

PROPUESTA DE METODOLOGÍA PARA EL ANÁLISIS DEL ERROR DE POSICIÓN EN BASES DE DATOS ESPACIALES MEDIANTE ESTADÍSTICA CIRCULAR Y MAPAS DE DENSIDAD

POLO GARCÍA, M.E.¹ y FELICÍSIMO PÉREZ, A.M.²

¹Centro Universitario de Mérida, Universidad de Extremadura
C/ Santa Teresa Jornet, 38. 06800 Mérida (España)

mepolo@unex.es

²Escuela Politécnica, Universidad de Extremadura
Avda. de la Universidad, s/n. 10071 Cáceres (España)

amfeli@unex.es

RESUMEN

El error de posición en las bases de datos espaciales se establece comparando la posición real con la medida, ambas definidas por sus coordenadas en un sistema de referencia. Con la diferencia entre ambas posiciones se genera un vector con nodo inicial en la posición real y nodo final en la posición medida. El vector de error tiene, por tanto, dos propiedades métricas: módulo y acimut. Las pruebas estándar para evaluar la exactitud posicional en una base de datos usan sólo el módulo y obvian la magnitud angular. En este trabajo proponemos el uso de ambas magnitudes para un análisis completo del error de posición y presentamos los métodos pertinentes y un ejemplo de su uso mediante la comparación de mapas de la misma zona a diferente escala.

Palabras clave: bases de datos espaciales, exactitud planimétrica, error de posición, vector de error, datos circulares, estadística circular.

POSITIONAL ERROR ANALYSIS IN SPATIAL DATABASES USING CIRCULAR STATISTICS AND DENSITY MAPS: A METHODOLOGICAL PROPOSAL

ABSTRACT

The positional error in spatial databases is established by comparing the actual position with the measure, both defined by their coordinates in a reference system. A vector is generated with the difference between the two positions. This vector has the initial node in the actual position and the end node in the measured position. The error vector therefore has two metric properties: module and azimuth. The standard tests to assess the accuracy positional in a database use only the module and obviate the angle magnitude. In this paper we propose the use of both magnitudes for a complete analysis of the positional error and present the relevant methods and an example of its use by comparing two maps from the same area at different scales.

Keywords: spatial database, planimetric accuracy, positional error, vector of error, circular data, circular statistic.

1. Introducción y objetivos

La demanda creciente de información sobre el territorio hace cada vez más importante establecer una garantía de calidad de los datos espaciales, ya que este tipo de datos es la forma en la que se materializa la información territorial (Ariza López, 2002). Las componentes de calidad de los datos espaciales son múltiples e incluyen, entre otros, las incertidumbres de posición, temporal o temática, la falta de consistencia interna de los datos o su integridad.

Las normas actuales sobre control de calidad posicional de los datos espaciales son conocidas (ASCI, 1983; ASPRS, 1989; FGDC, 1998; USGS, 1997) y siguen, en general, unas pautas similares que, esencialmente, son: a) evaluar la calidad de los datos originales con otros procedentes de fuentes de mayor exactitud, b) realizar la comparación sobre un número mínimo de una veintena de pares de datos, c) determinar las distancias entre los pares y d) aplicar tests estadísticos a estas distancias considerando, a efectos de significación, que se ajustan a una distribución normal. Estos análisis se realizan, por tanto, sobre magnitudes escalares (distancias) y de forma global, sin definir subzonas dentro del área de estudio.

Los procedimientos mencionados se aplican rutinariamente, pero, en nuestra opinión, presentan problemas debido tanto a la metodología como a las asunciones estadísticas implicadas. Aunque actualmente se están desarrollando técnicas para la mejora de la calidad posicional de los datos espaciales, como es la aplicación de estimadores robustos para determinar valores atípicos (Atkinson Gordo *et al.*, 2007), planteamos en este trabajo un tratamiento complementario, proponiendo los métodos para hacerlo y mostrando resultados a través de un ejemplo cartográfico concreto. Las mayores diferencias entre los métodos que proponemos y los convencionales residen en el tratamiento del error de posición como una magnitud vectorial, en el uso de estadística circular para el análisis de datos y en la propuesta de un método no paramétrico para la detección de datos anómalos que no asume la normalidad de las variables ni su ajuste a cualquier otra posible distribución estándar.

El objetivo de este trabajo es por lo tanto esencialmente metodológico: es posible cambiar la forma de tratamiento del error de posición y, además, es necesario hacerlo porque los métodos actuales no explotan la información adecuadamente y pueden pasar por alto problemas importantes que no se manifiestan en las pruebas convencionales.

2. Métodos

Cuando hablamos de datos espaciales hacemos referencia a datos sobre localizaciones definidas mediante valores de coordenadas en un sistema de referencia. El error espacial de un dato se define como la diferencia entre las coordenadas reales o valor verdadero (X_v , Y_v) y el valor medido (X_m , Y_m). Este error es una magnitud vectorial donde el vector queda definido por el nodo

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

inicial (valor verdadero) y el nodo final (valor medido) adquiriendo las propiedades métricas de módulo o magnitud y acimut.

Trabajando en dos dimensiones, módulo (m) y acimut (θ) quedan definidos por los valores de las coordenadas:

$$m = \sqrt{(X_v - X_m)^2 + (Y_v - Y_m)^2} \quad \theta = \text{arc tg} \left(\frac{X_v - X_m}{Y_v - Y_m} \right)$$

El análisis del error de posición en bases de datos espaciales, una vez que se ha definido como un vector, puede realizarse en tres etapas en función de las componentes: a) tratamiento de los módulos de los vectores (magnitud escalar), b) tratamiento de los acimuts (magnitud angular) y c) estudio de los módulos y acimuts de forma conjunta. Cada análisis requiere una estadística específica y ofrece resultados complementarios.

2.1. Estadística de módulos

El tratamiento estadístico de magnitudes lineales es el que se utiliza en la mayoría de los tests sobre control de calidad posicional. Algunos de estos tests ya fueron referenciados en el primer apartado, siendo la bibliografía al respecto conocida dentro del ámbito del tratamiento del error de posición, por lo que no incidiremos sobre el tema.

2.2. Estadística de acimuts

Al proponer en este artículo la importancia de analizar el error de posición como un vector, necesitamos estudiar la componente angular. Los datos angulares los trataremos, de entrada, como vectores unitarios denominándolos datos circulares. Los datos circulares son habituales en muchos ámbitos de la ciencia como, por ejemplo, meteorología (viento), geología (buzamientos) o biología (dispersión o movimientos), aunque hasta la fecha las referencias de los mismos en cartografía son escasas (Ariza López, 2002).

Como la estadística de datos unidimensionales no es la adecuada para este tipo de datos, aplicaremos la denominada estadística circular o estadística de datos circulares (Fisher *et al.*, 1987). Esta estadística tiene en cuenta algunas características diferenciales de este tipo de información (Batschelet, 1981) que son relevantes para nuestro trabajo: a) el origen de los datos circulares es una dirección arbitraria (Norte, eje X...), b) no existe una relación de magnitud entre los datos (185° no es mayor que 175° en el sentido convencional del término) lo que lleva a que c) las operaciones algebraicas con ángulos deben reducirse al intervalo $0-360^\circ$ (o los valores equivalentes en la escala que se use, por ejemplo $0-2\pi$) mediante operadores tipo "módulo". La estadística circular está diseñada exclusivamente para el análisis de direcciones o acimuts, considerándose todos los vectores unitarios, lo cual limita, lógicamente, su aplicación a vectores de magnitudes diferentes.

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

Los estadísticos descriptivos básicos para una muestra de mediciones circulares o acimuts ($\theta_1, \theta_2, \dots, \theta_n$) son los siguientes (Fisher, 1995):

- Acimut medio ($\bar{\theta}$), obtenido mediante la suma vectorial de todos los vectores de la muestra. Representa el acimut medio del vector resultante de la suma vectorial citada. Se obtiene a partir de las ecuaciones siguientes, donde C y S son la suma de todos los cosenos y senos de los acimutes de la muestra, respectivamente.

$$\bar{\theta} = \arctg \frac{S}{C} \quad ; \quad C = \sum_{i=1}^n \cos \theta_i \quad ; \quad S = \sum_{i=1}^n \sen \theta_i$$

El sentido de medición de $\bar{\theta}$, y de todas las observaciones circulares, se considera igual que un acimut topográfico, esto es, con origen en la dirección del eje Y y creciendo en sentido horario.

- Módulo del vector resultante (R), calculado a partir de la ecuación siguiente.

$$R = \sqrt{C^2 + S^2}$$

- Módulo medio (\bar{R}), obtenido al dividir R entre el número de observaciones. Lógicamente, al trabajar con vectores unitarios, su valor estará en el rango 0-1. Si todos los vectores coinciden en una misma dirección el valor será la unidad. Por el contrario, un valor pequeño de este estadístico significa que los vectores se anulan pero no implica necesariamente una distribución circular uniforme. Siendo n el número de observaciones, su valor es:

$$\bar{R} = \frac{R}{n}$$

- Varianza circular de la muestra (V), estadístico de dispersión angular con rango de valores 0-1, donde el valor crece con la dispersión:

$$V = 1 - \bar{R}^2$$

- Desviación circular estándar de la muestra (v), es el resultado de una raíz cuadrada, por analogía a lo que ocurre en datos lineales, y está definida por:

$$v = \sqrt{-2 \log(1 - V)}$$

- Parámetro de concentración (κ) que mide la desviación de la distribución respecto a una distribución circular uniforme; está entre los valores 0 (distribución uniforme) e ∞ (máxima concentración). Podemos considerar que κ indica concentración si $\kappa > 2$. Depende del valor del módulo medio y las expresiones para calcular su valor son las siguientes.

$$\kappa = \begin{cases} 2\bar{R} + \bar{R}^3 + 5\bar{R}^5 / 6 & \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + 0.43/(1 - \bar{R}) & 0.53 \leq \bar{R} \leq 0.85 \\ 1/(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}) & \bar{R} \geq 0.85 \end{cases}$$

El tratamiento de los errores como datos circulares no debe limitarse, obviamente, a estadísticos descriptivos sino que deben aplicarse pruebas que analizan las distribuciones de los datos (Jammalamadaka and SenGupta, 2001; Mardia and Jupp, 2000).

En nuestro análisis de datos de errores vamos a usar la estadística circular en tres pasos para comprobar propiedades de su distribución:

- aleatoriedad del acimut del error, lo cual se analiza comprobando si la distribución de acimutes es aleatoria o, lo que es lo mismo, se ajusta a una distribución circular uniforme. En esta distribución, la probabilidad de que un error tenga un acimut concreto es la misma con independencia de su valor.
- si la prueba anterior es negativa (los datos no se ajustan a una distribución circular uniforme), se hará la prueba de ajuste a una distribución de von Mises, unimodal y simétrica, equivalente en cierto sentido a la distribución normal o de Gauss en datos lineales.
- finalmente, si los ajustes a ambas distribuciones, uniforme y von Mises, no son significativos puede ser indicador de distribuciones bimodales, multimodales o con formas más complejas. Estas distribuciones no estándares se analizarán por métodos distintos, tal como comentaremos en el apartado siguiente.

Para responder a estas cuestiones se aplican tests de uniformidad como los de Rayleigh, Rao, Watson o Kuiper (Fisher, 1995) o bien modalidades de los tests de Watson y Kuiper (Fisher, 1995) para comprobar si los datos se adaptan a otro tipo de distribuciones circulares, como la von Mises.

Los tests citados se desarrollan, en esencia, de la siguiente forma:

- Test de Kuiper: La hipótesis nula del test de Kuiper establece que la muestra de valores ha sido obtenida a partir de una distribución concreta, ya sea la uniforme o la von Mises.
- Test de Rayleigh. Se plantea esta prueba en el caso específico de querer detectar una única dirección modal de una muestra de vectores contra la alternativa de aleatoriedad. La hipótesis de uniformidad se rechaza si la longitud media del vector resultante de la muestra es demasiado grande. Este test supone que un valor del módulo medio más largo implica una concentración mayor en torno a la media y por tanto menos probabilidad de que los datos estén uniformemente distribuidos.
- Test de Rao. Esta prueba considera también como hipótesis nula que los datos están distribuidos uniformemente, con la salvedad de analizar si el espaciado entre puntos adyacentes es

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

aproximadamente igual en todo el círculo. Para una distribución uniforme, el espaciado entre puntos debería ser $360^\circ/n$. Si el espaciado existente se desvía mucho de este valor, la probabilidad de que los datos pertenezcan a una distribución uniforme se reduce. Este test se presenta como más robusto que el de Rayleigh, pudiendo ser un buen complemento en cualquier caso.

- Test de Watson. Determina la bondad del ajuste del conjunto de datos a una distribución concreta, como puede ser la uniforme o la von Mises. Para ello calcula las desviaciones entre los datos y la distribución elegida; si estas desviaciones son muy grandes, el estadístico a calcular también lo será y se rechazará la hipótesis nula, que supone que los datos se adaptan bien a la distribución elegida.

Una vez analizados los módulos y los acimuts por separado cabe preguntarse si existe correlación entre la distribución de ambas magnitudes, lo que implicaría que a una determinada dirección le corresponde un módulo de determinada magnitud. Para ello se calcula el coeficiente de correlación entre magnitudes lineales y circulares (Mardia and Jupp, 2000) que parte con la premisa de que las variables lineales siguen una distribución normal.

2.3 Estadística conjunta de acimuts y módulos

Para ofrecer más información de la distribución de las dos variables a la vez, necesitamos recurrir a métodos no paramétricos, sobre todo si las variables lineales no siguen una distribución normal, y no se puede usar el coeficiente de correlación como parámetro de estudio conjunto. Como hasta la fecha no se ha contemplado la casuística más compleja de combinar las diferentes distribuciones lineales con las circulares, nuestra propuesta es transformar los vectores de error en una nube de puntos para encontrar la función de densidad de distribución de dicha nube. Para ello hacemos una traslación de todos los vectores, sin modificar el acimut respectivo, a un origen común (0,0), lo que origina que la localización de los nodos finales forme una nube de puntos cuya distribución nos proporciona información de ambas variables a la vez. A partir de la nube de puntos construiremos un "mapa de densidad" como el número de puntos que se corresponden por unidad de área. Estos mapas de densidad se configuran a modo de curvas de nivel cuya forma será indicativa de la distribución de los vectores de error (ver [figura 3](#)).

Los *outliers* u observaciones discordantes pueden localizarse mediante un proceso propuesto por Dixon y Chapman (1980) consistente en determinar el valor de la media armónica (H_m) para cada uno de los puntos, catalogando como *outliers* potenciales un determinado porcentaje que serán los que presenten un valor mayor de este estadístico. La media armónica, para un punto en concreto, se obtiene como el cociente entre el número de observaciones (n) y la suma de las inversas de las distancias entre un punto i y el resto de localizaciones x , tal como se indica en la ecuación siguiente. Este cálculo se realizará para todos los puntos, de tal forma que los puntos más alejados de la masa central tendrán un valor de H_m mayor y se considerarán como observaciones discordantes.

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

$$Hm = \frac{n}{\sum_{i=1}^n \frac{1}{r_{ix}}}$$

El uso de mapas de densidad permite la detección de errores groseros o valores anómalos sin asumir requisitos propios de las distribuciones estándar contando, además, con la ventaja de que cada mapa de densidad se adapta al caso específico que tratemos en cada momento. La utilidad de estos resultados es determinar dónde pueden estar concentrados los problemas de mayores errores en nuestro conjunto de datos. Los mapas de densidad de puntos permiten determinar las distribuciones reales de los datos en dos dimensiones lo que libera de "usar" círculos o elipses de probabilidad, ya que la forma y densidad de las distribuciones se decide localmente, no globalmente.

3. Material

Los métodos mencionados se han aplicado a varias muestras de vectores de error construidas mediante la localización de puntos comunes en mapas de diferentes escalas y la extracción de sus coordenadas respectivas: el vector i tiene su nodo inicial en las coordenadas del punto i según el mapa de mayor escala y su nodo final en las coordenadas respectivas según el mapa de menor escala. Los elementos para identificar los puntos homólogos proceden de puntos claramente definidos en vías de comunicación (carreteras de diferente orden y vías de tren), edificaciones, e infraestructuras de la zona regable, principalmente. En principio, cabe esperar errores debidos exclusivamente a la diferencia de escala entre las fuentes.

Las muestras utilizadas provienen de cartografía de la provincia de Badajoz y difieren en número de vectores y escalas de las fuentes:

- 1 muestra nombrada '10-50' de 500 vectores generados mediante la comparación de puntos entre dos mapas a escala 1/10000 y 1/50000 (MTN50). La hoja del MTN50 es la denominada 'Montijo', mientras que la cartografía a 1/10000 se corresponde con la misma superficie del citado MTN50.
- 1 muestra denominada '25-50' de 500 vectores generados mediante la comparación de puntos entre la zona común de un mapa a escala 1/25000 (MTN25) y otro a 1/50000 (MTN50), correspondientes ambos a 'Montijo'.
- 4 muestras ('Mérida', 'Villafranca', 'Villanueva' y 'Almendralejo') de 250 vectores cada una generados mediante la comparación de la cartografía digital del MTN25 del Instituto Geográfico Nacional (IGN) obtenida de un servidor de mapas web (*WMS*) y el MTN50 en formato digital. En concreto cada muestra resultó del análisis de la zona común entre las siguientes cartografías:
 - 'Mérida', las hojas 776, 777 y 778 del MTN 50 digitalizado y las correspondientes a la Base Cartográfica Numérica a 1/25000 del IGN obtenida a través de *WMS*.

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

- 'Villanueva', las hojas 779, 780 y 781 del MTN 50 digitalizado y las correspondientes a la Base Cartográfica Numérica a 1/25000 del IGN obtenida a través de *WMS*.
- 'Villafranca', las hojas 827, 828 y 829 del MTN 50 digitalizado y las correspondientes a la Base Cartográfica Numérica a 1/25000 del IGN obtenida a través de *WMS*.
- 'Almendralejo', las hojas 777, 803 y 829 del MTN 50 digitalizado y las correspondientes a la Base Cartográfica Numérica a 1/25000 del IGN obtenida a través de *WMS*.

4. Resultados

Obviamos en este apartado los resultados obtenidos del análisis de los módulos de los vectores por ser un tema conocido. Sólo indicar que la distribución de los módulos no sigue una distribución normal, excepto en dos muestras ('Villafranca' y 'Almendralejo') que sí presentaron evidencias de ajuste a esta distribución.

4.1. Estadísticos básicos para datos circulares

Como hemos comentado, los estadísticos básicos calculados para vectores unitarios son el acimut medio ($\bar{\theta}$), el módulo medio (\bar{R}), la desviación estándar circular de la muestra (ν) y el parámetro de concentración (κ). El acimut medio ($\bar{\theta}$) es el acimut del vector resultante de la suma vectorial de todos los elementos en cada muestra. Los resultados obtenidos se muestran en la [tabla 1](#). Observamos que los valores obtenidos para el módulo medio (\bar{R}), que puede oscilar entre 0 y 1, son bajos, por debajo de 0.30 con una sola excepción de 0.43. Con estos valores, la primera conclusión es que *descartamos la existencia de una dirección preferente de datos*. De forma coherente, el parámetro de concentración (κ) que es un indicador de aglutinación de datos presenta como mayor valor 0.96, cuando se establece que este parámetro es significativo a partir de 2 aproximadamente. Del mismo modo, la desviación estándar es muy elevada corroborando la gran dispersión de los vectores en las muestras tratadas.

Esta primera prueba exploratoria nos dice que no existe dirección preferente pero eso no significa que la distribución de los vectores sea uniforme, aspecto que se debe analizar mediante prueba para determinar la distribución de los vectores o descartar distribuciones concretas.

4.2. Análisis de la distribución de datos circulares

Se realizan en primer lugar las pruebas para determinar si los datos se ajustan a una distribución uniforme. Los tests utilizados (Kuiper, Rao, Rayleigh y Watson) son parecidos pero se usaron las cuatro modalidades para mayor fiabilidad. Los resultados se muestran en la [tabla 2](#) donde todos los valores de P confirman, sin excepción, que *los vectores de error no se distribuyen de acuerdo con una distribución circular uniforme*.

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

En este momento nos encontramos con que están descartadas las opciones de dirección preferente única y de distribución circular aleatoria. Aunque ya es previsible que los datos vectoriales no se ajusten a una distribución unimodal como la de von Mises, en la [tabla 3](#) se calculan los ajustes a esta distribución. Los resultados son coherentes y nos indican que los errores tampoco se ajustan a esta distribución.

Dentro del análisis exploratorio es importante la representación gráfica de la distribución de datos brutos ([figura 1](#)), especialmente cuando se manifiestan distribuciones complejas como se apunta en estos casos (no aleatoria pero no dirección preferente). En los gráficos de esta figura se representa para cada muestra la distribución de frecuencias de los acimuts, con indicación del acimut medio y su intervalo de confianza al 95% definido en función del error estándar circular. Los gráficos a) y b) corresponden a las muestras '10-50' y '25-50' y los c) a f) a las muestras 'Mérida', 'Villafranca', 'Villanueva' y 'Almendralejo'.

La observación más importante es que *existen varias direcciones concretas donde los vectores de error se acumulan*. En el caso de a) coinciden con los ejes principales (Norte-Sur y Este-Oeste) y con los secundarios a 45° de estos; en b) es manifiesta la acumulación para un acimut de 315° (Noroeste) y, aunque siguen siendo importantes los ejes principales, los acimuts se acumulan en el cuadrante NO. Lo importante es que *este tipo de distribuciones no se detectan analizando solamente los módulos de los vectores de error* como se hace habitualmente. Asimismo, pueden dar pistas sobre problemas subyacentes a la propia generación de los datos. Por ejemplo ¿cuál es el motivo de que los errores se agrupen en determinadas direcciones preferentes mientras que en otras zonas la acumulación de errores es menor? En la discusión final propondremos alguna posible causa.

Los gráficos anteriores muestran la agrupación de acimuts de error en múltiples direcciones preferentes pero no sabemos la magnitud de esos vectores de error. En la [figura 2](#) se ha hecho la representación de módulos y acimuts referidos a un mismo origen para evaluar no sólo su orientación sino su importancia. En las figuras se ven solamente los mayores errores ya que los pequeños quedan ocultos en la zona central del gráfico. Por referirnos a dos diferentes, el b) muestra que no sólo los acimuts son preferentemente los relativos al cuadrante NO sino que los módulos parecen también seguir la misma norma: los mayores errores aparecen también en ese sentido. Los errores son, sin embargo, relativamente modestos comparados con los cuatro últimos casos.

El siguiente paso es utilizar estos gráficos para construir los que hemos llamado "mapas de densidad" con los cuales tendremos una herramienta útil para detectar datos anómalos (*outliers*) y potenciales zonas con problemas específicos.

4.3. Mapas de densidad

Los mapas de densidad ofrecen la utilidad de determinar dónde pueden estar concentrados los problemas de mayores errores en el conjunto de datos. En la [figura 3](#) se muestran los mapas para los casos analizados. Los dos primeros casos a) y b) corresponden a las comparaciones '10-50' y '25-50', mientras que el resto se corresponden con las comparativas entre los mapas 1/25000 y

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

1/50000. Los mapas de densidad nos permiten sintetizar la información conjunto acimut-módulo y ver realmente la concentración de los vectores de error por zonas.

Los casos a) y b) muestran distribuciones predominantemente circulares aunque ya vimos que debe descartarse la distribución circular uniforme. Los casos c) y d) muestran que los errores de mayor entidad se acumulan en el eje NE-SO y el caso f) es un ejemplo de distribución predominante sólo hacia el sector NE. Un paso que no hemos realizado, que debería ser el siguiente en el análisis, es la localización geográfica sobre los mapas originales de los vectores de error en estos sectores críticos. En esa localización pueden presentarse dos distribuciones básicas. La primera es que los errores estén distribuidos de forma dispersa sobre el área de estudio en cuyo caso podemos plantearnos que los problemas son generales y no afectan a una zona concreta. La segunda es que encontremos dichos errores localizados en zonas concretas lo cual nos indica que la cartografía en esas zonas es deficiente en comparación con el resto. En el primer caso deberíamos analizar si los mayores errores se relacionan con puntos de control de algún tipo concreto, por ejemplo, esquinas de construcciones o cruces de caminos lo cual sería indicador de peor localización de esos elementos que del resto (no hay razón para asumir que todos los elementos del mapa están localizados con igual exactitud).

Debemos enfatizar que estas figuras se construyen con los datos reales y no reflejan distribuciones estándar más o menos deducidas de las muestras, sino que describen la realidad de los errores. Como en el caso anterior, las figuras dan lugar a preguntas interesantes, por ejemplo ¿cuál es la razón o mecanismo subyacente que provoca una distribución diferente entre las dos primeras muestras y las cuatro restantes?

4.4. Determinación de observaciones discordantes

Una utilidad más de los mapas de densidad es la detección de los *outliers* u observaciones discordantes ya que tenemos reflejado gráfica y numéricamente el "grado de pertenencia" de cada error individual al conjunto de errores. En este caso se etiquetaron como *outliers* los errores con un valor más alto de media armónica, lo que implica un mayor alejamiento al grupo general. En la [figura 3](#), y como primera aproximación, aparecen en color verde el 2% de los puntos de mayor valor de media armónica. Puede comprobarse que una observación puede ser discordante o no para una misma distancia al punto de origen, en función de la distribución de los errores reflejada en el mapa de densidad. Este procedimiento es mucho más potente que el más habitual trazado de elipses de probabilidad ya que este último asume distribuciones estándar que no son reales.

Las observaciones discordantes se sitúan en los extremos de las distribuciones en los casos c) a f), mientras que en las dos primeras muestras las hay en varias direcciones. Aquí debemos hacer énfasis en que el término "observación discordante" no es completamente correcto y que no se trata necesariamente de errores groseros sino de aquellas observaciones que están en las zonas marginales de la distribución y que nos dan una información muy valiosa sobre dónde se localizan los mayores problemas potenciales añadiendo un valor espacial al conjunto de estadísticos.

5. Conclusión y discusiones

Hemos planteado a lo largo de este trabajo lo evidente que resulta la necesidad de analizar la exactitud de los datos espaciales como un problema vectorial y no sólo de magnitud. La consideración del error como vector permite análisis nuevos para determinar, por ejemplo, la existencia o no de anisotropía o de problemas zonales. Plantear estas pruebas nos ha parecido algo necesario, a pesar de que parezca evidente, porque tanto en la bibliografía existente como en los trabajos cartográficos cotidianos, el desarrollo de estas técnicas aplicadas al control de calidad de posición es muy escaso.

Las pruebas son simples de implementar dentro de un proceso de control estándar y no exigen la toma de más información que las posiciones "real" y medida. Habitualmente, la toma de puntos de control se realiza en mucha menor cantidad que las que hemos usado en nuestras pruebas. Ese problema (una muestra pequeña) no surge, sin embargo, por realizar estadísticos como los que proponemos sino que es general e incluso en el caso de analizar exclusivamente la magnitud del error, un bajo número de puntos de control condiciona muy fuertemente la fiabilidad de los estadísticos resultantes. De esta consideración surge un trabajo que es necesario hacer en el futuro: determinar el número de puntos de control necesarios para obtener estadísticos con una fiabilidad concreta, determinada por las normas de calidad. Lamentablemente no parece que esta sea una inquietud general y esos valores de fiabilidad seguirán sin determinarse por el momento.

Un problema más surge cuando nos preguntamos cómo introducir los resultados de estos análisis en los metadatos del mapa o, en general, de la base de datos espacial. Los estándares actuales de metadatos, como la norma ISO 19115, no incluyen mucha información sobre los criterios de calidad de posición. En cualquier caso, esos criterios están basados siempre en la concepción de un error escalar, no vectorial. Entendemos que esta es otra asignatura pendiente.

¿Cómo pueden interpretarse los resultados obtenidos en estas pruebas? ¿Por qué aparecen direcciones preferentes en algunas de ellas? El análisis de los mapas originales y de los puntos usados han mostrado resultados interesantes. Por ejemplo, los errores agrupados en determinadas direcciones preferentes (ejes X o Y o, por ejemplo, a 135° en la primera muestra), se han extraído de elementos lineales, carreteras secundarias y elementos de la zona regable (acequias), mayoritariamente. En cambio, los errores "dispersos" tienen su origen mayoritario en elementos puntuales como edificaciones. La interpretación es difícil sin un análisis más profundo pero parece que un error de localización de un elemento lineal genera un efecto de arrastre o propagación al menos sobre una zona de influencia, mientras que los elementos puntuales no tienen ese efecto. La revisión de estos errores aparentemente no independientes deberá realizarse mediante análisis de autocorrelación espacial, otro posible efecto muy poco estudiado que puede condicionar la calidad de una mapa.

Referencias bibliográficas

Ariza López, F.J. (2002): *Calidad en la producción cartográfica*. Madrid. Editorial RA-MA.

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

ASCI (1983): "Map uses, scales and accuracies for engineering and associated purposes", *American Society of Civil Engineers, Committee on Cartographic Surveying, Surveying and Mapping Division*. New York.

ASPRS (1989): "Accuracy standards for large scale maps", *Photogrammetry and Remote Sensing*, 56.

Atkinson Gordo, A.D.J., Ariza López, F.J., García Balboa, J.L. (2007): "Estimadores robustos: una solución en la estimación de valores atípicos para el control de calidad posicional", *Geofocus*, 7, pp. 171-187.

Batschelet, E. (1981): *Circular Statistics in Biology*. London, Academic Press.

Dixon, K.R., Chapman, J.A (1980): "Harmonic mean measure of animal activity areas", *Ecology*, 61, pp 1040-1044.

FGDC (1998): "National standards for spatial data accuracy", *Federal Geographic Data Committee*.

Fisher, N.I. (1995): *Statistical analysis of circular data*. Cambridge, Cambridge University Press.

Fisher, N.I., Lewis, T., Embleton, B.J.J. (1987): *Statistical analysis of spherical data*. Cambridge, Cambridge University Press.

Jammalamadaka, S.R., SenGupta, A. (2001): *Topics in circular statistics*. Singapore, World Scientific Publishing.

Mardia, K.V, Jupp, P.E. (2000): *Directional statistics*. Chichester, Wiley.

USGS (1997): "Standards for digital elevation models". *United States Geological Survey*.

TABLAS

Tabla 1. Resultados de los estadísticos básicos para vectores unitarios en las muestras de comparativa entre diferentes cartografías.

Muestras	$\bar{\theta}$	\bar{R}	ν	κ
10-50	277.6°	0.15	111.8°	0.30
25-50	307.5°	0.43	74.1°	0.96
Mérida	324.9°	0.25	95.5°	0.50
Villafranca	98.6°	0.07	132.5°	0.14
Villanueva	130.7°	0.10	122.01°	0.21
Almendralejo	276.2°	0.19	103.8°	0.39

Tabla 2. Resultados de las pruebas de ajuste a la distribución uniforme de las muestras de la comparativa entre diferentes cartografías.

Muestras	T. Kuiper	T. de Rao	T. de Rayleigh	T. de Watson
10-50	P<0.01	P<0.01	P = 0	P<0.005
25-50	P<0.01	P<0.01	P = 0	P<0.005
Mérida	P<0.01	P<0.01	P = 0	P<0.005
Villafranca	P<0.01	P<0.01	P= 0.31	P<0.005
Villanueva	P<0.01	P<0.01	P=0.07	P<0.005
Almendralejo	P<0.01	P<0.01	P = 0	P<0.005

Se muestra la probabilidad de ajuste a la distribución uniforme en cada prueba.

Tabla 3. Resultados de las pruebas de ajuste a la distribución von Mises de las muestras de la comparativa entre diferentes cartografías.

Muestras	T. Kuiper	T. de Watson
10-50	0.1>P>0.5	P<0.005
25-50	0.15>P>0.10	0.1>P>0.05
Mérida	P<0.01	P<0.005
Villafranca	P<0.025	P<0.005
Villanueva	P>0.15	P<0.005
Almendralejo	0.10>P>0.05	P<0.005

Se muestra la probabilidad de ajuste a la distribución von Mises en cada prueba.

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

FIGURAS

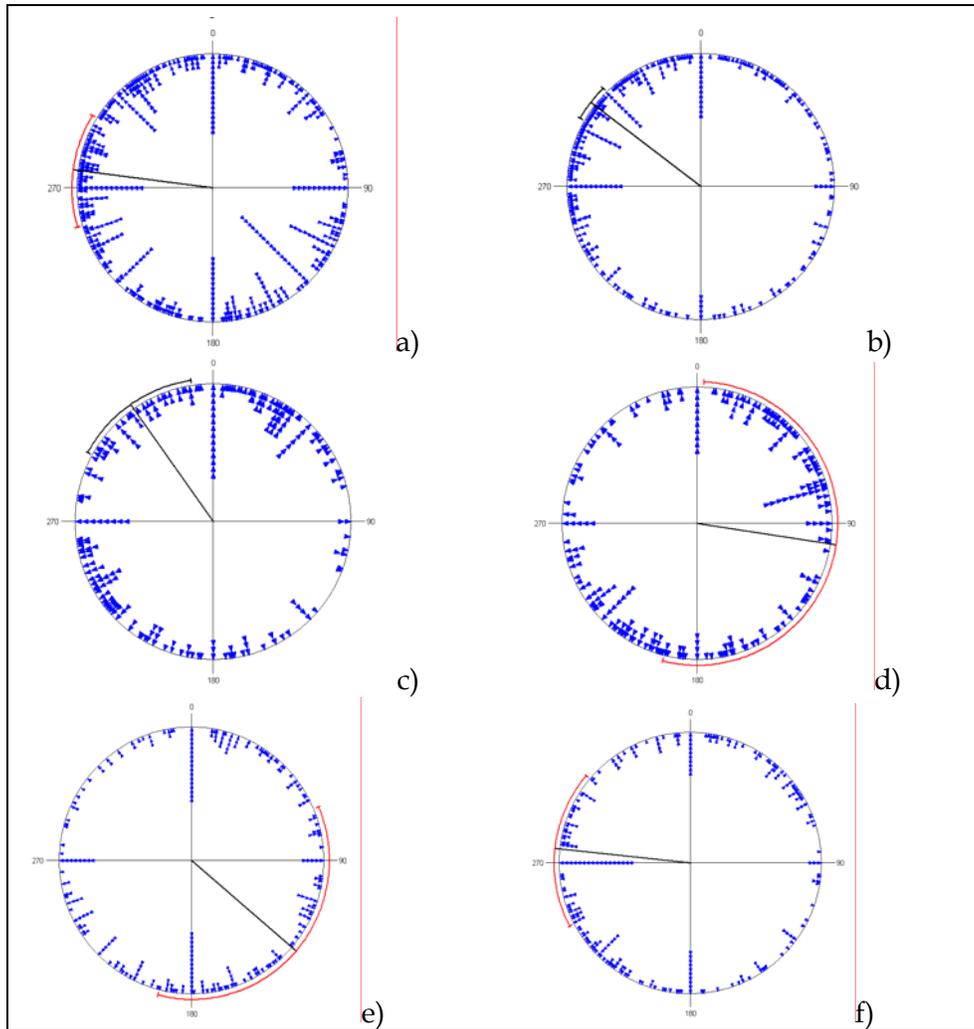


Figura 1. Distribución de datos brutos en la comparativa de diferentes cartografías.

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

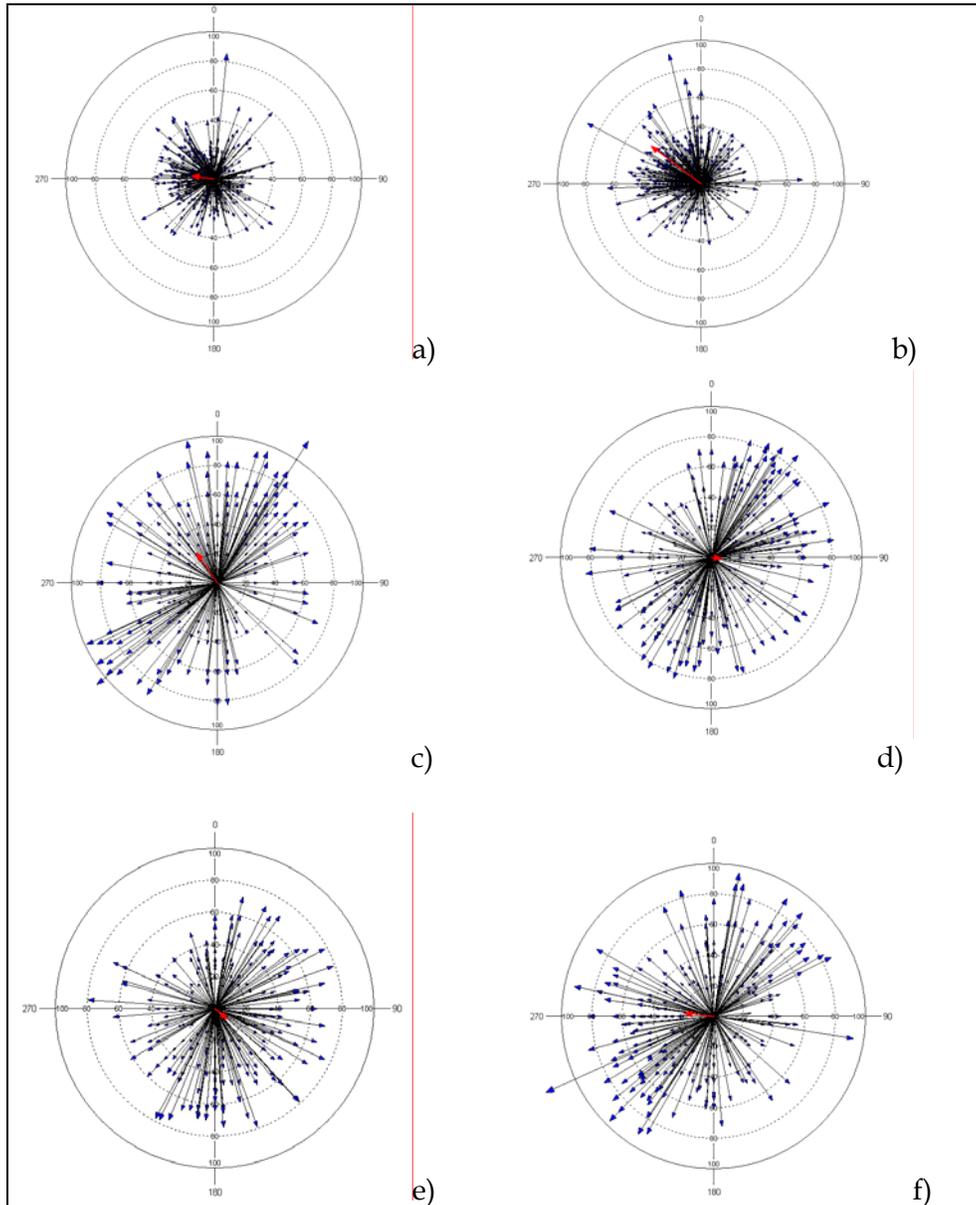


Figura 2. Gráfico de distribución de magnitudes lineales y angulares.

Polo García, M. E. y Felicísimo Pérez, A.M. (2008): "Propuesta de metodología para el análisis del error de posición en bases de datos espaciales mediante estadística circular y mapas de densidad", *GeoFocus (Artículos)*, n° 8, p. 281-296. ISSN: 1578-5157

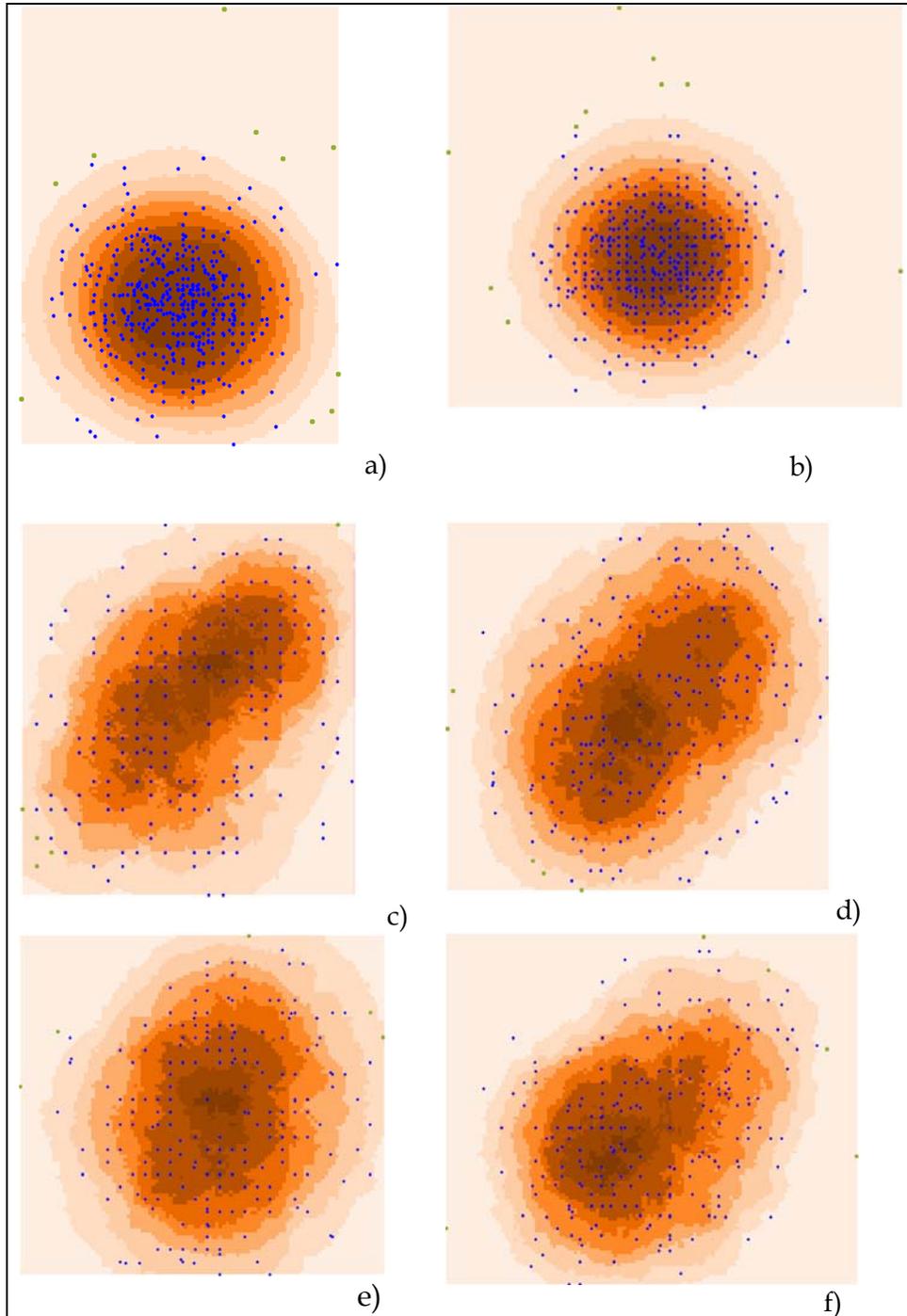


Figura 3. Mapas de densidad y distribución de observaciones discordantes en la comparativa entre diferentes cartografías.