica y con forma de campana.

En el caso del método descriptivo, el cociente de verificación da 1,426, no exactamente 1,30 como lo pide el procedimiento. Sin embargo y considerando que en la bibliografía se hace referencia a un caso en el que cociente da 1,210, se decide continuar con el resto del método. Los resultados posteriores, es decir la graficación de la curva de probabilidad normal (figura 5), muestran que el juego de datos normalizados claramente tiene una tendencia lineal.

Al respecto, en Mendenhall y Sincich (1997) se dice que las verificaciones de normalidad son técnicas fáciles y de gran utilidad, pero son descriptivas, y literalmente dicen que *“Es posible (aunque poco probable) que los datos no sean normales a pesar de las verificaciones se satisfacen razonablemente”*. Se puede decir por lo tanto que es razonable que los datos provengan de una distribución normal, tal como lo requiere la aplicación del estándar de la NSSDA.

Los cálculos anteriores y los resultados permiten estimar una media de 14,287 m y una desviación estándar de 0,129 m, para el conjunto de 20 ondulaciones del geode del vector  $X$ .

$$X \sim N(14,287; 0,129)$$

## ERRORES GROSEROS

El conjunto de datos por evaluar fue sometido adicionalmente a dos pruebas para

localizar datos aberrantes: a) una primera denominada Criterio de Chauvenet y b) el Test de Grubbs.

El criterio de Chauvenet, descrito en Kennedy y Neville (2000), rechaza a una observación de una muestra  $n$  si su desviación de la media es mayor que la que corresponde a una probabilidad de  $1/(2n)$ . Esta prueba se aplica suponiendo una distribución normal de los datos y una varianza obtenida con los mismos datos por evaluar.

En primer lugar se calcula la probabilidad con respecto al total de datos, en este caso 0,025, que corresponde a un valor de 2,241 en una distribución normal. Luego, se calcula el valor de prueba como el valor absoluto de la diferencia de cada valor  $x_i$  menos el promedio  $\bar{x}$ , entre la desviación estándar  $s$  ( $(x_i - \bar{x})/s$ ). Los resultados se muestran en el cuadro 6 y se aprecia como ningún valor “normalizado” es mayor que 2,241.

Se aplicó también el Test para detección de errores groseros de Grubbs, que se describe en Grubbs (1969). La prueba asume una distribución normal de los datos y consiste en detectar posibles errores groseros de uno en uno.

El Test de Grubbs plantea dos hipótesis: la nula  $H_0$  en la que se asume que no hay datos groseros en el conjunto de datos, y la hipótesis alternativa  $H_A$  en la que se asume que al menos existe un dato afectado. El estadístico o valor de prueba se plantea como

$$G = \frac{\max_{i=1,2,3,\dots,n} |Y_i - \bar{Y}|}{s}$$

El estadístico se compara con el límite de la región de aceptación o rechazo, la cual se calcula con base en la cantidad de datos  $N$  y la distribución  $t$  de Student con una probabilidad de  $(\alpha/2N)$  y  $N-2$  grados de libertad:

$$G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{t^2_{\frac{\alpha}{2N}, N-2}}{N-2 + t^2_{\frac{\alpha}{2N}, N-2}}}$$

En el caso del test de Grubbs, se utilizó adicionalmente una rutina programada en ambiente MatLab® (Holtzberg, 2008).

El valor límite de este test para un nivel de incertidumbre del 5%, es de 2,708 y en el cuadro 6 se presentan los resultados en donde se puede ver que ningún valor de  $G_i$  sobrepasa el valor límite.

**Cuadro 6. Resultados del Criterio de Chauvenet y del Test de Grubbs**

Punto	Criterio de Chauvenet	Test de Grubbs
	$ x_i - \bar{x} /s$	$G_i =  Y_i - \bar{Y} /s$
BARV	0,981	0,981
BENE	0,250	0,250
CARB	1,090	1,090
COMA	0,584	0,584
IRA2	0,131	0,131
KATU	0,841	0,841
MENO	1,447	1,447
PG02	1,121	1,121
PINO	0,512	0,512
PIRO	0,227	0,227
PISI	0,219	0,219
QUIZ	1,422	1,422
ROSA	0,683	0,683
ROSI	1,181	1,181
SOCO	2,043	2,043
SRC0	1,160	1,160
TANQ	0,395	0,395
TOMI	1,530	1,530
UNAB	0,227	0,227
VALE	0,248	0,248

**Cuadro 7. Resultados del NSSDA**

Punto	$N_{MGH44}$ [m]	$N_{MGH36}$ [m]	dif [mm]	dif <sup>2</sup> [mm <sup>2</sup> ]
BARV	14,203	14,161	<b>41,9</b>	1753,93
BENE	14,256	14,255	<b>1,0</b>	1,02
CARB	14,164	14,147	<b>17,3</b>	299,29
COMA	14,226	14,212	<b>13,7</b>	187,69
IRA2	14,306	14,304	<b>1,7</b>	2,89
KATU	14,168	14,179	<b>-10,9</b>	119,25
MENO	14,129	14,101	<b>27,7</b>	770,06
PG02	14,107	14,143	<b>-35,7</b>	1274,49
PINO	14,313	14,353	<b>-40,0</b>	1600,80
PIRO	14,282	14,258	<b>24,0</b>	576,00
PISI	14,288	14,259	<b>28,6</b>	817,96
QUIZ	14,451	14,470	<b>-19,4</b>	374,81
ROSA	14,375	14,375	<b>0,4</b>	0,16
ROSI	14,441	14,439	<b>1,8</b>	3,31
SOCO	14,546	14,550	<b>-4,2</b>	17,81
SRC0	14,094	14,138	<b>-43,6</b>	1903,58
TANQ	14,317	14,338	<b>-21,3</b>	453,69
TOMI	14,493	14,484	<b>8,5</b>	72,25
UNAB	14,257	14,258	<b>-0,6</b>	0,40
VALE	14,341	14,319	<b>22,3</b>	498,63
<i>Comprobada para una exactitud vertical de ±0,045 m al 95% de nivel de confianza</i> Exactitud posicional vertical, Ariza (2002), página 164.			<b>Suma</b>	<b>10728,02</b>
			<b>Promedio</b>	<b>536,40</b>
			<b>RMSEz</b>	<b>23,16</b>
			<b>NSSDA</b>	<b>45,39</b>

De acuerdo con los resultados en este apartado se puede concluir el conjunto de datos no presenta errores groseros, por lo que los 20 valores de ondulación del geoide para los puntos de control se puede utilizar en el estándar de la NSSDA.

**EXACTITUD VERTICAL DEL MODELO MGH44**

En este apartado se muestran los resultados al aplicar el estándar de la NSSDA sobre

los 20 puntos de control. El procedimiento de cálculo es relativamente sencillo y se describe en Ariza (2002).

En el cuadro 7 se presenta un resumen del cálculo correspondiente y se concluye que la exactitud vertical del Modelo Geométrico del geoide para Heredia, zona de 4 X 4 minutos (MGH44) es de ± 45,4 mm al 95% de nivel de confianza. Esta exactitud vertical es relativa a todo el modelo MGH44 y hasta cierto punto esperable considerando que la exactitud de

las alturas ortométricas en la generación del modelo MGH44, fue del orden de 3 cm a 5 cm (Moya y Dörries, 2005).

## BOOTSTRAP

De acuerdo con López (2008), la técnica bootstrap se puede usar como un complemento a los resultados después de la aplicación del estándar NSSDA. El remuestreo se puede utilizar cuando no se tienen suficientes muestras de una población o cuando se tiene certeza de la distribución que sigue el conjunto de datos en estudio. El bootstrap se puede utilizar para obtener diferentes parámetros, pero en este trabajo se utilizó para determinar el intervalo de confianza de la media del vector de observaciones  $X$ .

Según Alonso (2002), el método de bootstrap tiene como pieza clave la utilización "extrema" del principio de analogía, que constituye uno de los métodos más simples utilizados para obtener un estimador de un parámetro poblacional.

La idea básica del bootstrap, según Guerrero (2002), "*...es tratar la muestra como si fuera la población, y aplicar el muestreo Montecarlo para generar una estimación empírica de la distribución muestral del estadístico*". Para aplicar el procedimiento de Montecarlo, se requiere del total de la población, que generalmente no se tiene. Solo se cuenta con una muestra y esto hace necesaria la inferencia.

En el bootstrapping, tratamos la muestra como si fuera la población y realizamos un procedimiento del estilo Montecarlo sobre la muestra. Esto se hace extrayendo un gran número de "remuestras" de tamaño  $n$  de la muestra original aleatoriamente y con reposición.

Así, aunque cada remuestra tendrá el mismo número de elementos que la muestra original, mediante el remuestreo con reposición cada remuestra podría tener algunos de los datos originales representados en ella más de una vez, y algunos no aparecerán Guerrero (2002).

Por lo tanto, cada una de estas remuestras probablemente será levemente y aleato-

riamente diferente de la muestra original. Y como los elementos en estas remuestras varían levemente, un estadístico, calculado a partir de una de esas remuestras probablemente tomará un valor ligeramente diferente de los otros y del original.

La afirmación fundamental del bootstrapping consiste en que una distribución de frecuencias relativas de esos estadísticos es calculada a partir de las remuestras, es una estimación de la distribución muestral del estadístico original Guerrero (2002).

Los conceptos y las demostraciones matemáticas referentes al bootstrap no serán tratados aquí, sin embargo, se cuenta con la herramienta de cálculo test1b.m programada en ambiente de MatLab® para el cálculo del intervalo de confianza (Holtsberg, 2008).

En el cuadro 8 se muestran los límites izquierdo (LI) y derecho (LD) del intervalo de confianza para vectores de 1000 y 20 números aleatoriamente generados con MatLab® con el objetivo de conocer a priori la amplitud del intervalo de confianza. Del cuadro se desprende que para un grupo de 20 valores normalmente distribuidos. El tamaño del intervalo de confianza sería de 0,90 aproximadamente.

**Cuadro 8. Intervalos de confianza**

n	LI	M	LD	Amplitud
<b>1000</b>	-0,101	-0,043	0,017	<b>0,118</b>
<b>20</b>	-0,510	-0,055	0,380	<b>0,890</b>
<b>X</b>	14,233	14,287	14,353	<b>0,120</b>

Al someter el vector  $X$  al test, el tamaño del intervalo es prácticamente 7,5 veces más pequeño. Aquí, se decide generar un tercer vector de números aleatorios de tamaño (20,1) y se le calcula su intervalo cuya amplitud es de nuevo 0,890 coincidente con la primera determinación.

Estas pruebas llevan a varias conclusiones: en primer lugar, la rutina trabaja bien ya que al ver los valores medios obtenidos y sus desviaciones estándar, prácticamente coinciden con 0 y 1, respectivamente.

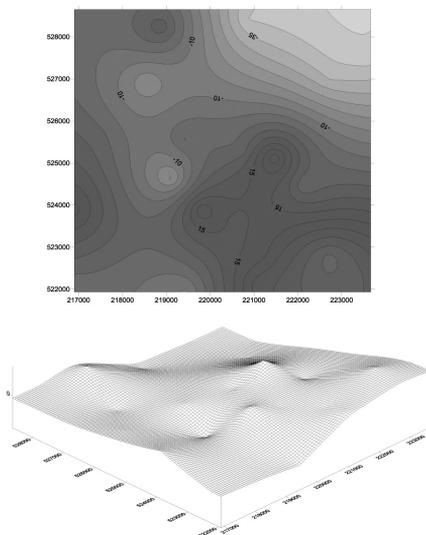
Con respecto al conjunto de datos en estudio, se puede decir que el intervalo de confianza obtenido por medio de la rutina de bootstrap es válido, en el sentido de que al comparar su amplitud con dos conjuntos de valores del mismo tamaño, el intervalo de confianza del vector  $X$  es más pequeño. Esto refleja que el valor central del conjunto de datos de ondulaciones, está contenido entre dos límites cuya amplitud es un 13,5% menor de lo esperado.

## CONCLUSIONES

Con base en los resultados en el tratamiento de este conjunto de datos, se pueden obtener las siguientes conclusiones:

- La distribución de los datos es de forma acampanada, algo sesgada a la izquierda y aplanada, de acuerdo con los parámetros de simetría, curtosis y las de medidas de centralidad.
- Las pruebas estadísticas aplicadas para verificar la normalidad de los datos, indican que no se tiene suficientes elementos como para descartar esta posibilidad. A excepción del test de Kolmogorov-Smirnov, el resto de pruebas tienen a indicar que la muestra posiblemente pertenezca a una población normalmente distribuida, tal como requiere la NSSDA.
- El conjunto de datos adicionalmente está exento de errores groseros con un 95% de probabilidad, de acuerdo con los resultados al aplicar el Criterio de Chauvenet y del Test de Grubbs. Esto también es requerido por el estándar de la NSSDA.
- Las ondulaciones del geoide de los 20 puntos de control sirvieron para determinar la exactitud vertical del modelo geométrico del geoide MGH44 en  $\pm 0,045$  m con un nivel de confianza del 95%. Es importante señalar que este parámetro de exactitud se ha determinado de acuerdo con una serie procedimientos normados y de reconocimiento internacional.

- El valor de casi 5 cm en la exactitud del modelo MGH44 es consistente con la exactitud relativa de las ondulaciones de cada uno de los puntos de la red original (Moya y Dörries, 2005). Este aspecto de nuevo confirma que desde el punto de vista del origen de los datos, la exactitud producto del proceso de nivelación convencional es la que más influye porque es del orden de 10 veces más baja que la obtenida con GPS. Sin embargo a efectos prácticos es suficiente.
- Mediante la aplicación del bootstrap, se pudo determinar con un 95% de probabilidad el intervalo de confianza del valor promedio de la ondulación del geoide de los 20 puntos de control. Este parámetro nos brinda los límites en los cuales se encuentra el valor medio, y en este caso la amplitud del intervalo es de 0,120 m (ver cuadro 8).



**Figura 6. Planta y perspectiva de la diferencias de ondulación del geoide**

- Al hacer la diferencia entre el límite derecho y el promedio; y la diferencia entre el promedio y el límite izquierdo, se obtienen dos valores cuya diferencia es de 0,012 m, lo que permite decir que el intervalo de confianza es prácticamente simétrico.

- Este resultado permite ahora ofrecer un parámetro de calidad adicional a los usuarios del modelo geométrico MGH44. Se cuenta ahora con un valor de exactitud relativa del modelo, es decir para la ondulación del geode determinada precisamente con el modelo MGH44.
  - Se puede ahora, con la exactitud relativa del modelo MGH44, hacer estimaciones de los errores que se puedan alcanzar al efectuar nivelaciones GPS o convencionales dentro del área de influencia del MGH44.
  - Ya desde 2005, Moya y Dörries señalaban como una de las recomendaciones del proyecto “Geoide”, la densificación de algunas de las zonas del área de estudio. Es importante insistir ahora aún más en esa dirección, pues con una densidad mayor de puntos se podría eventualmente volver a aplicar el estándar de la NSSDA y hacer una nueva verificación de la exactitud del modelo MGH44.
  - En la figura 3 se puede ver que las superficies del MGH44 y MGH36 presentan una gran similitud. Sin embargo, se observa que la superficie del MGH36 presenta curvas de iso-ondulación del geode más suavizadas producto de la falta de detalle debida a 20 puntos en la generación de su grilla.
  - En la figura 6 se tiene una planta con curvas de iso-diferencia de ondulación del geode (datos de la columna dif del cuadro 7). Se aprecian diferentes “abultamientos” en toda el área, los cuales son producto precisamente de las dos diferentes grillas utilizadas. Ahora, se ha demostrado que esas diferencias son aceptables al ser menores que la exactitud brindada por el estándar de la NSSDA.
- Ariza, F. J. (2002). *Calidad en la producción cartográfica*. RAMA Editorial. Madrid, España. 389 páginas.
- Biosystems Data Analysis. *Additional Statistical Routines. Grubbs outlier test*. Recuperado en mayo de 2008 de [http://www.bdagroup.nl/downloads/bda\\_downloads.html](http://www.bdagroup.nl/downloads/bda_downloads.html)
- Federal Geographic Data Committee. (1998). *Part 3: National Standard for Spatial Data Accuracy*. US Geological Survey. Estados Unidos. 27 páginas.
- Gámez, A. M. y L. M. Marín. (2000). *Estadística para ingenieros técnicos*. Servicio de publicaciones. Universidad de Cádiz. Cádiz, España. 474 páginas.
- Guerrero, C. (2002). *Aplicación de la simulación regenerativa y la técnica bootstrap, para mejorar la calidad del estimador. Capítulo 4. Introducción a la técnica Bootstrap*. Monografía. Universidad Nacional Mayor de San Carlos. Lima, Perú.
- Grubbs, F. (1969). *Procedures for Detecting Outlying Observations in Samples*. *Technometrics* 11 (1): 1-21.
- Holtsberg, A. *Statistics toolbox for Matlab, Octave, and Matcom/Mideva*. Recuperado en mayo de 2008 de <http://www.maths.lth.se/matstat/stibox/>
- Kennedy, J. B. y A. M. Neville. (1982). *Estadística para ciencias e ingeniería*. Editorial HARLA S.A. México DF, México. 468 páginas.
- López, C. (2008). *Material del curso de doctorado: Calidad de Datos Geográficos*. ETSI en Topografía, Geodesia y Cartografía. Universidad Politécnica de Madrid. Madrid, España.
- Mendenhall, W. y Sincich, T. (1997). *Probabilidad y Estadística para Ingeniería y Ciencias*. 4 Edición. Traducción del título original en inglés “Statistics for Engineering and the Sciences”. Editorial Prentice-Hall Hispanoamericana S.A. México D.F., México. 1182 páginas.
- Moya, J. y Dörries, E. (2005). *Determinación empírica de la ondulación del geode en los alrededores del vértice ETCG*. Escuela de Topografía, Catastro y Geodesia. Informe final de proyecto de investigación. Universidad Nacional. Heredia, Costa Rica.
- National Geodetic Survey NGS. Carib97. Recuperado en mayo de 2008 de <http://www.ngs.noaa.gov/GEOID/CARIB97/>

## BIBLIOGRAFÍA

Alonso, A. M. (2002). *Un ejemplo de bootstrap suavizado*. Revista *Lecturas Matemáticas*. Volumen 23, páginas 11 a 24. Universidad Autónoma de Madrid. Madrid, España.