

Evaluación de la incertidumbre en el cálculo de áreas sobre una determinada cartografía

Ing. Agrim. Rodolfo MENDEZ BAILLO, Uruguay

Palabras claves: método de Monte Carlo, simulación, incertidumbre, nivel de confianza, variograma

SUMMARY

El área de un territorio relativamente extenso, como la de un departamento o de todo el país, se calcula a partir de la cartografía existente.

Existen trabajos al respecto en donde se consignan las áreas del todo el territorio nacional calculadas por diferentes autores, en base a diferentes tipos de cartografía o mapas de base, (Reyes, 1859; Chebataroff, 1960, etc., todos citados en Lazo, 2010). Hay diferencias importantes entre las magnitudes, incluso entre trabajos relativamente modernos, por lo que correspondería dilucidar si las mismas son debidas a los datos, o simplemente están dentro de lo admisible. Se observa, no sólo en esos trabajos, sino también en los mapas oficiales, que no hay estimaciones de los errores cometidos. El aporte de este trabajo consiste en ilustrar cómo calcular las áreas con un nivel de confianza prefijado a partir de un cierto mapa de base, y teniendo en cuenta como única fuente de errores la incertidumbre posicional de las fronteras.

Se tomó una parcela ilustrativa y de forma arbitraria sobre la cartografía. Se tuvieron en cuenta conceptos geoestadísticos para considerar la correlación entre las variables y se hizo una simulación de Monte Carlo perturbando la frontera del polígono.

El resultado de las simulaciones mostró un conjunto de valores del área que no tiene distribución normal, donde mayormente el área se “expande” respecto de un valor de referencia. Se determinó el rango para el cual hay un 95% de nivel de confianza para que el área pertenezca a ese entorno. Para trabajos futuros, se prevé aplicar la misma metodología a gran escala, en particular para el cálculo de las áreas departamentales y/o del país por completo.

Introducción

El conocimiento del área de un país es importante desde el punto de vista de ordenamiento territorial y estadístico, e incluso de defensa nacional.

Tradicionalmente, el cálculo del área se realizaba en base a la mejor cartografía en papel disponible, definiendo sobre la misma un polígono que representa el territorio a medir.

La tecnología ha permitido que los mapas de base hayan evolucionado desde el formato papel al digital, lo que facilita la tarea de cálculo, pero no necesariamente mejorando la exactitud del resultado.

Evidentemente estos procesos no están exentos de error y los mismos tienen diferentes fuentes y causas.

Para Uruguay, los cálculos de la superficie del territorio se vienen realizando desde 1863. Varios autores han emprendido la tarea del cálculo, utilizando diferentes datos de base y procedimientos con resultados que se presentan en las Figuras 1 y 2.

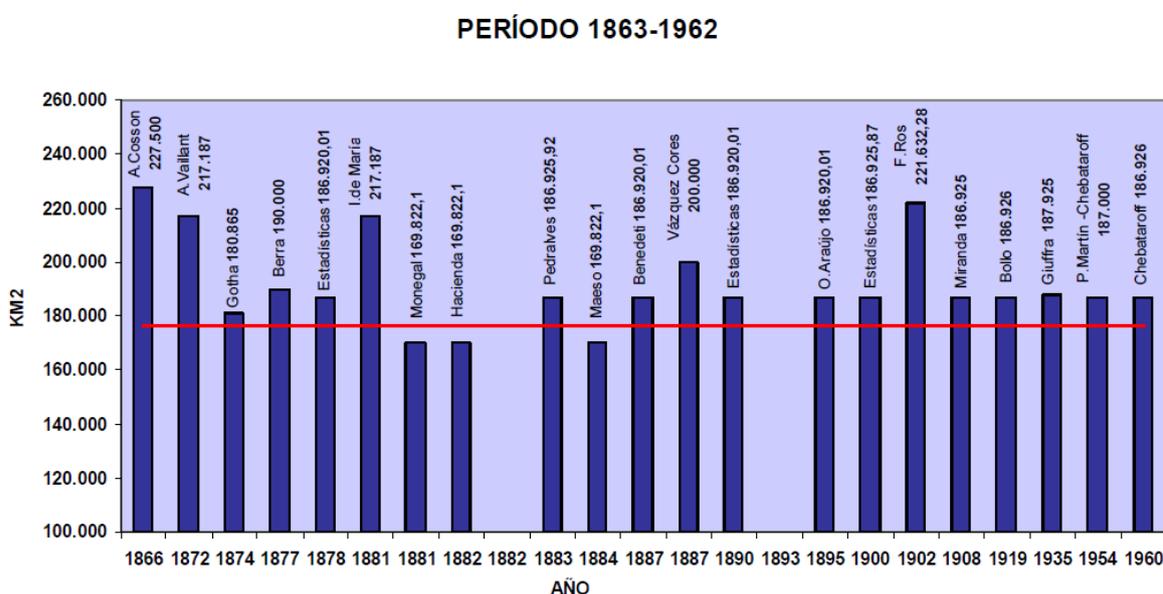


Figura 1. Áreas calculadas en kilómetros cuadrados para todo el territorio nacional publicadas entre 1863 y 1962. En rojo el último valor disponible de 2010. (Tomado de Lazo, 2010)

Tal cual se observa en la Figura 1 existen importantes discrepancias entre los valores, con un rango absoluto de $(277500-169822.1)=107678$ kilómetros cuadrados. Un rango de esa magnitud probablemente no sólo sea consecuencia de los métodos utilizados para el cálculo, sino también de la definición ambigua de los puntos que son efectivamente los límites del territorio.

Son llamativas las expresiones de las áreas calculadas en 1881 (Monegal), 1882 (Hacienda), y 1884 (Maeso), fundamentalmente por la coincidencia en los valores, y además por expresar el área a la décima de kilómetro cuadrado.

Los resultados de los cálculos publicados entre 1963 y 2010 se muestran en la Figura 2.

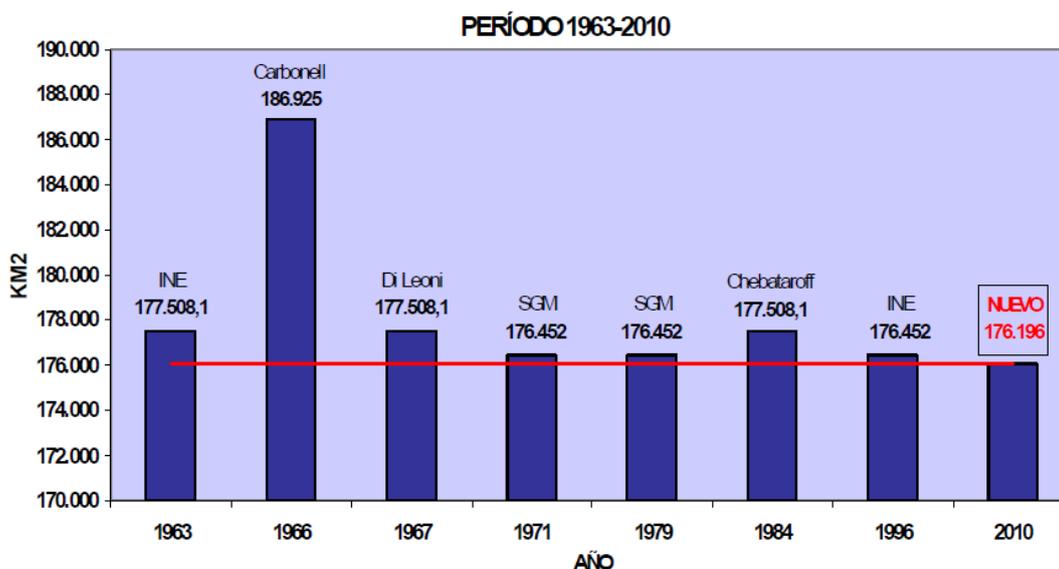


Figura 2. Áreas calculadas en kilómetros cuadrados para todo el territorio nacional publicadas entre 1963 y 2010. (Tomado de Lazo, 2010)

La tecnología en general, y las cartografías de base eran sensiblemente diferentes en dichas épocas. Es por ello seguramente que el rango disminuye considerablemente hasta quedar en el entorno de los 10000 kilómetros cuadrados, (aproximadamente 6%).

Los últimos cálculos datan de 2010 utilizando una versión digital del Plan Cartográfico Nacional (PCN), escala 1:50000, formato vectorial. A pesar de ser muy reciente, este resultado también carece de un intervalo de confianza.

Desde el punto de vista práctico, el cálculo de áreas con alta exactitud tiene mucha importancia, incluso por factores económicos. En Europa, la *European Union Common Agricultural Policy* (CAP), provee de subsidios a propietarios de establecimientos agrícolas en función de las áreas de dichos campos. Quienes desean aplicar a esos subsidios pueden representar 50 millones de parcelas. Se calculan las superficies de las mismas sobre imágenes satelitales, ortofotos, o midiendo directamente en campo por medio de GNSS. Claramente, para la CAP, la exactitud en el área calculada es de sumo interés y los resultados tienen implicancias económicas (Bogaert *et al.* 2005).

Causas posibles de inexactitudes en el cálculo de áreas

Se tratará exclusivamente el impacto de los errores posicionales de los vértices en el cálculo del área de un polígono.

A continuación se mencionan, algunas de las ambigüedades que afectan los cálculos de las áreas y que no serán consideradas en el resto de este trabajo:

- Referencias físicas difusas que determinan los vértices, por ejemplo en los cursos de aguas que son limítrofes (medio cauce, talweg, etc).

- Para los límites políticos interiores al territorio, definiciones poco claras de las jurisdicciones, y las modificaciones que históricamente han sufrido las mismas.
- Distintas proyecciones cartográficas determinan diferentes valores de áreas.
- En general, cuando se habla del área de un territorio, ¿está claramente estipulado a que tipo de área se quiere hacer referencia?. En algunos casos podría necesitarse el área geodésica, en otros el área plana después de haber elegido un proyección, incluso podría ser necesario saber la superficie real física, considerando la pendiente del terreno.

Existen diversos trabajos respecto del error posicional de los vértices y cómo los mismos influyen en el cálculo del área.

Varios autores (Chrisman y Yandell, 1988; Ghilani, 2003; Navratil, 2003 entre otros) llegan a determinar analíticamente una expresión para el error en área a partir de los errores en las coordenadas de los vértices:

$$S_{area} = \sqrt{\left(\left(\frac{\partial_{area}}{\partial X_A} S_{X_A} \right)^2 + \left(\frac{\partial_{area}}{\partial Y_A} S_{Y_A} \right)^2 + \left(\frac{\partial_{area}}{\partial X_B} S_{X_B} \right)^2 + \left(\frac{\partial_{area}}{\partial Y_B} S_{Y_B} \right)^2 + \dots + \left(\frac{\partial_{area}}{\partial X_n} S_{X_n} \right)^2 + \left(\frac{\partial_{area}}{\partial Y_n} S_{Y_n} \right)^2 \right)} \quad (1)$$

La expresión analítica es válida bajo ciertas hipótesis. Es necesario asumir que los errores en las coordenadas son pequeños, y que no están correlacionados espacialmente. Ello a su vez implica que en cada vértice se tiene la misma elipse de error, lo cual constituye una simplificación extrema. Sin embargo, los propios autores reconocen que investigaciones adicionales deberán tomar en cuenta la distribución y correlación espacial de los errores.

En este trabajo se descarta esta metodología por varias razones. En general, para cualquier estudio, cuando se recurre a la formulación anterior, se está asumiendo que en el experimento se cumplen algunas condiciones:

- Los errores son pequeños
- Se puede truncar el desarrollo de Taylor en el primero término
- Las elipses de error son iguales en cada punto.

Estas hipótesis no se comprueban en la práctica. Es por ello necesario basarse en otros procedimientos para estimar la variabilidad en el área, como los denominados métodos geoestadísticos.

En la década de los 60, Matheron formalizó y generalizó matemáticamente un conjunto de técnicas que explotaban la correlación espacial para hacer predicciones en reservas de oro. Surge así la geoestadística y la define como “la aplicación del formalismo de las funciones aleatorias al reconocimiento y estimación de fenómenos naturales”, (Samper y Carrera, 1990). La geoestadística plantea que no sólo se debe dar estimaciones de un valor, sino que también importa una medida de la incertidumbre.

Un análisis geoestadístico se compone por las siguientes etapas:

- Análisis exploratorio de datos
- Análisis estructural
- Predicciones (krigeado o simulaciones)

Respecto de la predicción, cuando el conocimiento de la incertidumbre es importante, y teniendo en cuenta lo difícil de las aproximaciones analíticas, se recurre a técnicas de simulación para obtener realizaciones plausibles (Samper y Carrera, 1990)

La idea fundamental de la simulación es construir un recurso experimental que “se comporte como” el sistema que se está estudiando.

Los modelos de simulación se utilizan para analizar decisiones bajo incertidumbre, ya que el comportamiento de las variables se representa mediante una distribución de probabilidad. A este tipo de simulación se le denomina Método de Monte Carlo y se emplea cuando se trabaja con variables aleatorias (Scheid y Di Constanzo, 1991).

Datos y métodos

Si bien el trabajo en curso apunta a caracterizar la incertidumbre en el área del territorio nacional, se decidió realizar un cálculo en una zona piloto representativa del procedimiento a seguir. Se definió un polígono de 20 lados en la carta J-28 del PCN 1:50.000 y se calculó el área plana determinada por sus vértices en coordenadas de carta. Ello será denominado como área de referencia (A_r) y tiene un valor de aproximadamente 353 Km^2 .

Para caracterizar el error planimétrico de esta carta, el Servicio Geográfico Militar midió 38 puntos de control, cartográficamente identificables y distribuidos homogéneamente en la zona abarcada por la misma carta. De esos mismos puntos, se leyeron sobre la carta papel, las coordenadas de los puntos de control. Las mediciones de campo fueron realizadas con receptores GPS (L1, L2), calculando las coordenadas mediante post-proceso en el siguiente sistema de referencia:

Elipsoide de Hayford
Proyección Gauss- Krugger
Meridiano de Contacto: $55^\circ 48'$
 $X_o=500 \text{ Km}$ al W del Meridiano de Contacto
 $Y_o= \text{Polo Sur}$
Datum horizontal_ I-Yacaré

Dicho sistema es el utilizado por la cartografía en formato papel a escala 1:50000 de todo el territorio nacional

(Los resultados de la caracterización del error, se encuentran en Méndez Baillo y López, 2010).

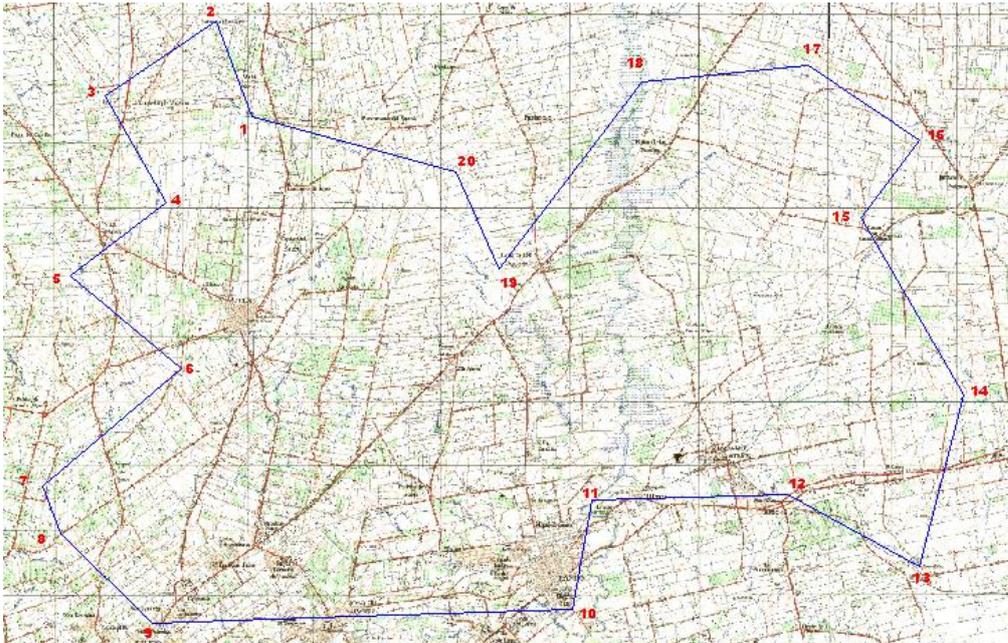


Figura 3 – Imagen de la Carta J-28. En azul se indica el polígono arbitrario de 20 vértices sobre el cual se realiza el cálculo de área

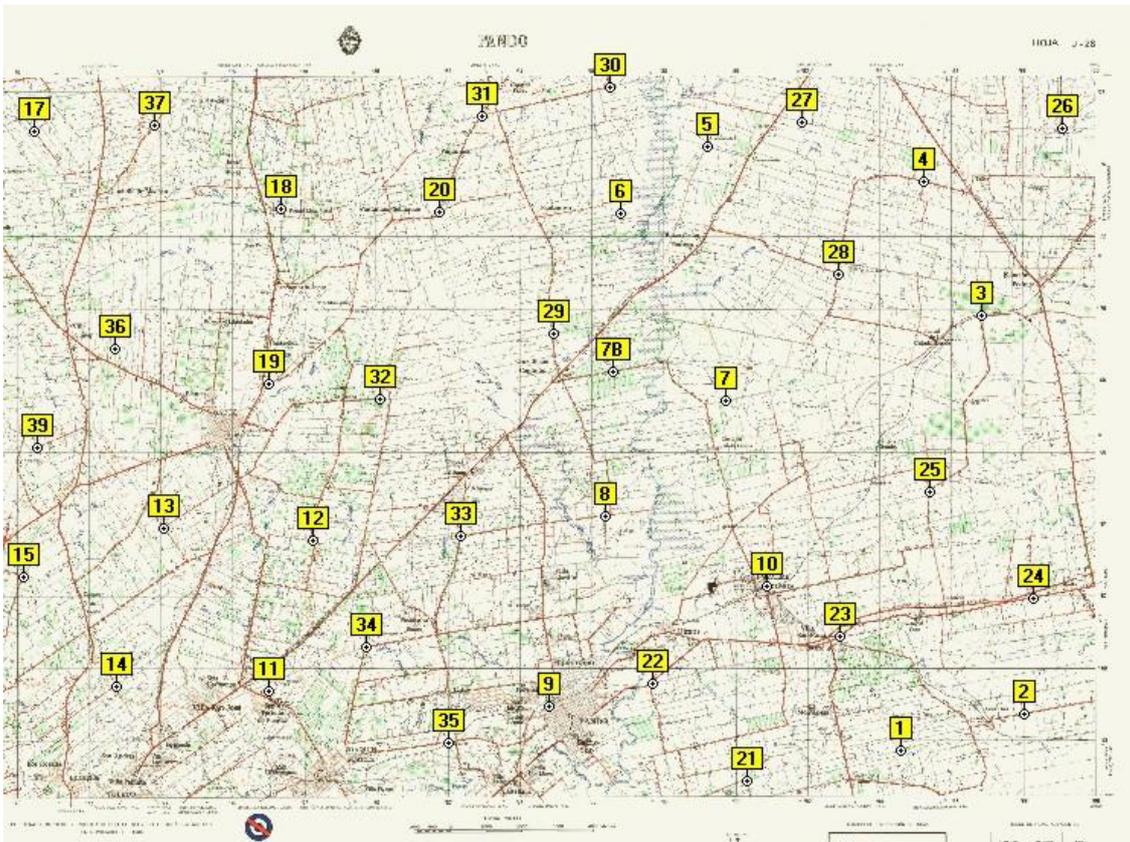


Figura 4 - Localización de los 38 puntos de control sobre la Carta J-28

A partir de las coordenadas de los puntos de control, las cuales surgen de las mediciones de campo y también de la lectura de coordenadas de la carta, se forma la siguiente matriz:

$$\begin{bmatrix} X_1 & Y_1 & \Delta E_1 & \Delta N_1 \\ X_2 & Y_2 & \Delta E_2 & \Delta N_2 \\ \dots & \dots & \dots & \dots \\ X_n & Y_n & \Delta E_n & \Delta N_n \end{bmatrix}$$

siendo las X_i, Y_i las coordenadas planas de carta de los n puntos de control, ΔE_i y ΔN_i es la diferencia en cada punto de control (en abscisas y ordenadas respectivamente), entre los valores leídos de la carta y los medidos con GNSS. Cabe señalar que las abscisas se miden como crecientes en la dirección Este, y que las ordenadas se miden crecientes en la dirección Norte.

El campo del error no es conocido salvo en los puntos de control, por lo que es necesario realizar una estimación del mismo mediante interpolación. Para ello se utilizarán métodos geoestadísticos (Samper y Carrera, 1990) que permiten no solamente estimar el valor del error en cualquier punto de la carta, sino también su incertidumbre. Existen otros métodos de interpolación (vecino más cercano, inverso del cuadrado de la distancia, etc.) pero ellos no proveen elementos para estimar su propio error. El cálculo geoestadístico se basa en la especificación del variograma, el cual a su vez puede estimarse a partir de valores conocidos en puntos experimentales (Samper y Carrera, 1990). En este trabajo, y por simplicidad, se asumirá que el campo del error es estacionario de segundo orden tanto para los desplazamientos ΔE y ΔN , y que los mismos son independientes entre sí.

Con un software geoestadístico se estimaron el alcance y la meseta de los respectivos variogramas (uno para cada dirección)

Como modelo teórico de variograma fue adoptado el esférico, cuya expresión es:

$$(2) \quad \begin{cases} \gamma(h) = (S/2)(3(h/a) - (h/a)^3) & \text{si } h \leq a \\ \gamma(h) = S & \text{si } h > a \end{cases}$$

Siendo a el alcance y S la meseta.

h es la matriz de distancias mutuas entre los vértices del polígono

El siguiente paso fue simular una perturbación en la posición de las coordenadas de los 20 vértices del polígono obteniendo un vector de error Z cuyo valor esperado es Z^* y cuya matriz de covarianzas es C , la cual se calcula a partir del variograma. Se supone también que los componentes de Z siguen una distribución normal.

Las simulaciones pueden ser no condicionadas o condicionadas. Que la simulación sea condicionada implica que en los puntos de control, los valores se ajustan exactamente a los valores de las coordenadas de dichos puntos. En el presente trabajo se realiza una simulación condicionada tomando Z^* igual a la estimación por el método del krigado. La siguiente expresión permite obtener las realizaciones:

$$Z_s = Z^* + M.u \quad (3)$$

Siendo:

u un vector con componentes aleatorios e independientes, media nula y varianza unidad; $(N(0,1))$ y M una matriz $n \times n$ tal que $M.M^T = C$. Ver Samper y Carrera (1990) por los detalles del método matricial descrito.

Cálculo de C y de M

Asumiendo que el sistema es estacionario de segundo orden:

$$C = Meseta - \gamma(H) \quad (4)$$

Obteniendo C se obtiene M aplicando a la matriz C la factorización de Cholesky.

Teniendo M , generando aleatoriamente u y realizando krigado puntual en los vértices del polígono para generar un vector Z^* , se está en condiciones de obtener las realizaciones Z_s .

Para llegar al valor de cada "X" simulada, se debe sumar la abscisa de cada punto del polígono, quedando:

$$VX = Xi + Mu + Z^* \quad (5)$$

El cálculo es análogo para obtener VY , aunque es importante aclarar que ya que el variograma se calcula independientemente para las X y para las Y, no serán iguales las matrices M y C en cada caso

Cada simulación genera una nueva coordenada para cada vértice del polígono, y en consecuencia un valor del área de dicha figura. Habrá tantos valores del área como número de simulaciones se realice. Estos valores se organizan por conveniencia en un único vector. Para facilitar la interpretación de los resultados, es conveniente normalizar el vector (en relación al A_r) y expresar las variaciones de área en porcentaje. Para ello se aplica entonces a cada área simulada A_{s_i} , la expresión:

$$((A_{s_i} - A_r) / A_r) * 100 \quad (6)$$

A_r es el área de referencia calculada en base a las coordenadas de los vértices del polígono elegido arbitrariamente sobre la carta

Resultados

Concretamente para el experimento se tienen los siguientes valores:

Número de simulaciones: 10000

Número de puntos de control: 38

Número de vértices del polígono: 20

Tiempo requerido para las 10000 simulaciones: 10 segundos (Intel ® Core(TM) Duo CPU, 1,7GHz, 1GB de RAM).

Para el cálculo del variograma, y de la meseta y el alcance, se utilizó el software **Easydem** obteniendo los siguientes resultados:

Incrementos ΔE en los puntos de control

Alcance: 14196 m

Meseta: 1036 m

Incrementos ΔN en los puntos de control

Alcance: 14198 m

Meseta: 875 m

La programación de la simulación se realizó en **Matlab**

Originalmente se obtuvo un histograma de los valores de las áreas obtenidos, y con ciertas modificaciones, se logra el siguiente gráfico que es una aproximación de la función de densidad de probabilidad del área con las dispersiones en porcentaje:

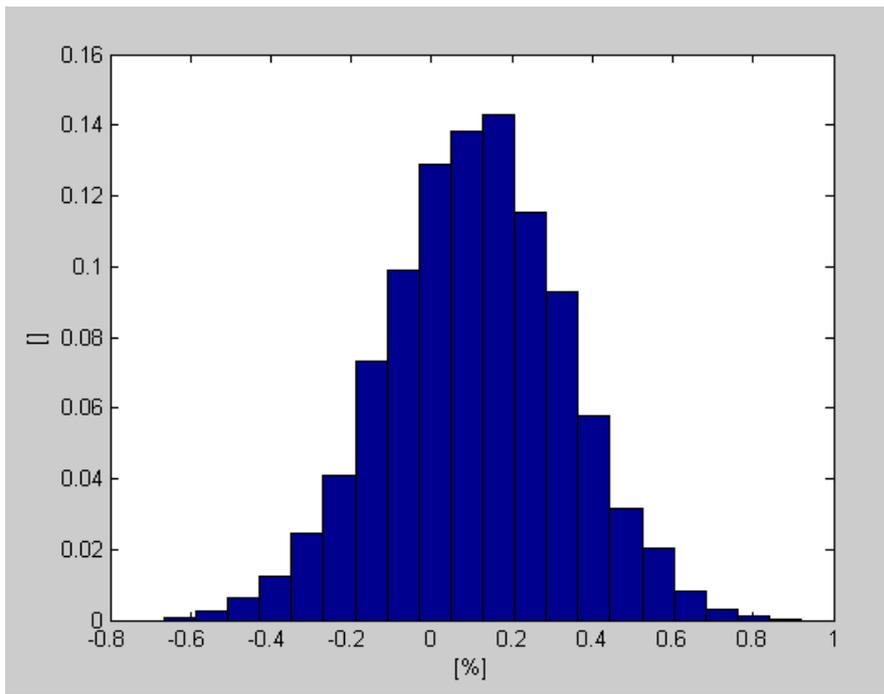


Figura. 5 –Aproximación de la función de densidad de probabilidad del área, con las dispersiones en porcentaje

El gráfico muestra una población bastante simétrica, no centrada en 0.0 y que no pasó el Test de normalidad de Kolmogorov-Smirnov. Por otra parte, en la Figura 6 se presenta un gráfico QQ-Plot que debería ser una recta si la población fuera normal. Se puede ver que hay leves discrepancias, fundamentalmente en las colas de la distribución, lo que confirma el resultado del Test.

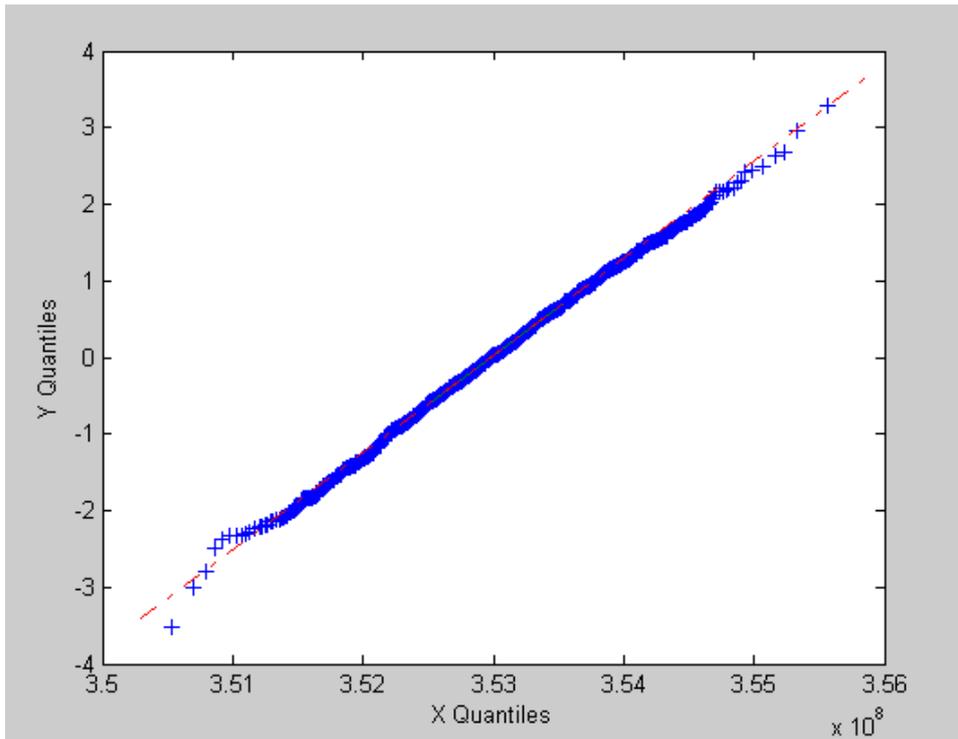


Figura 6 –Diagrama QQ-Plot ilustrando las desviaciones respecto a una población normal.

VALORES DE DISPERSION de Ar (en kilómetros cuadrados)

Valor máximo	Valor mínimo	Valor más probable	Mediana	Desviación estándar
355,9	350,3	353,02	353,04	0,78

Tabla 1. Dispersiones de los valores del área simulada

No hay una relación lineal entre los errores posicionales de los vértices y los errores en el área. Sin pérdida de generalidad, se puede considerar un triángulo ABC, (ver Figura 7). Si A y B son fijos, y C experimenta una variación en su posición dada pudiendo encontrarse en cualquier parte de la circunferencia, se aprecia como el desplazamiento en la dirección Este (desplazamiento en azul), mantiene invariable el área ya que se mueve en una paralela a AB. Sin embargo, si se observa el desplazamiento en color rojo en la dirección Norte, se observa que si C se aleja de AB, el triángulo aumenta en superficie, mientras que si se acerca, el área disminuye.

Generalizando al polígono definido en la carta, se puede afirmar que no hay linealidad entre los errores planimétricos de los vértices, y los errores superficiales en el cálculo del área. Se puede decir entonces, que existe una anisotropía con una dirección preferente.

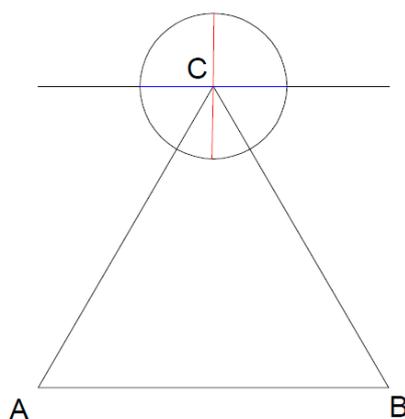


Figura 7. Desplazamiento de un vértice de un triángulo e influencia en el área.

Conclusiones

La simulación generó 10000 realizaciones del área y se verificó que el conjunto de esos valores no tiene una distribución normal, aunque es bastante próxima.

Al quedar demostrada la no normalidad de la distribución, y para poder dar un rango de valores del área con un nivel de confianza determinado, se recurrió a percentiles, pudiendo afirmar lo siguiente:

en el 95% de los casos el área se encuentra entre 351.45 y 354.57 Km². (10)

Si se observa la Figura 5, se puede concluir que el área se “expande”, ya que hay una cantidad mayor de valores que superan al del área de referencia, (ver expresión 9).

Si se debe proporcionar un único valor del área, y teniendo en cuenta una desviación estándar de 0,78 kilómetros cuadrados, se expresaría

$$\text{Área}=353 \text{ Km}^2$$

(el área es el valor más probable y por la magnitud de la desviación estándar se redondea al kilómetro cuadrado. El “353” es el valor más próximo a la moda, y cae dentro del intervalo de confianza).

En el caso de la expresión (10) los valores de los extremos pueden redondearse en razón de la magnitud de la desviación standard, pero teniendo en cuenta que esto implicaría recalcular el valor exacto del nivel de confianza

Si el mismo polígono fuera calculado en otra carta, probablemente los valores del área, así como los variogramas, y en definitiva el resto de los cálculos, darían valores diferentes. Es importante tener en cuenta que existe estrecha relación entre la incertidumbre y el linaje de la carta sobre la cual se calculen las áreas.

El trabajo cumple con el objetivo de presentar un método de cálculo dando una

estimación de incertidumbre y niveles de confianza para el área, apuntando a cambiar el concepto de expresar magnitudes de manera determinística, llevando la expresión a una forma probabilística.

A futuro, y aplicando la metodología expuesta en el presente trabajo, se calcularán completamente las áreas de las divisiones administrativas (denominadas departamentos en Uruguay) y del territorio nacional, expresando los resultados con los correspondientes niveles de confianza y cifras significativas que correspondan. Se podrá también aplicar la metodología expuesta, como herramienta para resolver simulación de inundaciones, cálculo y expresión de volúmenes de tierra con su incertidumbre, cálculo de longitudes de tendidos de cables, entre otras tareas.

Bibliografía

Bogaert, P., Delincé, J., y Kay, S., 2005. "Assessing the error of polygonal area measurements: a general formulation with applications to agriculture". *Measurement Science and Technology*. Vol. 6, N° 5.1170-1178 pp.

Chrisman, N.; Yandell, B., 1988. "Effects of point error on area calculations: a statistical model", *Surveying and Mapping*, Vol. 48, N° 4. pp 241-246

Ghilani, C., 2000. "Demystifying area uncertainty: more or less" *Surveying and Land Information Systems*. Vol. 60, N° 3. pp.183-189.

Lazo, J. 2010. "Superficies del Uruguay: medirlas, una tarea sin final." 22 pp. Montevideo. Comunicación personal

Méndez Baillo, R. y López, C.2010. "Evaluación de exactitud posicional horizontal y vertical de la Cartografía Oficial a escala 1:50.000". En *Anales del I Congreso Uruguayo de Infraestructura de Datos Espaciales* 25-26 Octubre del 2010, Montevideo, Uruguay.

http://www.thedigitalmap.com/~carlos/papers/rep10_3/EvaluacionDeExactitudPosicionalHorizontalYVerticalDeLaCartografiaOficial.pdf

Navratil, G. 2003. "Precision of area computation". ESRI 2003 – European User Conference. 11pp. Innsbruck, Austria.
downloads2.esri.com/campus/uploads/library/pdfs/68597.pdf

Samper, F.J. y Carrera, J., 1990. "Geoestadística. Aplicaciones a la Hidrología subterránea.". Centro Internacional de Métodos Numéricos en Ingeniería. Universitat Politècnica de Catalunya. 484 pp.

Scheid, F; Di Constanzo, R., 1991. "Métodos Numéricos".709 pp. Ed. McGraw-Hill. México

