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.
Introduction

But, these methods have problems:

• Control elements $\rightarrow$ points

• Operational procedure $\rightarrow$ 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

\[ CM = BiM(n, \pi \mid \pi \sim BaM) \]

A general view

- **Measures**
  - Distances (Euclidean, Hausdorff, Frèchet, Mahalanobis...)

- **Base Model**
  - Observed and non-parametric models (distribution free)
  - Parametric Models
    - Normal, Log-normal, Weibull, Raleigh...

- **Binomial Model**
  - Error behavior
  - Control in the population
  - Results

- Pass/Fail decision
- Risks

F.J. Ariza López, University of Jaén
The proposal

Observed errors

\[ E_i = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{ij}^R)^2} \rightarrow \text{BaM} \]

\[ \pi = P[E_i > \text{Tol}] \]

Correct

Positional defective

(fail event)

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

Ec.1

F.J. Ariza López, University of Jaén
The proposal

The test

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

The null hypothesis is:

- \( H_0 \): The SDS is adequate. Given a signification value (\( \alpha \)) (type I error or producer's risk), it means that errors are distributed according to the BaM and only \( \pi \)% of cases are greater than \( Tol \).

Versus

- \( H_1 \): 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.
- Decision. Determine if p-value ≥ α or p-value ≤ α in order to make the pass/fail decision.

\[ E_i = \sqrt{\sum_{j=1}^{n} (x_j - x_j^r)^2} \]

\[ E_i > Tol \]

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

Example 1

Examples

Case 2D

Example 2

Examples

Examples

A numerical example

Now consider the following, \( n=20 \) and \( \alpha = 5\% \) for all cases, and:

- **Case 1.** \( Tol_{C1}=0.5m \) \( \rightarrow \pi_{C1} = 0.1 (n_{C1} = 20, \alpha = 5\%). \)
- **Case 2.** \( Tol_{C2}=93m^2 \) \( \rightarrow \pi_{C2} = 0.1 (n_{C2} = 20, \alpha = 5\%). \)
- **Case 3.** \( Tol_{C3}=10.7m \) \( \rightarrow \pi_{C3} = 0.1 (n_{C3} = 20, \alpha = 5\%). \)
Examples

A numerical example

Now consider the following, n=20 and $\alpha = 5\%$ for all cases, and:

Case 1. $Tol_{C1}=0.5m \rightarrow \pi_{C1} = 0.1 (n_{C1} = 20, \alpha = 5\%)$.
Case 2. $Tol_{C2}=93m^2 \rightarrow \pi_{C2} = 0.1 (n_{C2} = 20, \alpha = 5\%)$.
Case 3. $Tol_{C3}=10.7m \rightarrow \pi_{C3} = 0.1 (n_{C3} = 20, \alpha = 5\%)$.

$\pi_{C1}$ $\pi_{C2}$ $\pi_{C3}$

$Tol_{C1}$ $Tol_{C2}$ $Tol_{C3}$

$\alpha = 5\%, n=20, \pi=0.1$

By Ec.1:
If $P$-value $\geq \alpha \rightarrow$ Accept
If $P$-value $< \alpha \rightarrow$ Reject

<table>
<thead>
<tr>
<th>Positional defectives</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8784</td>
</tr>
<tr>
<td>2</td>
<td>0.6082</td>
</tr>
<tr>
<td>3</td>
<td>0.3230</td>
</tr>
<tr>
<td>4</td>
<td>0.1329</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positional defectives</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.0431</td>
</tr>
<tr>
<td>6</td>
<td>0.0112</td>
</tr>
<tr>
<td>7</td>
<td>0.0023</td>
</tr>
<tr>
<td>8</td>
<td>0.0004</td>
</tr>
</tbody>
</table>
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

Case 2D

Other examples