

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

ESTIMADORES ROBUSTOS: UNA SOLUCIÓN EN LA UTILIZACIÓN DE VALORES ATÍPICOS PARA EL CONTROL DE LA CALIDAD POSICIONAL

ALAN D.J. ATKINSON GORDO¹, FRANCISCO J. ARIZA LÓPEZ², JOSÉ L. GARCÍA-BALBOA³

⁽¹⁾ Escuela Politécnica, Universidad de Extremadura
Av. Universidad s/n, 10.071-Cáceres

^(2,3) Escuela Politécnica Superior, Universidad de Jaén
P. Las Lagunillas s/n, 23.071-Jaén. España

⁽¹⁾ atkinson@unex.es, ⁽²⁾ fjariza@ujaen.es, ⁽³⁾ jlbaloa@ujaen.es

RESUMEN

Los principales estándares que verifican la calidad posicional de una Base de Datos Geográfica (BDG), tienen como común denominador la detección y eliminación previa de los valores considerados como atípicos (*outliers*). Ello puede llegar a suponer un importante sesgo en la información final que se transmita al usuario de la cartografía, por lo que aquí presentamos una herramienta capaz de solventar este problema: los estimadores robustos. Éstos nos permiten ponderar aquellos valores que se encuentran más alejados de los centrales en una serie. Así, hemos analizado algunos de los principales estimadores (*M* de Huber, método Danés, Geman y McClure...) y los hemos aplicado a la detección y ponderación de valores atípicos en un control de exactitud planimétrico sobre poblaciones sintéticas contaminadas artificialmente. Los resultados son claramente esperanzadores para su aplicación a corto plazo en el control de la calidad posicional.

Palabras clave: Estimador robusto, cartografía, calidad, exactitud posicional, *outlier*, error planimétrico

ROBUST ESTIMATORS: A SOLUTION FOR USING OUTLIERS IN POSITIONAL ACCURACY ASSESSMENT

ABSTRACT

All positional accuracy assessment standard methodologies are based on a previous detection and elimination of the values considered as atypical (*outliers*). This could imply an important lost of information when reporting the assessment results to the users. The robust estimators are a capable tool for dealing with outliers by means of weighting most extreme values from the mean. In this work we have analyzed some robust estimators (*M* from Huber, the Danish method, etc.) when applying positional accuracy assessment standards. A simulation process is

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

carried out in which synthetic populations of errors, of known parameters, are artificially contaminated. Our results point out that the Danish Method performs better than others.

Keywords: Robust estimator, cartography, quality, positional accuracy, outlier, planimetric error.

1. Introducción

A la hora de realizar un control de calidad, la estadística es una herramienta básica y, casi, imprescindible para poder obtener resultados válidos. En el caso de la exactitud posicional de una BDG, la práctica totalidad de los estándares a nivel internacional (EMAS, ASPRS, NSSDA ...) (Ariza, 2002) se fundamentan en un análisis desde un punto de vista de la estadística clásica, desechando valores atípicos o extremadamente grandes y considerando los datos como componentes unidimensionales e independientes entre sí (ASCE, 1983; ASPRS, 1989; FGDC, 1998).

Frente a ello, la estadística robusta es una rama de la estadística que pretende estudiar el comportamiento de diferentes procedimientos cuando existe una pequeña variación en los supuestos iniciales, o cuando existe la posibilidad de que el modelo esté contaminado por ciertas observaciones conocidas por el nombre de *outliers* (o valores atípicos) que producen influencias en los resultados que conllevan a resultados erróneos (Gento *et al.*, 2004). Así, la experiencia y los estudios de simulación sobre los procedimientos propuestos en estadística robusta, aconsejan el uso de estas técnicas en toda modelización estadística. Como el diseño de estos procedimientos, en el caso en el que no existan valores atípicos en los datos del estudio (aspecto imposible de confirmar), el resultado que proporcionan estos métodos son muy similares a los dados por el estimador clásico al que sustituye o complementa. En cambio, en presencia de valores atípicos, las buenas propiedades de las que gozan estos procedimientos ante ellos (ver Ortega, 2000), nos aseguran una estimaciones más fiables de los parámetros en estudio (Gento *et al.*, 2004).

Los valores atípicos de una muestra son aquellos excesivamente grandes o pequeños tal que, tras su comprobación, no pueden considerarse como equivocaciones o errores groseros del proceso. Por lo anterior, aunque tengan poca probabilidad de ocurrencia, estos valores deben tomarse como pertenecientes a la función de distribución del fenómeno o característica, pero su inclusión en el análisis condiciona extremadamente los resultados que se derivan de la estadística tradicional que se aplica. Frente a esta circunstancia, común en los controles posicionales, los métodos de control posicional (MCP) no dan ninguna indicación al respecto. Aquí la situación es interesante: por un lado la inclusión en los cálculos estadísticos afectaría mucho a los valores estimados y, por otro, la eliminación de los mismos es como cerrar los ojos frente a una realidad que nos incomoda. Por ello, la solución común suele consistir en un cuidadoso proceso de eliminación, para no tomarlos en cuenta en el análisis numérico, junto con su inclusión en los informes finales dando aviso de esta circunstancia. Ejemplos de este hacer, y que pueden servir de guía para el tratamiento de estos casos, pueden encontrarse en alguna información complementaria como MPLMIC (1999).

En un control de calidad posicional se toma una muestra reducida de puntos de control (p.e. 20) (FGDC, 1998) para minimizar los costes de su captura sobre el terreno. Cabe la posibilidad de que una parte de ellos resulte con valores atípicos tal que si se desechan se pierde representatividad

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

al perder elementos, pero también una parte importante de la información, dado que al no entrar en el análisis ya no sería transmitida al usuario del producto cartográfico. Llegados a este punto cabe preguntarse, ¿qué representa realmente un valor atípico?, ¿no se puede encontrar el usuario del mapa con valores atípicos?, ¿en qué medida y magnitud?, ¿cómo pueden influir en la toma de decisiones sobre una BDG?

Muy probablemente, la mayoría de estos aspectos podrán solventarse empleando estimadores robustos. Así, analizando los datos desde el punto de vista de la estadística robusta, ésta permite emplear aquellos valores que puedan ser considerados como *outliers* o valores atípicos, interviniendo en los cálculos y el análisis final del producto, sin que su influencia sea determinante.

Junto a esta introducción, el trabajo se organiza en cinco apartados. En el primero se realiza una presentación general de algunos estimadores robustos, a lo que sigue una explicación del procedimiento de ajuste para su aplicación en el control de la exactitud posicional. Posteriormente se realiza el análisis comparativo del funcionamiento de los estimadores frente a poblaciones sintéticas con cierto grado de contaminación con valores extremos. Para mostrar la validez del proceso se incluye un ejemplo práctico de aplicación sobre dos hojas a escala E10k en las que se ha tomado un número abundante de puntos de control. Finalmente se presentan las conclusiones. El trabajo presentado en este artículo se fundamenta en los estudios previos y la investigación realizada por Atkinson, Ariza y Rodríguez (2002), Atkinson, Ariza y García (2005) y Atkinson (2005).

2. Estimadores robustos en el control de la calidad posicional

Cuando se va a realizar un control de la exactitud posicional, quien desee aplicar un determinado estándar ha de saber que éste, normalmente, se fundamenta en determinadas hipótesis previas que han de cumplir los datos a analizar. Tal es el caso del comportamiento de los errores según una distribución normal $N(\mu, \sigma)$ (¹). En el caso de existir alguna duda respecto al modelo estadístico, la calidad de la muestra o incertidumbre acerca de las decisiones adoptadas, se hace necesario el emplear este tipo de métodos. Si la suposición del modelo es incorrecta, puede no estar demasiado claro y cuestionar si este tipo de procedimientos es necesario. Así, podría ser suficiente un análisis en dos pasos: limpieza de los datos aplicando algún tipo de método para rechazar observaciones atípicas; y utilización de la estimación clásica y procedimientos de comprobación en los datos restantes. No obstante, esto resulta bastante complicado ya que: es difícil separar claramente los dos pasos, el conjunto original de observaciones suele estar formado por datos normales mezclados con valores atípicos y, por último, los mejores procedimientos de rechazo no alcanzan las funciones de los mejores procedimientos robustos (Domingo, 2000).

Por otra parte, los actuales estándares, además de emplear muestras relativamente pequeñas (FGDC (1998): "...al menos 20 puntos de control..."), no tienen en cuenta aquellos valores (*outliers*), que se encuentran alejados del valor medio una determinada magnitud. Estos valores se consideran atípicos por lo que previamente han de ser eliminados.

Existen diferentes formas para detectarlos. Aunque pueden existir serios problemas en datos multidimensionales (Rocke y Woodruff, 1996), para nuestro caso de variables bidimensionales

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

(X,Y), la más empleada y una de las más simples es localizando aquellos valores que se distancian más de $3 \cdot \sigma$ veces del valor medio: estos valores no se tienen en cuenta para los cálculos en el estándar de exactitud posicional. Otra posibilidad para detectarlos es mediante diagramas Box-Whiskers, mediante atípicos o atípicos medios (aquellos valores que distancian más de $1.5 \cdot RQ$ [2]), y atípicos extremos (distancian más de $3.0 \cdot RQ$).

Los estimadores robustos, menos sensibles a las observaciones extremas que los estimadores clásicos (p.e.: la media o la desviación típica), permiten emplear la totalidad de los valores al ponderar los valores extremos. Así, un estimador robusto es eficiente si proporciona una menor estimación de la varianza que S^2 para distribuciones con grandes colas. De esta forma, Mosteller y Tuckey (1977) comparan la desviación media con S en una distribución normal contaminada, de modo que el 1% de los datos proviene de una población con desviación típica 3 veces mayor. Para este caso, encontraron que la desviación media es un 144% más eficiente que S.

En los 4 subapartados siguientes se presentan algunos estimadores robustos que podrían aplicarse en procesos para el control de la calidad posicional. Al aplicar los estimadores robustos sobre el control de la exactitud posicional, en la [tabla 1](#) se muestra la parte del proceso en la que deberían incluirse. De esta forma, se minimiza el efecto producido por los valores atípicos y, a un mismo tiempo, nos permite trabajar con todos los datos sin eliminar ninguno de ellos. Asimismo, para presentar su aplicación de una forma más clara y concisa, en cada estimador hacemos referencia a la tabla en la que se expone su procedimiento de aplicación.

2.1. Estimador robusto: MVT – MLT

Se incluyen dos estimadores que podrían considerarse como aceptables (Atkinson, 2005): MVT (Multi Variate Trimming) y MLT (Maximum Likelihood Trimming) (Devlin *et al.*, 1981). El MVT, o recorte multivariante, es un proceso iterativo, basado en la distancia de Mahalanobis (d_i^2), eliminándose aquellos valores que se encuentren más alejados. El segundo de ellos, el estimador recortado de máxima verosimilitud (MLT) es un caso particular de estimador M de Huber, en el que, en vez de recortes, se utilizan pesos basados en la distancia de Mahalanobis. Este segundo es el estimador analizado para el control posicional dado que, al no eliminar ningún dato, permite emplear todos los puntos de control. El procedimiento de aplicación sería el indicado en la [tabla 2](#).

2.2. Estimador robusto: suavizado gaussiano

De forma similar al caso anterior, se emplea un suavizado gaussiano sobre aquellos valores considerarlos como atípicos. El proceso de ponderación, que sigue una distribución normal, se realiza en función de la distancia a la que se encuentra el error en un punto " i " (considerado como *outlier*) con respecto al valor central de la campana de Gaus o error medio ($e_i - \bar{e}$). La función gaussiana viene dada por la expresión:

$$g_{\sigma}(e_i) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\bar{e}-e_i)^2}{2\sigma^2}}$$

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, n° 7, p. 171-187, ISSN: 1578-5157

siendo: $g(e_i)$ = función gaussiana, σ = desviación típica, e_i = error para el punto de orden i , $\bar{e}$ = error medio.

El procedimiento de aplicación sería el indicado en la [tabla 3](#), siendo el valor ponderado el que ha de aplicarse para cualquier otro cálculo posterior (errores medios, desviaciones típicas...).

2.3. Estimador robusto: método danés

Este método, desarrollado a partir de los trabajos de Krarup *et al.* (1967), ha sido utilizado por el Instituto Geodésico de Dinamarca para detectar valores extremos en soluciones por mínimos cuadrados (mm.cc.). El punto de partida es un ajuste convencional por mínimos cuadrados y, a partir de los residuos del primer ajuste, el cálculo de nuevos pesos para cada medida de forma individual en base a la siguiente función de pesos:

$$p = \begin{cases} 1 & \text{para } |v| \leq 2 \cdot \sigma \\ \text{proporcional a } \exp(-cv^2) & \text{para } |v| > 2 \cdot \sigma \end{cases}$$

Tal y como se indica en su procedimiento de aplicación (ver [tabla 4](#)), utilizando los valores ponderados, se realiza un nuevo ajuste por mm.cc. y se repite el proceso de reponderación y ajuste hasta que se logre la convergencia (siendo muy efectivo con *outliers* en ajustes por mm.cc.). Tras los buenos resultados que describe Domingo (2000) en fotogrametría, se ha realizado una adaptación del mismo para su aplicación en la detección y ponderación de atípicos en el control de calidad posicional cartográfico.

2.3. Estimador robusto: Geman y McClure

En el método de Geman y McClure (1992) la función de pesos viene definida tanto por una constante como por el valor asociado en parte del divisor de la ecuación de ponderación. De forma muy similar al caso anterior, hemos esquematizado su procedimiento de aplicación en la [tabla 5](#).

3. Procedimiento de ajuste de los estimadores robustos para su aplicación en el control de la exactitud posicional

Partiendo de que sólo una pequeña parte de los puntos de control pueden ser objeto de un valor atípico (Rencher, 1998), se ha comparado el comportamiento de los estimadores robustos frente a los clásicos. Para ello se han analizado dos de los principales estadísticos que nos definen el comportamiento de la muestra de puntos de control: el error medio (μ) y su desviación (σ). Así, se trata de generar una población sintética bajo unas condiciones determinadas a priori, proceder a contaminarla y, posteriormente, tratar de minimizar el efecto de dicha contaminación sobre el cálculo de μ y σ . Esto se hace necesario si se desean emplear todos los puntos en el control de calidad, dado que los principales estándares (v.g. EMAS) se fundamentan en dichos estadísticos.

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

Hemos de tener en cuenta que los principales estándares (ASCE, 1983; ASPRS, 1989; FGDC, 1998) se fundamentan en la aleatoriedad e independencia de los errores en X e Y . Asimismo, dichos estándares se fundamentan en que la distribución de errores se asemeja a una función normal. Para verificar estos extremos, hemos realizado numerosos ensayos prácticos sobre datos reales (AEV-DUOT, 2004 y 2006) tomados sobre el terreno en cartografía a diferente escala, empleando contrastes sobre las hipótesis de aleatoriedad (Wald-Wolfovitz), normalidad (Kolmogorov-Smirnov) e independencia entre componentes (Chi-Cuadrado) (Ariza, 2002; Atkinson, 2005; Atkinson *et al.*, 2005). Al confirmarse la aleatoriedad e independencia en la distribución de los errores, el análisis del variograma no presenta ningún resultado óptimo con respecto a su comportamiento, por lo que su utilización en este caso no ofrece información adicional.

Si las hipótesis previas no se cumplen no debe aplicarse el estándar para el control de la exactitud posicional, por lo que la opción de trabajo más adecuada ha sido generar poblaciones sintéticas que cumplan con dichas hipótesis. Así, para realizar el experimento, se ha generado una población aleatoria de 100 puntos con errores en dos componentes (X e Y) tipificados según una función normal. De los múltiples experimentos realizados, aquí presentamos algunos de los resultados más significativos, como es el caso de 3 poblaciones: $N(0,1)$, $N(0,2)$ y $N(0,10)$. Seguidamente, se ha procedido a contaminar dicha población con otras, de menor número de elementos (5%, 10% y 20% del total de la muestra principal), distribuidas de diferente forma (v.g. $N(0, 1.1)$, $N(0, 1.5)$, $N(1.5, 1.5)$, $N(0, 6)$...). Partiendo de estos datos, a modo de supuestos puntos tomados para un control cartográfico, se ha realizado el estudio del comportamiento de μ y σ para aplicaciones iterativas del estimador MLT, y de forma comparativa con el resto de estimadores: gaussiano, método danés y Geman y McClure. A continuación se muestran los procesos realizados para lograr el ajuste de los estimadores en su aplicación al control de calidad posicional.

3.1. Ajuste del estimador: MVT – MLT

Uno de los principales inconvenientes del estimador MLT es lo rápidamente que altera los estadísticos al aplicarlo de forma iterativa en la ponderación de la media y la desviación típica muestral (μ y S respectivamente). A modo de ejemplo, en la [figura 1a](#) se representa el valor del error (μ y σ) frente al número de iteraciones del estimador MLT. Como se puede apreciar, para una muestra $N(0,1)$ contaminada en un 10% con una distribución $N(1.5,1.5)$, se observa cómo el valor de μ se reduce ligeramente. No obstante, la mayor variación se produce en la σ para ambas componentes ([figura 1b](#)): la reducción es de tal envergadura que, para una supuesta muestra original de $\sigma=1$, en la iteración nº 12 ésta queda reducida a 0.76. Sin embargo, los resultados en la primera iteración no son totalmente incorrectos tal y como podrá observarse más adelante.

3.2. Ajuste del estimador: suavizado gaussiano

Uno de los aspectos más determinantes a la hora de aplicar la ponderación gaussiana, es el definir qué valores han de considerarse como *outliers*. El criterio puede ser a partir de: los atípicos medios; los atípicos extremos³; aquellos que se encuentren a más de $3 \cdot \sigma_p$ (siendo σ_p la desviación típica estimada a priori para la población); u otros (v.g. a más de $2.5 \cdot \sigma_p$). Los experimentos se han

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, n° 7, p. 171-187, ISSN: 1578-5157

realizado sobre poblaciones normales de distinta variabilidad: $\sigma_p=1, 2 \dots 10$ para cada una de los experimentos de simulación.

3.3. Ajuste del estimador: método danés

Para poder aplicar el método danés, es necesario determinar previamente el coeficiente constante k (fase 2 en la [tabla 4](#)). Para ello, se ha realizado un exhaustivo estudio previo con muestras de diferente tamaño y grado de contaminación, tanto en magnitud, como tipo (sobre μ y σ) y porcentaje. En cuanto al ajuste de la constante, la finalidad ha sido el que μ y σ de la población contaminada se aproximasen al máximo a los datos de la población sin contaminar. De esta forma, los valores contaminados se minimizan frente a la población original, sin tener que eliminar ningún punto de control. La ecuación ajustada queda definida por:

$$W_i = e^{-0.05 \cdot e_{Xi}^2}$$

siendo 0.05 = el valor constante (k) ajustado empíricamente al 95 % de nivel de confianza mediante procesos de simulación.

De esta forma, el error asociado al punto i en la componente X (de forma análoga se actuaría en la componente Y), vendría dado por la ecuación de la [tabla 4](#) (fase 3). El método no necesita de ninguna iteración dado que la ponderación final ajusta adecuadamente el resultado sobre μ y σ .

3.4. Ajuste del estimador: Geman y McClure

De igual forma sucede el estimador de Geman y McClure, ha de obtenerse el valor del coeficiente constante a aplicar en la ecuación. La ecuación ajustada se define por:

$$W_i = 5 \cdot \frac{1}{1 + e_{Xi}^2}$$

siendo 5 = el valor constante (k) ajustado empíricamente al 95 % de nivel de confianza mediante procesos de simulación.

El error asociado al punto i en la componente X (de forma análoga se actuaría en la componente Y), viene dado en la fase 3 de la [tabla 5](#). El método tampoco precisa de iteración alguna.

4. Análisis comparativo de los estimadores

Tras realizar el pertinente ajuste los diferentes estimadores robustos (número de iteraciones, nivel de corte en la detección de *outliers* y constantes en las ecuaciones de ponderación), se ha procedido con un análisis comparativo de su comportamiento ante distintos niveles de contaminación poblacional (ver apartado 3). Se ha de mencionar que todos los experimentos se

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, n° 7, p. 171-187, ISSN: 1578-5157

basan en una distribución normal de los errores dado que ésta es la hipótesis previa de trabajo en la mayoría de los estándares.

El principal objetivo es minimizar la influencia de los valores atípicos sobre el comportamiento de μ y σ en la población. Los resultados de algunos de los experimentos realizados (ver [tabla 6](#)) se muestran de la siguiente forma: en las dos primeras columnas las características de la muestra original, así como el porcentaje y el tipo de contaminación aplicada; las siguientes columnas presentan los valores de μ y σ : para estimación clásica empleando todos los puntos y eliminando aquellos que superen $3 \cdot \sigma_p$; estimación por MLT con una única iteración, estimación gaussiana por atípicos medios (en función del RQ), estimación gaussiana a más de $3 \cdot \sigma_p$, y estimación gaussiana a más de $2.5 \cdot \sigma_p$; estimación aplicado el método danés con una ponderación $3 \cdot \sigma_p$ y $2.5 \cdot \sigma_p$; y estimación aplicado el método de Geman y McClure con ponderación sobre $3 \cdot \sigma_p$ y $2.5 \cdot \sigma_p$. Por último, las filas muestran los resultados obtenidos para cada una de las componentes (X, Y) en cada uno de los experimentos.

El mejor resultado de cada experimento (el valor de μ y σ de la población contaminada se aproximan más al de la población sin contaminar), ha sido marcado en color verde, seguido del marcado en color amarillo y en color naranja (2º y 3º mejor resultado). A pesar de ello, en algunos casos las diferencias son mínimas, por lo que es importante detenerse a observar las magnitudes de los valores de la [tabla 6](#).

A continuación, se puede observar cómo en el caso de las dos primeras filas se ha trabajado con la población sin contaminar con errores en X e Y según una $N(0,1)$. Obviamente, los mejores resultados se obtienen empleando la totalidad de los datos, mediante estadística clásica. Algo similar sucede al aplicar el método danés o el de Geman y McClure (no se realiza ponderación alguna).

Si se observan los resultados del experimento **A**, los mejores valores se alcanzan con una ponderación gaussiana (considerando valores atípicos a partir de $3 \cdot \sigma_p$) y con una estimación clásica (eliminando aquellos datos que superan $3 \cdot \sigma_p$).

De forma muy similar se ha actuado en el resto de los experimentos. Por ejemplo, en el experimento **O**, los mejores resultados se obtienen con la ponderación por el método danés ($3 \cdot \sigma_p$) y con la ponderación de Geman y McClure ($3 \cdot \sigma$ y $2.5 \cdot \sigma_p$). En el experimento **R**, el mejor resultado se obtiene con la ponderación a partir de $2.5 \cdot \sigma_p$ de Geman y McClure, por el método danés ($2.5 \cdot \sigma_p$) y con el de Geman y McClure a partir de $3 \cdot \sigma_p$. Por otra parte, al analizar los resultados del estimador MLT, se observa que no resulta conveniente su aplicación en este tipo de trabajos dado que, incluso con una única iteración, los resultados no se aproximan sustancialmente a los valores deseados. Tan sólo en el caso de los experimentos **D** y **E** obtiene los mejores valores finales. En cuanto a la ponderación gaussiana con diferentes niveles de "corte", si bien los resultados son aceptables, no llega a alcanzar los niveles del método danés o el de Geman y McClure. Respecto a los estimadores clásicos, prácticamente resulta desaconsejable el emplear la totalidad de los datos sin realizar un estudio previo de posibles atípicos. Además, el hecho de eliminar aquellos valores que superan $3 \cdot \sigma_p$, aparte de desechar posible información útil, sólo ofrece resultados óptimos con índices muy bajos de contaminación muestral.

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

Por último, si se analizan los resultados de forma global, el estimador que obtiene los mejores resultados es el método danés con la ponderación sobre aquellos datos que se encuentren a más de $2.5 \cdot \sigma_p$. Como puede apreciarse, el intervalo en el que se encuentran los resultados proporcionados por él, es el que más se aproxima al valor de la población sin contaminar. Por otra parte el estimador de Geman y McClure con $2.5 \cdot \sigma_p$ y con $3 \cdot \sigma_p$ es el que obtiene el segundo y tercer mejor resultado respectivamente en los experimentos realizados ([tabla 6](#)).

5. Aplicación a un caso práctico

Para verificar los resultados obtenidos sobre poblaciones sintéticas, hemos realizado un control posicional planimétrico sobre 2 hojas a escala E10k elaboradas por la Agencia Extremeña de la Vivienda (AEV-DUOT, 2006). Sobre una de ellas, la hoja 704-06 (T.M. de Cáceres), se ha realizado un análisis muy exhaustivo con 127 puntos de control tomados sobre el terreno mediante GPS (con resolución de ambigüedades). En la segunda, la hoja 704-08 (T.M. de Sierra de Fuentes), se ha realizado un control exhaustivo con 61 puntos. Ambas hojas se encuentran al sur de la ciudad de Cáceres ([figura 2](#)).

En dichas hojas, se ha aplicado un control de calidad según el estándar NSSDA (FGDC, 1998) bajo 3 perspectivas diferentes: aplicarlo sin considerar la detección/supresión de posibles valores atípicos, aplicarlo con la supresión previa de dichos valores y aplicarlo con una ponderación mediante el método danés. En la tabla 7, se muestran los resultados obtenidos.

En este caso, el porcentaje de valores atípicos es mayor en la hoja con menor número de puntos de control, por lo que su influencia será determinante en el resultado final. Como se puede observar, la peor de las opciones, el obviar la detección de *outliers*, al aumentar considerablemente la exactitud horizontal ($Exactitud_r$) estimada por el NSSDA. Si seguimos las recomendaciones del propio estándar (con la supresión de los valores atípicos), los resultados son adecuados, aunque hayamos tenido que "pagar" un precio muy alto al tener que desechar entre un 7% y un 10% de los puntos tomados en campo. El problema, en este caso, no es únicamente el coste de la captura y cálculo de dichos puntos de control, sino el hecho de que esta información no llegue al receptor de la BDG. El usuario de la cartografía podrá encontrarse con este tipo de errores, por lo que si se aplica el estimador robusto recomendado en el apartado 4 (método danés con ponderación de aquellos valores que estén a más de $2.5 \cdot \sigma_p$), los resultados obtenidos son más que aceptables: prácticamente no se altera el valor final del estándar NSSDA con respecto a la propuesta de éste (suprimir los valores atípicos) y, sin embargo, se ha utilizado la totalidad de los puntos tomados en campo, por lo que el usuario final dispone de una información mucho más exhaustiva y completa.

De esta forma, creemos que se hace necesario incluir un apartado en el que se indique el porcentaje de valores atípicos detectados y su ponderación en el cálculo final. Así, creemos muy conveniente incluir algún tipo de leyenda en el que se informe de esta circunstancia (p.e.: "Se ha detectado un 7 % de puntos de control considerados como valores atípicos. Dichos valores han sido ponderados mediante el método danés").

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

Como se puede observar, las ventajas de no desechar los valores atípicos son notables: tenemos un mayor número de datos, los resultados son mucho más exhaustivos, tanto el usuario como el productor tienen un mayor conocimiento de la calidad final del producto, etc.

6. Conclusiones

En cartografía, la gran mayoría de los estándares para el control de la exactitud posicional emplean técnicas de estadística clásica, optando por no considerar en los cálculos aquellos datos con valores considerados como "atípicos". Esto supone una información incompleta al usuario (en la mayoría de los casos ni siquiera es preciso informar sobre la presencia/ausencia de atípicos) y una importante pérdida de información al suprimir puntos de control. Dado que su adquisición sobre el terreno conlleva un cierto coste, su eliminación supone una pérdida de representatividad estadística y un incremento del coste económico, al haber invertido recursos en la toma de datos que posteriormente han de ser detectados y eliminados. Frente a esta problemática, los estimadores robustos permiten trabajar con valores atípicos y utilizar la totalidad de los puntos de control, ofreciendo una información mucho más completa y exhaustiva.

La metodología empleada ha resultado ser la idónea en este tipo de estudios: un entorno controlado de simulación, generando poblaciones sintéticas que luego han sido contaminadas en diverso grado de presencia (0%, 5%, 10%, 20%...), de sistematismo y de variabilidad (1,1x, 1,2x, 1,5x, 2x, 3x, 4x, 5x...) respecto a la varianza de los datos base, todo ello al objeto de asimilar la presencia de valores atípicos. Prueba de ello son los buenos resultados obtenidos en su aplicación a un caso práctico: se pueden emplear todos los puntos de control y se ofrece al usuario una información detallada de los valores atípicos que podrá encontrarse al utilizar la cartografía.

De entre los estimadores robustos analizados (MLT, Gausiano, Danés y Geman & McClure), el que mejores resultados ofrece es el método danés con un valor de "corte" en la ponderación sobre aquellos valores que superan $2.5 \cdot \sigma_p$. Este método reduce considerablemente la influencia de los valores atípicos procedentes de un proceso de contaminación sobre la muestra principal, aproximándose los valores de media y desviación globales a los datos originales.

Por último, se recomienda que los estándares informen sobre la presencia/ausencia de valores atípicos al realizar el control de la calidad posicional y que dichos valores sean empleados en los cálculos. De esta forma, sería necesario incluir un apartado en el que se indique el porcentaje de valores atípicos detectados y su ponderación en el cálculo final, insertando algún tipo de leyenda en el que se informe sobre esta circunstancia, como p.e.: "Se ha detectado un P % de puntos de control considerados como valores atípicos. Dichos valores han sido ponderados mediante el método danés" (siendo P el porcentaje de puntos de control considerados como atípicos).

Por otra parte, nuestros futuros trabajos de investigación probablemente se centren en estudiar la adaptación de nuevos estimadores robustos (como los "high breakdown" o el propuesto por Al-Othmana A.K. e Irvingb, M.R., 2007) al problema del control de la exactitud posicional sobre bases de datos geográficas.

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): “Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional”, *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

Agradecimientos

A la Junta de Extremadura y la Agencia Extremeña de la Vivienda (Dirección de Urbanismo y Ordenación del Territorio), por su colaboración con el Grupo de Investigación en Ingeniería Geomática y Patrimonio Urbano (IGPU: <http://www.unex.es/igpu>).

Al Ministerio de Educación y Ciencia por la financiaron parcial con cargo a la beca de Colaboración con el Departamento de Expresión Gráfica de la Universidad de Extremadura (“Control de Calidad Cartográfico – CalCar 2005”, realizado por el alumno de Ingeniería en Geodesia y Cartografía, D. Manuel Blanco Murillo) (Blanco, 2006), y la financiaron parcial con cargo al proyecto “Control Posicional en Cartografía por Elementos Lineales” (BIA2003-02234), también del Ministerio de Educación y Ciencia.

Referencias bibliográficas

- AEV-DUOT (2004): *Cartografía de la C.A. de Extremadura*, hojas nº 704-05, 704-07, 704-11, 704-12, 704-15 y 779-11. Agencia Extremeña de la Vivienda, Dirección de Urbanismo y Ordenación del Territorio, Junta de Extremadura, Mérida, Badajoz.
- AEV-DUOT (2006): *Cartografía de la C.A. de Extremadura*, hojas nº 704-06 y 704-08. Agencia Extremeña de la Vivienda, Dirección de Urbanismo y Ordenación del Territorio, Junta de Extremadura, Mérida, Badajoz.
- Al-Othmana A.K. e Irvingb, M.R. (2007): “Robust state estimator based on maximum constraints satisfaction of uncertain measurements”, *Measurement*, 40, 3, pp 347-359
- Ariza, F.J. (2002): *Calidad en la producción cartográfica*. Ra-MA, Madrid. ISBN:84-7897-524-1
- ASCE (1983): *Map uses, scales and accuracies for engineering and associated proposes*. American Society of Civil Engineers, Committee on Cartographic Surveying, Surveying and Mapping Division, Nueva York.
- ASPRS, (1989): “Accuracy standards for large scale maps”, *PE&RS*, vol. 56, nº 7.
- Atkinson, A.D.J.; Ariza, F.J. y Rodríguez, J. (2002): “Empleo de estimadores robustos en control posicional cartográfico”, en *Actas de III Conferência Nacional de Cartografia e Geodesia*, Aveiro.
- Atkinson, A.D.J, Ariza, F.J. y García-Balboa, J.L. (2005): “Positional accuracy control using robust estimators”, en *XXII International Cartographic Conference*, by The International Cartographic Association, A Coruña.
- Atkinson, A.D.J. (2005): “Control de calidad posicional en cartografía: análisis de los principales estándares y propuesta de mejora”, Tesis doctoral, Universidad de Jaén, Jaén.
- Blanco, M. (2006): *Control de calidad cartográfico a escala 1/10.000 en la zona de Cáceres*. PFC de Ingeniería en Geodesia y Cartografía, Escuela Politécnica, Universidad de Extremadura, Cáceres.
- Devlin, S. J., Gnanadesikan, R. y Kettenring, J. R. (1981) “Robust estimations of dispersion matrices and principal components”, *Journal of the American Statistical Association*, 76.
- Domingo, A. (2000): *Investigación sobre los métodos de estimación robusta aplicados a la resolución de los problemas fundamentales de la fotogrametría*. Tesis doctoral, Dep. IGTEG, Universidad de Cantabria, Santander.
- FGDC, (1998): *Geospatial positioning accuracy standards*. Minnesota Planning, 1999. FGCD-STD-007-1998. Virginia.
- Geman S. y McClure, D.E. (1992): “A nonlinear filter for film restoration and other problems in image processing”. *CVGIP: Graphical Models and Image Processing*, vol 54, nº 4.

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): “Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional”, *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

Gento, P., Ortega, J.F. y García-Donato, G. (2004): “Alternativas estadísticas al cálculo del Valor en Riesgo”, en *Estadística Española*, vol. 46, nº 155, pp. 119-148

Krarup, K., Kubik, K y Juhl, J. (1967): “Götterdämmerung over least squares adjustment”, en *Proceedings XIV Congress of ISP*. Hamburgo.

Mosteller, F. y Tuckey, J. W. (1977): *Data analysis and regression*. Reading, MA: Addison – Wesley.

MPLMIC (1999): Positional accuracy handbook. Minnesota Planning, Land Management Information Center, Minnesota. EE.UU.

Ortega, J.F. (2000): *Nuevas familias de estimadores robustos y detección de observaciones atípicas en modelos lineales*. Tesis Doctoral. Univ. de Castilla-La Mancha.

Rencher, A.C. (1998): *Multivariate statistical inference and applications*. Wiley.

Rocke, D. M. y Woodruff, D. L. (1996): “Identification of outliers in multivariate data”, *Journal of the American Statistical Association*, vol. 91, nº 435, p. 1047-1061.

TABLAS

Tabla 1. Aplicación de estimadores robustos en los estándares de control de calidad posicional.

Control posicional con estimadores clásicos	Control posicional con estimadores robustos
Selección del estándar a aplicar: EMAS, ASPRS, NSSDA ...	Selección del estándar a aplicar: EMAS, ASPRS, NSSDA ...
Selección de puntos de control.	Selección de puntos de control.
Detección de <i>outliers</i> a más de $3 \cdot \sigma$ veces del valor medio.	Detección de <i>outliers</i> : a más de $3 \cdot \sigma$ veces del valor medio, $2.5 \cdot \sigma$ veces o por atípicos medios y/o extremos
Eliminación de los puntos considerados como <i>outliers</i> .	Ponderación de los puntos considerados como <i>outliers</i> empleando un estimador robusto.
	Incluir en la metainformación de la BDG la información referente al número de valores atípicos ponderados.
Continuar aplicando el estándar con los puntos restantes.	Continuar aplicando el estándar con los puntos ponderados.

Fte. Elaboración propia

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, n° 7, p. 171-187, ISSN: 1578-5157

Tabla 2. Procedimiento de aplicación: MLT.

Fase	Descripción
1	Obtención, mediante estadística clásica, de los valores iniciales para la media ($\tilde{\mu}$) y la desviación (S)
2	Proceso de iteración:
	2.1 Cálculo de la distancia de Mahalanobis para el punto i: $d_i^2 = (y_i - \tilde{\mu})^t \cdot \tilde{S}^{-1} \cdot (y_i - \tilde{\mu})$ en donde: d_i^2 es la distancia de Mahalanobis en el punto i, y_i es la matriz de errores en el punto i (para el análisis de la componente horizontal (X e Y), se trata de una matriz de dos dimensiones [2×2]), $\tilde{\mu}$ es la matriz con los errores medios en la iteración ($i-1$), $\tilde{S}$ es la matriz varianza-covarianza de los errores en la iteración ($i-1$).
	2.2 Detección de valores atípicos en función de las distancias de Mahalanobis frente al coeficiente de una distribución χ^2 con p grados de libertad (siendo p el número de componentes analizadas, en este caso $p = 2$): - Atípicos medios: son aquellos cuya $d_i^2 > \chi_{2,0.05}^2$ ($= 5.9915$). El contraste se realiza al 95% de nivel de confianza. - Atípicos extremos: son aquellos cuya $d_i^2 > \chi_{2,0.01}^2$ ($= 9.2103$). El contraste se realiza al 99% de nivel de confianza⁴.
	2.3 Cálculo de los estadísticos ponderados a emplear en el control de calidad: $\tilde{\mu} = \frac{\sum_{i=1}^n \omega_i \cdot y_i}{\sum_{i=1}^n \omega_i}; \quad \tilde{S} = \frac{1}{n} \sum_{i=1}^n \omega_i (y_i - \tilde{\mu})(y_i - \tilde{\mu})^t; \quad \text{en donde: } \omega_i = \frac{p+1}{d_i^2 + 1}$ siendo: ω_i matriz de pesos, p número de variables (para un control planimétrico en X e Y, $p = 2$), d_i^2 distancia de Mahalanobis en el punto i.
	2.4 Determinar si se cumple el criterio de convergencia. Devlin, Gnanadesikan y Kettenring (1981) sugieren comenzar con $\tilde{\mu}$ y $\tilde{S}$, y finalizar con 25 iteraciones o cuando la transformada de Fisher de los $\tilde{r}_{jk}$ varíe menos de 0.001. Ésta viene dada por: $z_{jk} = \frac{1}{2} \ln \left(\frac{1 + \tilde{r}_{jk}}{1 - \tilde{r}_{jk}} \right) = \tanh^{-1} \tilde{r}_{jk}$ de forma que los $\tilde{r}_{jk}$ se pueden obtener a partir de: $\tilde{R} = D^{-1} \cdot \tilde{S} \cdot D^{-1}$; en donde D es la matriz diagonal cuyos elementos son las raíces cuadradas de los elementos de la diagonal de $\tilde{S}$.
3	Si no se cumple el criterio de convergencia, realizar la siguiente iteración en el cálculo de la distancia de Mahalanobis (fase 2).

Fte. Elaboración propia

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, n° 7, p. 171-187, ISSN: 1578-5157

Tabla 3. Procedimiento de aplicación: suavizado gaussiano.

Fase	Descripción
1	Detección de valores atípicos, considerando como tales aquellos que, por ejemplo, superen en 3 veces la desviación poblacional. Otra posibilidad, es mediante la diferenciación en atípicos medios y extremos en función de a qué distancia se sitúen del valor promedio y el recorrido intercuartílico.
2	Ponderación mediante la función gaussiana de aquellos puntos en los que alguna de sus componentes pueda considerarse como valor atípico. Dicho valor vendrá ponderado según la ecuación: $e_{Xi}(w) = e_{Xi} \cdot g_{\sigma}(e_{Xi})$

Fte. Elaboración propia.

Tabla 4. Procedimiento de aplicación: método danés.

Fase	Descripción
1	Detección de valores atípicos, considerando como tales aquellos que, por ejemplo, superen en 3 veces la desviación poblacional (v.g. $[\bar{e}-e_{Xi}] > 3 \cdot \sigma_{mapa}$).
2	Cálculo de los pesos a aplicar, empleando la ecuación: $W_i = e^{-k \cdot e_{Xi}^2}$ siendo: W_i = el peso a aplicar en el punto i , e_{Xi} = el valor del error en la componente X para el punto i , k = un valor constante a determinar empíricamente.
3	Ponderación de los valores atípicos, siendo el error asociado al punto i en la componente X : $e'_{Xi} = e_{Xi} \cdot W_i$
4	En caso de que fuera necesario, repetir el proceso de ponderación (fase 2).

Fte. Elaboración propia.

Tabla 5. Procedimiento de aplicación: método de Geman y McClure.

Fase	Descripción
1	Detección de valores atípicos, considerando como tales aquellos que, por ejemplo, superen en 3 veces la desviación poblacional (v.g. $[\bar{e}-e_{Xi}] > 3 \cdot \sigma_{mapa}$).
2	Cálculo de los pesos a aplicar, empleando la ecuación: $W_i = k \cdot \frac{1}{1 + e_{Xi}^2}$ siendo: W_i = el peso a aplicar en el punto i , e_{Xi} = el valor del error en la componente X para el punto i , k = un valor constante a determinar empíricamente.
3	Ponderación de los valores atípicos, siendo el error asociado al punto i en la componente X : $e'_{Xi} = e_{Xi} \cdot W_i$
4	En caso de que fuera necesario, repetir el proceso de ponderación (fase 2).

Fte. Elaboración propia.

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, n° 7, p. 171-187, ISSN: 1578-5157

Tabla 6. Resultados obtenidos tras aplicar diferentes estadísticos para calcular μ y σ : comparativa con diferentes poblaciones y niveles de contaminación

Pobla.c.Orig.	Experimento	Componente	Contaminación			CLÁSICOS			MLT			Función GAUSIANA						Método DANÉS						Geman & Mc-Clure								
			TODOS			SOLO < 3 cp						CUARTILES			3 cp			2.5 cp			3 cp			2.5 cp			3 cp			2.5 cp		
			(%)	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ			
$\mu = 0$ $\sigma = 1$	X	0	1	0.000	1.000	0.000	1.000	0.000	1.000	0.004	0.916	0.000	1.000	0.000	1.000	-0.027	0.962	0.000	1.000	0.008	0.980	0.000	1.000	0.000	1.000	0.011	0.976					
	Y	0	1	0.000	1.000	0.000	1.000	0.039	0.874	-0.025	0.967	0.000	1.000	-0.025	0.967	0.000	1.000	-0.025	0.967	0.000	1.000	0.007	0.985	0.000	1.000	0.008	0.983					
	A	X	5	0	1.1	-0.007	1.010	0.024	0.923	0.007	1.010	-0.020	0.973	-0.048	0.954	-0.023	0.988	-0.048	0.954	-0.023	0.988	0.006	1.048	0.007	1.013	0.010	0.998					
	Y	0	1.1	0.023	0.988	0.015	0.885	-0.048	0.954	-0.023	0.988	-0.048	0.954	0.023	0.988	0.006	1.048	0.007	1.013	0.010	0.998	0.006	1.048	0.007	1.013	0.010	0.998					
	B	X	10	0	1.5	-0.006	1.076	-0.038	1.032	-0.004	0.977	-0.025	1.029	-0.026	0.988	0.006	1.048	0.007	1.013	0.010	0.998	0.006	1.048	0.007	1.013	0.010	0.998					
	Y	0	1.5	0.044	1.047	-0.055	0.986	-0.017	0.906	-0.077	0.870	-0.075	0.997	-0.100	0.962	0.056	1.016	0.061	0.988	0.069	0.990	0.056	1.016	0.061	0.988	0.069	0.990					
	C	X	10	1.5	1	-0.006	1.076	-0.038	1.032	-0.004	0.977	-0.025	1.029	-0.026	0.988	0.006	1.048	0.007	1.013	0.010	0.998	0.006	1.048	0.007	1.013	0.010	0.998					
	Y	1.5	1	0.044	1.047	-0.055	0.986	-0.017	0.906	-0.077	0.870	-0.075	0.997	-0.100	0.962	0.056	1.016	0.061	0.988	0.069	0.990	0.056	1.016	0.061	0.988	0.069	0.990					
	D	X	10	1.5	1.5	0.133	1.146	-0.093	1.102	-0.099	1.027	-0.067	1.050	-0.098	1.093	-0.067	0.977	0.105	1.082	0.105	1.045	0.095	1.067	0.094	1.021	0.084	1.021					
	Y	1.5	1.5	0.164	1.161	-0.090	1.048	-0.066	0.981	-0.089	0.911	-0.088	1.037	-0.137	0.971	0.125	1.066	0.125	1.037	0.113	1.050	0.112	1.050	0.112	1.050	0.112	1.050					
	E	X	10	0	2	-0.017	1.124	-0.043	1.053	-0.001	1.002	-0.024	1.046	0.002	0.935	0.007	1.061	-0.002	1.006	0.013	1.052	0.002	0.984	0.013	1.052	0.002	0.984					
	Y	0	2	0.014	1.144	0.000	1.038	0.030	0.979	-0.036	0.996	-0.011	1.028	-0.036	0.996	0.011	1.063	0.018	1.049	0.011	1.045	0.019	1.029	0.019	1.029	0.019	1.029					
$\mu = 0$ $\sigma = 1$	F	X	10	0	3	0.014	1.335	-0.007	1.080	-0.007	1.141	-0.014	1.055	-0.041	1.019	-0.016	1.091	-0.008	1.074	-0.006	1.076	0.005	1.054	0.005	1.054	0.005	1.054					
	Y	0	3	0.011	1.289	0.027	1.027	0.028	1.066	-0.017	0.967	0.039	1.046	-0.042	0.932	-0.025	1.087	0.000	1.031	-0.029	1.085	0.004	0.995	0.004	0.995	0.004	0.995					
	G	X	10	0	4	-0.017	1.631	-0.003	1.075	0.013	1.187	-0.015	0.982	-0.045	1.025	-0.042	0.942	0.033	1.046	0.031	1.002	0.035	1.038	0.031	0.981	0.031	0.981					
	Y	0	4	-0.031	1.533	-0.011	1.095	0.036	1.245	-0.027	1.053	-0.027	1.053	-0.001	1.021	0.015	1.108	0.014	1.079	0.018	1.083	0.017	1.048	0.017	1.048	0.017	1.048					
	H	X	10	4	4	0.383	1.953	-0.011	0.948	-0.109	1.421	0.057	0.984	0.003	0.909	0.009	1.029	0.026	0.992	0.001	1.012	0.023	0.965	0.023	0.965	0.023	0.965					
	Y	4	4	0.369	1.957	0.038	1.023	-0.086	1.412	0.000	0.974	-0.031	1.023	-0.048	0.907	0.056	1.027	0.063	1.012	0.051	1.006	0.059	0.988	0.059	0.988	0.059	0.988					
	I	X	10	0	5	-0.058	1.108	0.055	0.979	0.059	1.109	0.054	1.000	0.054	1.000	0.011	0.928	-0.053	1.036	-0.038	1.002	-0.052	0.976	-0.052	0.976	-0.052	0.976					
	Y	0	5	-0.016	1.795	0.039	1.041	0.086	1.336	0.014	1.027	0.014	1.027	0.014	0.961	-0.043	1.093	-0.043	1.063	-0.030	1.059	-0.031	1.024	-0.031	1.024	-0.031	1.024					
	J	X	20	0	1.2	-0.034	0.980	0.034	0.980	0.039	0.897	0.034	0.980	0.034	0.980	0.034	0.980	-0.034	0.980	-0.034	0.980	-0.034	0.980	-0.034	0.980	-0.034	0.980					
	Y	0	1.2	-0.074	1.065	0.074	1.065	0.106	0.923	0.074	1.065	0.074	1.065	0.074	1.065	-0.074	1.065	-0.074	1.065	-0.074	1.065	-0.057	1.031	-0.057	1.031	-0.057	1.031					
	K	X	20	1.2	1.2	0.245	1.046	-0.248	1.051	-0.236	0.967	-0.245	1.046	-0.245	1.046	-0.188	0.975	0.245	1.046	0.201	0.977	0.245	1.046	0.219	0.983	0.219	0.983					
	Y	1.2	1.2	0.268	1.127	-0.235	1.083	-0.222	0.995	-0.199	0.973	-0.233	1.078	-0.199	0.973	0.214	1.041	0.193	1.010	0.231	1.063	0.219	1.031	0.231	1.063	0.219	1.031					
$\mu = 0$ $\sigma = 2$	L	X	20	0	3	-0.052	1.548	-0.026	0.979	0.018	1.242	-0.076	1.017	-0.047	0.975	0.106	1.021	0.083	0.982	0.041	1.065	0.026	1.035	0.026	1.035	0.026	1.035					
	Y	3	2	0.595	1.779	-0.182	1.169	-0.289	1.457	-0.269	1.280	-0.256	1.197	-0.284	1.076	0.172	1.111	0.158	1.007	0.244	1.147	0.223	1.078	0.223	1.078	0.223	1.078					
	M	X	10	0	4	-0.015	2.180	-0.010	1.918	0.044	1.863	-0.054	1.844	-0.039	1.898	0.026	1.901	0.016	1.931	0.022	1.901	0.012	1.931	0.012	1.931	0.012	1.931					
	Y	0	4	-0.112	2.189	0.060	2.096	0.020	1.999	0.043	2.081	0.043	2.081	0.050	1.929	-0.050	2.082	-0.059	1.978	-0.050	2.082	-0.058	1.975	-0.058	1.975	-0.058	1.975					
	N	X	10	4	4	0.331	2.536	-0.011	1.966	-0.134	2.173	-0.024	2.035	-0.054	2.035	0.052	2.047	0.005	1.996	0.060	2.048	0.009	1.986	0.009	1.986	0.009	1.986					
	Y	4	4	0.388	2.530	-0.127	2.072	-0.308	2.138	-0.125	1.988	-0.186	2.078	-0.176	1.916	0.159	1.960	0.117	1.944	0.153	1.960	0.125	1.930	0.125	1.930	0.125	1.930					
	O	X	10	0	6	-0.069	2.628	0.138	2.056	0.145	2.156	0.125	1.896	0.135	2.042	0.077	1.959	-0.127	2.034	-0.098	1.973	-0.128	2.045	-0.067	1.973	-0.067	1.973					
	Y	0	6	0.046	2.633	-0.020	2.085	-0.100	2.225	-0.009	1.995	-0.059	1.951	-0.004	1.842	0.042	1.987	-0.005	1.933	0.026	1.989	0.021	1.906	0.021	1.906	0.021	1.906					
	P	X	10	0	8	-0.033	3.066	0.048	1.932	0.103	2.401	0.113	1.846	-0.029	2.026	0.027	2.028	0.027	2.028	0.023	2.030	0.023	2.030	0.023	2.030	0.023	2.030					
	Y	0	8	-0.024	3.101	0.052	2.041	0.015	2.471	0.045	2.008	0.045	2.038	0.049	1.856	-0.055	2.030	-0.042	1.907	-0.038	2.032	-0.033	1.895	-0.033	1.895	-0.033	1.895					
	$\mu = 0$ $\sigma = 10$	Q	X	10	0	20	-0.577	11.173	0.841	9.992	0.634	9.715	0.571	10.160	0.571	10.160	0.566	9.505	-0.571	10.160	-0.571	10.160	-0.571	10.160	-0.571	10.160	-0.571	10.160				
		Y	0	20	-0.634	11.133	0.170	10.384	0.745	9.591	0.256	10.478	0.256	10.478	0.502	10.173	-0.256	10.478	-0.256	10.478	-0.257	10.478	-0.257	10.478	-0.257	10.478	-0.257	10.478				
R		X	10	25	30	2.084	15.216	0.484	10.365	-0.814	11.940	0.063	10.523	0.063	10.523	-0.135	9.486	-0.063	10.523	-0.094	9.763	-0.059	10.523	-0.091	9.763	-0.091	9.763					
Y		25	30	2.435	15.227	0.442	10.189	-0.774	12.316	0.415	9.876	0.415	9.876	0.415	9.876	-0.415	9.876	-0.415	9.876	-0.409	9.876	-0.409	9.876	-0.409	9.876	-0.409	9.876					
S		X	10	0	40	-0.134	15.416	0.149	10.127	0.157	12.420	0.382	9.804	0.382	9.804	0.121	9.455	-0.382	9.804	-0.121	9.455	-0.382	9.804	-0.121	9.455	-0.121	9.455					
Y		0	40	0.337	15.482	-0.374	10.237	-0.283	12.485	-0.084	10.207	-0.084	10.207	-0.084	10.207	-0.339	9.876	0.339	9.876	0.338	9.876	0.338	9.876	0.338	9.876	0.338	9.876					
T		X	10	15	50	1.836	18.492	-0.434	10.620	-0.976	13.547	-0.476	10.261	-0.737	9.906	0.476	10.261	-0.476	10.261	0.478	10.261	0.478	10.261	0.478	10.261	0.478	10.261					
Y		25	30	2.274	15.275	-0.164	11.050	-0.716	12.670	-0.476	11.061	-0.157	11.061	0.171	10.014	-0.476	11.061	-0.476	11.061	-0.478	11.061	-0.478	11.061	-0.478	11.061	-0.478	11.061					

Fte. Elaboración propia.

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, n° 7, p. 171-187, ISSN: 1578-5157

Tabla 7. Ejemplo de control posicional planimétrico sobre 2 hojas a escala 1:10000, aplicando el estándar NSSDA.

		Hoja 704-06			Hoja 704-08		
		Todos	Cásica	M.Danés	Todos	Cásica	M.Danés
Datos	Nº total de puntos	127	127	127	61	61	61
	Valores atípicos	-	9	13	-	6	10
	Nº puntos en cálculo	127	118	127	61	55	61
	% de pérdida	0%	7%	0%	0%	10%	0%
Errores	Error medio X (m)	-0.25	0.26	-0.19	1.02	-0.12	0.11
	Error medio Y (m)	-0.01	-0.40	0.43	-2.17	0.55	-0.51
	Desviac. X (m)	6.87	1.54	1.53	5.14	1.39	1.32
	Desviac. Y (m)	4.75	1.79	1.78	13.12	1.15	1.27
NSSDA	Exactitud (m)	14.40	4.15	4.12	24.55	3.24	3.26

Fte. Elaboración propia.

FIGURAS

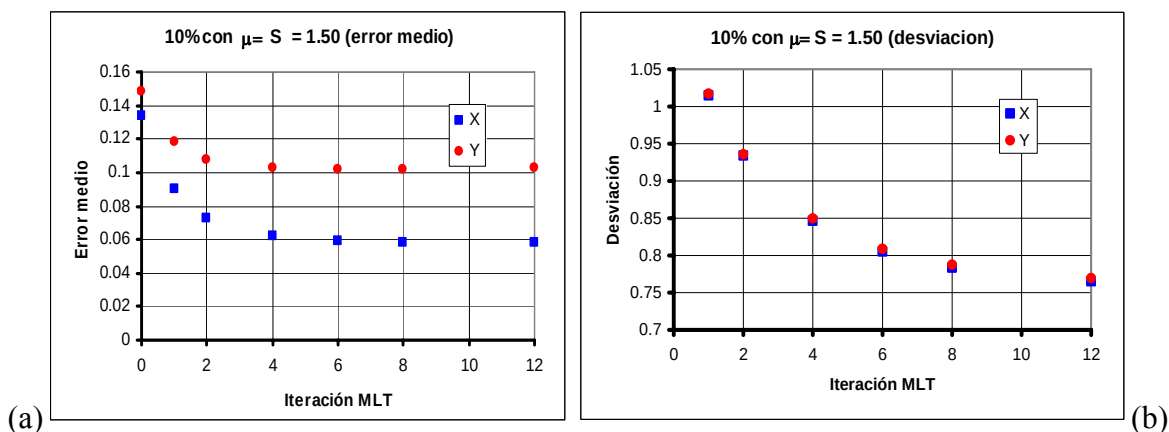


Figura 1. Valores del error medio resultante (μ) en función del número de iteraciones del estimador MLT (a); y valores de la desviación resultante (σ) en función del número de iteraciones del estimador MLT (b).

Fte: elaboración propia.

Atkinson Gordo, A. D. J., Ariza López, F. J. y García-Balboa, J. L. (2007): "Estimadores robustos: una solución en la utilización de valores atípicos para el control de la calidad posicional", *GeoFocus (Artículos)*, nº 7, p. 171-187, ISSN: 1578-5157

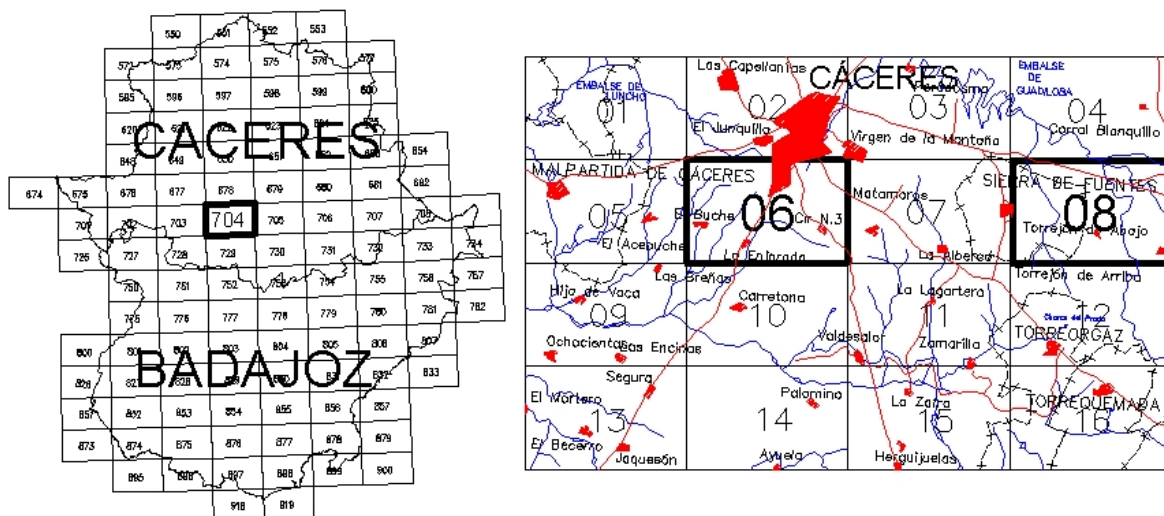


Figura 2. Localización de las hojas analizadas: 704-06 y 704-08.

Fte: Agencia Extremeña de la Vivienda (AEV-DUOT, 2006).

¹ Con la nomenclatura $N(\mu, \sigma)$, se presenta una población con distribución Normal de media μ y desviación σ .

² Se entiende por RQ el recorrido o rango intercuartílico: la variación alrededor del centro de una distribución de categorías ordenadas. Este valor mide la variabilidad en el espacio que va entre Q1 (el primer cuartil) y Q3 (el tercer cuartil). Como se aprecia, RQ no es otra cosa que una medida de variación alrededor de la mediana.

³ Esta opción no se ha considerado en el estudio al ser relativamente pocos los valores considerados como atípicos extremos que intervienen en el cálculo.

⁴ El valor de χ^2 es la medida teórica de la distribución Chi-Cuadrado. Para un análisis tridimensional (X,Y,Z), los valores límite son: 7.8147 ($p = 3$; $1-\alpha = 0.05$) para los atípicos medios, y 11.3449 ($p = 3$; $1-\alpha = 0.01$) para los atípicos extremos.