

Mapping Matters

Your Questions Answered

By Qassim A. Abdullah, Ph.D., PLS, CP**

The layman's perspective on technical theory and practical applications of mapping and GIS

Question 1: Sometimes we receive requests for corrective actions on products that we delivered to clients despite the fact that the product is meeting the promised accuracy expressed in Root Mean Squares Error (RMSE). After investigating the problem, most of the time we find that clients are pointing to a few localized and sporadic occurrences where the data deviates from the promised RMSE. In most of the cases, clients expect perfect data, meaning that no part of the data may deviate from the required RMSE accuracy. What would be the best way to solve these kinds of situations without extensive production and editing resources, while still satisfying our client's expectations?

*Evgenia Brodyagina
Frederick, Maryland, USA*

Dr. Abdullah: The question brings forward common concerns among geospatial data providers whether it is lidar data, GIS vector database, or ortho-rectified maps. Many end-users judge the final delivered products with a standard that may differ from the standard that they agreed upon with the data providers. For example, if a user requested a lidar data set with an accuracy to meet class 1 of ASPRS standards for 0.3 m (or 1 ft) contours, the data provider in this case has to guarantee that the delivered product shall have a vertical accuracy of 10 cm as RMSE. According to the concern raised in this question, the user may reject the data if he or she experienced a localized occurrence where the data did not meet the 10 cm criteria. This is a legitimate concern and one that can be addressed by examining the end-user's accuracy requirement more carefully.

"When we say an RMSE of 10 cm, it refers to an acceptance criterion that requires only about 68% of the data to meet the 10 cm and the remaining 32% of the data could have an error of up to 2.5 to 3 times the RMSE (or 25 to 30 cm in this instance)."

When the user awards a contract to the data provider under the accuracy requirements we mentioned earlier, it needs to be understood that by requesting an RMSE of 10 cm, the end-user accepts the statistical reality of the specified vertical accuracy. When we say an RMSE of 10 cm, it refers to an acceptance criterion that requires only about 68% of the data to meet the 10 cm and the remaining 32% of the data could have an error of up to 2.5 to 3 times the RMSE (or 25 to 30 cm in this instance). On the other hand, if the user wanted a guarantee that most of the data meets the 10 cm, then he or she needed to ask for different statistical criteria, such as 95%. The problem with the latter alternative of 95% is reflected in the cost of achieving such tight accuracy. For the data set to meet 10 cm vertical accuracy at 95%, the RMSE should be around 5.1 cm.

I would like to emphasize that the criteria of RMSE is a sound statistical approach and it is widely employed by surveyors and engineers. The issue is not in the statistical approach we use but in the proper understanding of these approaches. The vertical accuracy statement of 10 cm as RMSE is another way of saying that about 68% of the data should be within 10 cm and it is the same as saying 95% of the data should be within 19.6 cm. If the user understood the previous interpretation of the RMSE measure, then he or she would not be concerned when a small portion (less than 32%) of the data deviated by more than 10 cm. Education is crucial here and it is the responsi-

bility of the guardian of the geospatial data accuracy standard(s) to educate the public on the intent of the standard and all the statistical concepts behind it. However, it is also the responsibility of the user and data providers to educate themselves on understanding the standard and how it applies to their data. Without proper understanding and education, the cost of geospatial data will be inflated as data providers strive to achieve perfection in the product they provide in order to meet the user exaggerated expectations.

Similar problems are also common among users of ortho photo maps because most users expect their data set to be perfect to the degree that they will not accept even a small shift of a few pixels between two tiles in the dataset, despite the fact that the data set meets all the routine accuracy testing as set by the National Standard for Spatial Data Accuracy (NSSDA). The lack of clear and detailed guidelines on how to test and accept different geospatial data types often causes confusion. None of the popular map accuracy standards or guidelines, for example, address the quality of seamlines and the shift that may occur between data from adjacent flight lines whether it is imagery or lidar. Many users expect the data set to be seamless despite the fact that the data meet all geometrical testing criteria called for by the different map standards.

The common understanding among the community of geo-spatial data users for the word "seamless" is a zero tolerance or no tolerance for any shift between tiles or flight lines even if it is only a pixel or two for an ortho map and a few centimeters if it is elevation data. Such high quality requirements force the data providers to Photoshop the data so adjacent ortho tiles perfectly fit each other or to apply some kind of polynomial fitting if it is elevation data. Such enforcement, while it results in aesthetically appealing data, sometimes causes unwanted warp in the data rendering it less accurate.

"None of the popular map accuracy standards or guidelines, for example, address the quality of seamlines and the shift that may occur between data from adjacent flight lines whether it is imagery or Lidar."

Another example on the weakness of the testing procedure of the NSSDA standard is the check points requirement. The NSSDA standard calls for 20 well-defined, high-fidelity points in order to perform accuracy verification. This requirement often causes confusion among data users and providers alike. For large areas, such as an entire county or state, the project is usually divided into multiple

continued on page 206

production blocks and users are divided on whether the 20 points requirement is for each block or for the entire state. Even if we assume that the points are for the entire state or county and the data set meets the accuracy criteria on these 20 points, none of the other geometric anomalies that may occur in areas outside of the 20 locations of the ground control check points are identified or addressed in any standard. The seamline issue that I discussed above falls in this category, especially if it is located far from the 20 check point locations. Furthermore, asking for seamless data may have a negative impact on the geometrical accuracy, as I explained earlier, and it further increases the project's budget.

“None of the map standards address the issue or provide guidelines on how to deal with situations where the end user finds out that 6% of the breaklines in one tile are in disagreement with the lidar elevation by more than the RMSE value.”

Most of the time, in a lidar dataset, a minor stepping effect of 5 cm or less in a stretch of road that is only a few hundred meters in length does not affect the overall quality of the data. However, it may cause a rejection of the delivery by the users despite the fact that the data set meets the accuracy requirements of RMSE = 10 cm when checked against ground check points. Similar concerns arise when users evaluate lidar-based digital terrain model data that con-

tains photogrammetric break lines and lidar points. None of the map standards address the issue or provide guidelines on how to deal with situations where the end user finds out that 6% of the breaklines in one tile are in disagreement with the lidar elevation by more than the RMSE value. For acceptable non-seamless data that meet the NSSDA accuracy check, a standard needs to be developed to clearly provide guidelines on details of testing and to clearly define what is acceptable and what is not outside the ground control locations. Such guidelines, once published, will become the industry standard for generations to come.

Please send your question to Mapping_Matters@asprs.org and indicate whether you want your name to be blocked from publishing.

Answers for all questions that are not published in *PE&RS* can be found on line at [www.asprs.org/Mapping Matters](http://www.asprs.org/Mapping_Matters).

***Dr. Abdullah is the Chief Scientist at Fugro EarthData, Inc, Frederick, Maryland. He is the 2010 recipient of the ASPRS Photogrammetric (Fairchild) Award.*



The contents of this column reflect the views of the author, who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the American Society for Photogrammetry and Remote Sensing and/or Fugro EarthData, Inc.