Analysis of Some Positional Accuracy Assessment Methodologies

Francisco Javier Ariza López¹ and Alan David Atkinson Gordo²

Abstract: This work presents an analysis of some standard methodologies for positional accuracy assessment of geographic data bases, taking into account aspects like the statistical formulation, the size of the control sample, the distribution and typology of the control elements, etc. Here we point out some weaknesses of the majority of standards: scarce formalism, inappropriate terminology for dealing with uncertainty, minimum recommended sample size, no assessment of the base hypothesis of the statistical model being applied, no information about the statistical behavior and reliability of the method, etc. The analysis developed here can serve as a starting point for the development of improved methodologies.

DOI: 10.1061/(ASCE)0733-9453(2008)134:2(45)

CE Database subject headings: Standards and codes; Quality control; Accuracy; Methodology; Surveys.

Introduction

The positional accuracy of cartographic products has always been of great importance. It is, together with logical consistency, the quality element of geographic information most extensively used by the national mapping agencies (NMAs), and also the more commonly evaluated quality element option (Jakobsson and Vauglin 2002). Positional accuracy is a matter of renewed interest because of the capabilities offered by the global positioning system (GPS) and the need of a greater spatial interoperability for supporting the spatial data infrastructures. Different positional behaviors of geographic data sets means the existence of an interproduct positional distortion and a barrier to interoperability (Church et al. 1998). This barrier is not only for the positional and geometric aspects, but also for thematic ones which are greatly affected by position (Carmel et al. 2006). For these reasons many NMAs are currently involved in the development of positional accuracy improvement programs (EuroSDR 2004).

In a geographic data base (GDB) the position of a real world entity is described with values in an appropriate coordinate system. Positional accuracy represents the nearness of those values to the entity’s “true” position in that system. The positional accuracy requirements for a GDB are directly related to its intended use(s). Positional accuracy is determined by means of a statistical evaluation of random and systematic errors (DOD 1990) and specified by means of the root-mean-square error (RMSE) or by the mean value of errors (μ) and their standard deviation (σ).

Comparison with an independent source of higher accuracy is the preferred method for assessing positional accuracy (ANSI 1998). Since positional accuracy is essential in cartographic production, all NMAs have used statistical methods for its control, which we call positional accuracy assessment methodologies (PAAMs). Many of them have been established as national or international standards and can be used for specifying spatial data products but also the resultant positional accuracy assessment compliance criteria. Standards should be taken into account when seeking an economic optimization of the quality of geographic information (Krek and Frank 1999): with a quality standard the producer provides the product according to the known specification and characteristics, as defined in the standard. This assures a certain level of reliability and certainty, allowing the acquirer to avoid excessive measuring of the quality and thus reducing the measuring cost and shortening the decision-making process.

The International Organization for Standardization (ISO) considers positional accuracy to be one of the quantitative quality elements of geographic information, as stated in its International Standard 19113 (ISO 2002), which is a general framework for describing and reporting the quality of geographic information. ISO also proposes a general quality evaluation methodology for geographic information in the International Standard 19114 (ISO 2003), which must be applied to the positional aspect of geographic information. But this International Standard is also a general framework, and so there is a need to clearly define aspects such as the computation of errors, sample size and schema, acceptance/rejection criteria, and so on. International Standard 19114 refers to other well-known ISO international standards which are applied to industrial control/acceptance processes that specify these aspects when dealing with attributes (ISO 2859-1) (ISO 1999) or variables (ISO 3951-1) (ISO 2005). We believe that the future of positional accuracy assessment must be resolved within ISO standards, but prior to this we need to know actual methods and their improvement possibilities in order to develop appropriate assessment methods for the positional aspect of geographic information.

In this work we analyze some standards that have been, or are...
being, used for positional accuracy assessment (planimetric or altimetric, but not bathymetric). So a general view is presented that covers recent decades and shows a clear evolution in methodologies. This implies a constant and continuous need for updating and accommodating these methodologies to different scenarios and realities (fitness for use), for example, the assessment methodologies recently developed for digital elevation models (DEMs) derived from light detection and ranging techniques (LIDAR), as considered by FEMA (2003). In chronological order, the standards analyzed in this study are as follows:

1. National map accuracy standard (NMAS) (USBB 1947): standard of very simple application and broadly used in the entire world from its publication date;
2. Engineering map accuracy standard (EMAS) (Committee on Cartographic Surveying 1983): robust statistical method developed by the ASCE, within its Committee on Cartographic Surveying, Surveying and Mapping Division;
3. Accuracy standards for large scale maps (ASLSM) (APR9 1990): method of the American Society of Photogrammetry and Remote Sensing broadly used for large-scale works (g.t. 1/20,000) by the civil sector and military administration (USACE 2002);
4. MIL-STD-600001 (DOD 1990): military standard which more than a norm, with specifications and control methods, is a guide for mathematical-statistical calculations;
5. National standard for spatial data accuracy (NSSDA) (FGDC 1998): nowadays the compulsory standard for all agencies producing GDBs. From its approval has reached a wide diffusion and acceptance. At the moment is under revision (FGDC 2003); and
6. STANAG 2215 (STANAG 2002): standard agreement from the North Atlantic Treatment Organization (NATO), a rigorous method and a well-structured document.

As can be observed, all of these standards come from the United States, with the exception of the one from NATO, an international organization where the United States also plays a predominant role. The initiatives coming from the United States play an important role because: (1) they are many; (2) they show a great dynamism and a clear orientation to practical mapping; (3) they also have great worldwide importance; and (4) of the abundance and accessibility of public and private sources that don’t exist in any other cases. Nevertheless, this work is also based on wider international research developed by means of inquiry and other indirect procedures (Atkinson 2005; Nero and Cintra 2005), as well as specific research within the scope of positional accuracy (Atkinson 2005; Nero 2005; Nogueira, 2003). These sources are not analyzed here, but are used in order to enrich our comments.

This document is organized under two main sections. The first develops the analysis and critique. It is a large section subdivided into many subsections dealing with aspects such as: sample size, control elements, reporting mechanism, etc. The second section presents the summarized conclusions of the work.

Critical Analysis

This section presents the analysis of the main characteristics of the PAAMs under study. For this reason the section is divided into a group of subsections, each one presenting the considerations of a specific aspect. This analysis is partially summarized and presented in Table 1, which shows a global vision of the different aspects compared. In this table the blank spaces indicate a negative or indefinite result.

Formal Aspects

These aspects refer to the presentation, order, and structure of the documents in which methodologies are described. The formal aspects are important because they give us an idea of the rigor of the development process. We can consider that formality is an appreciable external quality attribute of a standard. All over the world these documents adopt different forms: directives or laws as in France (2003), technical documents, and, less frequently, templates of standards. In our case a clear tendency of improvement over time can be observed, from the NMAS to the STANAG 2215 formulation. In general, few of the analyzed documents present as rigorous a formulation as expected for a standard, which could be defined, for instance, by the “Rules for the structure and drafting of international standards” (ISO 2004) or another similar document. Another formal aspect is the way in which measurements and uncertainty are referred to and expressed. The nomenclature and statistical formulation used are mainly not in accordance with the recommendations of the international vocabulary of metrology (ISO 1993), and with the guide to the expression of uncertainty in measurement (ISO 1995). Formal aspects are easy and inexpensive to adopt, and this would give alignment to other standards and profits coming from a more rigorous formulation. Table 1 includes a column with our general evaluation of this topic.

Planimetry and/or Altimetry

The differential behavior of these two components, as well as the possibility that a product could be solely planimetric or altimetric, means that all PAAMs must consider, in an independent way, the evaluation of both planimetry and altimetry. This independent evaluation means that some products are only evaluated for planimetry or altimetry, or that one can demand different accuracy levels for each component in the same product.

The PAAMs show two basic ways of working with the planimetric components: (1) independent analysis for each component (X and Y separated); and (2) joined or planimetric analysis (X and Y together). This actually implies the assumption of two basic hypotheses that will condition the statistical analysis: linear errors for each component (X and Y separated), or a circular or global error (X and Y together). The altimetric case is clearly linear. Most PAAMs consider the altimetry to be independent of the planimetry, although some methods allow us to play with the planimetric position in order to adjust the altimetric one. This is not considered in later proposals. The boom of the production and use of DEMs has generated a special interest in the control of these products, but not all the analyzed PAAMs include a specific reference to them or to their control. Here there are some contributions of special interest like USGS (1997) which include three levels of accuracy, and the guides of the FEMA (FEMA 2003) and of the National Digital Elevation Program (NDEP 2004), which assume the NSSDA formulation with some particularities.

Isolated or Flow Production Control

Today in cartography there exists more and more massive production processes, making it closer to industrial processes. Nevertheless PAAMs do not show the refinement level reached in control and acceptance processes of the industrial environment. Because there are many differences among the industrial and cartographic processes, the wider experience of the first should be applied, as much as possible, to the second. We find that the industrial pro-
Table 1. Summary of Reviewed Aspects of Analyzed Positional Accuracy Assessment Methodologies

<table>
<thead>
<tr>
<th>Name: name of the standard</th>
<th>NMAS (1a)</th>
<th>EMAS (2a)</th>
<th>ASLMS (3a)</th>
<th>MILSTD 60001 (4a)</th>
<th>NSSDA (5a)</th>
<th>STANAG (6a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin: organization</td>
<td>USBB (1b)</td>
<td>ASCE (2b)</td>
<td>ASPRS (3b)</td>
<td>DOD (4b)</td>
<td>FGDC (5b)</td>
<td>NATO (6b)</td>
</tr>
<tr>
<td>Formal aspects: adherence of the document to a conventional structure of a standard. The following levels are assigned: null=0; very scarce=1; scarce=2; medium=3; appropriate=4; very complete=5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Sheet, lot of sheets or a series: indication of the type of element to which the control is applied</td>
<td>Map</td>
<td>Area</td>
<td>Sheet</td>
<td>Area</td>
<td>Area</td>
<td>Area</td>
</tr>
<tr>
<td>Isolated/flow: whether the control is supposed on isolated elements (e.g., sheet or lot) or in a continuous process (e.g., lot by lot process)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scales: if there is a distinction or recommendation based on scales</td>
<td>All</td>
<td>&gt;20,000</td>
<td>&gt;20,000</td>
<td>All</td>
<td>All</td>
<td>&lt;25,000</td>
</tr>
<tr>
<td>Control elements: if there is an explicit indication to use point elements as control elements</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Sample size: recommended minimum size for the control sample</td>
<td>—</td>
<td>20</td>
<td>20</td>
<td>—</td>
<td>20</td>
<td>167</td>
</tr>
<tr>
<td>Typology of control points: if there is an indication of the typology of the control elements</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Spatial distribution: if there is a guide about the appropriate spatial distribution of the control sample</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Sub-regions: if there is a proposal for using subregions in the event of diverse accuracies</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
</tr>
<tr>
<td>Precision of the sampling: suggested precision needed for the control sample</td>
<td>—</td>
<td>3x</td>
<td>3x</td>
<td>—</td>
<td>3x</td>
<td>5x</td>
</tr>
<tr>
<td>Absolute accuracy: consideration of this quality subelement</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Relative accuracy: consideration of this quality subelement</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Basic hypothesis testing: if there is an indication to perform these tests</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Implicit normality: if the normality (Gaussian distribution) is assumed</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Outliers: if there is an indication concerning their elimination or how to deal with them</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bias: if there is an indication of how to deal with it</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
</tr>
<tr>
<td>RMSE: if the root mean squared error is the proposed uncertainty measure</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mean and deviation: if the mean and standard deviation are the proposed measures</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
</tr>
<tr>
<td>Planimetry: if the application to the horizontal component is considered</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Altimetry: if the application to the vertical component is considered</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3D: if the application to 3D cases is considered</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
</tr>
<tr>
<td>XY displacement for Z: if a planimetric shifting for altimetric adjusting is possible</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Circular/linear: if planimetric components (XY) are analyzed together (circular) or independently (linear)</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Categories: if the use of accuracy categories is proposed</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>DEM: if there is an explicit consideration of the application of the method to digital elevation models</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>Expression of results: type of report to indicate results (pass/fail=1; classification=2, value=3)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>General information: an indication of the information included in the specification and in the presence of recommendations, examples, calculation examples, etc. The following scale is used: null=0; very scarce=1; scarce=2; medium=3; appropriate=4; very complete=5</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Probability of the result: the probability level considered for the result of the assessment</td>
<td>90</td>
<td>≥90</td>
<td>≤90</td>
<td>90</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>Uncertainty of the method: if there is an indication of the uncertainty of the control methodology itself</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Yes (10%)</td>
</tr>
<tr>
<td>Global valuation: a global score evaluation following the scale: very bad (1); bad (2); functional (3); good (4); and very good (5), assigned by the authors of the work as a subjective global evaluation of the method</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

Note: Blank spaces indicate that the standard analyzed did not deal with this aspect. 1a=National Map Accuracy Standard, 1b=U.S. Bureau of the Budget, 2a=Engineering Map Accuracy Standard; 2b=American Society of Civil Engineers; 3a=accuracy standard for large scale maps; 3b=American Society of Photogrammetry and Remote Sensing; 4a=military standard; 4b=U.S. Department of Defense; 5a=National Standard for Spatial Data Accuracy; 5b=Federal Geographic Data Committee; 6a=standard agreement; 6b=North Atlantic Treatment Organization.
processes distinguish whether the acceptance is carried out on a lot by lot process or in an isolated one. Continuous measurement is demonstrated to be more informative than isolated data when more information is provided for the same sample size or a smaller sample size is required for the same information. Continuous measurements provide distinct advantages when the conformance probability is high. There are also several inspection levels (more or less rigorous), simple and double sampling processes, and so on, all in order to obtain better economies within an acceptable risks assumption. So this is a great issue whose importance comes from obtaining better controls and better economies in the positional control process. Classical cartographic positional controls are not as competitive as industrial ones and here there are clear possibilities for innovation.

For PAAMs the general circumstance is to indicate nothing in relation to all the above-mentioned questions that we consider critical. A special case is the rule presented in the manual EM-1110-1-1000 (USACE 2002): When a series of sheets are involved in a mapping project, the existence of errors (i.e., map test failure) on any individual sheet will constitute prima facie evidence of deficiencies throughout the project (i.e., all the sheets are assumed to have similar deficiencies) and field map testing will cease. PAAMs in general do not specify if the control should be applied to every one sheet of a series, to a sample of them, or to a relaxation factor for the proposed categories. There are also several inspection levels (e.g., NMAS as a function of the scale (0.5 mm for scales <1/20,000 and 0.85 mm for scales >1/20,000), Nero and Cintra (2005) present a review of such values for many national standards (Austria, Belgium, Canada, Colombia, France, etc.).

In the digital era we use the concept of resolution to refer to the precision of geographic data themselves, not in a portrayal, being expressed in terrain units. This evolution is very clear in the documents analyzed. Here the possibilities are many and hence there exist two basic options adopted by PAAMs: reporting of the estimated value through the application of the assessment methodology, or a classification scheme. Classification schemes generally use a limited set of precisions (e.g., ASPRS (2001) or GC (1996)), but this option can be expanded for any result value when appropriately tested (France 2003). In other proposals like the ASLSM (ASPRS 1990), categories are defined as a set of limiting errors corresponding to a set of scales. In this way products are defined as being of first, second, and third classes, second-class products being two times less precise than first-class products, and third-class products being three times less precise than first-class products. In the ASPRS case limit errors are a direct translation of the visual perception limit to the scale, considering the possibility of a relaxation factor for the proposed categories. There are also some specific classifications for DEM like that of the USGS (1997). The use of categories is interesting for the cartographic sector, for example one of the proposals for the revision of NSSDA is to adopt a classification scheme (Tilley 2002).

Sample Size of Control Points

Sample size is one of the key aspects of any estimation methodology based on sampling and this is due to the fact that, ceteris paribus, the greater the sample size the greater the precision of the estimation. Moreover, sampling in the field is always expensive, so there is a need for balancing these two opposing forces. Normality is usually the assumed hypothesis for positional error behavior. From this hypothesis one can derive the expression for the sample size $n$ as a function of the population size $N$, the precision of the sampling $e$, the population’s variance $\sigma^2$, and the significance level $\alpha$, of the $t$ statistic. So it is actually a closed matter from the statistical point of view. However PAAMs, and other examples found, present diverse recommendations for sampling sizes: $n=20$ (NMAS, EMAS, NSSDA), $n=50$ (Newby 1992; NJUG 1988), $n=167$ (STANAG 2215). For us this is one of the most critical aspects of the majority of PAAMs analyzed, since the recommended minimum sample size is usually very low. In coincidence with the works of Li (1991) and the prescriptions of the STANAG 2215, our opinion (Ariza and Atkinson 2008; Ariza et al. 2008) is that sizes greater than 100 are needed in order to limit the variability of the estimation process to an acceptable level of reliability.

Nevertheless, when assessing a product we must consider an economic function linking the sampling costs, the total investment in the product, and the reliability of the product in the form of costs of bad decisions (risks) derived from data of poorer quality than expected. This kind of analysis is already applied as presented in the works of De Bruin et al. (2001) and Hunter and De Bruin (2006).

Accuracy Levels

Here we refer to the establishment of a level of quality (standard accuracy or standard precision) for cartographic products. This is a very important issue because it is one of the most visible aspects of the standards. The use of different standards with different standard accuracies generates, in some way, a barrier to interoperability because the comparison is not immediate.

Traditionally, standard accuracy has been considered in map units because interpretative and cartometric uses take place on maps. As one deduces from Nero and Cintra (2005), there is no agreement between standards that use this option. The stated values for accuracy are related to the probability level that is assigned to the maximum probable error, and most cases assume 90% probability. Some PAAMs set this limit to 0.2 or 0.25 mm, while in other cases this value is increased to 0.28 mm when considering 95% probability. Others distinguish several levels (e.g., NMAS) as a function of the scale (0.5 mm for scales <1/20,000 and 0.85 mm for scales >1/20,000). Nero and Cintra (2005) present a review of such values for many national standards (Austria, Belgium, Canada, Colombia, France, etc.).

Spatial Distribution of Sample of Control Points

The spatial distribution of the sample can condition the validity of a statistical sampling assessment. A bad spatial distribution affects the representativeness of the sample. This means that the sample does not capture adequately the structure of the population being sampled, resulting in an erroneous estimation. A few PAAMs
present explicit criteria for a suitable spatial distribution of the control elements, although in a brief way. In some cases the need for an agreement between the producer and the user (USACE 2002) is indicated, which is also common practice in other countries (France 2003).

If it is supposed that the area to be controlled is homogeneous regarding the error behavior and the interest of the features to map, it can be considered that a random and well-spaced distribution is the most appropriate. This is the recommendation given by several PAAMs.

According to the above, when the GDB being controlled covers a rectangular area, an ideal distribution of test points allows for at least 20% of the points to be located in each quadrant. Test points should be spaced at intervals of at least 10% of the diagonal distance across the rectangular area. If elements of interest exist (e.g., an infrastructure), the sampling should be intensified around them. Nevertheless the previous is only a recommendation that is also subjected to agreement between the client and the producer.

For the control of DEM, in some cases it is specified that control points must be selected in terrain zones that are flat or uniformly sloped, where the uniform slope must not exceed 20% (NDEP 2004). Additional conditions are stated by FEMA (2003): control points should be at a certain distance from terrain structuring elements and zones where control points are taken should have a size determined in function of the step of mesh of the DEM. Also, in the case of using data obtained from LIDAR it is required that the control points be distributed among the different categories of the vegetable cover. In other cases (USGS 1997) sample size rises up to 28 points, which is composed of a single test using 20 interior points and eight edge points. Edge points are those which are located along, at, or near the borders of the area and are deemed to be useful for evaluating the accuracy of the edges of the DEM.

In general, none of the documents state how to realize the sample by a random process. The best way to generate a random sample should be by a computerized method such as that used in NJUG (1988). So the introduction of bias in the sample distribution by the operator, who designs the control, is avoided. Undoubtedly this automated selection should be revised later on.

As one deduces from the previous, in no case is any metric, measure, or index mentioned, used, or proposed to objectify this circumstance in order to assess, or evaluate, the goodness of the spatial distribution in a more quantitative and objective way. Also there is no requirement of a graphical reporting mechanism to show the distribution of control points.

**Statistical Aspects**

All the PAAMs are statistically based processes with an underlying statistical base model. For this reason special care should be taken for everything (formulation, hypothesis, etc.), that can affect the accomplishment of the base model and thus the validity of the derived estimations. PAAMs don’t share the same formulations, thus the results and obtained reliabilities of applying different PAAMs to the same GDB can be different. In other sections we have presented aspects like sample size, spatial distribution, etc. that have statistical importance. However, and without the aim of being exhaustive, here we will center our interest on the following: basic hypothesis, precision of the sampling surveys, estimation of the uncertainty, treatment of bias and atypical values, and finally, the information given on the method itself.

**Basic Hypothesis**

Statistical methods are usually based on some previous hypotheses that must be satisfied in order to obtain valid conclusions. Therefore, it is important to verify this fulfillment in order to know the actual statistical relevance of the results. These hypotheses are not clearly enunciated in some PAAMs, and even less in the way to contrast them. An example of the previous is the generally adopted normality of errors (Gaussian model) as the base model (DOD 1990; Li 1991; Leung 1998; Shi and Liu 2000). So randomness and normality are supposed for applying PAAMs. In most cases these hypotheses will be satisfied, but there are cases in which they are not. The last are interesting cases to be detected in order to analyze the causes and even to use alternative models or methodologies. We consider that the cost of carrying out this kind of statistical tests is now really low with the computers and software tools available. So that from the perspective of a NMA there is an opportunity, that should not be lost, to know better and to improve processes and control methods.

**Precision of Sampling Surveys**

Precision is another key aspect because it is directly related to the cost of the field survey. Greater precision means greater cost due to more expensive equipment, more time-consuming methodology, etc. But precision is also important from a statistical point of view, in order to minimize the impact of bias and deviations in an estimation processes, and assure the representativeness of the estimation. In order to obtain this representativeness, and following the Nyquist–Shannon sampling theorem, the control sampling needs a minimum precision of 2× that of controlled elements. The preferred test for positional accuracy is a comparison to an independent source of higher accuracy (ANSI 1998), in most cases by means of field surveys. Here the question is what higher accuracy of control works means. Proposals are diverse and they go from 2× (France 2003) to 5× (STANAG 2215) the accuracy of the GDB being controlled. In our case, the majority of standards indicate a control with an accuracy 3× better than the product, as suggested by Merchant (1982). For example, let us consider how a relative bias of 1/3 σ affects the proportion of a control sample included within an interval of ±3σ of the controlled population when accuracy of sample and population are of the same order (σ_control ≈ σ_population), or the control sample is 3× accurate (σ_control ≈ σ_population/3). For the first case the common proportion is about 65.36%, and 97.72% for the second. This implies a much better estimation derived from the second case because the distribution of the control sample is mainly included within the distribution of the controlled population. It is interesting to note that in some proposals (e.g., France 2003) the accuracy of control surveys is introduced in the model as a security index. This gives more versatility without loss of reliability.

**Estimate of Level of Uncertainty**

Basically two options exist in the estimate of the positional uncertainty level for a GDB. These are as follows:

1. **Root mean squared error**: the presence of the uncertainty is summarized under this unique index; and
2. **Mean and deviation**: in this case, the behavior is divided into two indexes: (1) a central measurement that quantifies the mean displacement (bias); and (2) a dispersion measure called deviation, but also standard deviation or standard error.
Both alternatives are used by PAAMs, but the second option gives more information than the first because of the two indexes. This is an aspect of great importance for the case of evaluating positional interoperability, since the bias can be corrected with transformations, but deviations (variability) is an aspect with a more difficult treatment. Advanced PAAMs also prefer the second since it distinguishes possible causes that can affect noncompliance with the standard, and it allows us to detect if it is necessary to look for an attributable cause of the bias or the increment of the variability.

**Bias**

Bias is a typical effect present in sampling with important consequences. A biased sample causes problems because any statistic computed from that sample has the potential to be erroneous, leading to an over- or underrepresentation of its corresponding parameter in the population. In practice, all samples are biased because it is practically impossible to ensure a perfect sample. The presence of bias in an estimation supposes undesirable behavior that can be attributed to a concrete cause that should be discovered and removed. In some PAAMs (e.g., NMAS) this behavior is not detectable and nothing is indicated. Neither is this possible with the NSSDA, but the standard itself indicates that previous to its application all biases will have been eliminated. In the EMAS there is a specific statistical test to detect the presence of bias. What is not clear in any case is the treatment of the bias, that is to say, if the references to the elimination refer to the causes or to an analytic treatment for it. This aspect is very important for the EMAS, because it is very different to consider the test carried out on bias as previous to the analysis or as a part of the standard itself. In other cases, like STANAG 2215, statistical formulation allows us to consider bias in an explicit way in the analysis and determination of uncertainty.

**Atypical Values**

The atypical values, or outliers, of a sample are those that are excessively large or small, but which are not mistakes or coarse errors. Although they have little occurrence probability, these values can appear and for this reason should be considered as belonging to the distribution function of the phenomenon or characteristic. The problem here is important because the inclusion of atypical values in a traditional statistical analysis strongly conditions the derived estimation results by introducing bias and greater deviations. Their inclusion biases the sample by overrepresenting these values in relation to others in the population. To overscope this situation robust statistics can be used (Atkinson 2005). Nevertheless PAAMs do not commonly give any indication of their statistical treatment. Here the situation is interesting, on the one hand because their inclusion in the computations would greatly affect the estimation of values and, on the other, because their elimination is to close our eyes to an inconvenient reality. Hence the common solution usually consists of a careful elimination process, not taking the outliers into account in the numeric analysis but including them in the reports and giving a warning of this circumstance. Examples of this can be in found in MPLMIC (1999) and ASPRS (2001). An example of a method for dealing with atypical values is described in France (2003): the standard base model considers probability thresholds on the order of 1 and 0.01% for atypical cases.

**Information on Method**

In general PAAMs are formulated in a very simple way and there is no information given about their statistical behavior or other relevant statistical aspects. Observe the following examples:

1. Reliability of the method: all PAAMs are formulated without giving information about their reliability. The case of STANAG 2215 is different and this document repeats several times that to achieve the expected reliability a sample of 167 points is required, otherwise alternative formulations should be applied. However, in other methodologies like NSSDA, which has been demonstrated to possess great variability (Ariza and Atkinson 2008) when using small sample sizes (as recommended by the standard itself), nothing is indicated about the relation between sample size and expected reliability; and

2. Peculiarities of the method: the EMAS is formulated for the horizontal case as a battery of four statistical tests that must be passed altogether to consider that a GDB satisfies the standard. The application of four tests, each one with a significance level of $\alpha$ and under the hypothesis of independence between them, results in a global significance level of $4\alpha$. As indicated by Ariza et al. (2008), the last supposes that Type I error increases greatly and thus Bonferroni corrections should be applied in order to obtain 5% of global significance. Otherwise, rejection of good positional products occurs.

We consider it necessary that PAAS standards should be accompanied, either by annexes in the same document, explanatory guides, or by any other possible option, with information addressing hypothesis, limitations, critical aspects, reliability, convenience, and so on, in order to improve their adequate interpretation and use, avoiding all possible errors derived from the lack of relevant information.

**Recommendations and Examples**

The vocation of standards is their use. Thus their knowledge and popularization are indispensable to the agents of our sector. Since the practical and rigorous application of a standard can be difficult when starting from the text of the standard, its use and attractiveness can be questioned. We consider that they should be accompanied by annexes or other types of documents in order to explain in detail their philosophy and their application by means of examples. This has been understood by standardizations bodies, for instance ISO/TC 211, which develops the family of standards ISO 19100, but also by other sectorial bodies like the Federal Geographic Data Committee. In this way a guide should be available with pertinent recommendations for applying the standard.

The situation described in the previous paragraph is increasingly common. In this way, we can find that for the oldest PAAM standards no guide exists, while for recent ones annexes to the standard are included with examples (e.g., NSSDA) but also that explanatory manuals have been developed, for instance the *Positional accuracy handbook* (MPLMIC 1999). In some cases programs or computer codes are included in the same document (e.g., STANAG 2215) or are available on the Internet, as is the case of NSSDA (LMIC 2006).

**Absolute or Relative Accuracy**

Until recently relative accuracy has been more important than absolute (Rönsdorf 2004). Due to surveying limitations one can assume that the absolute accuracy of the traditional GDB is always worse than the relative accuracy. Nevertheless, thanks to GPS absolute perspective has nowadays gained a greater importance. All the analyzed PAAMs are oriented towards absolute
accuracy control, and only one of them makes explicit reference to the control of relative accuracy. Undoubtedly the absolute perspective is primordial from the point of view of a NMA since it determines how a specific GDB fits into the national positional framework where all cartographic products are put together. Nevertheless, for some engineering applications the interest is more on the relative positional behavior among elements, which can be assessed by means of an estimation of the relative positional accuracy. Only one standard considers this circumstance leaving the rest of them, from our point of view, without covering an important field of application.

We consider that given the current available computational capacity, once we have a sample of control points for analyzing absolute accuracy it is not problematic to derive an estimation of the relative. Having absolute and relative estimations is not redundant and can help, in some cases, to detect the existence of problems with the application of the positional framework.

The above-mentioned is evidenced in some modern standards like STANAG 2215, France (2003), and ICSM (2004), where the use of both measures is proposed. Also ISO 19138 (ISO 2006) standard includes measures to apply in the evaluation of relative positional accuracy. We understand that in the immediate future a wider use of this perspective will be common.

Control Elements

None of the standards analyzed mention classes of elements other than points, for instance, curves, surfaces, or volumes, which are important elements of digital GBs following a vector data model. Of course points are the base elements for the others, but with these more complex elements there is a functional behavior that cannot be captured well by points because different criteria for matching primitives will affect positional accuracy estimations. So there is a certain degree of discordance between control elements (only points) and controlled elements (all kind of elements of the GDB). Until now almost all the PAAMs are based on the use of points as control elements, and practically all of them include the expression “well-defined points” or “elements easily identifiable” to refer to features that can be sharply identified as discrete points. This means using a group of homologous points, both in the GDB under control and in the field survey, that don’t offer doubts in their identification, and that present a geometry that evidences their position in a clear way. There are some types of elements which allow an unequivocal identification, for example a corner or a crossing, while they can also present differences from the point of view of their positional behavior. The types of elements used should be a faithful reflection of all the existing elements in the GDB, and not only of a limited and very specific subset that could result in an optimistic estimation and also introduce some degree of dependence from an external observer. The exploitation of a GDB is commonly carried out using many typologies of elements and not only the limited subset used for the control. In this way it can be imagined that the control elements should be conformned as a stratified sample of what exists in the GDB, otherwise the representativeness of the sample can be questioned: a biased sample resulting in a biased estimation. Thus, if we use specific typologies for control elements the estimation is only strictly valid for these typologies. The determination of the typologies is an important issue for producers and users of the GDB, and this must be considered in light of the uses of the GDB and its particular elements of interest (fitness for use). In this case a clear stratification of the sample is needed but also a weighing of the results or a multireporting mechanism.

Some modern GDB specifications propose different positional accuracy levels for each element typology, justifying our previous comments.

The control elements must always be independent and this is specified in some cases (STANAG 2215): the control sample should be independent of all elements used in photogrammetric block adjustments and triangulations. Other aspects that have a great influence on positional accuracy assessment are the cartographic generalization and symbolization processes. In this way it should be left clear, as in some product specifications, the hierarchy applied in the generalization processes, since this affects possible displacements of the elements of the GDB.

From our point of view, the standards analyzed here follow a classic perspective. There are some cases (France 2003) that consider the use of other kinds of points (not only well defined) and elements (lines and surfaces). There are also some NMAs (e.g., Instituto Geográfico Nacional of Spain) that evaluate and report positional accuracies of different linear elements (Lucas and Rodríguez 2004) and others that specify different accuracy levels for different kinds of linear features (NLSF 1995). These new possibilities should be taken into account, for instance the use of linear elements (Goodchild and Hunter 1997; Teveite and Langaas 1999), because of the increasing importance of the use of navigation systems, and also for supporting conflation processes within interoperability mechanisms.

When assessing the positional accuracy of DEMs there are no well-defined points; they don’t exist and it would be very fortunate to have a ground control point corresponding to a node of the mesh. In this case, with the whole statistical base that supports PAAMs being valid, it should be specified how to carry out these controls attending to this particularity, which brings with it the need to interpolate. In some references (USACE 2002; FEMA 2003) the use of profiles for this case is also considered, for example, a profile for each sheet or three for a photogrammetric model with a length greater or similar to 127 mm (at map scale) and that cuts a minimum of ten contours. In the case of DEMs the use of the visual checks (USGS 1997; Rodríguez and Lumbrras 1997) is also indicated as a method for controlling this type of product.

Report of Results of Evaluation

Supposing that the applied methodology assures a standard reliability for the positional accuracy assessment the reporting of the evaluation result or results carried out by the PAAMs, can basically follow one of these options:

1. Pass/fail: it is indicated if the GDB whose positional accuracy is assessed has or has not passed the level of exigency stated. It is a very common and easy form of expressing the result of an assessment, but fails in the sense that there is a loss of information because no numeric value is provided. This loss of information is very important because of the imprecision it entails (the degree of pass/fail), and therefore we do not know the effort required in order to produce a more precise GDB (Giordano and Veregin 1994). An example is the NMAS standard: “This map complies with National Map Accuracy Standard.”

2. Categories: a special pass/fail case where a multichoice is possible. The accuracy assessed is translated into a category or class of accuracy by means of a set of pass/fail filters. There should be a standard in which a classification system is proposed. This is also an easy method that has the advantage of considering different well-defined predefined levels of ac-
accuracies. This clarifies the situation, also giving enough flexibility if classes have been chosen in an appropriate way. The lack of classes is one of the critics of the NSSDA (Tilley 2002), and an example of this option is the ASPRS standard: “This map was checked and found to meet the ASPRS standard for Class 1 map accuracy,” and

3. Values: accuracy is given by means of an uncertainty index or indexes which characterize the positional aspect of a GDB. From the producer’s side we have more information and its quantitative character allows a better statistical analysis of the production process. From the user’s side this option gives a better understanding of actual possibilities of the use of the GDB, but it is also left to the user to consider whether the accuracy is suitable or not. The NSSDA follows this approach: “Tested meters horizontal/vertical accuracy at 95% confidence level.”

Each standard adheres to one of the previous options. When necessary two reports are given: one for the horizontal case and another for the vertical.

In our opinion, the most advantageous way of reporting is to proceed by means of a hybrid form: giving a classification and a numeric value. Although seemingly redundant, the combined use of a value and a class means: (1) we do not lose information (the numerical value) that can be of interest to some specialized users; and (2) we have an easy way to refer to and standardize accuracy levels. We considered that the report of the result of the assessment cannot be limited to the estimation of the accuracy of the GDB. The report should also include, in a standardized way, information about the other tests (e.g., randomness, normality) and processes (e.g., bias treatment) carried out for the assessment.

As deduced from above, the reporting mechanisms are always textual but could be improved by means of graphic expressions (Mackaness et al. 1994; Matos and Gonçalvez 1998). This is indispensable when two homogeneous subregions have been detected in the product and different accuracy reporting is included for each one instead of using the worst accuracy for the complete GDB. Some standards consider this situation: MIL-STD-6000001, STANAG 2215, and IHO (1998).

Conclusions

The positional aspect is a primordial component of the quality of a GDB. Nowadays NMAs are very concerned with positional accuracy, with many of them developing positional accuracy improvement programs (EuroSDR 2004).

The analysis carried out in this work has been focused on numerous aspects: formalism, sample size, estimation method, bias and outliers management, and so on. The study reveals the existence of great differences within the standards considered but also great coincidences, mainly in weak aspects such as recommended minimum sample size, treatment of bias and outliers, statistical base model compliance, etc. In conjunction with a general overview, this analysis also provides ideas that will be of use to mapping administrations if they are considering reviewing positional assessment standards. In a brief way some of the main conclusions about analyzed PAAMs are as follows:

1. The majority of cases, they don’t present great formalism in their definition and don’t use the standardized terminology when referring to uncertainty;
2. They are statistic-based processes but differ a lot in their estimation methods and present disparity in the minimum recommended size for the control sample;
3. Usually proofs are not required for any of the previous basic hypotheses of the statistical base model, and also no information is given about the statistical behavior and reliability of the method itself;
4. Nothing is usually indicated, in an explicit way, in relation to the treatment for bias and outliers;
5. Well-defined points are the base elements for the controls, but positional accuracy is not only a matter of these features. All the features included in a GDB should have a positional accuracy evaluation. In the case of dealing with DEMs, where there are no identifiable points, specific methodologies are needed;
6. They don’t usually incorporate the relative positional accuracy perspective;
7. The results can be expressed in different textual ways (pass/fail, category, value) but graphic expression is considered only a few times; and
8. Manuals or explanatory documents don’t usually exist.

We consider that the majority of PAAMs, although operative methods, have problems of nondefinition, statistical character, lack of explanatory information, etc. All of these can lead to bad interpretations of the results and so false expectations. We understand that the cost of some of the improvements which we have indicated here is actually low thanks to the computational and reporting possibilities available at the present time. Here is an opportunity that should not be rejected for better understanding and for improving the processes.

The cartographic and normative institutions should join forces in the development of specific positional accuracy assessment standards with a high definition level. This development should contemplate different sampling options, risk, and protection of users and producers, etc. It should also include, alongside its prescriptive part, a descriptive one with practical examples, allowing the elimination of doubts and providing an example of good working procedures. We must bear in mind that the absence of prescription, or guidance, in a standard in relation to the most important aspects of its application can cause a contradictory effect to the idea of standardization itself. A lack of definition of a given standard gives rise to different uses of it. However, these applications can be interpreted as being similar.

Now there is also available the ISO 19100 family framework, specifically the ISO 19113 (ISO 2002), 19114 (ISO 2003), and 19138 (ISO 2006) standards, which gives versatile but also detailed possibilities that should be used for the development of measures, evaluation, and reporting methods.

Acknowledgments

This work has been partially funded by the National Ministry of Sciences and Technology of the Kingdom of Spain under Grant No. BIA2003-02234. The writers also acknowledge the Regional Government of Andalusia (Spain) (Department of Education, Science and Technology) for financial support from 1997 to their research group (PAI-TEP-164).

References

American Society of Photogrammetry and Remote Sensing (ASPRS).


Inter-Governmental Committee on Surveying and Mapping (ICSM). (2004). *Standards and practices for control surveys (SP1) V1.6*, Canberra, Australia.


Mackaness, W., Beard, M. K., and Buttenfield, B. P. (1994). *Selected annotated bibliography on visualization of the quality of spatial information, National Centre for Geographic Information and Analysis, Santa Barbara, Calif.*


mento: Visão geral das normas de diversos países.” Proc., 22nd Congresso Brasileiro de Cartografia, Sociedade Brasileira do Cartografia, Macaé, Brazil, 1–16.