

## UN MÉTODO ESTADÍSTICO GENERAL PARA EL CONTROL POSICIONAL DE DATOS ESPACIALES

FRANCISCO JAVIER ARIZA LÓPEZ <sup>(1)</sup>, JOSÉ RODRÍGEZ AVI <sup>(2)</sup>

<sup>1</sup>Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría

<sup>2</sup>Departamento de Estadística e Investigación Operativa

Universidad de Jaén

<sup>1</sup> [fjariza@ujaen.es](mailto:fjariza@ujaen.es), <sup>2</sup> [jravi@ujaen.es](mailto:jravi@ujaen.es),

### RESUMEN

Se presenta un método de control de la calidad posicional adecuado a cualquier geometría y dimensión. Se aplica un contraste estadístico de hipótesis y, para ello, se propone el uso de dos modelos estadísticos: un modelo binomial (BiM) y un modelo de base (BaM). El BaM representa la hipótesis del comportamiento del error en la población, que puede ser una función de distribución paramétrica o con un modelo de distribución libre. El BiM se aplica sobre el BaM y consiste en contar el número  $F$  de eventos que superan una tolerancia. El parámetro  $\pi$  del BiM se deriva del BaM por medio de la tolerancia deseada. Mediante la comparación de las probabilidades asociadas a  $F$  y  $\pi$  se decide la aceptación/rechazo. Este método permite conocer los riesgos de usuario y productor. Se presentan varios ejemplos y, con el fin de facilitar su aplicación, se incluyen tablas que relacionan  $\pi$  con  $F$  y diversos los tamaños de las muestras de control.

Palabras clave: calidad, control posicional, modelos paramétricos, modelos de distribución libre.

### A GENERAL STATISTICAL METHOD FOR POSITIONAL QUALITY CONTROL OF SPATIAL DATA

#### ABSTRACT

We present a positional-quality-control method adequate for any geometry and dimension. It is based on a statistical test of hypothesis and two statistical models are applied: a Binomial Model (BiM) is applied over a Base Model (BaM). By means of the BaM the method can be applied to any parametric or distribution free error model. The BaM represents the hypothesis about the error behavior. The BiM consists of counting the number  $F$  of fail events. The  $\pi$  parameter of the BiM is derived from the BaM by means of a desired tolerance. By comparing the probabilities associated to  $F$  and  $\pi$  a statistical acceptance/rejection decision is achieved. This method allows us

to know the user's and producer's risk. Some examples are included and in order to facilitate the application of the method some tables linking  $\pi$  with  $F$  and the control sample sizes are presented.

Keywords: quality, positional accuracy control, parametric models, distribution free models.

## 1. Introducción

La calidad posicional es una de las características más deseables de los datos espaciales y se determina por la exactitud posicional, que es uno de los elementos cuantitativos de calidad de la información geográfica, como establece la Norma Internacional 19157 (ISO, 2013). Exactitud posicional es un asunto de interés renovado debido a las capacidades ofrecidas por los Sistemas Globales de Navegación por Satélite (GNSS) y la necesidad de una mayor interoperabilidad espacial para el apoyo a las Infraestructuras de Datos Espaciales. La presencia de comportamientos distintos en cuanto a la posición de dos conjuntos de datos espaciales significa la existencia de una barrera para la interoperabilidad (Church *et al.*, 1998). Pero esta barrera no sólo afecta a los aspectos geométricos y de posición, también afecta a la componente temática (Carmel *et al.*, 2006).

En un conjunto de datos espaciales (CDE) la posición ( $X$ ) de una entidad del mundo real se registra con valores de posición de una primitiva geométrica (por ejemplo, punto, línea, superficie) en un sistema de coordenadas determinado. La exactitud posicional representa la cercanía de esos valores a la posición "real" ( $Q$ ) de la entidad en ese mismo sistema. En general,  $Q$  se obtendrá por un método más preciso que el que ha generado  $X$ ; el nivel de exigencia usual es que la fuente de referencia  $Q$  (por ejemplo, otro producto o un trabajo de levantamiento en campo) sea, al menos, tres veces más precisa.

La exactitud posicional de CDE y productos cartográficos se ha evaluado tradicionalmente utilizando puntos de control. Estos puntos se definen como "*puntos bien definidos y fácilmente identificables*" (ASCE, 1983; FGDC, 1998), y su uso venía condicionado por los métodos clásicos de levantamiento preciso. La exactitud posicional se ha determinado por medio de una evaluación estadística de los errores aleatorios y sistemáticos (DOD, 1990) y ha sido usual expresarla por medio del error cuadrático medio ( $ECM$ ), o por el valor medio de los errores ( $\mu$ ) y su nivel de desviación ( $\sigma$ ). Siguiendo esta idea, hay muchos métodos de evaluación estadística de la exactitud posicional basadas en puntos (MCPxP), entre ellas: *NMAS* (USBB, 1947), *EMAS* (ASCE, 1983), *ASPRS* (1990), *NSSDA* (FGDC, 1998), *STANAG 2215* (STANAG 2002). Algunos de estos métodos se han comparado de manera general por Ariza López y Atkinson-Gordo (2008). Muchos investigadores, tales como Dutton (1999), Chrisman (1982), Joao (1998) o Keefer *et al.* (1988), han criticado estos métodos por limitarse a puntos bien definidos y también por no hacer frente a los elementos más complejos, como lineales y los de área. Muchos de estos métodos se basan en la hipótesis subyacente de la normalidad de los errores posicionales, que no siempre es correcta (Zandbergen, 2008).

En el mundo real hay muchas entidades (carreteras, autopistas, caminos, corrientes de agua, parcelas, casas, etc.) que no son puntos, de modo que la línea y los elementos de área son muy comunes en los CDE, donde son registrados en forma de poligonales (por ejemplo, carreteras o

camino) o polígonos (por ejemplo, parcelas o construcciones). Estos elementos son los más abundantes en una base de datos general o topográfica (Cuenin, 1972). En este caso los MCPxP no son aplicables y por ello se necesitan métodos de control posicional por elementos lineales (MCPxL). Hasta la fecha ya se han propuesto algunas ideas para el control por elementos lineales, entre ellas: el *método de la distancia de Hausdorff* (MDH) (Abbas *et al.*, 1995), el *método de la distancia media* (MDM) (Skidmore y Turner, 1992), el *método del orlado simple* (MOS) (Goodchild y Hunter, 1997) y el *método del orlado doble* (MOD) (Tveite y Langaas, 1999). Es importante destacar que estas propuestas sólo vienen a ofrecer un método de análisis de la semejanza geométrica de elementos lineales y que, en ningún caso, se realiza la propuesta de un método de control posicional completo (tamaño de muestra, reglas de aceptación o rechazo, etc.). Para mayor detalle, en Ariza López y Mozas Calvache (2012) se puede encontrar un resumen y comparación de estos cuatro métodos.

Además, hoy en día el aspecto 3D es un tema muy importante en Geomática, así que hay un gran interés en el control posicional 3D de puntos (Li *et al.*, 2009), líneas (Cai y Rasdorf, 2009), volúmenes (Sargent *et al.*, 2007; Elberink y Vosselman, 2011). Este interés no es teórico, sino que se relaciona con los nuevos usos de los datos espaciales, como en los sistemas inteligentes de transporte, las ciudades inteligentes, ciudades en 3D, realidad virtual y aumentada, etc.

Según lo indicado, existe una clara limitación en la aplicación de los MCPxP (puntos bien definidos, normalidad, etc.) y todavía no existen verdaderos MCPxL, sólo las ideas base sobre la comparación de las geometrías. Con el objetivo de ofrecer un marco unificado al control posicional de cualquier geometría (punto, línea, área, etc.) y para cualquier dimensión (1D, 2D, 3D...), se propone en este trabajo un método de base estadística basado en el contraste de hipótesis, cuya finalidad es ofrecer un criterio de aceptación o rechazo del CDE bajo control. El método ha sido presentado en Ariza López y Rodríguez Avi (2014) para el caso de modelos paramétricos basados en la normalidad de los errores, pero en este artículo nos centramos en su aplicación a modelos observados (por ejemplo, elementos lineales).

Además de esta introducción, el documento consta de seis secciones principales. La siguiente sección trata las principales ideas sobre el control estadístico en la aceptación de productos; le sigue una sección centrada en presentar el método propuesto. La sección cuarta resume los principales pasos para la aplicación del método. En la sección quinta se presentan cuatro casos que abarcan controles por puntos y elementos lineales en 1D, 2D y 3D. Finalmente, se presentan unas conclusiones generales.

## 2. Control estadístico en la aceptación de productos

El objetivo del control de calidad es establecer y mantener la conformidad de los productos con los requisitos del diseño. Uno de los aspectos más importantes del control estadístico de la calidad es el control de aceptación de los productos. En estos casos, el producto final se analiza sin tener en cuenta el proceso de producción en un análisis del tipo "caja negra" en el que, tanto la aceptación como el rechazo, se deciden con criterios estadísticos.

Los controles de exactitud posicional pueden ser entendidos como procesos de aceptación basados en muestreos y pueden ser aplicados tanto por los productores como por los adquirentes o usuarios de un CDE. Cuando se aplica un control en la aceptación de un producto en la industria se requiere lo que se denomina un plan de muestreo, con el fin de establecer y hacer explícitas, tanto al producto como al usuario, las condiciones del proceso de control y la toma de la decisión de aceptación/rechazo. Aspectos que se definen son: lote (por ejemplo, un lote puede ser un conjunto de hojas de mapas producidos bajo las mismas condiciones); el tamaño del lote; el método de muestreo, el tamaño de la muestra; los riesgos de productor y usuario, etc. (AEC, 1990). La decisión de aceptar/rechazar una muestra de control puede llevarse a cabo a través de un contraste estadístico de hipótesis. Una hipótesis estadística es una declaración acerca de una función de distribución de probabilidad o sobre los valores de los parámetros de una función de distribución de probabilidad (caso paramétrico). En este contraste o prueba estadística siempre hay dos alternativas complementarias: a)  $H_0$ , la llamada hipótesis nula y b)  $H_1$ , la llamada hipótesis alternativa.

Dado que todo muestreo es un proceso al azar, se pueden dar dos tipos de errores al probar una hipótesis usando datos de un muestreo. Si se rechaza  $H_0$  cuando es verdadera, entonces se ha producido un error de tipo I ( $\alpha$ ). Si la  $H_0$  no se rechaza cuando es falsa, entonces ( $\beta$ ) se ha producido un error de tipo II. El error de tipo I se denomina riesgo del productor, ya que denota la probabilidad de que un lote/producto bueno sea rechazado. El error de tipo II se denomina riesgo del usuario o adquirente e indica la probabilidad de aceptar un lote/producto de mala calidad. A veces es más conveniente trabajar con la potencia del contraste (Montgomery, 2001), que es la probabilidad de rechazar  $H_0$  ( $Potencia = 1 - \beta = P \{rechazar H_0\}$ ).

En relación con el tamaño  $n$  de la muestra de control, el procedimiento estadístico en los contrastes de hipótesis consiste en especificar un valor para  $\alpha$  y, luego, diseñar un procedimiento de prueba, de manera que se obtenga un valor pequeño para  $\beta$ . Por lo tanto, el riesgo del productor  $\alpha$  está directamente controlado o elegido, mientras que el riesgo del consumidor es función de  $n$  y se controla de forma indirecta. Cuanto mayor sea el tamaño de la muestra el riesgo del consumidor será menor, pero también será mayor la capacidad de rechazo frente a pequeñas variaciones, lo que puede hacer que el control sea excesivamente sensible.

Este nuevo paradigma que considera los riesgos de usuario y el productor ya se ha aplicado a otros métodos de control posicional, como son los métodos ASPRS (Ariza López *et al.*, 2010) y EMAS (Ariza López *et al.*, 2008).

### 3. Propuesta de un método estadístico general para el control posicional de datos espaciales

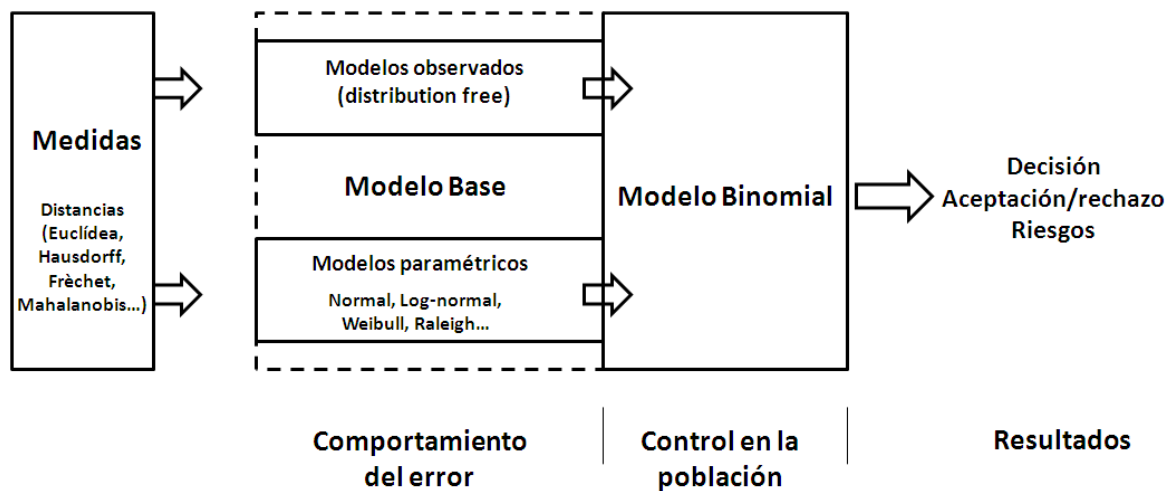
El método que se propone se resume gráficamente en la Figura 1. El método se basa en lo que denominamos un *modelo acoplado* (MC), pues considera la aplicación secuencial de dos modelos estadísticos:

- El Modelo Base (BaM). El BaM representa convenientemente el comportamiento del error en el CDE que se quiere controlar, por ello puede adoptar la forma de cualquier función de distribución, ya sea paramétrica (por ejemplo, Normal) u observada o de distribución libre. El

BaM puede basarse en cualquier tipo de medida (por ejemplo, distancia Euclídea, de Hausdorff, Mahalanobis, etc.).

- El Modelo Binomial (BiM). Este modelo permite el control sobre la población mediante un contraste de hipótesis que resulta en una decisión de aceptación o rechazo, donde los riesgos de productor y usuarios son conocidos y pueden estar controlados.

El método se puede resumir como:  $MC = BiM(BaM)$ .



**Figura 1. Visión general del método de control propuesto.**

Como en todo control de calidad, se deben considerar unos límites de control. En este caso, puesto que el problema es que los errores sean mayores que un umbral, solo se requiere uno de los límites, que denominamos *Tol* (de tolerancia). Esta tolerancia se expresa en las mismas unidades de medida que considera el BaM. La *Tol* genera una partición de la población del CDE en dos conjuntos disjuntos, aquellos elementos  $i$  que cumplen con la tolerancia, es decir, cuyos errores  $E_i \leq Tol$ , y aquellos que tienen errores mayores que la tolerancia. El porcentaje de elementos con errores mayores que la tolerancia se denota por  $\pi$ .

Según lo indicado, si el BaM es adecuado al tipo de realidad que se quiere controlar (por ejemplo, error posicional en GNSS, error posicional en datos LIDAR, error posicional en geocodificaciones, etc.) y está bien estimado, si se extrae una muestra de control del CDE cabrá esperar que la proporción de elementos de control con errores posicionales mayores que la tolerancia sea congruente con  $\pi$ . Por tanto, la base del control es justamente analizar esta situación.

En una muestra de control de tamaño  $n$  definimos como "evento-fallo" aquellos elementos de control  $i$ , tales que el error  $E_i$  en ese elemento verifica que excede a la tolerancia establecida

( $E_i > Tol$ ). El proceso de control se basa en contar el número  $F$  de casos del evento-fallo. Debido a la aleatoriedad del proceso de muestreo se puede considerar (Johnsorn *et al.*, 2005) que el número de casos  $F$  del evento-fallo sigue el modelo de una distribución binomial de parámetros  $n$  y  $\pi$ , es decir:

$$P[F > mc | F \rightarrow B(n, \pi)] = \sum_{k=mc+1}^n \binom{n}{k} \pi^k (1-\pi)^{n-k} \quad (1)$$

Donde:

- $F$       Número de casos del evento fallo.
- $mc$      Número de casos determinado.
- $n$         Tamaño de la muestra.
- $\pi$         Probabilidad de que el error sea mayor que la tolerancia ( $\pi = P[E_i > Tol]$ )

De esta forma, se puede establecer un contraste de hipótesis como forma de tomar una decisión de base estadística, donde los riesgos de productor y usuario pueden estar controlados. En este caso, la hipótesis nula es:

- $H_0$ . El CDE es adecuado y sus errores se ajustan al BaM y sólo un  $\pi\%$  de los elementos posee error mayor que la tolerancia. Como todo contraste estadístico, esta hipótesis se asume con una significación  $\alpha$  (el riesgo de equivocación, error de tipo I o riesgo del productor).

Frente a:

- $H_1$ . El CDE no es adecuado.

Las Tablas 1 a 3 presentan los resultados de calcular los  $p$ -valores para distintos tamaños de muestra  $n \in [10, 20, 30, 40, 50]$ , y posible cantidad de eventos fallo  $F \in [1 \dots 8]$  para tres valores concretos de  $\pi$ . Estas tablas se pueden calcular de manera sencilla para diversas combinaciones de parámetros, facilitando así la aplicación del método y la uniformidad en los valores.

Por ejemplo, el  $p$ -valor 0.2237 que presenta la Tabla 1 para  $n=10$  y  $F=1$  se calcula en *Excel* 2010 con la siguiente instrucción: " $=1-DISTR.BINOM.N(0;10;0,025;FALSO)$ ". En general, el  $p$ -valor se obtiene como  $1-DISTR.BINOM.N(F-1; n; \pi; FALSO)$ , en donde  $F$ ,  $n$  y  $\pi$  son los indicados en la Tabla y FALSO es el parámetro para que se obtenga la distribución acumulada en vez de la puntual, lo que se corresponde con la Ec.1.

**Tabla 1.  $p$ - valores para distintos tamaños de muestra (n) y elementos de control fuera de tolerancia (F) para  $\pi=0.025$**

| Número de elementos de control con error fuera de tolerancia<br>F | Caso $\pi = 0.025$<br>Tamaño de muestra (n) |        |        |        |        |
|---|---|--------|--------|--------|--------|
|   | n=10  | n=20   | n=30   | n=40   | n=50   |
| 1   | 0.2237                                      | 0.3973 | 0.5321 | 0.6368 | 0.7180 |
| 2   | 0.0246                                      | 0.0882 | 0.1722 | 0.2642 | 0.3565 |
| 3   | 0.0016                                      | 0.0130 | 0.0384 | 0.0779 | 0.1294 |
| 4   | 0.0001                                      | 0.0014 | 0.0064 | 0.0174 | 0.0362 |
| 5   | 0.0000                                      | 0.0001 | 0.0008 | 0.0031 | 0.0081 |
| 6   | 0.0000                                      | 0.0000 | 0.0001 | 0.0005 | 0.0015 |
| 7   | 0.0000                                      | 0.0000 | 0.0000 | 0.0001 | 0.0002 |
| 8   | 0.0000                                      | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 9   | 0.0000                                      | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 10  | 0.0000                                      | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Tabla 2.  $p$ -valores para distintos tamaños de muestra (n) y elementos de control fuera de tolerancia (F) para  $\pi=0.05$**

| Número de elementos de control con error fuera de tolerancia<br>F | Caso $\pi = 0.05$<br>Tamaño de muestra (n) |        |        |        |        |
|---|--|--------|--------|--------|--------|
|   | n=10                                       | n=20   | n=30   | n=40   | n=50   |
| 1   | 0.4013                                     | 0.6415 | 0.7854 | 0.8715 | 0.9231 |
| 2   | 0.0861                                     | 0.2642 | 0.4465 | 0.6009 | 0.7206 |
| 3   | 0.0115                                     | 0.0755 | 0.1878 | 0.3233 | 0.4595 |
| 4   | 0.0010                                     | 0.0159 | 0.0608 | 0.1381 | 0.2396 |
| 5   | 0.0001                                     | 0.0026 | 0.0156 | 0.0480 | 0.1036 |
| 6   | 0.0000                                     | 0.0003 | 0.0033 | 0.0139 | 0.0378 |
| 7   | 0.0000                                     | 0.0000 | 0.0006 | 0.0034 | 0.0118 |
| 8   | 0.0000                                     | 0.0000 | 0.0001 | 0.0007 | 0.0032 |
| 9   | 0.0000                                     | 0.0000 | 0.0000 | 0.0001 | 0.0008 |
| 10  | 0.0000                                     | 0.0000 | 0.0000 | 0.0000 | 0.0002 |

**Tabla 3.  $p$ - valores para distintos tamaños de muestra (n) y elementos de control fuera de tolerancia (F) para  $\pi=0.1$**

| Número de elementos de control con error fuera de tolerancia<br>F | Caso $\pi = 0.1$<br>Tamaño de muestra (n) |        |        |        |        |
|---|---|--------|--------|--------|--------|
|   | n=10                                      | n=20   | n=30   | n=40   | n=50   |
| 1   | 0.6513                                    | 0.8784 | 0.9576 | 0.9852 | 0.9948 |
| 2   | 0.2639                                    | 0.6083 | 0.8163 | 0.9195 | 0.9662 |
| 3   | 0.0702                                    | 0.3231 | 0.5886 | 0.7772 | 0.8883 |
| 4   | 0.0128                                    | 0.1330 | 0.3526 | 0.5769 | 0.7497 |
| 5   | 0.0016                                    | 0.0432 | 0.1755 | 0.3710 | 0.5688 |
| 6   | 0.0001                                    | 0.0113 | 0.0732 | 0.2063 | 0.3839 |
| 7   | 0.0000                                    | 0.0024 | 0.0258 | 0.0995 | 0.2298 |
| 8   | 0.0000                                    | 0.0004 | 0.0078 | 0.0419 | 0.1221 |
| 9   | 0.0000                                    | 0.0001 | 0.0020 | 0.0155 | 0.0579 |
| 10  | 0.0000                                    | 0.0000 | 0.0005 | 0.0051 | 0.0245 |

#### 4. Guía para la aplicación del método

Desde un punto de vista práctico, los pasos para aplicar este método son los siguientes:

- Determinar un BaM adecuado a los datos a controlar. En general, ésta es labor del productor. El productor debe conocer su producto, tener los procesos de producción bajo control y realizar seguimientos periódicos de su producción. Todo ello proporciona evidencias más que suficientes para determinar el BaM de una manera fiable y estable, bien en la forma de una función de distribución observada o como una función de distribución paramétrica. El BaM es la Función de Distribución estadística de los errores en la población de interés, esto quiere decir que para cada valor de error, por ejemplo  $E$ , la curva expresa la probabilidad acumulada hasta ese valor, es decir,  $BaM(E) = P(X \leq E)$ . La representación visual más usual de los errores de una población es en forma de histograma (función de densidad), pero su conversión en Función de Distribución es directa, tanto para el caso paramétrico (por integración) como para el caso no paramétrico; así, en una distribución observada para un valor de error  $E$  basta con ir acumulando las frecuencias de las ocurrencias hasta ese valor:  $BaM(E) = \sum f(e) / \forall e \leq E$ . Otro aspecto de la definición del BaM es la medida utilizada y la métrica que se utiliza. En la mayoría de los MCPP el BaM es una hipótesis subyacente, pero en nuestro método es una hipótesis clara y evidente. Esto es así pues en los MCPP al uso (por ejemplo, NSSDA, EMAS, etc.) los cálculos y contrastes que se realizan se centran en los datos y no se comprueba el modelo base subyacente (por ejemplo, el NSSDA o el EMAS no exigen comprobar la normalidad), que se da por supuesto. En el método propuesto el modelo base sí se considera siempre, aparece en todos los cálculos como BaM.
- Determinar una tolerancia acorde a unos requisitos de uso concretos. La tolerancia debe expresarse con la misma métrica y medida que el BaM. En cuanto a su valor, la tolerancia debe ser sensata en cuanto a que ha de tener en cuenta la voz del proceso. Es decir, a un producto no se le puede pedir más allá de lo que es capaz de proporcionar su proceso productivo (por ejemplo, no podemos exigir errores centimétricos a unos datos tomados con un receptor GNSS de navegación). Es interesante indicar que, dado que el método se basa en el conteo de casos, la tolerancia que se proponga no tiene por qué ser un valor medio de error, o un valor máximo de error, puede ser cualquier tipo de valor y, como ya se ha indicado, puede utilizarse cualquier métrica.
- Determinación de un tamaño de muestra. En un contraste estadístico de hipótesis siempre están presentes el nivel de significación ( $\alpha$ ) o error de tipo I y el riesgo de usuario ( $\beta$ ) o error de tipo II. En los contrastes siempre se establece el nivel de significación, valores usuales en estadística son 10%, 5% y 1%. En control de calidad, el riesgo del productor se suele acotar al 5%. En cualquier caso, el tamaño de muestra no es relevante frente al error de tipo I, lo que significa que, si no se plantea restricción alguna sobre  $\beta$ , este tamaño puede ser realmente reducido. Por tanto, es el riesgo de usuario el que se encuentra ligado al tamaño de muestra. La fijación del tamaño de muestra para acotar  $\beta$  se puede realizar a partir de la determinación del valor de  $\pi_1$  (probabilidad de que un punto tenga una tolerancia mayor que la especificada mínima para la que se desee rechazar el contraste en caso que  $H_0$  sea falsa). De esta manera la curva correspondiente a un nivel dado de error de tipo I se calcula:

$$P[F < mc | F \rightarrow B(n, \pi^a)];$$

para  $\pi^a > \pi_1$  y distintos valores de  $n$ .

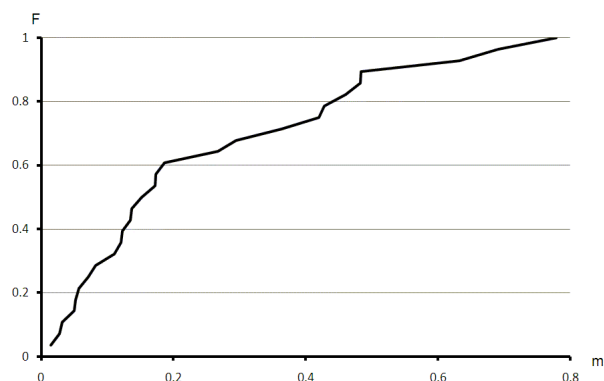
- Identificación de la muestra de control. La muestra de control se obtendrá siempre por procedimientos de muestreo aleatorio sobre el CDE  $X$ . Estos procedimientos deben ser automatizados para que no intervenga el sesgo humano en la selección de los elementos de control.
- Materialización de la muestra y cálculo de discrepancias. La muestra se deberá materializar tanto en el CDE  $X$  como en la fuente de mayor exactitud ( $Q$ ). Se calcularán las discrepancias entre parejas de elementos homólogos de  $X$  y  $Q$  utilizando la misma definición de métrica y medida que en el BaM.
- Determinación del número de casos  $F$  con error mayor que la tolerancia. Se contará el número de casos cuyo valor de error excede la tolerancia establecida ( $E_i > Tol$ ).
- Cálculo del  $p$ -valor del contraste. Aplicación de la Ec.1 o la lectura de las Tablas 1, 2 y 3 permitirá conocer el  $p$ -valor correspondiente al conjunto de datos  $\{F, n, \pi\}$ .
- Decisión. Para la decisión de aceptación o rechazo se compara el  $p$ -valor con  $\alpha$ . Si  $p < \alpha$  se rechaza y si  $p \geq \alpha$  no existen evidencias para el rechazo.

## 5. Ejemplos de aplicación a casos reales

En este apartado se presentan varios ejemplos de aplicación del método a conjuntos de casos reales con distintas dimensiones (1D, 2D, 3D). Se ha elegido una casuística diversa con el fin de demostrar que la propuesta es válida para casos en los que los MCPxP al uso (NMA, EMAS, NSSDA, etc.) no son satisfactorios dado que no se cumpliría el modelo subyacente de tipo Normal. En todos los ejemplos se supondrá que el BaM es conocido, por ello, en primer lugar, para cada caso, se presentan los datos del BaM utilizando una referencia a su origen y, posteriormente, se plantean los restantes pasos del método propuesto. Los BaM que se presentan son todos modelos observados, pero el método propuesto se puede aplicar sobre modelos paramétricos (por ejemplo, el gaussiano) (Ariza López y Rodríguez Avi, 2014).

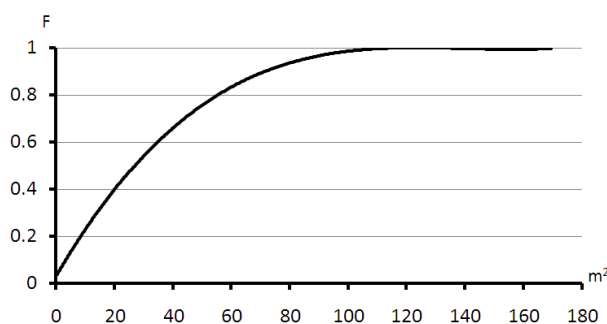
Los BaM se muestran en forma de Funciones de Distribución, en todos los gráficos el eje horizontal representa los errores, pero se desea advertir que no en todos los casos la definición de distancia utilizada es la misma (por ejemplo, distancia euclídea, distancia de Hausdorff), y que el rango de valores depende de cada caso. Por su parte, el eje vertical contabiliza frecuencias acumuladas, es decir, el número de casos, expresados en tanto por uno.

Caso 1. Este caso se refiere a datos 1D de controles LIDAR altimétricos. La particularidad de este caso es que se sabe que los errores posicionales verticales LIDAR en terrenos no-desnudos no son normales. Para el ejemplo se tomará la Figura 2 como BaM. Esta figura muestra la función de distribución de este tipo los errores absolutos procedentes del informe de Dewberry (Dewberry, 2004) para el caso de una cubierta de cultivos herbáceos. La distancia utilizada para la medida del error es la euclídea con dimensión de metros [m].



**Figura 2. Función de distribución de probabilidad observada de errores altimétricos LIDAR (Caso 1).**

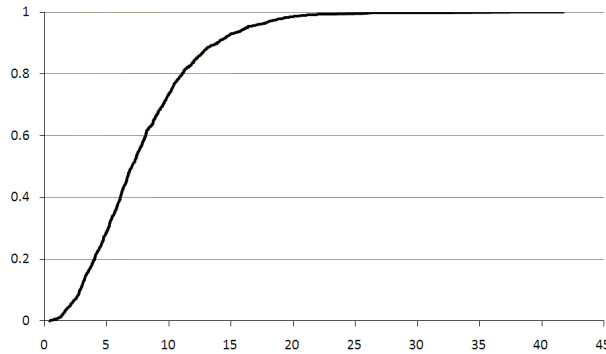
Caso 2. Este caso se refiere a datos 2D de controles planimétricos por puntos. La particularidad de este caso es que el método se aplica directamente a los datos observados, no se usa ninguna hipótesis subyacente sobre la naturaleza del modelo. Como fuente del BaM se toman los datos procedentes del Apéndice 3-B de la norma NSSDA (FGDC, 1998). Es interesante indicar que el NSSDA se basa en la normalidad de los datos y al aplicarse sobre sumas de cuadrados de errores el modelo subyacente es una  $\chi^2$ , no obstante insistimos que esta hipótesis no se utiliza en nuestro método, nosotros usamos el modelo observado (Figura 3). La distancia utilizada para la medida del error es la Euclídea, con dimensión de metros al cuadrado [ $m^2$ ].



**Figura 3. Función de distribución de probabilidad observada de errores cuadráticos planimétricos (Caso 2).**

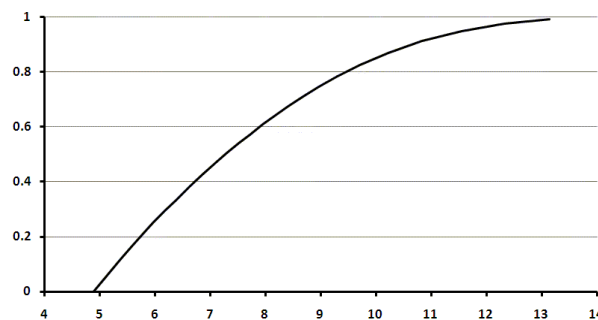
Caso 3. Este caso se refiere a datos 2D de controles planimétricos por elementos lineales. La particularidad de este caso es la evaluación de la posición 2D en elementos lineales donde la distancia euclídea no funciona adecuadamente. En este caso se ha utilizado la distancia de Hausdorff que, frente a la distancia euclídea, presenta propiedades más ventajosas en el caso de su aplicación a geometrías. El BaM que se adopta es el que se presenta en la Figura 4. En este caso los

datos proceden del trabajo referido en Ariza López *et al.* (2011). La distancia utilizada para la medida del error es la Hausdorff, con dimensión de metros [m].



**Figura 4. Función de distribución de probabilidad observada del error 2D en líneas (Caso 3).**

Caso 4. Este caso se refiere a datos 3D de controles planimétricos por elementos lineales. La particularidad de este caso es la evaluación de la posición 3D en elementos lineales. El BaM que se adopta es el que se presenta en la Figura 5. En este caso los datos son resultados preliminares del proyecto E3DELING (Ariza López *et al.* 2011). La distancia utilizada para la medida del error es la Hausdorff, con dimensión de metros [m].



**Figura 5: Función de distribución de probabilidad observada del error 3D en líneas (Case 4).**

Ya que se han presentado los cuatro conjuntos de datos, considérese ahora lo siguiente para cada uno de estos casos:

Caso 1. Se toma una  $Tol_{C1}=0.5m$ ,  $n_{C1} = 20$ ,  $\alpha = 5\%$ .  $Tol_{C1}$  implica que  $\pi_{C1} = 0.1$ .

Caso 1. Se toma una  $Tol_{C2}=93m^2$ ,  $n_{C1} = 20$ ,  $\alpha = 5\%$ .  $Tol_{C2}$  implica que  $\pi_{C2} = 0.1$ .

Caso 3. Se toma una  $Tol_{C3}=18m$ ,  $n_{C3} = 20$ ,  $\alpha = 5\%$ .  $Tol_{C3}$  implica que  $\pi_{C2} = 0.1$ .

Caso 4. Se toma una  $Tol_{C4}=10.7m$ ,  $n_{C4} = 20$ ,  $\alpha = 5\%$ .  $Tol_{C4}$  implica que  $\pi_{C4} = 0.1$ .

Donde las tolerancias, los tamaños de muestra y los niveles de significación son decisiones previas y donde los valores de  $\pi$  son el resultado de entrar en cada uno de los BaM con la tolerancia considerada para ese caso.

Dado que se trata de ejemplos dispares, para simplificar la comparación y entender mejor la generalidad del método, se han tomado distintos valores de tolerancia, pero se han mantenido iguales en todos los casos el tamaño de muestra, la significación del contraste y  $\pi$ . Obsérvese que las tolerancias establecidas implican que  $\pi_{C1} = \pi_{C2} = \pi_{C3} = \pi$ . De esta forma, según lo ya indicado, con  $\pi$  y  $n$ , por medio de la Ec.(1) se puede determinar la probabilidad de  $F$  eventos en el BaM, pero también se puede utilizar la Tabla 3.

Por tanto, asumiendo los BaM expuestos para cada uno de los casos y los datos indicados en los párrafos anteriores, y según lo indicado en el apartado 4, lo que realmente quedaría es la ejecución material del control, el cálculo de los errores, el conteo de los elementos  $F$  y la toma de la decisión de aceptación o rechazo. Así, para cualquiera de los cuatro casos presentado, si  $F=2$  se tiene que  $p=0.6083$  (Tabla 3 para  $n=20$ ) y, por tanto, se acepta el suministro de producto. Echando un vistazo a la Tabla 3, en su columna de  $n=20$  se observa que se aceptará siempre que  $F \leq 4$  y que se rechazará para todo  $F \geq 5$ . Como se ha indicado, los  $p$ -valores son los riesgos del productor, así si el número de elementos que superan la tolerancia fuera ahora  $F=3$  la Tabla 3 ofrece el valor  $p=0.3231$ , lo que viene a significar que no aceptar el producto conlleva, en este caso, la probabilidad de equivocarnos el 32% de las veces.

## 6. Conclusiones

Se ha presentado un método de control posicional válido para los casos 1D, 2D y 3D y aplicable sobre cualquier geometría (punto, línea, área). El control se basa en un contraste estadístico de hipótesis que se plantea sobre dos modelos de manera acoplada, primero un modelo base (BaM) y, sobre él, un modelo binomial (BiM). El BaM debe reflejar el comportamiento de la población de errores a controlar. Este comportamiento puede quedar caracterizado por una función de distribución paramétrica (por ejemplo, Normal, Chi2, etc.) o por cualquier modelo observado (distribución libre). El BiM es fijo y la idea de su aplicación es contar el número de errores que superan un valor de tolerancia previamente establecido. El control estadístico se establece sobre la distribución binomial en la que su parámetro  $\pi$  se deriva del BaM a partir de la tolerancia. De esta forma, el esquema propuesto es adecuado para cualquier BaM y, por ello, puede cubrir casuísticas en las que los métodos de control posicional tradicionales no son operativos.

El control propuesto es no paramétrico en el sentido de que no se requiere resumir la muestra en un reducido número de parámetros (por ejemplo, media, desviación) y tomar la decisión sobre ellos. El contraste se realiza directamente sobre la población, lo que libera del uso de modelos subyacentes (por ejemplo, normalidad, propiedad del uso de la media, etc.). Por ello, el contraste es robusto al realizarse sobre la muestra en su totalidad

Se han presentado cuatro ejemplos de aplicación que muestran la sencillez y uniformidad de aplicación de la parte estadística del método en casos donde el tipo de geometría (punto, línea), la dimensión espacial (puntos 1D y 2D y líneas 2D y 3D) y la métrica utilizada (euclídea, Hausdorff) son distintas. También se ha presentado un caso de aplicación a un control con dos tolerancias. Para facilitar la aplicación se han proporcionado tres tablas que muestran los *p-valores* esperados para los tamaños de muestra  $n=[10, 20, 30, 40, 50]$ , los parámetros  $\pi=[0.025, 0.05, 0.10]$  del BiM y el conteo de errores en  $F=[1 \dots 10]$ . Estas tablas se pueden generalizar de manera sencilla por medio de la aplicación de la Ec.(1).

La clara ventaja de este método es que ofrece un marco común para el control posicional de todo tipo de datos espaciales, por lo que, siempre que se esté utilizando la misma métrica y medida, el método ofrece una misma forma de control para geometrías de datos diversas y, por tanto, la posibilidad de obtener resultados totalmente comparables .

### Agradecimientos

Este trabajo ha sido financiado parcialmente por el Ministerio de Economía y Competitividad del Reino de España y por el Fondo Europeo de Desarrollo Regional (FEDER), con la ayuda de referencia BIA2011-23271 correspondiente al proyecto de investigación "Evaluación 3D de elementos lineales de información geográfica (E3DLING)". Igualmente, se agradece a la Junta de Andalucía la financiación económica del Grupo de Investigación Ingeniería Cartográfica (PAIDI-TEP-164) desde 1997 hasta la fecha.

### Referencias bibliográficas

- Abbas, L.; Grussenmeyer, P. y Hottier, P. (1995): "Contrôle de la planimétrie d'une base de données vectorielles: une nouvelle méthode basée sur la distance de Hausdorff: la méthode du contrôle linéaire", *Bul. Société Française de Photogrammétrie et Télédétection* 137, pp. 6-11.
- AEC (1990): *Técnicas de control de calidad*. Asociación Española para la Calidad, Madrid.
- Ariza López, F.J.; Atkinson-Gordo, A.D.; Rodríguez Avi, J. y García-Balboa, J.L. (2010): "Analysis of user and producer risk when applying the ASPRS Standards for Large Scale Maps", *Photogrammetric Engineering and Remote Sensing*, 76(5), pp. 625-632.
- Ariza López, F.J. y Atkinson-Gordo, A.D.J. (2008): "Analysis of some positional accuracy assessment methodologies", *Journal of Surveying Engineering* 134 (2), pp. 404-407.
- Ariza López, F.J.; Atkinson-Gordo, A.D.J. y Rodríguez Avi, J. (2008): "Acceptance curves for the positional control of geographic data bases", *Journal of Surveying Engineering* 134 (1), pp. 26-32.
- Ariza López, F.J. y Mozas Calvache, A.T. (2012): "Comparison of four line-based positional assessment methods by means of synthetic data", *J GeoInformatica*, 16(2), pp. 221-243.

Ariza López, F.J. y Rodríguez Avi, J. (2014): "Un método estadístico general para el control posicional de datos espaciales", *GeoFocus (Artículos)*, n° 14, p. 105-119. ISSN: 1578-5157

---

Ariza López, F.J.; Mozas Calvache, A.T.; Ureña Cámara, M.A.; Alba Fernández, V.; García Balboa, J.L.; Rodríguez Avi, J. y Ruiz-Lendínez, J.J. (2011): "Sample size influence on line-based positional assessment methods for road data", *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(5), pp. 708–719.

Ariza López, F.J. y Rodríguez Avi, J. (2014): "A statistical model inspired by the National Map Accuracy Standard", *Photogrammetric Engineering & Remote Sensing*, doi: 10.14358/PERS.80.3.000

ASCE (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, USA.

ASPRS (1990): "Accuracy standards for large scale maps", *Photogrammetric Engineering and Remote Sensing* 56(7), pp. 1068-1070.

Cai, H. B. y Rasdorf, W. (2009): "Accuracy evaluation and sensitivity analysis of estimating 3D road centerline length using LIDAR and NED", *Photogrammetric Engineering and Remote Sensing*, 75(6): 657-665.

Carmel, Y.; Flather C. y Dean, D. (2006): "A methodology for translating positional error into measures of attribute error, and combining the two error sources", *Proceedings of Accuracy 2006. 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Lisbon, pp. 3-17.

Cuenin, R. (1972): *Cartographie générale*, Eyrolles, Paris, 1972.

Chrisman, N.R. (1982): "A theory of cartographic error and its measurement in digital data bases", *Proc. AutoCarto 5*, Crystal City, Virginia, USA, pp. 159-168.

Church, R.; Curtin, K.; Fohl, P.; Funk, C.; Goodchild, M.; Kyriakidis, P. y Noronha, V. (1998): *Positional distortion in geographic data sets as a barrier to interoperability*. Technical Papers ACSM. American Congress on Surveying and Mapping. Bethesda, Maryland.

Dewberry, L.L.C. (2004): *Worcester County LIDAR 2002 Quality Assurance Report*. Maryland Department of natural Resources.

DOD (1990): *MIL STD 60001: Mapping, charting and geodesy accuracy*. U.S. Department of Defense. Washington, D.C.

Dutton, G.H. (1999): "Scale, sinuosity and point selection in digital line generalization", *Cartography and Geographic Information Science* 26 (1), pp. 33-53.

Elberink, S.O. y Vosselman, G. (2011): "Quality analysis on 3D building models reconstructed from airborne laser scanning data", *ISPRS Journal of Photogrammetry and Remote Sensing*, 66 (2), pp. 157-165,

FGDC (1998): *FGDC-STD-007: Geospatial Positioning Accuracy Standards, Part 3. National Standard for Spatial Data Accuracy*, Federal Geographic Data Committee, Reston, USA.

Goodchild, M.F. y Hunter, G. (1997): "A simple positional accuracy measure for linear features", *International Journal of Geographical Information Science*, 11 (3), pp. 299-206.

Ariza López, F.J. y Rodríguez Avi, J. (2014): "Un método estadístico general para el control posicional de datos espaciales", *GeoFocus (Artículos)*, n° 14, p. 105-119. ISSN: 1578-5157

---

ISO (2013): *ISO 19157: Geographic information - Data quality*. International Organization for Standardization. Geneva.

Joao, E.M. (1998): *Causes and consequences of map generalisation*, Taylor & Francis, London, UK.

Johnsorn, N.L.; Kemp, A.W. y Kotz. S. (2005): *Univariate discrete distributions*. Wiley.

Keefer, B.J.; Smith, J.L. y Gregoire, T.G. (1988): "Simulating manual digitizing error with statistical models", *Proc. of GIS/LIS 88*, San Antonio, Texas, pp. 475-483.

Li, R.; Niu, X.; Liu, C. y Wu, B. (2009): "Impact of Imaging Geometry on 3D Geopositioning Accuracy of Stereo IKONOS Imagery", *Photogrammetric Engineering and Remote Sensing*, 75(9), pp. 1119-1125.

Montgomery, D.C. (2001): *Introduction to statistical quality control*. 4th Ed. John Wiley & Sons, New York.

Sargent, I.; Harding, J. y Freeman, M. (2007): "Data quality in 3D: Gauging quality measures from users' requirements", *Proceedings of 5th International Symposium on Spatial Data Quality 2007*. ITC, Enschede, The Netherlands.

Skidmore, A. y Turner, B. (1992): "Map accuracy assessment using line intersect sampling", *Photogrammetric Engineering and Remote Sensing* 58 (10), pp. 1453-1457.

STANAG (2002): *Standardization Agreement 2215: Evaluation of land maps, aeronautical charts and digital topographic data*. North Atlantic Treaty Organization. Bruxelles, Belgium.

Tveite, H. y Langaas, S. (1999): "An accuracy assessment method for geographical line data sets based on buffering", *International Journal of Geographical Information Science* 13 (1), pp. 27-47.

USBB (1947): *United States National Map Accuracy Standards*. U.S. Bureau of the Budget, Washington, D.C.

Zandbergen, P.A. (2008): "Positional accuracy of spatial data: non-normal distributions and a critique of the National Standard for Spatial Data Accuracy", *Transactions in GIS* 12(1), pp. 103-130.