A statistical model for positional quality control of spatial data

Ariza-López, FJ[1]*; Rodríguez-Avi, J*[2]

[1] Universidad de Jaén (España).

Abstract

A statistical method for positional-quality-acceptation control is presented. This method can be applied to any kind of geometry and any parametric or non-parametric model. Two statistical models are applied together: a fixed Binomial is applied over a Base Model. The Base Model represents the hypothesis about the error behavior. The conceptual bases and the application procedure are presented. Actual examples for 1D, 2D and 3D spatial data are shown.

Keywords: positional accuracy, quality control.

1. Introduction

Positional accuracy of spatial data sets (SDS) has traditionally been evaluated using control points. Following this idea there are many statistical Positional Accuracy Assessment Methodologies (PAAM) like: the National Map Accuracy Standards, the Accuracy Standards for Large Scale Maps or the National Standard for Spatial Data Accuracy. Many of the existing PAAM are based on the hypothesis of Gaussian error. But there are studies that indicate other behaviors (Raleigh distribution, log-normal distribution, Chi2, etc.) for the errors of SDS.

The PAAM go in two directions:

- Estimation of a parameter with some statistical significance.
- Acceptance or rejection of the SDS is taken by a statistical contrast.

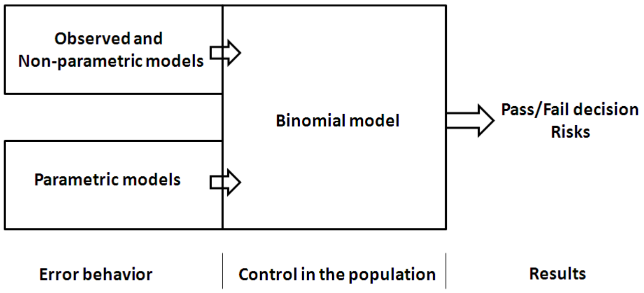
We propose a positional-quality-acceptation control which is valid for 1D, 2D and 3D data and any kind of geometries (e.g. points, line strings, etc.). The most important features are:

- It is available for arbitrary error models.
- The test is run on the population and not on a parameter.

2. A proposal for positional accuracy control

We propose a control based on two statistical models. The first is the Base Model (BaM), which can be any parametric or non-parametric model but adequate to represent the behavior of the errors we want to control. The second is the Binomial Model (BiM) and is applied over the former by means of counting positional defectives in order to get a pass/fail decision. This idea is illustrated in Figure 1.

Figure 1: General idea of the proposed statistical control.



The BaM represents the error behavior in the population, so given a tolerance *Tol* the percentage of cases greater than the desired *Tol* is □. In a control sample of size *n*, we define the fact that the error E_i in element *i* verifies $E_i>Tol$ as a fail event. The *BiM* consists of counting the number $\#E_i$ of fail events. This test follows a Binomial *B(n,□)* distribution and the probability of fail events is (Johnson et al. 2005):

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

Where $\pi = P[E_i > Tol]$

The null hypothesis is:

¬ H0: The SDS is adequate. Given a signification value (α) (type I error or producer's risk), it means that errors are distributed according to the BaM and only π% of cases are greater than *Tol*.

Versus

¬ H1: The SDS is not adequate.

The essential steps are:

1. A BaM is needed. It must be previously determined.
2. Selection of the *Tol* in order to satisfy the quality requirements.
3. Realization of the random sample of size *n*.
4. Calculation of errors and counting of *F* events.
5. Decision. Determine if $p\geq\alpha$ or $p\leq\alpha$ in order to make the pass/fail decision.

3. Examples of application

Three examples are presented where different BaMs are considered in order to demonstrate the general applicability. First of all, the BaMs are stated graphically and afterwards the remaining steps are presented. In all cases figures representing the BaMs show the observed error distribution where frequency *f* of observed positional errors is in the Y-axis and the size of such errors is on the X-axis.

Case 1 (1D Lidar altimetric control). Let us suppose that the BaM is the observed error distribution of Figure 2. This distribution corresponds to the Weeds/Crop class of a report by Dewberry (Dewberry, 2004).

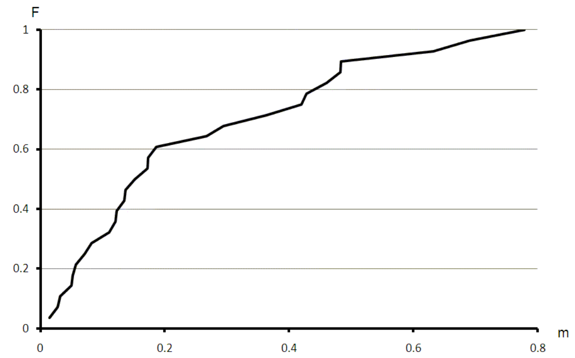


Figure 2: Distribution for altimetric Lidar errors (Case 1).

Case 2 (2D planimetric control). Now the BaM is a Chi2 (sum of two independent squared standard normal distributed errors). Figure 3 is derived from the data of Appendix 3-B of the NSSDA (FGDC, 1998).

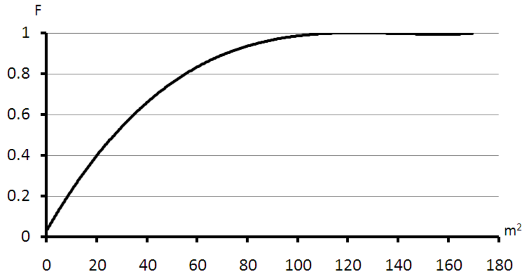


Figure 3: Distribution of squared planimetric errors (Case 2).

Case 3 (3D line-string control). A 3D line-string control is performed using the Hausdorff distance. Figure 4 shows the BaM observed for some results of the E3DELING project (Ariza-López et al., 2012).

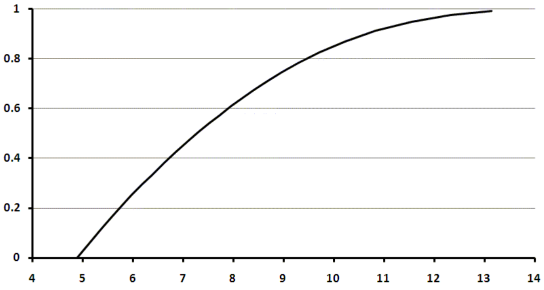


Figure 4: Distribution of distance errors in 3D line-strings (Case 3).

Now consider the following:

¬ Case 1. $Tol_{C1}=0.5m$, $n_{C1} = 20$, $\alpha = 5\%$. Tol_{C1} implies that $\pi_{C1} = 0.1$.

¬ Case 2. $Tol_{C2}=93m^2$, $n_{C2} = 20$, $\alpha = 5\%$. Tol_{C2} implies that $\pi_{C2} = 0.1$.

¬ Case 3. $Tol_{C3}=10.7m$, $n_{C3} = 20$, $\alpha = 5\%$. Tol_{C3} implies that $\pi_{C3} = 0.1$.

Observe that the tolerances selected implies that $\pi_{C1} = \pi_{C2} = \pi_{C3} = \pi$. With π and n, by Equation (1), we can calculate the probability of F events in the sample. Table 1 shows p values and F values in the interval [1, 8] when tolerance is exceeded by $\pi\%$ of cases and $n=20$.

Table 1: *p* values for diverse count of control elements (F) out of tolerance when $n=20$ and $\pi=0.1$.

F	P	F	P
1	0.8784	5	0.0431
2	0.6082	6	0.0112
3	0.3230	7	0.0023
4	0.1329	8	0.0004

For instance, if we find that $F\leq 4$ we do not reject H_0 because $p=0.13>\alpha$. On the contrary, if $F\geq 5$ we can reject H_0.

As has been observed $BaM_{C1}\neq BaM_{C2}\neq BaM_{C3}$ but as $\pi_{C1}=\pi_{C2}=\pi_{C3}$ the final test behavior is the same. In other words, Table 1 is valid for any BaM and the result of the test only depends on π.

3. Conclusion

A new statistical method for positional control has been presented. This method can be applied to any kind of geometry (e.g. points, line strings, etc.) and any error model (parametric or non parametric).

Some examples demonstrate the general applicability of the proposal. The main strength of the proposal is that it is not linked to any specific statistical hypothesis on errors.

Acknowledgments

This work has been funded by the Ministry of Science and Technology (Spain) and the European Regional Development Fund under grant no. BIA2011-23217. The authors also acknowledge the Regional Government of Andalusia (Spain) for the financial support since 1997 for the research groups with code PAIDI-TEP-164 and PAIDI-FQM-245.

References

Ariza-López F.J, García-Balboa J.L, Ureña-Cámara M.A, Reinoso-Gordo F.J. (2012). Metodología para la evaluación de la calidad de elementos lineales 3D. En X Congreso TOPCART 2012, 16-19 Octubre, Madrid.

Dewberry, (2004). Worcester County LIDAR 2002 Quality Assurance Report. Maryland Department of Natural Resources.

FGDC (1998). FGDC-STD-007: Geospatial Positioning Accuracy Standards, Part 3. NSSDA. FGDC, Reston, USA.

Johnson, NL, Kemp, AW, Kotz, S. (2005). Univariate Discrete distributions (2005). Wiley.