

QUANTIFICATION OF SYSTEMATIC ERRORS IN POSITIONAL QUALITY CONTROL

Alan D.J. Atkinson Gordo¹, Francisco J. Ariza López², José L. García-Balboa³

⁽¹⁾ Universidad de Extremadura, atkinson@unex.es

⁽²⁾ Universidad de Jaén, fjariza@ujaen.es

⁽³⁾ Universidad de Jaén, jlbalboa@ujaen.es

ABSTRACT

We present the results of employing an old tool with a new application: the affine transformation adjustment by least squares is used for detecting and quantify bias in a geographic data base. In this way, we are able to quantify displacement, scale factor and rotation bias, so that it is possible to use results as a control methodology for accepting or rejecting data bases, but also for correcting such errors and improving the positional accuracy of the data.

The work carried out is based on a simulation process. First results point out that it is possible to estimate bias which supports the control and correction uses of proposed methodology. The major drawback is the large number of control points that are needed for a precise estimation.

1. INTRODUCTION

From the point of view of positional quality of a Geographic Data Base (GDB), several tests or methods exist that permit investigators to monitor the positional accuracy X, Y and/or Z on the map. The three components must be taken into account for a correct examination of positional accuracy, the differential characteristics of the altimetric component must be known: methodology employed in the capturing of data, greater uncertain altimetric representation of the data... Although a great portion of these aspects is directly applicable to the cartography in analogical format, it is also true that an elevated number of products exist in digital format that do not contemplate the altimetric component or that appear represented as the analogical cartography that it deals with (eg. using contour lines in 2D). Likewise, the analysis of the altimetric accuracy is realized in a differentiated form on digital land models. Because of all this, in this project, the realization of the treatment of data is planimetric, although it could also be applied to the component Z.

The tests commonly employed for positional control are based in the comparison of the map with a Higher Accuracy Source (HAS). This could be a map of higher accuracy (cartography at a significantly superior scale and/or with a higher positional quality contrasted previously), or data taken on the ground with a higher accuracy to the cartography that is being analysed (a source 3 times more accurate than the product to be contrasted is recommended). In this last aspect, the use of current Global Position System (GPS) reduces the costs considerably in this phase of the process.

As for the errors that affect the geographic data, they can be classified in three types:

- Gross errors. Those values that are considered atypical (outliers).
- Random errors. These are the product of random variations produced in the conditions of the project, given that by measuring the same parameter several times under identical conditions, different values are obtained. If the dispersion in the measured values is random, the errors could be treated using statistical techniques and obtaining, by way of a combination of measurements, a value representative of all them (Volkman, 2002). These types of errors, also known as accidental, generate random, individually small errors, and do not usually present a fixed behaviour which is generally partially compensated by increasing the number of data observed (Ariza, 2002).
- Systematic Errors. These are associated with the conditions in which the experiment is realized. They do not have statistical variation and the treatment for their posterior correction requires a careful revision of the experimental assembly. The habitual sources of these types of errors are found in the use of incorrectly calibrated measurement instruments, or the erroneous assumption of conditions, such as the atmospheric pressure, temperature, etc. (Volkman, 2002). However, on the cartographic products, systematic errors (constant or variable) occur with the same sign and, generally, with the same magnitude in a consecutive number of observations. They are generated by permanent causes in such a way that they can be eliminated if they are detected and quantified (Ariza, 2002). To these definitions, we must add the fact that they could come to produce systematic errors of a local form. As an

example, a systematic error could exist that produces a displacement towards the northeast in a mountainous zone of the GDB, and a displacement of different magnitude towards the south in another zone of the GDB because of errors in the supporting photogrammetry.

Systematic errors must be corrected and eliminated before the application of the majority of the control standards of cartographic quality (ASPRS, NSSDA). Therefore, in this project, we propose the application of a tool commonly used in photogrammetry and geodesy: the affine transformation adjusted by least squares (LS) (Richardson, 1957).

2. THE AFFINE TRANSFORMATION

The transformation of the conforming bi-dimensional coordinates, also known as 2D Affine Transformation, has the characteristics of the shapes maintained after the transformation. Cartographically, a systematic positional error could be a displacement (in any X, Y, Z component) a factor of scale (in X, Y, and/or Z), or a rotation, with a total or partial combination of these factors. If the planimetric component is the only one worked, the type of systematic error is reduced to those mentioned in table 1. In this way, all these errors could be detected using a LS adjustment of an affine transformation in 2 dimensions. Said technique is frequently employed in geodesy and photogrammetric processes, but not in the quality control of GDB.

Systematic error	Magnitude of error
Displacement X	X_0
Displacement Y	Y_0
Scale factor X	λ_x
Scale factor Y	λ_y
Rotation	α

Table 1: Possible systematic errors: errors in planimetry.

In order to solve the system, a minimum of 3 control points are necessary with their planimetric coordinates known in both systems. So, there will be 6 equations with 5 unknowns that permit the application of the LS. Logically, the greater the number of control points, the greater redundancy the adjustment will have.

Concerning the coordinates of the GDB that define the position of point P , the following transformations or corrections should be applied in order to relate them to the HAS system:

- 2 Factors of scale (λ_x , λ_y). To make longitudes of the axes in both systems of coordinates coincide, it is necessary to multiply each one of the coordinates by said factor of scale:

$$\begin{aligned} X &= \lambda_x \cdot X_{map} \\ Y &= \lambda_y \cdot Y_{map} \end{aligned} \quad (\text{Ec. 1})$$

- 1 Rotation (α). The angle between the axis of coordinates X of the GDB and the earth must be turned to reach parallelism between both systems.
- 2 Displacements (X_0 , Y_0). To make the origin of coordinates coincide in both systems.

For each of the components, the calculation of the equations that define the transformation will be that which appears in table 2. Not taken into consideration is the inclusion of weight because the error to quantify in any of the components has the same importance and therefore, weighting.

In this way, the 5 parameters of the transformation are defined in function of the coefficients of the adjustment. In order to solve them, the system of at least 6 equations and 5 unknowns that can be solved using an adjustment by LS should be considered.

Component X	Component Y
$X' = X_0 + \overline{X_0 X''} - \overline{X' X''}$ $\text{Cos}\alpha = \frac{\overline{X_0 X''}}{X \cdot \lambda x} \rightarrow \overline{X_0 X''} = X \cdot \lambda x \cdot \text{Cos}\alpha$ $\text{Sin}\alpha = \frac{\overline{X' X''}}{Y \cdot \lambda y} \rightarrow \overline{X' X''} = Y \cdot \lambda y \cdot \text{Sin}\alpha$ $X' = X_0 + X \cdot \lambda x \cdot \text{Cos}\alpha - Y \cdot \lambda y \cdot \text{Sin}\alpha \text{ (Ec. 2)}$	$Y' = Y_0 + \overline{Y_0 Y''} - \overline{Y' Y''}$ $\text{Cos}\alpha = \frac{\overline{Y_0 Y''}}{Y \cdot \lambda y} \rightarrow \overline{Y_0 Y''} = Y \cdot \lambda y \cdot \text{Cos}\alpha$ $\text{Sin}\alpha = \frac{\overline{Y' Y''}}{X \cdot \lambda x} \rightarrow \overline{Y' Y''} = X \cdot \lambda x \cdot \text{Sin}\alpha$ $Y' = Y_0 + X \cdot \lambda x \cdot \text{Sin}\alpha + Y \cdot \lambda y \cdot \text{Cos}\alpha \text{ (Ec. 3)}$
Variable change:	
$X' = a \cdot X + b \cdot Y + X_0 ; Y' = c \cdot X + d \cdot Y + Y_0 \text{ (Ec. 4)}$	
Coefficients:	
$a = \lambda x \cdot \text{Cos}\alpha ; b = -\lambda y \cdot \text{Sin}\alpha ; c = \lambda x \cdot \text{Sin}\alpha ; d = \lambda y \cdot \text{Cos}\alpha \text{ (Ec. 5)}$	
Transformation parameters:	
$\lambda x = \sqrt{a^2 + c^2} ; \lambda y = \sqrt{b^2 + d^2} ; \alpha = \text{Arctg}\left(\frac{c}{a}\right) = \text{Arctg}\left(\frac{-b}{d}\right) \text{ (Ec. 6)}$	

Table 2: Calculations of the parameters of the affine transformation between the systems of coordinates GDB and land (HAS).

In order to obtain the values of the coefficients (Ec. 5), the coefficients a, b, c, d, X_0 and Y_0 must be determined so that V will be the remainder of the adjustment:

$$a \cdot X_1 + b \cdot Y_1 + X_{01} = X'_1 + Vx_1 ; c \cdot Y_1 + d \cdot Y_1 + Y_{01} = Y'_1 + Vy_1 ; a \cdot X_2 + b \cdot Y_2 + X_{02} = X'_2 + Vx_2 \dots \text{ (Ec. 7)}$$

Expressing it in matrix form, it would be:

$$A \cdot X = T + V \text{ (Ec. 8)}$$

where:

- A : the matrix of the observations, of dimensions $[2 \cdot n \times 6]$ (n being the number of control points used in the adjustment).
- X : is the matrix of unknowns, of dimensions $[6 \times 1]$
- T : is the matrix of the independent term, with the data of the HAS, and of dimension $[2 \cdot n \times 1]$
- V : is the matrix of the adjustment traces, of dimension $[2 \cdot n \times 1]$

In the adjustment by LS, the goal is for the sum of the squares of the remainders is minimal (Ec. 9), being the resolution of system Ec. 10, and the quality of the adjustment will be given by Ec. 11.

$$M = \sum_{i=1}^n (V_i)^2 \text{ (Ec. 9)} ; A \cdot X = T + V ; X = (A^t \cdot A)^{-1} \cdot A^t \cdot T ; V = A \cdot X - T \text{ (Ec. 10)}$$

$$\sigma_0^2 = \frac{V^t \cdot V}{2 \cdot n - 6} ; \sum \hat{x}\hat{x} = \sigma_0^2 \cdot (A^t \cdot A)^{-1} ; \sigma = \sqrt{\sum \hat{x}\hat{x}} \text{ (Ec. 11)}$$

Once the coefficients of the adjustment are calculated and the quality is verified, the coordinates can be obtained in another system in a different way. In table 3, the equations are shown in order to relate to the land coordinates to those of the GDB (Ec. 12 and 13).

In this way, a systematic error can be detected and quantified with sufficient precision and a more exhaustive knowledge of it can be obtained than with the techniques employed at present. Likewise, the knowledge of the magnitude of said error would permit a user to make a correction of said systematic error, improving the quality of the product at a relatively low cost.

This type of technique has been used profusely in other branches of geomatics (geodesy, photogrammetry, topography...), but in no case have references to cartography quality control been found, nor in the detection and elimination of systematic errors on GDBs.

However, if a priori is presented as an adequate tool for detecting and correcting systematic errors, then it is also true that given the characteristics of the geographic data, this comes affected by its own variability of the scale that it represents (positional uncertainty). This directly affects the typical deviation of the geographic element, so that the existing difference between the real position and the position that represents it will have two clear components:

- Systematic component, which can be detected and corrected using an affine transformation adjusted by least squares.
- Random component, intrinsic of the same geographic data in function of the scale of representation.

Component X	Component Y
$\begin{cases} X' = a \cdot X + b \cdot Y + X_0 \\ Y' = c \cdot X + d \cdot Y + Y_0 \end{cases} \Rightarrow \begin{cases} Y = \frac{X' - a \cdot X - X_0}{b} \\ Y = \frac{Y' - c \cdot X - Y_0}{d} \end{cases}$ $d \cdot X' - d \cdot a \cdot X - d \cdot X_0 = b \cdot Y' - b \cdot c \cdot X - b \cdot Y_0$ $X \cdot (b \cdot c - a \cdot d) = b \cdot Y' - d \cdot X' + d \cdot X_0 - b \cdot Y_0$ $X = \frac{b \cdot (Y' - Y_0) + d \cdot (X_0 - X')}{(b \cdot c - a \cdot d)} \quad (\text{Ec. 12})$	$\begin{cases} X' = a \cdot X + b \cdot Y + X_0 \\ Y' = c \cdot X + d \cdot Y + Y_0 \end{cases} \Rightarrow \begin{cases} X = \frac{X' - b \cdot Y - X_0}{a} \\ X = \frac{Y' - d \cdot Y - Y_0}{c} \end{cases}$ $c \cdot X' - c \cdot b \cdot Y - c \cdot X_0 = a \cdot Y' - a \cdot d \cdot Y - a \cdot Y_0$ $Y \cdot (b \cdot c - a \cdot d) = -a \cdot Y' + a \cdot Y_0 + c \cdot X' - c \cdot X_0$ $Y = \frac{a \cdot (Y_0 - Y') + c \cdot (X' - X_0)}{(b \cdot c - a \cdot d)} \quad (\text{Ec.13})$

Table 3: Calculation of the equations to relate the system of land coordinates (HAS) to the system of the GDB.

This double slope of the uncertainty makes the adjustment by least squares “contaminated” by the random component of the data. So, part of the random error is absorbed, increasing the residual values of the adjustment. This obligates differentiating both types of error and determining between what parameters this technique can be applied.

3. METHODOLOGY

When a systematic error is detected (table 1), the possibility exists to try to identify and quantify it. However, the errors that are produced by displacement in both axes are quite easy to observe using a study of the average error in its components. The project becomes complicated if combined with other systematic errors like variations of scale and rotations. In this case, the goal isn't detection of the systematic error, but to discover what type of error it is and quantify it (displacement, climbing, and/or rotation).

Given the great influence that exists between the sample size and the level of random contamination of the population, different sizes of population at different levels of normal contamination have been worked with. The simulation process could be summarized as:

1. Generate a population of M points (eg. in the form of a regular grid).
2. Contamination of the data:
 - a. Contaminate the data with a random distribution of errors according to $N(0, \sigma)$.
 - b. Contaminate the data with systematic errors: 2 displacements (X_0, Y_0), two factors of scale (λ_x, λ_y) and a rotation (α).
3. Check the data:
 - a. Apply the EMAS standard to detect systematic and random errors.
 - b. Perform an adjustment by least squares with the affine transformation. If the rest are acceptable it is necessary to validate the detected parameters in the affine transformation: $X_0, Y_0, \lambda_x, \lambda_y, \alpha$.
4. Correction of the systematic errors and verification of the corrected results applying the EMAS standard. Analysis of the results.
5. Follow the simulation modifying the systematic errors (return to step 2).

The reason for the decision to contaminate, in the first place, with a random distribution $N(0, \sigma)$, is attributed intrinsically to the fact that in normal conditions all the data will have a determined random error. Likewise, its variability will come bound by the accuracy at which the GDB object of analysis has been realized. In the second place, the contamination of the data with one or more systematic errors is because it is understood that these will be produced as a result of the treatment and manipulation of the data in the process of the cartographic production. Although it may seem to be irrelevant, this is of vital importance given that the results could vary substantially.

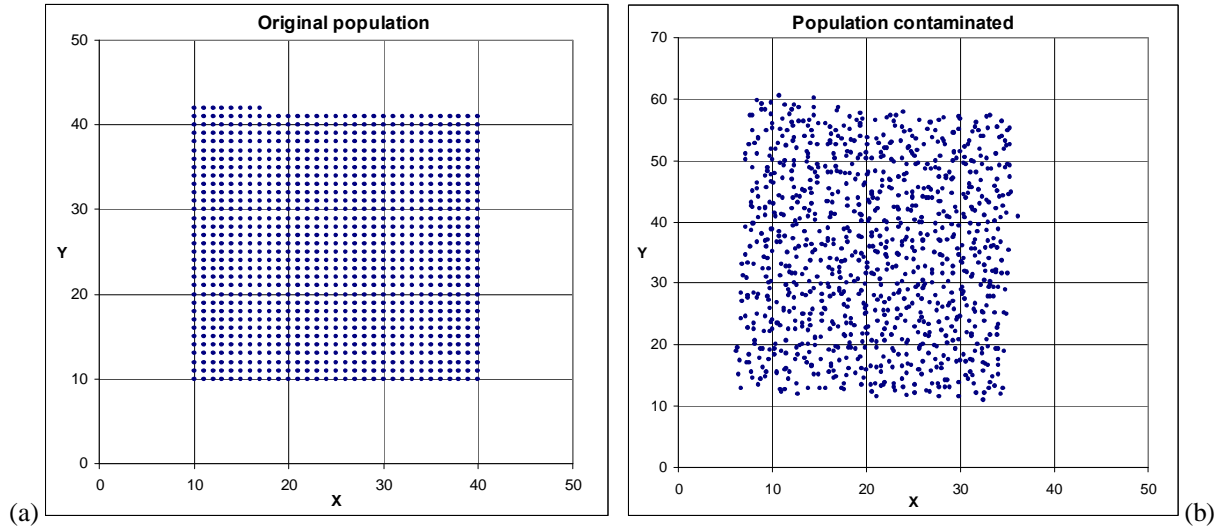


Figure 1: Distribution of the population of 1000 homogeneously simulated points (a) original population and (b) population contaminated by 2 displacements, two factors of scale, a rotation and random errors distributed in a $N(0, I)$.

4. QUANTIFICATION OF SYSTEMATIC ERRORS

The first experiments are developed on a population of 1000 points with no type of systematic error, only with an intrinsic random contamination so that $\sigma = 1\text{m}$ (which can be considered equivalent to a scale of 1/4000), distributing the points according to $N(0, I)$. In this case, the standard EMAS does not detect any systematic or random error. In order to study the behaviour of the affine transformation, it was then adjusted by LS, obtaining the results that appear in table 4 (experiment A). For each experiment (rows A – N), in the left margin of the table are indicated the values obtained before the adjustment for the average and typical deviation (columns P, Q), and whether it has been detected or not, some type of systematic (S_x, S_y) or random (C_x, C_y) error, marked with an “X” (columns R, S).

In the right margin, in the first row of each experiment, appearing in red colour, is the magnitude of systematic contamination (“Contamination \rightarrow ”) that has been applied to the displacement (X_0, Y_0), the factors of scale (λ_x, λ_y) and the rotation (α) (columns T, U, V). Under said values, the second row of each experiment (in blue), are the results of the adjustment by LS with the quantification that has been detected. Lastly, in the last two columns (W, X), a reached deviation is shown for the displacement (V_{x_0, y_0}) and the climbing and rotation ($V_{\lambda\alpha}$) (since the values of deviation for the two factors of scale and the rotation are very similar, in the table they appear averaged, just as is that which occurred with the displacement in X and Y).

The act of presenting the contamination in the first row of each experiment and the results obtained by LS in the second (columns T, U, V), facilitates the visualization of how much the adjustment has approached the actual contamination. In the entire simulation, after validating and performing the adjustment by LS very similar values are always obtained in the search in the original population (for the average and typical deviation).

Returning to the initial experiment, in row “A” of table 4, in spite of the non-existence of any systematic error, the adjustment detects a displacement on Y of 0.47m and a rotation of -0.3° . This is due to the same random errors of a $N(0, I)$. As can be seen more clearly in the displacement, this is inferior to the deviation typical of the GDB (1m), which is negligible. However, these false systematic errors are maintained constant during the entire process of the simulation, altering final results.

In order to know the influence of said error on the values of the final adjustment, a process of simulation was realized with more than 100 contaminated populations with $N(0, I)$, obtaining maximum values of 0.65m in the distortion of the adjustment in displacement, 0.97 as a factor of scale and 0.3° in the rotation. As far as the deviations of the adjustment by LS, these didn’t rise above 20% of the deviation of the map in the case of the displacements and 0.0065 in the case of the factors of scale and rotations.

So, we can say that precisions have been:

- Displacements: are detected with a precision of better than $\pm\sigma$ m (generally around $\pm 0.65 \cdot \sigma$).
- Factors of scale: reach an acceptable precision as far as the rounding off to the first or second decimal.
- Rotations: the precision in the estimation is of one degree

Original population: distribution $N(0,1)$																
Experiment	Values obtained on the contaminated population without correction				Error detection using EMAS				Contamination and results in the quantification of the systematic errors	Contamination by systematic errors and quantification using adjustment obtained by LS					Deviation obtained in the adjustment	
	Average		Deviation		Systm.		Rand.			Displacement		Factor of scale		Rot.	V_{x_0, y_0}	$V_{\lambda \alpha}$
	μ_x	μ_y	σ_x	σ_y	S_x	S_y	C_x	C_y		X_0	Y_0	λ_x	λ_y	α		
A	0	0	1	1					Contamination →	0	0	1	1	0		
									Adjustment →	0,137	0,478	0,991	0,988	-0,296	0,127	0,003
B	2	0,5	1	1	X	X			Contamination →	2	0,5	1	1	0		
									Adjustment →	2,120	0,958	0,991	0,988	-0,296	0,121	0,003
C	1	3,5	1	1	X	X			Contamination →	1	3,5	1	1	0		
									Adjustment →	1,141	3,928	0,991	0,988	-0,296	0,117	0,003
D	-2,75	2,33	1,47	1,24	X	X	X	X	Contamination →	0	0	0,9	1,1	0		
									Adjustment →	0,137	0,478	0,892	1,086	-0,296	0,127	0,003
E	4,15	7,32	1,73	2,75	X	X	X	X	Contamination →	0	0	1,2	1,4	0		
									Adjustment →	0,137	0,478	1,189	1,383	-0,296	0,127	0,004
F	-0,33	0,33	1	1	X	X			Contamination →	0	0	1	1	0,75		
									Adjustment →	0,137	0,478	0,991	0,988	0,454	0,127	0,003
G	-0,88	0,88	1,04	1,03	X	X			Contamination →	0	0	1	1	2		
									Adjustment →	0,137	0,478	0,991	0,988	1,704	0,127	0,003
H	2,15	1,84	1,26	1,55	X	X	X	X	Contamination →	5	-3	0,95	1,15	3,5		
									Adjustment →	5,078	-2,517	0,941	1,136	3,204	0,124	0,004
I	1,83	-1,25	1,02	1,18	X	X		X	Contamination →	2,5	-0,5	1	0,95	1,5		
									Adjustment →	2,611	-0,032	0,991	0,938	1,204	0,123	0,003
J	2,34	0,58	1,02	1,01	X	X			Contamination →	3	0	1	1	1,5		
									Adjustment →	3,109	0,458	0,991	0,988	1,204	0,120	0,003
K	3,63	0	1,24	1	X		X		Contamination →	1,5	0	1,1	1	0		
									Adjustment →	1,623	0,468	1,090	0,988	-0,296	0,124	0,004
L	1	1	1	1	X	X			Contamination →	1	1	1	1	0		
									Adjustment →	1,131	1,459	0,991	0,988	-0,296	0,123	0,003
M	0,51	0,27	1,03	1,05	X	X			Contamination →	1	-1	0,98	1,05	0		
									Adjustment →	1,124	-0,517	0,971	1,037	-0,296	0,127	0,003
N	0,28	2,38	1,04	1,05	X	X			Contamination →	1,2	0,8	0,98	1,05	1		
									Adjustment →	1,329	1,260	0,971	1,037	0,704	0,123	0,003
	P	Q	R	S						T	U	V	W	X		

Simulation performed on a population of 1000 points with random errors according to $N(0,1)$. On each experiment (rows A –N), a determined contamination has been introduced (first row in red of each experiment columns T,U,V) and the detection of systematic errors and randomness was studied using the standard EMAS (columns R and S). The results after quantifying the systematic errors by LS appear in blue (second row of each experiment, columns T,U,V) and the deviation obtained in the adjustment is presented in the last two columns (W and X).

Table 4: Quantification of systematic errors adjusted by LS.

After this first part of the experiment, the valid results obtained with different levels of contamination by systematic errors can be noted in table 4. In Experiments B and C it was altered only by means of a displacement, always obtaining valid results if the systematic errors are inferior to σ . In experiments F and G, it can be noted that the methodology is correct in rotations of 1° or more. Next, various types of systematic errors have been contaminated jointly and in different magnitude, with which the values of precision mentioned before can be confirmed. Also noteworthy are the deviations of the adjustment. These are found to be within the boundaries mentioned.

In the next place, different levels of random contamination were used (always under the assumption that this type of error is distributed in a normal manner). Some of the results reached (Atkinson, 2005), with different levels of systematic contamination are reflected in tables 5, 6, and 7. Random contamination in deviation of 0.125m, 0.50m and 2.00m (which corresponded with the scales 1/500, 1/2000 and 1/8000 respectively) were employed.

Original population: distribution $N(0, 0.125)$																
Experiment	Values obtained on the contaminated population without correction				Error detection using EMAS				Contamination and results in the quantification of the systematic errors	Contamination by systematic errors and quantification using adjustment obtained by LS					Deviation obtained in the adjustment	
	Average		Deviation		System.		Rand.			Displacement		Factor scale		Rot.	Vx_0, y_0	$V\lambda\alpha$
	μ_x	μ_y	σ_x	σ_y	S_x	S_y	C_x	C_y		X_0	Y_0	λ_x	λ_y	α		
A	0	0	0,125	0,125					Contamination → Adjustment →	0 0,092	0 0,250	1 0,999	1 0,983	0 0,097	0,116	0,005
B	0,25	0,06	0,125	0,125	X	X			Contamination → Adjustment →	0,25 0,342	0,06 0,308	1 0,999	1 0,983	0 0,097	0,116	0,005
C	-2,76	1,04	1,02	0,15	X	X	X	X	Contamination → Adjustment →	0 0,092	0 0,250	0,9 0,899	1,1 1,081	0 0,097	0,116	0,006
D	-0,39	0,87	0,13	0,32	X	X		X	Contamination → Adjustment →	0 0,092	0 0,250	1 0,999	1 0,983	2 2,097	0,116	0,005
E	-1,25	2,37	0,49	0,49	X	X	X	X	Contamination → Adjustment →	0,75 0,844	-0,5 -0,244	0,95 0,948	1,15 1,131	3,5 3,597	0,121	0,006
	P	Q	R	S						T	U	V	W	X		

Table 5: Quantification of systematic errors by means of simulation according to a distribution $N(0, 0.125)$.

Original population: distribution $N(0, 0.50)$																
Experiment	Values obtained on the contaminated population without correction				Error detection using EMAS				Contamination and results in the quantification of the systematic errors	Contamination by systematic errors and quantification using adjustment obtained by LS					Deviation obtained in the adjustment	
	Average		Deviation		System.		Rand.			Displacement		Factor of scale		Rot.	Vx_0, y_0	$V\lambda\alpha$
	μ_x	μ_y	σ_x	σ_y	S_x	S_y	C_x	C_y		X_0	Y_0	λ_x	λ_y	α		
A	0	0	0,5	0,5					Contamination → Adjustment →	0 0,001	0 0,065	1 0,998	1 0,998	0 -0,073	0,064	0,002
B	0,5	3	0,5	0,5	X	X			Contamination → Adjustment →	0,5 0,505	3 3,059	1 0,998	1 0,998	0 -0,073	0,060	0,002
C	2,7	2,3	1,1	1	X	X	X	X	Contamination → Adjustment →	0 0,001	0 0,065	0,9 0,898	1,1 1,098	0 -0,073	0,064	0,002
D	0,8	0,8	0,6	0,6	X	X	X	X	Contamination → Adjustment →	0 0,001	0 0,065	1 0,998	1 0,998	2 1,927	0,064	0,002
E	0,35	1,38	0,8	1,4	X	X	X	X	Contamination → Adjustment →	2,5 2,494	-1,2 -1,135	0,95 0,948	1,15 1,148	3,5 3,427	0,063	0,002
	P	Q	R	S						T	U	V	W	X		

Table 6: Quantification of systematic errors by means of simulation of an $N(0, 0.50)$.

As can be noted, the methodology serves to classify and correctly detect systematic errors. With respect to the level of random contamination and the size of the sample with which the adjustment by LS is realized, by increasing the typical deviation of the random errors, the parameters responsible for quantifying the quality of the adjustment (the remainders) that should be maintained near zero, increase considerably. This involves an invalidation of the results in the parameters of the affine transformation. On the contrary, if the typical deviation of the random errors is decreased, the number of control points needed to realize the adjustment of control points can be reduced.

After numerous trials, the graph in figure 2 was developed, in which the relationship between the size of the sample and the typical deviation estimated for the map is shown. The curve determines the zone of validity (above) for the application of the adjustment or an affine transformation by LS in the quantification of systematic error. The graph is defined in function of the sample size and the typical deviation estimated for the map. For those points that are found on the bottom part of the graph, the results obtained in the adjustment by LS are not admissible given that their remainders are elevated. On the contrary, those points that are found in the upper portion of the graph offer optimal results in the adjustment by LS

As an approximation, it can be confirmed that the number of control points that should be employed should not be less than the number proportioned by the curve of adjustment in figure 2, and that it is defined by means of the equation:

$$n = -150,02 \cdot \sigma^2 + 764,81 \cdot \sigma + 30,794 \quad (\text{Ec. 14})$$

with:

- n : minimum number of control points recommended.
- σ : estimated a priori value as the typical deviation of the map (in meters).

It must be noted that, not only for figure 2 but also for its adjustment curve (Ec. 14), only the random errors distributed normally have been taken into account. In the case that atypical values exist, they should be eliminated and/or weighed in such a way that they are diminished. It is also advisable to increase significantly the size of the sample of control points in order to obtain adequate remainders after the adjustment.

Original population: distribution $N(0, 2)$																
Experiment	Values obtained on the contaminated population without correction				Error detection using EMAS				Contamination and results in the quantification of the systematic errors	Contamination by systematic errors and quantification using adjustment obtained by LS					Deviation obtained in the adjustment	
	Average		Deviation		System.		Rand.			Displacement		Factor of scale		Rot		
	μ_x	μ_y	σ_x	σ_y	S_x	S_y	C_x	C_y		X_o	Y_o	λ_x	λ_y	α	V_{x_o}	$V_{\lambda \alpha}$
A	0	0	2	2					Contamination →	0	0	1	1	0		
									Adjustment →	1,079	1,194	0,958	0,957	-0,087	0,248	0,007
C	1	6	2	2	X	X			Contamination →	1	6	1	1	0		
									Adjustment →	2,029	6,934	0,958	0,957	-0,087	0,217	0,007
D	2,7	2,3	2,4	2	X	X	X		Contamination →	0	0	0,9	1,1	0		
									Adjustment →	1,079	1,194	0,862	1,053	-0,087	0,248	0,007
E	4,1	7,3	2,2	3	X	X	X	X	Contamination →	0	0	1,2	1,4	0		
									Adjustment →	1,079	1,194	1,149	1,340	-0,087	0,248	0,009
F	0,3	0,3	2	2	X	X			Contamination →	0	0	1	1	0,75		
									Adjustment →	1,079	1,194	0,958	0,957	0,663	0,248	0,007
G	0,9	0,9	2	2	X	X			Contamination →	0	0	1	1	2		
									Adjustment →	1,079	1,194	0,958	0,958	1,913	0,248	0,007
H	2,1	1,8	2,2	2,2	X	X	X	X	Contamination →	5	-3	0,95	1,15	3,5		
									Adjustment →	5,872	-1,699	0,910	1,101	3,413	0,240	0,007
	P	Q			R	S				T	U	V	W	X		

Table 6: Quantification of systematic errors by means of simulation of an $N(0,2)$.

So, to be able to quantify the magnitude of one or more systematic errors that have been detected previously (for example using a type *t-student* contrast), the scale of the map (or intrinsic accuracy of the map) and the admissible number of control points should be taken into account.

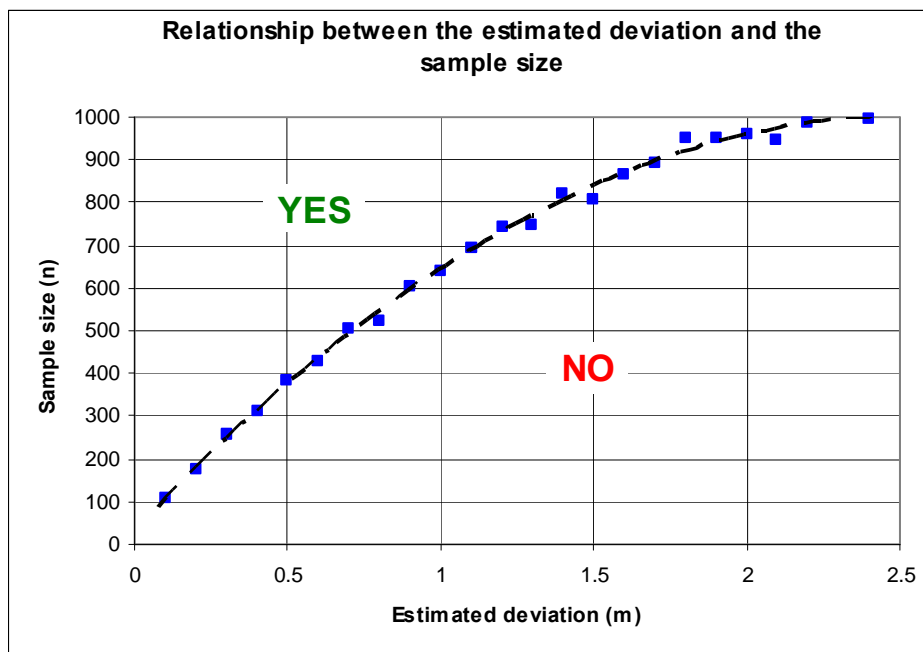


Figure 2: Relationship between the estimated deviation and the sample size for the systematic errors quantification by LS.

For example, if a systematic error has been detected in a cartography of the scale 1/500 ($\sigma = 0,125\text{m}$), at least 125 perfectly defined and homogeneously distributed control points should be taken. If, on the other hand, the cartography to be dealt with is of the scale 1/1000 ($\sigma = 0,25\text{m}$), approximately ≈ 225 points should be used. If the scale is less, it is advisable to consider the economic viability at the moment of capturing a greater number of control points, given that the relationship between the number of control points and scale would be:

- Scale 1/2.000 ($\sigma = 0,50\text{m}$) \rightarrow 380 control points
- Scale 1/5.000 ($\sigma = 1,25\text{m}$) \rightarrow 760 control points
- Scale 1/10.000 ($\sigma = 2,50\text{m}$) \rightarrow 1010 control points
-

For those cases in which the number of control points makes the use of the LS economically envious, the systematic errors in displacement can possibly be corrected by means of the correction by the average error of the coordinates of each component:

$$e_i' = e_i - \bar{e} \quad (\text{Ec. 15})$$

given that:

- e_i' : error for point i correction of displacement.
- e_i : error for point i without correction of displacement.
- \bar{e} : average error and detected displacement.

5. CONCLUSIONS

We have presented the results of employing the affine transformation adjustment by least squares for detecting and quantifying bias in a geographic data base. The work carried out has been based on a simulation process using synthetic populations and controlled contamination. First results point out that the methodology is able to:

- Detect and quantify the magnitude of any displacement produced on the map in either of its axes (X_0, Y_0). The precision is around the typical estimated deviation a priori for the map.
- Detect and quantify the magnitude of any change in scale of its axes (λ_x, λ_y). The precision oscillates around the first decimal in the factor of scale.
- Detect and quantify the magnitude of any rotation produced on the map (α). The precision reached in this type of systematic error is of one degree.
- Offer the possibility of correcting and eliminating all the systematic errors in a relatively quick and simple way: if the GDB in digital format is available, then all that is needed is to apply the affine transformation detected to the coordinates that it defines.
- The major drawback is the large number of control points that are needed for a precise estimation when working with scales inferior to 1/1000.

It must be taken into account up to date that no standard has been found that provides such exhaustive and reliable information on systematic errors which have an effect on certain cartography. It is essential for the hypothesis to have corrected systematic errors beforehand, because the new standards (FGDC, 1998) presume that mistakes have been solved.

Although the adjustment by least squares is an old tool, it is used in an innovative way in the field of cartographic quality using the quantification and elimination of systematic errors in positional control. This obligates the user to get accustomed to the results that it provides, and to try to discern when the remainders provided in the adjustment are or are not optimal in function of the scale and size of the sample. For this a first approximation shown in figure 2 is relied on.

6. REFERENCES

- Ariza, F.J. (2002). "Calidad en la Producción Cartográfica". Ra-MA, Madrid, (Spain).
- ASP, (1985). "Accuracy specification for large-scale Line Maps". En PE&RS, vol. 51, n° 2.
- ASPRS, (1989). "Accuracy standards for large scale maps". En PE&RS, vol. 56, n° 7.
- 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 (Spain).
- FGDC, (1998). "Geospatial Positioning Accuracy Standards". Ninnnesota Planning, 1999. FGCD-STD-007-1998. Virginia (U.S.A).
- Richardson, H.F. (1957). "Survey Adjustments and Last Squares". Constable & Co. Londres (U.K.).
- Volkmann, U. (2002). "Elementos básicos para procesamiento de datos experimentales". Fac. de Física, Pontificia Univ. Católica de Chile. Santiago (Chile).

7. BIOGRAPHY

7.1. Personal Information

Dr. Alan D.J. Atkinson Gordo; E-mail: atkinson@unex.es; web: www.unex.es/igpu/Alan
Address: Escuela Politécnica; Av. Universidad s/n; 10.190 Cáceres; Spain
Phone: (+34) 927.257.195 (Ext. 7578); Fax: (+34) 927.257.203

7.2. Academic information

Associate Professor T.C. at the University of Extremadura (UEX).

7.3. Investigative Activity

Author of the doctoral thesis “Positional Quality Control in Cartography: Analysis of the Main Standards and Proposal for Improvement”, within the doctorate program in Geodesic Engineering Cartography and Photogrammetry of the University of Jaén.

Secretary and founding member of the Investigative team in Geomatic Engineering and Urban Heritage (IGPU) of the UEX (website of IGPU: www.unex.es/igpu). Member of the editorial counsel of the magazine DATUM XXI (Official publication of the Association of Engineers in Geodesy and Cartography) since its foundation (2002). Member of the Directing Counsel of the Association of Engineers in Geodesy and Cartography (AIGC) since its foundation (website for AIGC: www.aigc.es).

Participation as investigator in the projects:

- Project of National Investigation (I+D): Control de calidad en cartografía por elementos lineales - ConPoCar.
- Project of National Investigation (I+D): Elaboración de un sistema integrado para la prevención de riesgos hidrovulcánicos
- Project of Regional Investigation (PRI): Influencia de las condiciones de bienestar animal en la optimización de la producción del cerdo ibérico.
- Project of Regional Investigation (PRI): Desarrollo de un restituidor digital para la realización de cartografía automática bajo Linex
- Project of National Investigation (I+D): Degradación de permafrost y significado geográfico y ambiental en sierra nevada. El caso del corral del veleta.
- Project of National Investigation (I+D): Cálculo de abundancias mediante redes neuronales a partir de imágenes hiperespectrales.
- Project of National Investigation (I+D): Los lahares del popocatépetl (méxico): control y prevención de riesgos.