Statistically Robust Detection and Evaluation of Errors in DTMs

Amirreza Saati , Hossein Arefi

Remote Sensing Technology Institute

German Aerospace Center (DLR)

Wessling, Germany

{Amirreza.Saati,Hossein.Arefi}@dlr.de

Michael Schmitt ,Uwe Stilla

Photogrammetry and Remote Sensing

Technische Universitaet Muenchen

Munich, Germany

Michael.Schmitt@bv.tum.de,   Stilla @tum.de

Abstract—**Digital Terrain Models (DTMs) have been an important topic in the study of ground surface landform, therefore precise evaluation of errors in DTMs production is a critical factor to assess the quality of DTM. In this paper the attribute of errors in DTMs are characterized and robust statistical methods are proposed as accuracy measure. A method based on robust statistical estimation is presented to detect gross errors in DTMs. For practical example a region in Catalonia, Spain, including city areas (Terrassa) as well as forest steep mountainous terrain (La Mola) is selected to evaluate the performance of DTM generation algorithm and to analyze the significance of errors for World view-1 satellite images.**

I.   INTRODUCTION

Digital Terrain Models have found wide applications in various disciplines such as mapping, remote sensing, civil engineering, mining engineering, geology, geomorphology, military engineering, land planning, and communications since their origin in the late 1950s, [1], [2] and [3].Today several techniques are available for generating elevation data such as photogrammetric techniques, SAR interferometry and airborne laser scanning as a powerful technology for automated elevation data collecting from the Earth's surface.

However, all of these corresponding techniques to generate DTMs imply random, systematic and gross errors and thus, include inherent errors in DTMs which constitute uncertainty to achieve the desirable precision in interested applications. Consequently some procedures or methodologies for quality management and control of the DTMs are desired. In this manner several methods have been developed to assess the quality of produced DTMs within the recent years. Root Mean Square Error (RMSE) is the most common way to quantify the difference between the generated DTM and ground truth. Additionally other statistical parameters such as arithmetic mean of height differences, terrain slope, standard deviation, covariant function for heights[4], autocorrelation analysis [5] as well as enhanced visual techniques can be utilized for quality assessment [6].

 Nevertheless for derivation of accuracy measurements it should be noted that outliers may exist and the distribution of error might not be normal. These facts are well known and mentioned in recently published textbooks and manuals for example [7], [8] and [9]. Therefore, as first step for DTM quality assessment normality of data and considering the existing blunders which introduce non-normal distribution and advocate robust statistical methods for accuracy assessment are taken into account.

The accuracy of a DTM is a result of many individual factors which are 1) - attributes of the source data as accuracy, density, and distribution. 2) - Characteristics of the terrain and 3)- finally the methods used for construction of DTM surfaces, i.e., interpolation techniques [10].

As result of above discussion the objectives of this paper are:

(a) Propose an algorithm based on robust statistical methods to detect the gross errors.

 (b) Analyze the accuracy of the DTMs with statistical methods. For that, the DTMs generated by Word view sensor for different regions (residential and hilly forest area) are processed.

II.   DTM GENERATION ALGORITHM

A novel algorithm for automatic DTM generation from high resolution satellite images has been developed at German Aerospace Center (DLR). It consists of two major steps: DSM generation and DTM generation. In the first step, Digital Surface Models (DSMs) are created from stereo scenes with emphasis on fully automated georeferencing based on semi-global matching [11]. The automatic georeferencing is used to derive a high quality rational polynomial coefficients (RPC) correction from lower resolution reference datasets, such as Landsat ETM+ Geocover and STRM C band DSM. Digital Surface Models (DSM) are derived from dense stereo matching. [12].

In the second step which is dedicated to DSM filtering, the DSM pixels are classified into ground and non-ground using the algorithm motivated from the gray-scale image reconstruction to suppress unwanted elevation pixels. In this method, non -ground regions, i.e., 3D objects are hierarchically separated from the ground regions [13].

III. ROBUST STATISTICAL METHODS

As mentioned before, in derivation of the DTM accuracy the fact of existing outliers should be considered and the normality of the error distribution should be analyzed. Normality of errors can be assessed by visualization of distribution of errors using histogram and Quantile-Quantile plot (Q-Q plot). A Q-Q plot is a graphical method for comparing two probability distributions by plotting their quantiles against each other. Herein, a point on the plot corresponds to one of the quantile of consideration area plotted against the same quantile of standard normal distribution. If the two distributions are linearly related, the points in the q-q plot lie on the line. Alternatively statistical parameters such as skewness, kurtosis and statistical tests might be employed as powerful methods for investigation of normality [14].

If the distribution of errors is significantly non-normal because of a considerable amount of outliers, another approach has to be taken into account for deriving accuracy measures. That is a sample quantile of distribution of errors. The quantile of a distribution is defined by inverse of its cumulative distribution function (CDF) (7):

$$Q(P) = F^{-1}(P) \qquad (1)$$

With $O < P < 1$

As an example a quantile 50% is equal to the median of the distribution. In addition to quantile, the Median Absolute Deviation (MAD) is introduced as a result of heavy tail of distribution of errors due to a large amount of outliers. The MAD is a measure of statistical dispersion and an alternative approach to estimate the scale of the dh (error height) distribution rather than the sample variance or standard deviation.

$$MAD = median\left(\left|\Delta hj - m\Delta h\right|\right), \qquad (2)$$

where Δh_j denotes the individual errors and $m \Delta h$ is the median of the errors. Table 4 summarizes the results of the robust statistical methods for these areas.

IV. ERROR TYPES IN DTMS

In general there are three types of error in generated DTMs which are random errors, systematic errors, and gross errors. The word of random error indicates that they are inherently unpredictable and tend to have null arithmetic mean when the measurement is repeated several times and they have a normal distribution. Random errors are also referred to as white noise in statistics. For such errors, improvement of DTM data quality is conducted by applying a low pass filtering in source data to reduce their effect. However necessity of applying filtering processes completely depends on magnitude of the random error occurring during the measurements. For example 0.05% of flying height for photogrammetric measurement might be appropriate threshold [15]. Herein based on preliminary accuracy of using source data, effect of random errors is not considered.

Systematic errors are bias in the measurements caused by the situation where the mean of many separated measurements are significantly different from the actual value of measured attributes. Systematic errors usually occur due to lack of adequate adjustment of instruments, physical reasons such as lens distortion in photogrammetric measurements and finally misalignment in georeferencing due to datum or processing errors.

DEMs misalignment has a significant contribution in systematic errors. In this paper the misalignment is calculated by matching the two surfaces. Global shifts (X0, Y0, Z0) are estimated between two DEMs. Deviations are measured using iterative least squares adjustment by minimization of:

$$\sum \left(\left(Z2(X,Y) + Z0 - Z1(X + X0, Y + Y0)\right)^2\right) \qquad (3)$$

Gross errors which are also called blunders, in fact can be any size in nature. Compared with random and systematic errors, they occur with small probability during measurements. In DEM generation Gross errors often occur in automatic image matching due to mismatching of image points. The methodology which is used here to detect the gross errors is based on the robust statistical measures and consists of four steps:

1. For each quantile the corresponding sample quantile is computed from the median sequencely, i.e. Q (50), Q (51) … Q (100).

2. Since linear least-squares estimators are badly affected by outliers, a robust regression estimator such as Least Median Square is utilized to acquire regression parameters based on computed quantiles and identify which quantiles are emerging as outlieres.

3. Recording the index of first quantile which is recognized as an outlier.

4. Computing an appropriate threshold to remove the outliers. In this manner a certain confidence interval is proposed showing that observations (dh) lie within the interval based on Vysochanskij–Petunin theory[16]. This theory gives an interval for the probability that a random variable lie within a certain number of standard deviations of variable mean:

$$p\left(\left|x - u\right| \geq \lambda\sigma\right) \leq \frac{4}{9\lambda^2} \qquad (4)$$

Corresponding to an upper probability bound which is derived from recorded index of first quantile as an outlier, the parameter of λ is computed and an appropriate threshold is built by λ times of MAD . Quick operation for very large datasets is the significant advantage of this method.

V. RESULT OF PRACTICAL EXAMINATION OF DTMS FOR SELECTED TEST AREA

The purpose of this section is to establish a sensible flowchart to detect and measure the magnitude of errors based on their attributions and make a comparison for different test areas.

The test region in Catalonia, near Barcelona has been selected due to availability of several stereo satellite data and a high quality reference dataset provided by the Institute

Cartografic de Catalunya(ICC). They consist of color orthoimages with a spatial resolution of 50 cm as well as an airborne laser scanning point cloud with approximately 0.3 points per square meter. The four ISPRS datasets are used for the test region. (ISPRS-Commission I, working group I /4, Benchmarking and quality analysis of DEM). The characteristics of these datasets and properties of selected test areas are described in table 1 and 2.

In this section DTMs are derived directly from generated DEMs by stereo matching of Word view images and compared with the ICC reference. The corresponding histogram and normal probability plots are shown in figure 1 and 2.

The histogram shows that the kurtosis of the errors distribution is positive, i.e. the distribution has a sharp peak around the mean and fatter tails than normal distribution. The value for skewness of La Mola region indicates that deviation from the mean by outliers is going to be positive.

The Q-Q plot diverges from straight line and it also can be seen there are more positive outliers than negative ones. Table 3 and 4 summarize skewness, kurtosis as well as chi square value computed for these investigation areas. In addition the bootstrap distribution of the standard deviation and MAD based on 100 bootstrap samples have been computed and shown in figure 3. It can be clearly seen that the distribution of standard deviation in comparison to MAD is erratic as a result of outliers. According to the values of statistical parameters shown in table 3 and preceding discussion for presented graphs it can be concluded that there is an excessive amount of outliers in the observations. Furthermore other approaches that are resistant to outlier such as robust accuracy measure have to be applied.

Table I. Properties of source dataset.

Dataset	Description	
	Image resolution(m)	*Generated DEM resolution(m)*
Worldview -1	0.5	1

Table2 Properties of selected test area

Area	Description			
	Height Range(m)	*Mean Slope (°)*	*Terrain Description*	*Area Size*
Terrassa	281-311	2.5	City ,Industrial	5KmX5km
La Mola	596-792	14.8	Mountainous forest	5kmX5km

Table 3. Statistical measures to describe the distribution of observed errors for investigation areas.

Area	Statistical Measures			
	Skewness	*Kurtosis*	*Computed chi-square value*	*Chi square distribution value α=%95, v=3*
Terrassa	0.55	13.75	4899900	0.352
La Mola	11.13	142.49	11446000	0.352

Table 4 .Accuracy measure of DTM generated by Word views images before removing systematic and gross errors.

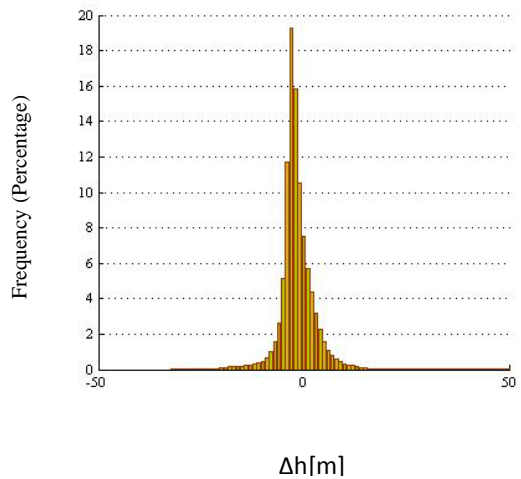| Area	Accuracy Measure			
	50% quantile |Δh| (m)	*MAD |Δh| (m)*	*90%quantile |Δh| (m)*	*95%quantile |Δh|(m)*
Terrassa	1.72	2.02	6.50	9.14
La Mola	6.56	13.45	20.27	26.47

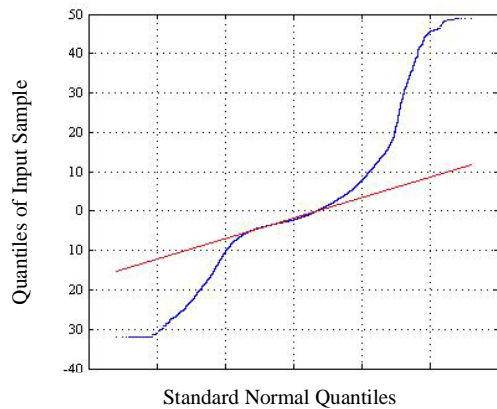Figure1. Histogram of the errors Δh. the vertical axes scaled to percentage.

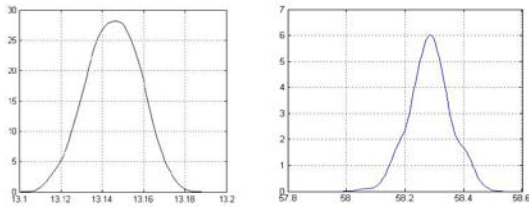Figure2. Q-Q plot for the distribution of dh

Figure 3. This figure shows the bootstrap distribution of the standard deviation (right) and median absolute deviation (left) based on 100 bootstrap samples for each estimator.

Computation of misalignment as the systematic error is done by minimization of height discrepancy between DSM generated by the semi-global matching algorithm and reference DSM produced by laser data. These differences are summarized in table 5. According to the values of statistical parameters in table 6 the median of distribution errors reduced slightly, in contrast to median, there is no change in MAD as the result of outliers.

Finally to detect the gross errors based on the proposed algorithm, the values of $\lambda=1.84$ and $\lambda=2.01$ computed for Terrassa and La Mola respectively and result are summarized intable7. Time computation for 25 million datasets is around 5 minutes. Compared to λ computed for Terrassa area the λ of La Mola is smaller which indicates greater outliers for this area.

VI. CONCLUSION

To measure the accuracy of DTMs and to improve it, the attributes of errors and their magnitudes have to be considered. Moreover we classified error characteristics as random, systematic and gross errors and evaluated significance of them. To avoid influence of outliers in error assessment robust statistical methods were proposed and based on them a method for gross error detection was proposed and finally according to final values obtained for two test areas it can be concluded that the performance of the algorithm is lower for mountainous areas than for urban areas.

Table5. .Misalignment calibration due to Georeferencing errors.

Area	shift parameters for misalignment		
	X(m)	Y (m)	Z(m)
Terrassa	3.53	8.42	0.77
La Mola	3.89	7.62	2.08

Table 6. Accuracy measure of theDTM after removing systematic before eliminating of gross errors.

Area	Accuracy Measure			
	50% quantile $\Delta h(m)$	MAD $\Delta h(m)$	90%quantile $|\Delta h|/(m)$	95%quantile $|\Delta h|/(m)$
Terrassa	1.53	2.02	6.32	8.95
La Mola	6.18	13.45	19.16	25.89

Table7. .Accuracy measure of the DTM after removing gross errors.

Area	Accuracy Measure			
	50% quantile $\Delta h(m)$	MAD $\Delta h(m)$	90%quantile $|\Delta h|/(m)$	95%quantile $|\Delta h|/(m)$
Terrassa	1.22	0.73	2.91	3.28
La Mola	5.78	4.81	16.42	19.72

REFERENCES

[1] D.R. Catlow, "The multu diciplinary application of DEMs", Auto –Carto London, 1986, pp. 447-454

[2] G. Petrie and T. Kennie, "Terrain modeling in surveying and civil engineering.",Whittles publishing, Caithness, England, 1990, pp. 85-111.

[3] D.F. Maune, L.C. Huff and G.C. Guenther, "Digital Elevation Model techniques and applications", In: The DEM user Manual, American Society for photogrammetry and Remote Sensing, Bethesda, MD, 2001, pp. 367-394.

[4] A. Osman, "Quality control of photogrammetrically sampled digital elevation models", In: Photogrammetric Record,12/69, 1987, pp. 331-341.

[5] J. Lee and L.K. Marion, "Analysis of Spactial Autocorrelation of USGS. 1:250,000 Digital Elevation Models, GIS/LIS,1994,pp. 504-513.

[6] T.Prodobnikar, "Methods for visual quality assessment of a digital terrain model", In:S.A.P.I.E.N.S vol2, N2, 2009.

[7] J.Heohle and M.Heohle, "Accuracy assessment of Digital Elevation Models by means of robust statistics methods", ISPRS Joirnal of Photogrammetry and Remote Sensing, 2009, vol.64, issue 4,pp.398-406.

[8] Z.Li, Q.Zhu, C.Gold, "Digital Terrain Modeling –Principle and Methodology", 2005, CRC Press, ISBN: 0-415-32462-9.

[9] D.F.Maune, "Digital Elevation Model Technologies and Application" In:The DEM user manual, 2007, 2nd edition, ISBN: 1-57083-082-7.

[10] Z.L.Li, "Variation of the accuracy of digital terrain models with sampling intervals", In: 1992, Photogrammetric record, 14(79), pp.113-128.

[11] H.Hirschmeuller, "Streo processing by semi-global matching and mutual information", In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008,30(2), pp.328-341.

[12] P.d' Angelo, P.Schwind, T. Krauss, B.Frithjof and P. Reinartz, "Automated DSM based Georeferencing of CARTOSAT-1 Stereo Scenes", In: ISPRS Hannover Workshop 2009- High-Resolution Earth Imaging for Geospatial Information.

[13] H.Arefi and M.Hahn, " A morphological reconstruction algorithm for separating off-terrain points from terrain points in laser scanning data", In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2005, Vol.36 (3/w19).

[14] R.B.D'Agostino, A.Belanger and R.B.D'Agostino Jr, "Asuggestion for using powerful and informative test of normality, The American Statistician 44(4), pp.316-321

[15] K.Kubic and R.Roy, "Digital Terrain Model Workshop proceeding", 1986, Colombus, OH.

[16] D.F.Vysochanskij and Y.I, "justification of the 3σ rule for unimodal distributions", Theory of Probability and Mathematical Statistics 21, 1980, PP.25-36.