A statistical model for positional quality control of spatial data

F.J. Ariza-López & J. Rodríguez-Avi

Universidad de Jaén
Escuela Politécnica Superior
A3-336
Tel (+34) 953 212469

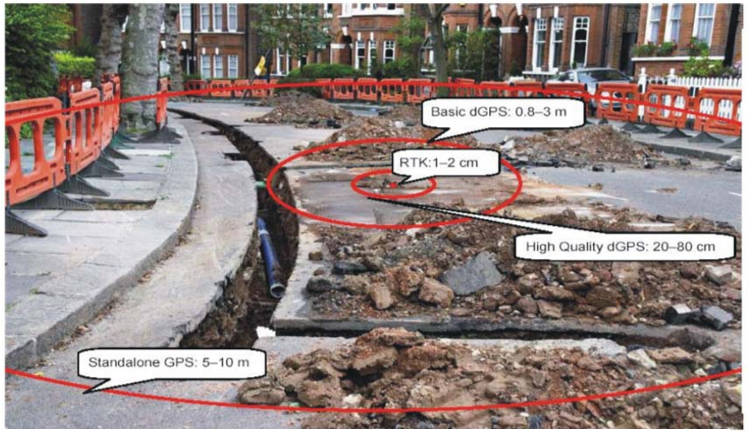
fjariza@ujaen.es / jravi@ujaen.es

Contents

- Introduction
- Objectives
- The proposal
- Examples
- Conclusions & Prospects

Introduction

Positional accuracy is of great importance



In general:

- •Increase of use of GI implies increasing demand of quality.
- •SDI need interoperability.
- •GNSS allow everybody to get coordinates.

Demanding applications:

- •Intelligence.
- •Military applications (eg weapons and missiles)
- •Unmanned vehicles (UA).
- •Navigation.
- •Precision farming.
- •Etc.

Introduction

There are many positional accuracy assessment methods (PAAMs) available:

- **National Map Accuracy Standard (1947) by USBB.**
- **Accuracy Standards for Large Scale Maps (1990) by ASPRS**
- **Engineering Map Accuracy Standard (1983) by ASCE**
- **National Standard for Spatial Data Accuracy (1998) by FGDC**
- **STANAG 2215 by NATO.**
- **Etc.**

United States National Map Accuracy Standards

With a view to the utmost economy and expedition in producing maps which fulfill not only the broad needs for standard or principal maps, but also the reasonable particular needs of individual agencies, standards of accuracy for published maps are defined as follows:

1. **Horizontal accuracy.** For maps on publication scales larger than 1:20,000, not more than 10 percent of the points tested shall be in error by more than 1/30 inch, measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50 inch. These limits of accuracy shall apply in all cases to positions of well-defined points only. Well-defined points are those that are easily visible or recoverable on the ground, such as the following: monuments or markers, such as bench marks, property boundary monuments; intersections of roads, railroads, etc.; corners of large buildings or structures (or center points of small buildings); etc. In general what is well defined will be determined by what is plottable on the scale of the map within 1/100 inch. Thus while the intersection of two road or property lines meeting at right angles would come within a sensible interpretation, identification of the intersection of such lines meeting at an acute angle would obviously not be practicable within 1/100 inch. Similarly, features not identifiable upon the ground within close limits are not to be considered as test points within the limits quoted, even though their positions may be scaled closely upon the map. In this class would come timber lines, soil boundaries, etc.
2. **Vertical accuracy,** as applied to contour maps on all publication scales, shall be such that not more than 10 percent of the elevations tested shall be in error more than one-half the contour interval. In checking elevations taken from the map, the apparent vertical error may be decreased by assuming a horizontal displacement within the permissible horizontal error for a map of that scale.
3. **The accuracy of any map may be tested** by comparing the positions of points whose locations or elevations are shown upon it with corresponding positions as determined by surveys of a higher accuracy. Tests shall be made by the producing agency, which shall also determine which of its maps are to be tested, and the extent of the testing.
4. **Published maps meeting these accuracy requirements** shall note this fact on their legends, as follows: "This map complies with National Map accuracy Standards."
5. **Published maps whose errors exceed those aforestated** shall omit from their legends all mention of standard accuracy.
6. **When a published map is a considerable enlargement** of a map drawing (manuscript) or of a published map, that fact shall be stated in the legend. For example, "This map is an enlargement of a 1:20,000-scale map drawing," or "This map is an enlargement of a 1:24,000-scale published map."
7. **To facilitate ready interchange and use of basic information for map construction** among all Federal mapmaking agencies, manuscript maps and published maps, wherever economically feasible and consistent with the uses to which the map is to be put, shall conform to latitude and longitude boundaries, being 15 minutes of latitude and longitude, or 7.5 minutes, or 3-3/4 minutes in size.

Issued June 10, 1941
Revised April 26, 1943
Revised June 17, 1947

U.S. BUREAU OF THE BUDGET

Introduction

But, these methods have problems:

- Control elements → points
- Operational procedure → control of isolated SDS

- Statistical problems:
  - The main underlying assumption is the normality of the errors. This assumption has disastrous results if not satisfied.
  - Another common underlying hypothesis is the equality (or near equality) of error distributions in all components (ie $\sigma x = \sigma y$).
  - As indicated in numerous studies, positional errors do not always follow a normal distribution.
  - Outliers should be removed as they affect the results, but do not exist uniform criteria.

Objective

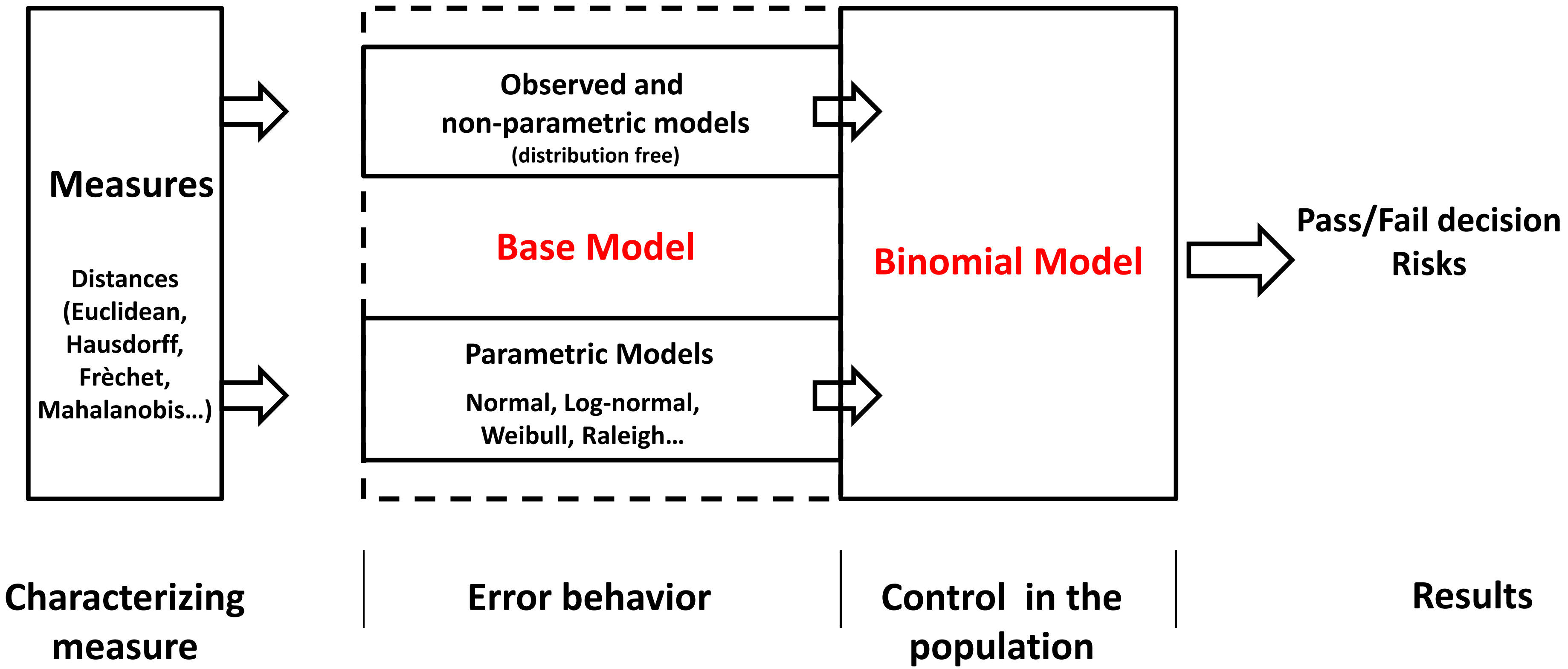
Our goals were to develop:

- A simple statistical method.
- A suitable method for any error model (parametric or non-parametric)
- A method that runs on the population and not on parameters of the population.
- A method valid for 1D, 2D and 3D data and any kind of geometries (e.g. points, line strings, etc.).

The proposal

A general view

$$CM = BiM(n, \pi | \pi \sim BaM)$$

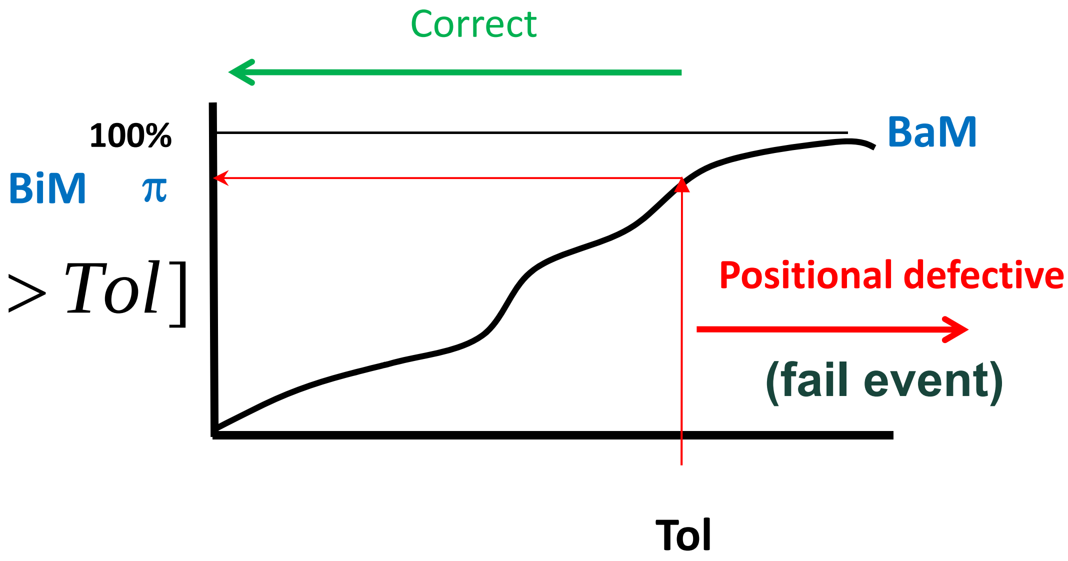


The proposal

Observed errors

$$E_i = \sqrt{\sum_{j=1}^{p} \left(x_{ij} - x_{ij}^{R}\right)^2} \rightarrow BaM$$

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



$$E_i > Tol$$

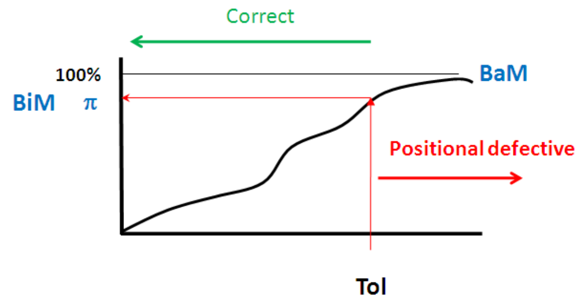
$$P[F > mc \mid F \rightarrow B(n,\pi)] = \sum_{k=mc+1}^{n} \binom{n}{k} \pi^{k} (1-\pi)^{n-k}$$

Ec.1

The proposal

The test

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



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

The null hypothesis is:

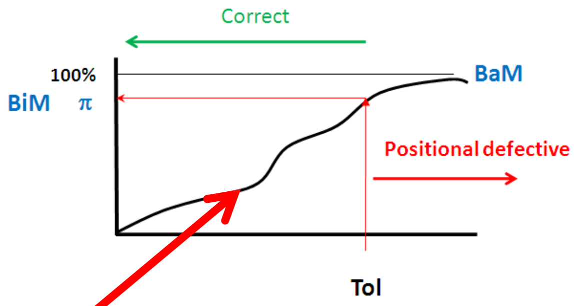
- **H0: The SDS is adequate. Given a signification value (α) (type I error or producer's risk), it means that errors are distributed according to the BaM and only π% of cases are greater than *Tol*.**

Versus

- **H1: The SDS is not adequate.**

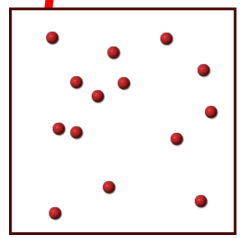
The proposal

The procedure



The steps are:

- **A BaM is needed. It must be previously determined.**
- **Selection of the Tol in order to satisfy the requirements.**
- **Realization of the random sample of size n.**



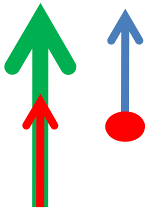
- **Calculation of positional errors and counting of positional defectives.**

$$E_i = \sqrt{\sum_{j=1}^{p} \left(x_{ij} - x_{ij}^{R}\right)^2} \qquad E_i > Tol$$

- **Decision. Determine if p-value≥α or p-value≤α in order to make the pass/fail decision.**

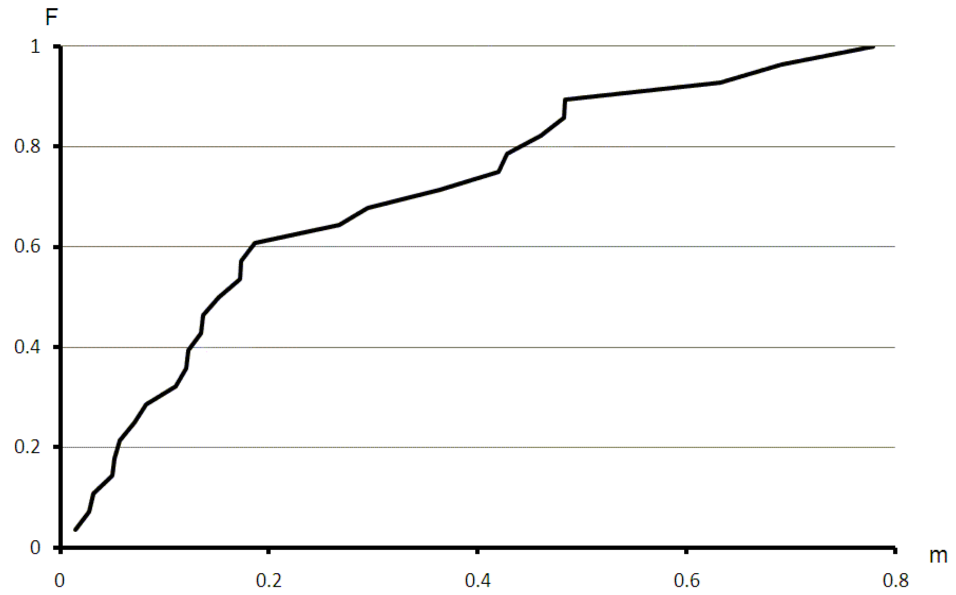
$$P[F > mc \mid F \rightarrow B(n,\pi)] = \sum_{k=mc+1}^{n} \binom{n}{k} \pi^k (1-\pi)^{n-k}$$

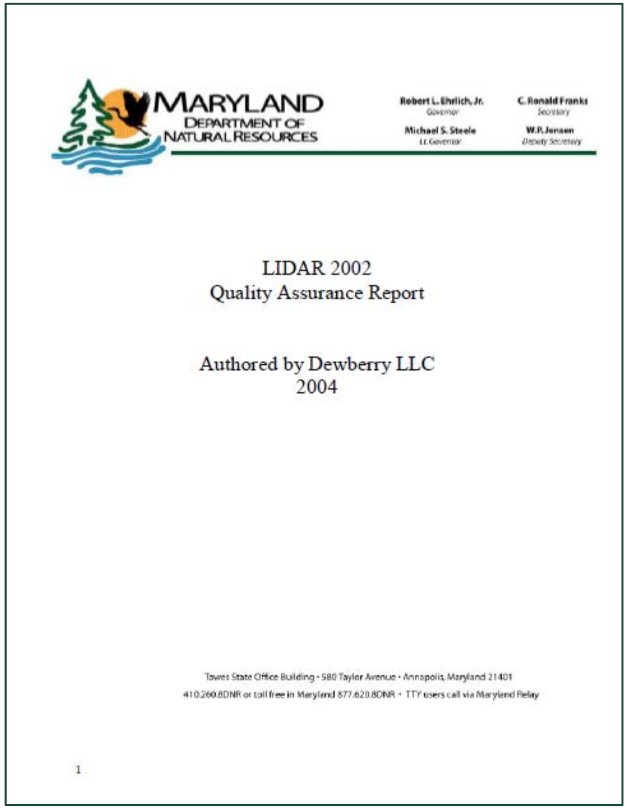
Examples



Case 1D

Example 1





MARYLAND DEPARTMENT OF NATURAL RESOURCES

Robert L. Ehrlich, Jr. Governor
Michael S. Steele Lt. Governor
C. Ronald Franks Secretary
W.P. Jensen Deputy Secretary

LIDAR 2002
Quality Assurance Report

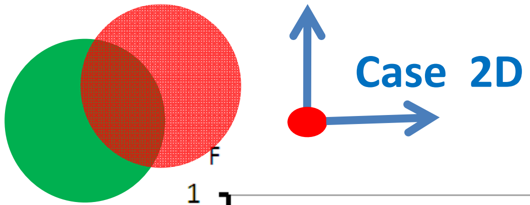
Authored by Dewberry LLC
2004

Tawes State Office Building - 580 Taylor Avenue - Annapolis, Maryland 21401
410.260.8DNR or toll free in Maryland 877.620.8DNR • TTY users call via Maryland Relay

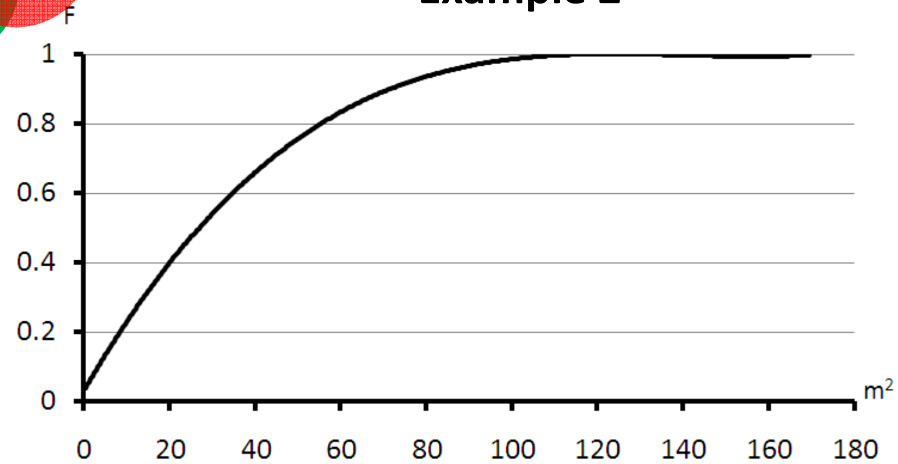
1

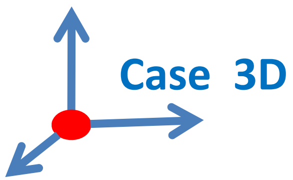
Source. Dewberry, (2004). Worcester County LIDAR 2002 Quality Assurance Report. Maryland Department of Natural Resources.

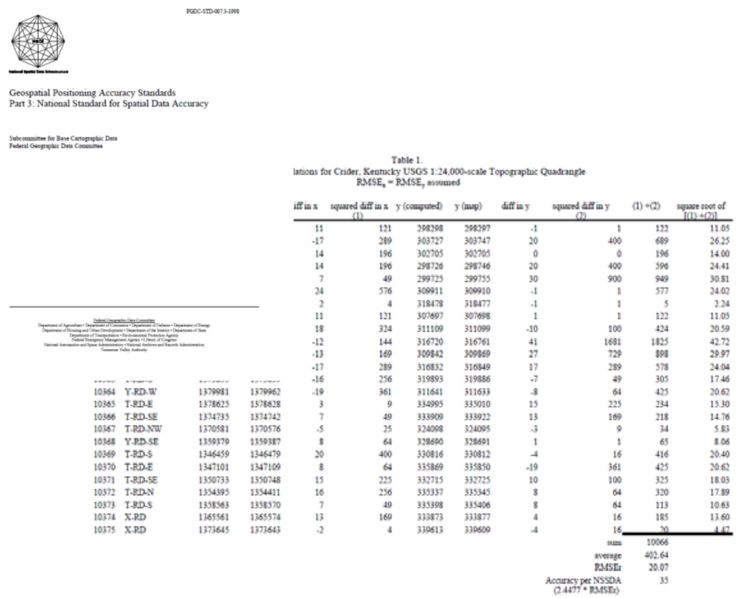
Examples



Example 2







FGDC-STD-007.3-1998

National Spatial Data Infrastructure

Geospatial Positioning Accuracy Standards
Part 3: National Standard for Spatial Data Accuracy

Subcommittee for Base Cartographic Data
Federal Geographic Data Committee

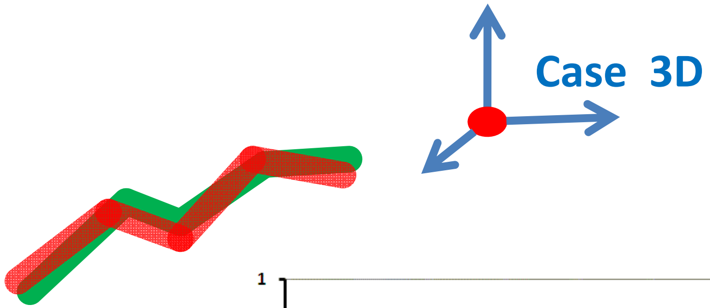
Federal Geographic Data Committee
Department of Agriculture • Department of Commerce • Department of Defense • Department of Energy
Department of Housing and Urban Development • Department of the Interior • Department of State
Department of Transportation • Environmental Protection Agency
Federal Emergency Management Agency • Library of Congress
National Aeronautics and Space Administration • National Archives and Records Administration
Tennessee Valley Authority

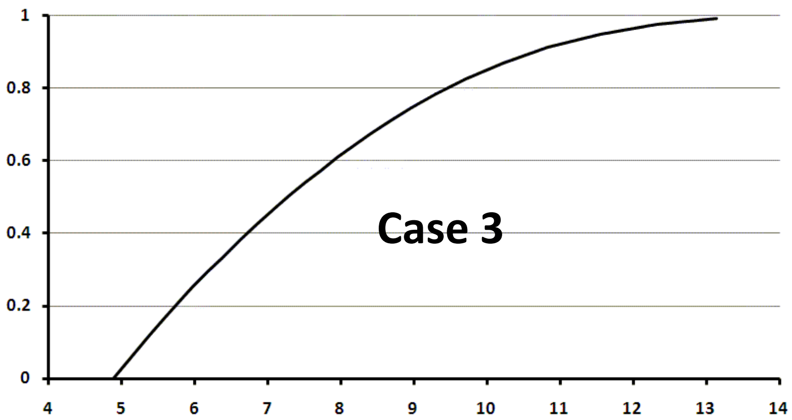
Table 1.
lations for Crider, Kentucky USGS 1:24,000-scale Topographic Quadrangle
$RMSE_x = RMSE_y$ assumed

				iff in x	squared diff in x (1)	y (computed)	y (map)	diff in y	squared diff in y (2)	(1) +(2)	square root of [(1) +(2)]
				11	121	298298	298297	-1	1	122	11.05
				-17	289	303727	303747	20	400	689	26.25
				14	196	302705	302705	0	0	196	14.00
				14	196	298726	298746	20	400	596	24.41
				7	49	299725	299755	30	900	949	30.81
				24	576	309911	309910	-1	1	577	24.02
				2	4	318478	318477	-1	1	5	2.24
				11	121	307697	307698	1	1	122	11.05
				18	324	311109	311099	-10	100	424	20.59
				-12	144	316720	316761	41	1681	1825	42.72
				-13	169	309842	309869	27	729	898	29.97
				-17	289	316832	316849	17	289	578	24.04
				-16	256	319893	319886	-7	49	305	17.46
10364	Y-RD-W	1379981	1379962	-19	361	311641	311633	-8	64	425	20.62
10365	T-RD-E	1378625	1378628	3	9	334995	335010	15	225	234	15.30
10366	T-RD-SE	1374735	1374742	7	49	333909	333922	13	169	218	14.76
10367	T-RD-NW	1370581	1370576	-5	25	324098	324095	-3	9	34	5.83
10368	Y-RD-SE	1359379	1359387	8	64	328690	328691	1	1	65	8.06
10369	T-RD-S	1346459	1346479	20	400	330816	330812	-4	16	416	20.40
10370	T-RD-E	1347101	1347109	8	64	335869	335850	-19	361	425	20.62
10371	T-RD-SE	1350733	1350748	15	225	332715	332725	10	100	325	18.03
10372	T-RD-N	1354395	1354411	16	256	335337	335345	8	64	320	17.89
10373	T-RD-S	1358563	1358570	7	49	335398	335406	8	64	113	10.63
10374	X-RD	1365561	1365574	13	169	333873	333877	4	16	185	13.60
10375	X-RD	1373645	1373643	-2	4	339613	339609	-4	16	20	4.47
									sum	10066	
									average	402.64	
									RMSEr	20.07	
									Accuracy per NSSDA (2.4477 * RMSEr)	35	

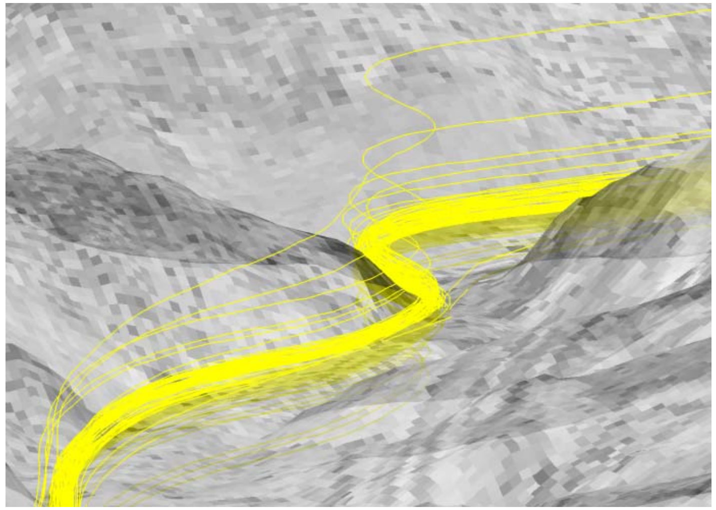
Source: FGDC (1998). FGDC-STD-007: Geospatial Positioning Accuracy Standards, Part 3. NSSDA. FGDC, Reston, USA.

Examples





Source: 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.

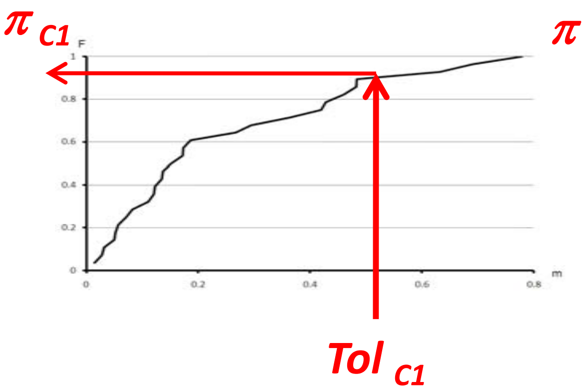


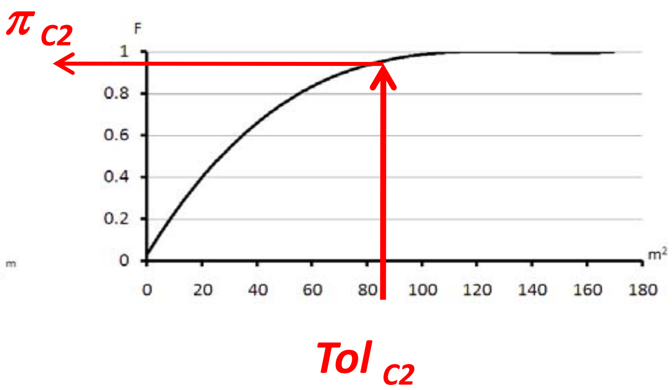
Examples

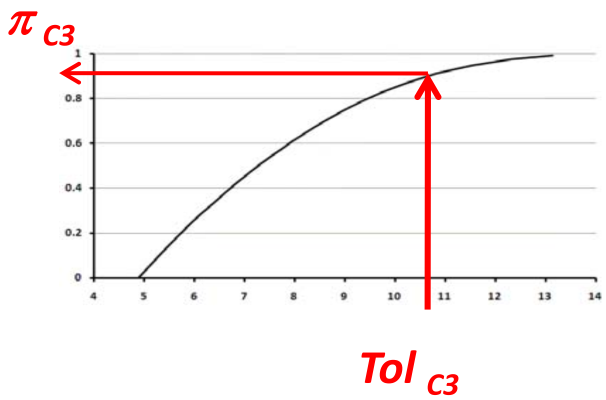
A numerical example

Now consider the following, n=20 and α = *5% for all cases, and*:

Case 1. *Tol_{C1}=0.5m → π_{C1} = 0.1 (n_{C1} = 20, α = 5%).*
Case 2. Tol_{C2}=93m^2 → π_{C2} = 0.1 (n_{C2} = 20, α = 5%).
Case 3. Tol_{C3}=10.7m → π_{C3} = 0.1 (n_{C3} = 20, α = 5%).





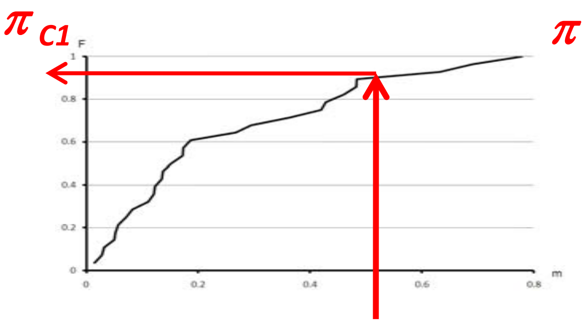


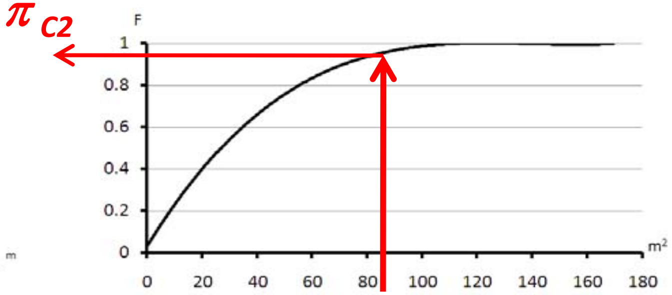
Examples

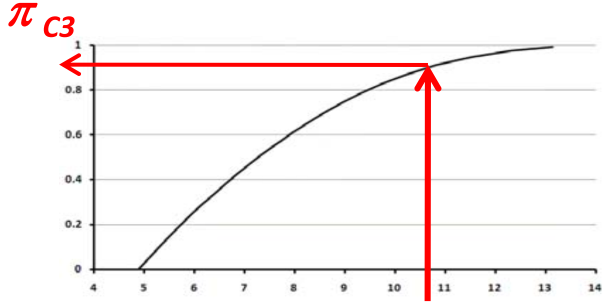
A numerical example

Now consider the following, n=20 and α = *5% for all cases, and*:

Case 1. Tol_{C1}=0.5m → π_{C1} = 0.1 (n_{C1} = 20, α = 5%).
Case 2. Tol_{C2}=93m² → π_{C2} = 0.1 (n_{C2} = 20, α = 5%).
Case 3. Tol_{C3}=10.7m → π_{C3} = 0.1 (n_{C3} = 20, α = 5%).







By Ec.1: α = 5%, n=20, π=0.1

If P-value ≥α → Accept
If P-value<α → Reject

Accept H0

Positional defectives	P-value
1	0.8784
2	0.6082
3	0.3230
4	0.1329

Reject H0

Positional defectives	P-value
5	0.0431
6	0.0112
7	0.0023
8	0.0004

Conclusions

- A new statistical method for positional control has been presented.
- The method is very simple.
- A statistical hypothesis test is applied.
- This method can be applied to any kind of geometry (e.g. points, line strings, etc.)
- This method can be applied to any kind of 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.**

Future work

This method can be liked to acceptance sampling standards:

•e.g. ISO 2859-1 and ISO 2859-2, Dodge-Roming or Philips procedures, zero defects .

- •This idea allows controlling the supply of isolated lots of spatial data.

- •This idea allows control of lot-by-lot supplies of spatial data.

This method allows controlling position and thematic attributes together.

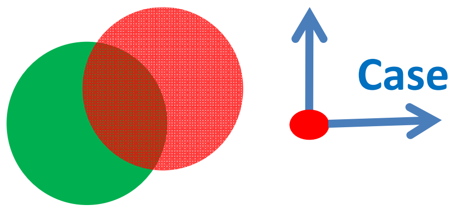
ACKNOWLEDGEMENTS
This work has been funded by the Ministry of Science and Technology (Spain) and the European Regional Development Fund under grant no. BIA2011-23217. The authors also acknowledge the Regional Government of Andalusia (Spain) for the financial support since 1997 for the research groups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

A statistical model for positional quality control of spatial data

THANK YOU VERY MUCH FOR YOUR ATTENTION

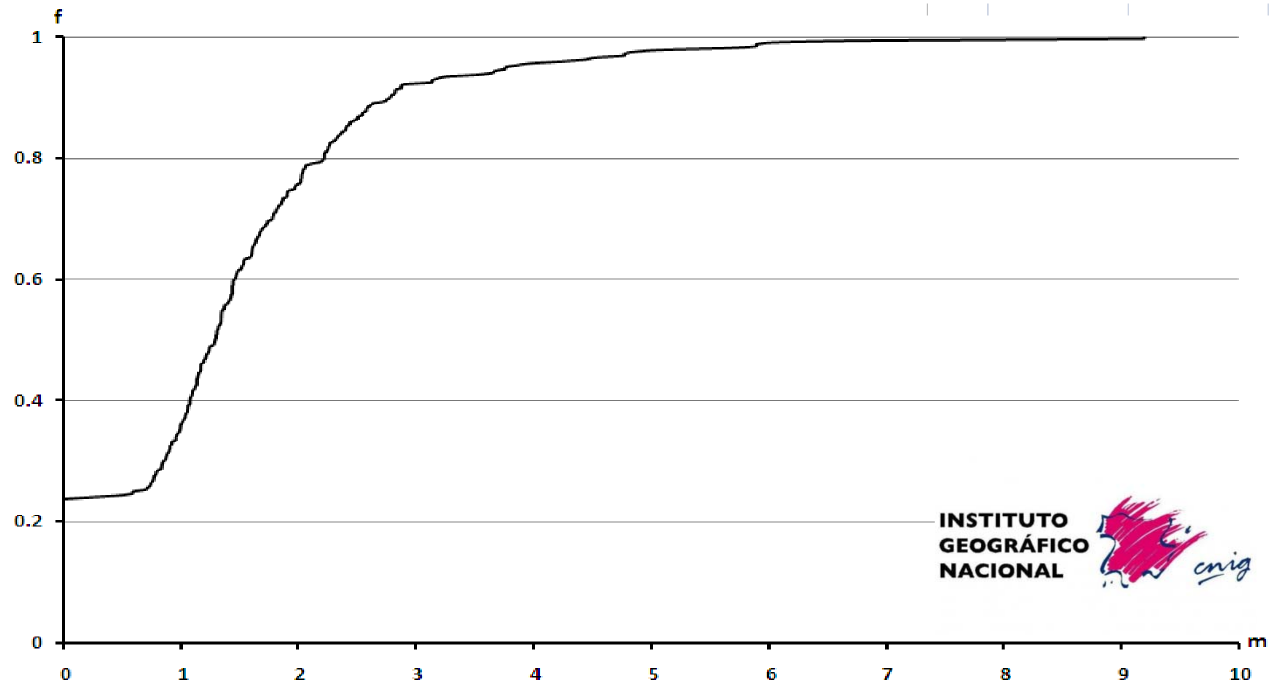
F.J. Ariza-López & J. Rodríguez-Avi
Universidad de Jaén

Examples



Case 2D

Other examples



Examples