ACCURACY ASSESSMENT OF DIGITAL SURFACE MODELS GENERATED BY SEMI-GLOBAL MATCHING ALGORITHM USING LIDAR DATA

A. R. Saati [a] , H. Arefi [b] , N. Aeini [c] , P. Reinartz [b]

[a] TUM, Institut für Astronomische und Physikalische Geodäsie,Arcisstrasse 21, 80333 Munich, Germany
amir_saati@mytum.de

[b] German Aerospace Center (DLR), Wessling, Germany
hossein.arefi@dlr.de, peter.reinartz@dlr.de

[c] University of Isfahan, Geomatic Department, Hezar jarib St, Isfahan, Iran
nazanin.aeini@gmail.com

KEY WORDS: LIDAR, DSM, DTM, Remote Sensing, Matching Algorithm

ABSTRACT:

To measure the accuracy of Digital Surface Models (DSMs) generated by high resolution satellite images (HRSI) using semi-global matching algorithm in comparison with LIDAR DSMs, two different test areas with different properties and corresponding attributes and magnitudes of errors are considered. Error characteristics are classified as systematic and gross errors and significance of them to measure the accuracy of DSMs are evaluated. In this manner and to avoid the influence of outliers in accuracy assessment robust statistical methods are proposed. According to final values obtained for two test areas it can be concluded that the performance of DSMs generated by stereo matching for mountainous wooden areas in respect to the accuracy of LIDAR DSM are poor. In contrast, in case of residential urban areas the quality of the DSM generated by HRSI is able to follow the accuracy of LIDAR data.

1. INTRODUCTION

A model of terrain surface is often a necessary requirement in identifying, analysing and mitigating problems in many fields including hydrology, geomorphology and environmental modelling. Nowadays several techniques are available for generating Digital Surface Models (DSMs) and corresponding Digital Terrain Models (DTMs) that represent the bare earth at some level of details. With the upcoming of new technologies for generation of DSMs and new development in the area of digital photogrammetry due to automatic image matching techniques and revolution of laser scanning for capture of topographic data the question of accuracy "how accurate is DSM?" has to be studied. Automated processing of the raw data to generate DSMs is not always successful and systematic errors and many outliers may still be present in the final product(Heohle and Heohle, 2009). Distribution of accuracy in DSMs depends on the spatial variation of the accuracy, density of the height data, suitability of the interpolation methods and finally the accuracy of the original observations (Karel et al., 2006).

While new techniques such as LIDAR are available for almost instant DSM generation, the use of stereoscopic high resolution satellite imagery coupled with image matching, affords cost-effective measurement of surface topography over large coverage area (poon et al, 2005).

However, all of these corresponding techniques to generate DSMs imply random, systematic and gross errors and thus, some procedures or methodologies for quality management and control of the DSMs are desired. For this purpose, several methods have been already proposed based on statistical methods or visual interpretation. Visual methods such as 2D raster rendering and bi-polar difference maps can be very important for the evaluation of DTMs and can balance some weakness of statistical methods. The usage of visual methods depends on the expertise and experience of the operator. Visual methods actually offer the first assessments of DTMs (Prodobnikar, 2009). Root Mean Square Error (RMSE) is the most common way for statistical methods to quantify the difference between the generated DTM and ground truth.

Many of the statistical procedures assume that errors are normally distributed. Unfortunately, when there are outliers in data, classical statistical methods often have very poor performance and large deviations from the normal distribution can cause problems (Heohle and Heohle, 2009).

As example, considering *n* independent measurements of the same quantity, the question arises which value should be taken as best estimate of the unknown true value. This question is answered if the error distribution is known and the arithmetic mean is accepted as a good estimator for unknown true value as long as normal distribution is considered as the distribution of errors.

However empirical investigations show that the distribution of errors is slightly but clearly longer tailed because of this fact that real data normally contain outliers. Therefore considering of outliers is crucial since they can play havoc with standard statistical methods.

Robust statistical measurements provide an alternative approach to classical statistical methods to produce estimators which are independent of error distribution (free distribution).

The objective of this paper is to analyze the accuracy of DSMs created using HRSI provided by semi-global matching algorithm (Hirschmueller, 2008) in comparison with LIDAR DSMs and employ the robust statistical methods as effective methods to diminish the influence of outliers in evaluation of corresponding DSMs.

2. SEMI-GLOBAL IMAGE MATCHING ALGORITHM (SGM)

In the past, DSM generation using satellite imagery at medium resolutions was associated with across-track stereo geometry and unreliable image matching due to large time lags between data acquisition of images. However at the present time with employing new techniques in imagery collected by high resolution satellite image sensors allows consistent imaging conditions and substantially increases image matching success (Poon et. al., 2005). Correlation or Image matching algorithm refers to the automatic identification and measurement of corresponding image points that are located on the overlapping area of multiple images. This method determines the correspondence between two image areas according to the similarity of their gray level values.

Semi-global image matching algorithm (Hirschmueller, 2008) avoids using matching windows, and is thus able to reconstruct sharp object boundaries. Instead of strong local assumption on the local surface shape, a global energy function E is minimized for all disparities (local shift between stereo pair) D. SGM performs a semi-global optimization by aggregation of costs from 16 directions and find an image D which lead to the low energy E :

$$E(D)=\sum_p \left(C(p,D_p)+\sum_{q\in N_p} D_C = \begin{cases} p_1, | D_p - D_q | <=1 \\ p_2, otherwise \end{cases} \right) \quad (1)$$

The function C defines the matching cost (mutual information) between the image pixels for each pixel location p and possible disparity D_p in the first image. These cost functions adapt to brightness changes in the stereo images and allow matching of images with large viewing angle differences. The second and third terms of E penalize disparity changes in the neighbourhood N_p at each position p. The penalty p_1 is added for all disparity changes equal to one pixel. At larger discontinuities, fixed cost p_2 is added. This cost function favours similar or slightly changing disparities between neighbouring pixels, and thus stabilizes the matching in image areas with weak contrast, but also allow large disparity jumps in areas with high contrast.

3. ACCURACY MEASUREMENT USING ROBUST STATISTICAL METHODS

Due to the fact of existing outliers, to determine the accuracy of DSMs initially the normality of error distribution has to be examined. This manner can be done by means of statistical test and visual statistical methods (Histogram and Q-Q plot) as a component of good data analysis for investigating normality.

1. Histogram: The distribution of errors can be visualized by a histogram of the sampled errors, where the number of errors (frequency) within certain predefined interval is plotted which is an estimate of the probability distribution of a continuous variable. Such a histogram gives a first impression of the normality of the error distribution. A better diagnostic to check the normality of error distribution is relied on two significant characteristics of histogram, namely *skewness* and *kurtosis.*

Skewness is referred to asymmetry of a distribution. A distribution with an asymmetric tail extending out to the right is referred to as positively skewed or skewed to the right, while a distribution with an asymmetric tail extending out to the left is referred to as negatively skewed or skewed to the left. Skewness can range from minus infinity to positive infinity.

Kurtosis is introduced as a measure of how flat is the top of a symmetric distribution when compared to a normal distribution of the same variance. It is actually more influenced by scores in the tails of the distribution than scores in the center of a distribution. Distribution with the positive kurtosis is fat in the tails. In contrast negative kurtosis depicts that distribution of errors is thin in the tails.

2. *Quantile-Quantile plot*(Aster et al., 2004).*:* A better diagnostic plot for checking a deviation from the normal distribution quantile-quantile (Q-Q) plot. The Q-Q plot provides a more precise graphical test of whether a set of data could have come from a particular distribution. The data points:

$$\mathbf{d} = [d_1, d_2, ... d_n]^T \qquad (2)$$

are first sorted in numerical order from smallest to largest into a vector y, which is plotted versus

$$x_i = F^{-1}((i-0.5)/n) \qquad (i=1,2,...,n) \qquad (3)$$

Where $F(x)$ is cumulative distribution function (CDF) of the distribution against which we wish to compare our observations. If we are testing to see if the elements of d could have come from the normal distribution, then $F(x)$ is the CDF for the standard normal distribution:

$$F_N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}z^2} dz \qquad (4)$$

If the element of d is normally distributed, the points (y_i, x_i) will follow a straight line.

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 CDF (Heohle and Heohle, 2009):

$$Q(P) = F^{-1}(P) \qquad (5)$$

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 error distribution rather than the sample variance or standard deviation.

$$MAD = median_i\left(\left|X_i - median_j\left(X_j\right)\right|\right) \qquad (6)$$

where x_i denotes the individual errors and $median_j(x_j)$ is the median of the errors.

4. STUDY AREAS AND DATA ACQUISITION.

Two test regions in Catalonia, near Barcelona have been selected due to availability of several stereo satellite data and a

Figure 1: Two test areas namely Terrassa (left) and Lamola (right) regions

high quality reference dataset provided by the Institute Cartografic de Catalunya (ICC). They consist of color orthoimages with a spatial resolution of 50cm as well as an airborne laser scanning point cloud with approximately 0.85 and 0.4 point per square meters for Lamola and Terrassa respectively.

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.

As reported in (Husing et al., 1998), the systematic errors of measured LIDAR points corresponding to the flat, flat gross, hilly, and hilly gross areas are 5-20, 20-200, 5-20, and 20-200 centimetre respectively. Accordingly, related random errors for LIDAR data point for these areas are 10-20, 10-50, 20-200 and 20-200 centimetres.

The DSMs generated using semi-global stereo matching of Worldview-1 satellite images are compared with the first pulse laser points. More than 19 and 10 million random LIDAR points for Lamola and Terrassa region respectively contribute to detect errors. It should be stated here that derived accuracy error for DSM is relative error respect to the accuracy of the LIDAR datasets. For evaluation of distribution errors, corresponding histograms and Q-Q plots for Terrassa and Lamola regions are shown in Figure 3 and 4.

Area	Properties of selected test area			
	Height Range(m)	*Mean Slope degree*	*Terrain Description*	*Area Km^2*
Terrassa	281-311	12	City, Industrial	5×5
Lamola	596-792	24.5	Mountainous forest	5×5

Table1: Properties of selected test area.

Dataset	Description of datasets	
	Image resolution(m)	*Generated DEM resolution(m)*
Worldview-1	0.5	2.5

Table2: Properties of source datasets

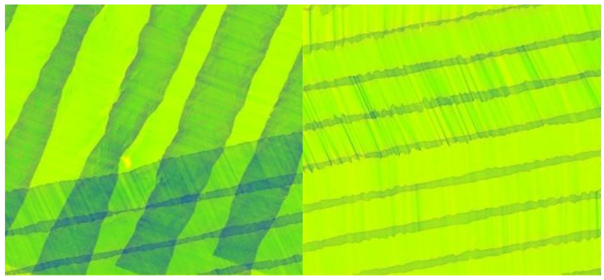

Figure 2: Two images showing the return point density of LIDAR data. Cells with point densities within the 0.5 to 1 point per square meter are coloured green, above this range are coloured blue and below are coloured red corresponding to Lamola (left) and Terrassa area (right).

Figure 4 shows the Q-Q plots for corresponding regions. Both Q-Q plots diverge from straight line and it also can be concluded that there are more positive outliers than negative ones for Lamola region. In addition the bootstrap distribution of mean and median based on 50 bootstrap samples have been computed and shown in figure 5. It can be clearly seen that the distribution of *mean* in comparison to *median* 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 robust statistical method which is resistant to outlier has to be applied.

Consequently median, 68.3% and 95% sample quantiles of absolute errors and Median Absolute Deviation are measured to estimate the accuracy of DSM respect to LIDAR datasets. The results are summarized in tables 4 and 5.

Area	Statistical Measures	
	Skewness	*Kurtosis*
Terrassa	1.000	16.30
Lamola	6.50	74.61

Table3 Statistical measures to describe the distribution of observed errors.

Area	Accuracy Measure		
	Mean(m)	*Standard Deviation (m)*	*RMSE(m)*
Terrassa	0.349	4.23	4.25
Lamola	-0.306	10.49	10.50

Table 4 Accuracy measures of DSM generated from Worldview images respect to LIDAR DSM

Area	Accuracy Measure			
	50% quantile Δh (m)	*MAD Δ (m)*	*68.3%quantile $\|\Delta h\|$ (m)*	*95%quantile $\|\Delta h\|$(m)*
Terrassa	0.116	2.43	2.16	9.72
Lamola	-1.59	5.08	4.966	14.07

Table 5 Accuracy measures of DSM generated from Worldview images respect to LIDAR DSM.

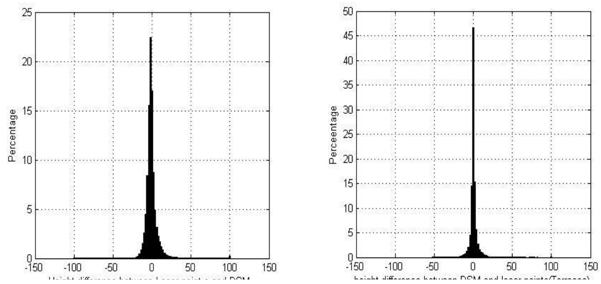

Figure 3 Corresponding normalized histograms of Δh between Laser points and DSM for Terrassa (Left) and Lamola(Right).

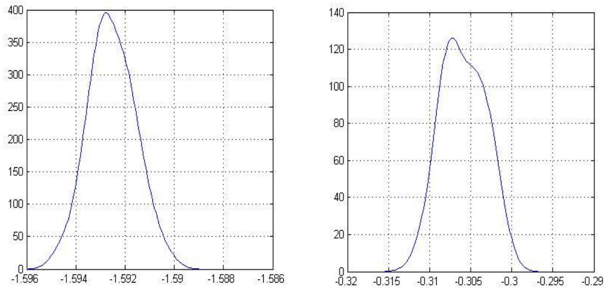

Figure5: The bootstrap distribution of the median (left) and mean (Right) based on 50 bootstrap samples for each estimator.

Area	Accuracy Measure			
	50% quantile Δh *(m)*	*MAD* Δ *(m)*	*68.3%quantile* $\lvert\Delta h\rvert$ *(m)*	*95%quantile* $\lvert\Delta h\rvert$*(m)*
Terrassa	0.099	2.004	2.013	7.54
Lamola	-1.636	4.02	4.81	12.522

Table 7 Accuracy measures of DSM; Outliers are classified by 3times of RMSE

Systematic errors are bias in the measurements caused by the situation where the *mean* of many separated measurements are different from the actual value of measured attributes. Systematic errors usually occur due to lack of adequate adjustment of instruments, and misalignment in georeferencing due to datum or processing errors. Herein *median* of differences with 0.116 and -1.59 meter for Terrassa and Lamola regions respectively are interpreted as systematic errors which are the values for systematic shift between the DSM and LIDAR datasets.

Gross errors which are also called blunders, in fact can be of any size in nature. Compared with random and systematic errors, they occur with small probability during measurements. In DSM generation, gross errors often occur in automatic image matching due to mismatching of image points.

It is obvious from Q-Q plot and also statistical measures from table 3, outliers exist and from table 4 and 5 this fact is deduced that they have a great influence on the estimated standard deviation. From these tables for both areas 68.3% quantile and median absolute deviation are very close. However it should be noted that the 95% quintile for both regions are greater than two times the 68.3 % quantile due to fat tails of both distribution that clearly show the non-normality of errors.

To classify the outlier for accuracy measurements initially 3 times RMSE is considered and results are tabulated in tables 6 and 7. As can be seen, the standard deviation and RMSE after removal of outliers are much lower as with outliers included. Additionally an improvement is observed in MAD and 68.3% quantle.

Area	Accuracy Measure		
	Mean	*Standard Deviation*	*RMSE*
Terrassa	0.206	3.124	3.131
Lamola	-1.13	5.892	6.004

Table 6 Accuracy measures of DSM; Outlier are classified by 3-times of RMSE

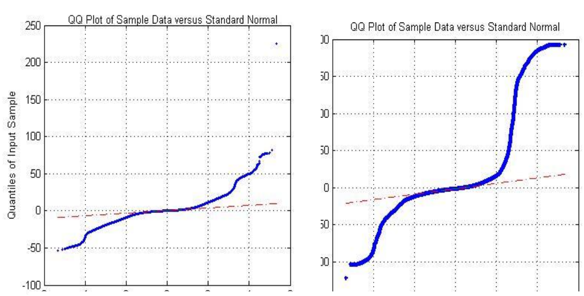

Figure4 Corresponding Q-Q plots for Lamola (right) and Terrassa (Left) regions.

5. CONCLUSION

The accuracy measures for generated DSM should not be influenced from outliers and non-normality of the error distribution. To avoid influence of outliers in error assessment robust statistical methods were considered

According to the final values obtained for two test areas it can be concluded that the performance of the DSM algorithm for mountainous wooden areas respect to the accuracy of LIDAR datasets is poor. Nevertheless, according to relative accuracy of urban area it can be concluded that DSM accuracy is able to follow the accuracy of LIDAR datasets.

Moreover in comparison to cost of using LIDAR system and according to this fact there is a possibility that some remained outliers remove by filter algorithms which are used for generating Digital Terrain Models (DTMs), therefore by employing an appropriate interpolation method, generating DTMs from high resolution satellite images in urban area can be an appropriate alternative for LIDAR systems.

REFERENCES

Aster R, Borchers B, Thurber C, 2005. *Parameter estimation and inverse problems.* Academic press, Burlington,pp.15-40

Heohle J., Heohle M., 2009. Accuracy assessment of Digital Elevation Models by means of robust statistics methods. ISPRS *Journal of Photogrammetry and Remote S*ensing, 64,pp.398-406

Hirschmueller H., 2008. Stereo processing by semi-global matching and mutual information. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2),pp.328-341

Poon J, Fraser C, Chunsun Z,Li Z,2005. Quality assessment of digital surface modelsgenerated from IKONOS imagery. *The Photogrammetric Record* 20(110):,pp.162–171

Prodobnikar T, 2009. Methods for visual quality assessment of a digital terrain model. Surveys And Perspectives Integrating Environment & Society (S.A.P.I.EN.S), vol2, N2Hirschmeuller