

How accurate and usable is GDEM?

A statistical assessment of GDEM using LiDAR data

Tomislav Hengl

ISRIC – World Soil Information
University of Wageningen
Wageningen, The Netherlands
tom.hengl@wur.nl

Hannes Reuter

ISRIC – World Soil Information
University of Wageningen
Wageningen, The Netherlands
hannes.reuter@wur.nl

Abstract—A methodological framework for assessing accuracy of the GDEM product is described using four small case studies in areas of variable relief and forest canopy (Booschord in the Netherlands, Calabria in Italy, Fishcamp in USA and Zlatibor in Serbia). Focus is put on evaluating the true accuracy of ASTER GDEM using LiDAR data aggregated to 30 m resolution. Three aspects of accuracy have been evaluated: (a) absolute accuracy of elevations (goodness of fit between true and GDEM elevations), (2) accuracy of stream networks (goodness of fit for buffer distance maps for stream networks), and (3) accuracy of surface roughness parameters (goodness of representation of nugget variation and residual errors). Results show that GDEM seems to be of little use in areas of low relief (st.dev. <20 m), as in such areas the difference between the topographic features will be statistically significant. Nugget variation in all cases is 3-8 times lower than in the LiDAR DEMs, which indicates that surface roughness is under-estimated. These results also suggest that an adjusted R-square of >.995 could be used as the threshold level for a satisfactory fit between LiDAR and GDEM (this R-square corresponds to RMSE of <10 m). For stream networks, an R-square of >.60 seems to be satisfactory. Analysis of the short-range variability allows determination of the effective grid cell size that more closely matches the true surface roughness. These results support previous work that indicates that a more suitable grid cell size for GDEM v1 is about 90 m.

I. INTRODUCTION

GDEM, produced by the Ministry of Economy, Trade and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA) from optical stereo data acquired by the Advanced Spaceborne Thermal Emission [1] and Reflection Radiometer (ASTER), is the first 30 m resolution global DEM [2]. It was created by stereo-correlating the 1.3 million-scene ASTER archive of optical images, covering almost 98% of Earth's land surface. GDEM has been released on June 29th 2009, and this is now (nominally) the most detailed global GIS layer with public access. The one-by-one-degree GDEM

tiles can be downloaded from NASA's EOS data archive¹ and/or Japan's Ground Data System². The download of DEMs for large areas is at the moment difficult and limited to 100 tiles. METI and NASA has decided to release it but then have emphasized that v1 should be viewed as an “*experimental*” data product.

The claimed accuracy of GDEM is 20 meters at 95% confidence for vertical data and 30 meters at 95% confidence for horizontal data. The ASTER GDEM Validation Team [2] has evaluated accuracy of GDEM using a large number of geodetic control points spread over USA and Japan and by comparing GDEM versus NED, SRTM1 and SRTM3. [3], [4], and [5] have run similar comparisons between GDEM and SRTM DEM. These results in general confirm that: (a), the mean RMSE for GDEM is about 9-11 m, which fits the 20 m interval indicated above, (b) GDEM contains significant anomalies and artifacts (due to clouds and poor spatial matching of scenes), which will affect its usefulness for certain user applications, (c) both SRTM DEM and GDEM report canopy elevations and need to be filtered before they can be used for e.g. hydrological modeling, and (d) absolute accuracy of GDEM in general increases in areas with higher number of scenes. Guth [5] also estimated that a large portion of GDEM (20%) contains anomalies that degrade its use for most applications.

We have downloaded GDEM data for four areas in different parts of the World: the Netherlands, Italy, Serbia and USA, and compared these with the most accurate airborne LIDAR-derived DEMs (dense point sampling) that we have aggregated to 30 m resolution. We were interested to see how accurate is the GDEM and what are the main limitations of using it for various mapping applications. For this purpose we have developed an original statistical framework for comparison of raster maps – a framework that focuses on both global and local assessment of accuracy of

¹ <https://wist.echo.nasa.gov>

² <http://www.gdem.aster.ersdac.or.jp>

relief. At the end, we report the opinion of geomorphometry.org visitors who have rated the value of GDEM in comparison to SRTM DEM.

II. STATISTICAL ASSESSMENT

Here we suggested that accuracy of representation of relief can be represented by examining (at least) the following three aspects of a DEM:

- Accuracy of absolute elevations (absolute error);
- Positional and attribute accuracy of hydrological features (streams, watersheds, landforms etc);
- Accuracy of surface roughness (i.e. representation of the short-range variation);

In this paper, and for practical reasons, we focus on evaluating only the most interesting accuracy measures: accuracy of elevations (Fig. 1), positional accuracy of stream networks (Fig. 2) and accuracy of short-range variation.

A. Absolute error

Absolute error of a DEM is most commonly derived as the mean difference between the target DEM and true elevation [5, 6, 7]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (z^{GDEM} - z)^2}{N - 1}} \quad (1)$$

Even more meaningful results can be obtained by fitting a regression model between the two DEMs (if both the ground truth and target DEM are available at all grid nodes). In this case we can assume that target elevation is a function of the true elevation:

$$z^{GDEM} = f(z) \quad (2)$$

The result of regression modeling can show us if the relationship is close to linear i.e. if there are significant over/underestimations at different elevations. In addition, regression modeling allows us to see how significant is the difference between the two elevations (target and true), which is usually visible via R-square (goodness of fit). R-square indicates how much of variability in true elevations (total sum of squares – SSTO) can be explained by the target DEM (residual sum of squares – SSE):

$$R^2 = 1 - \frac{SSE}{SSTO} \quad (3)$$

B. Accuracy of stream networks

Spatial accuracy of stream networks can be assessed by comparing the buffer distance maps for stream networks derived using the same settings but different DEMs as inputs. As with absolute elevations, the accuracy can be evaluated by fitting a regression model and then looking at R-square values and statistical significance for buffer distance maps. For more details about how to access accuracy of DEMs from the hydrological aspect see also: [6].

C. Accuracy of representation of surface roughness

Surface roughness is the local variation of values in a search radius. There are many ideas on how to derive this parameter (see [7] for discussion). Here, we have decided to evaluate surface roughness using two standard measures:

1. Difference in the variogram parameters – especially in the range and nugget parameters – fitted using (randomly) sampled elevation data. This is a global measure.
2. Difference in the local variance derived using a 7×7 search-window. This is a local measure.

D. Study areas

For GDEM accuracy assessment we use four small case studies spread over two continents: area of low relief – Booschord in the Netherlands (30 km²), area of high relief and without canopy – Calabria in Italy (16.1 km²), area of high relief and high canopy – Fishcamp in USA (2 km²) and area of medium relief and little mixed canopy – Zlatibor in Serbia (13.5 km²). For all these case studies both GDEM and LiDAR DEMs at 30 m resolution have been prepared. The LDEM and GDEM raster maps for the four case studies can be obtained from the geomorphometry.org website (look under “Data sets”). Many more sample LiDAR DEMs can be also obtained from the opentopography.org project.

E. Processing steps and implementation

We implement this methodology in R environment for statistical computing. Geographic operations such as derivation of stream networks and derivation of the local variance were implemented in SAGA GIS, which can be easily controlled from R [8]. The R script used to produce these results can be obtained from the geomorphometry.org website and adopted for any similar case study.

III. RESULTS

TABLE I. shows summary results of statistical comparison of GDEM and LiDAR DEM for four study areas. This shows that the average RMSE for elevations for the four data sets is: 18.7 m, which is somewhat worse than what has been reported by the ASTER GDEM Validation Team [2]. Although reported RMSE was the lowest in the area of low relief, R-square indicates that the two DEMs are statistically different for this area, i.e. they show spatial patterns which are uncorrelated.

TABLE I. RESULTS OF REGRESSION MODELING FOR THE FOUR CASE STUDIES.

Case study	St. dev. LDEM	Adjusted R-square	RMSE
Booschord	1.6 m	0.01 ^a	3.3 m
Calabria	171.3 m	.9941	45.9 m
Fishcamp	111.1 m	.9915	16.6 m
Zlatibor	37.7 m	.9656	8.9 m

^a Difference significant at 95% confidence level.

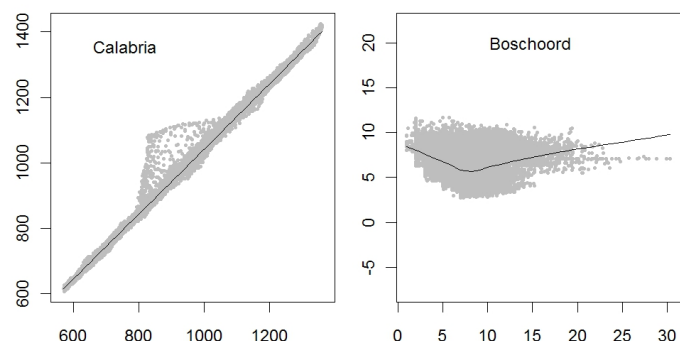


Figure 1. Visual comparison between GDEM (x-axis) and LiDAR DEM (y-axis) for a case study in Italy (Calabria; 25,754 grid nodes) and in area of low relief (Boschoord; 48,391 grid nodes).

Note from TABLE I. that difference between the two DEMs in an area of low relief is significant. The model explains only 1% of variability in the original signal, hence the DEMs for the case study in Booschord are basically two independent images (Fig. 1).

The average error of locating streams is between 60-100 m. RMSE for buffer distance maps for four case studies is, as expected, much smaller than for modeling of elevations: it ranges from 3.3 for Booschord case study (significant difference at 95% probability level) to 46.0 for Calabria case study. For both the Fishcamp and Zlatibor case study it seems that the difference in

stream networks is also not significantly different, although RMSE is still smaller than for the Calabria case study (RMSE of 16.6 and 8.9).

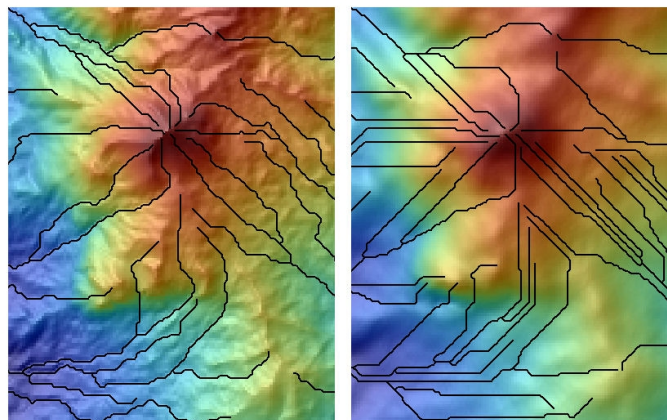


Figure 2. Comparison between stream network (case study Calabria) derived using LDEM (left) and GDEM (right) using exactly the same parameters. The R-square between the buffer distance to streams is statistically significant ($R^2=.46$), but from a practical point of view there are obvious differences. Notice also that surface roughness in LDEM is at the order of magnitude higher than in the GDEM. The size of the rectangle is 3.95 by 4.075 km.

The results for assessing surface roughness of GDEM indicate that surface roughness is typically under-represented. First, the nugget variation in the GDEM is 3-8 times smaller than in the LiDAR DEM (which means that the true roughness is under-estimated). Second, the local variance maps for GDEM under-estimates values obtained by using LiDAR DEM. All this indicates that the effective resolution of GDEM is possibly 2-3 times coarser than the nominal resolution of 30 m.

In addition, by visually comparing DEMs for the four case studies, one can notice that GDEM often carries some artificial lines and ghost-like features [5]. Because GDEM is based on images from visible and near and mid infrared bands, the elevations represent a mixture of land elevation and vegetation cover. In addition, because GDEM is produced by mosaicking large amount of scenes, also borders between tiles are often visible in GDEM.

IV. DISCUSSION

The results of this limited comparison using a small sample indicate two important things: (1) GDEM is of little use in areas of low relief, (2) effective resolution of GDEM is over-optimistic and should be generalized to e.g. 90 or 100 m. For the Boschord area GDEM looks to be of absolutely no use because two images are statistically different. For the same case study we have also

noticed many of, what the producers of GDEM call, “mole run” artifacts. As the producers of GDEM themselves indicated: “The ASTER GDEM contains anomalies and artifacts that will reduce its usability for certain applications, because they can introduce large elevation errors on local scales” [2].

These results also suggest that an adjusted R-square of $>.995$ could be used as the threshold level for a satisfactory fit between LiDAR and GDEM (this R-square corresponds to RMSE of <10 m in an area of medium relief e.g. with st.dev. in elevations of about 50-100 m). For stream networks, an R-square of $>.60$ seems to be satisfactory. The nugget variation should not be more than two times smaller in order to picture a similar surface roughness. Analysis of the short-range variability allows us to determine the effective grid cell size that more closely matches the true surface roughness. These results indicate that a more suitable grid cell size for GDEM is about 90 m.

GDEM usability on april 19 2011 (N=20)

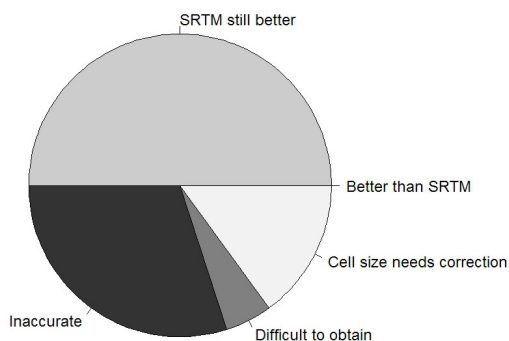


Figure 3. Opinion about the ASTER GDEM in comparison with SRTM (mostly agree that SRTM is still more valuable layer). Based on the limited number of replies collected via the geomorphometry.org website.

GDEM is probably not more detailed and accurate than the 90 m SRTM DEM, especially if one considers parameters such as the surface roughness and accuracy land surface objects. On the other hand, the horizontal accuracy of GDEM is more than satisfactory and GDEM has a near to complete global coverage, so that it can be used to fill the gaps and improve the global SRTM DEM [3], [4], [9]. In addition, the GDEM comes also with a quality assessment (QA) map. Each QA file pixel contains

either: (1) the number of scene-based DEMs contributing to the final GDEM value for each 30 m pixel (stack number); or (2) the source data set used to replace identified bad values in the ASTER GDEM.

A problem is that the creators of GDEM have decided to distribute the GDEM in high resolution (30 m), which is certainly not realistic and results in a decreased usability (as the results of this poll indicate; Fig. 3). For simple reasons that the users have to spend more time to download the data and then generalize it to more appropriate resolution/scale (90 m).

In summary, ASTER GDEM is a valuable new global layer, however, it still needs to be filtered and improved before it reaches the quality criteria of the latest version of SRTM DEM. Finally, one can anticipate that both GDEM and SRTM DEM will become redundant once the TanDEM-X (global coverage 12 m resolution; height accuracy better than 2 m) comes out sometime in 2012/2013. But this data might not be as accessible as the GDEM and SRTM DEM.

REFERENCES

- [1] Hirano, A., Welch, R., Lang, H., 2003. “Mapping from ASTER stereo image data: DEM validation and accuracy assessment”. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(5-6): 356-370.
- [2] ASTER GDEM Validation Team, 2009. “ASTER Global DEM Validation Summary Report”. METI & NASA, 28 p.
- [3] Hayakawa, Y.S., T. Oguchi and Z. Lin. 2008. “Comparison of new and existing global digital elevation models: ASTER G-DEM and SRTM-3”. *Geophys. Res. Lett.*, 35(17), L17404.
- [4] Reuter, H.I. and Nelson, A. and Strobl, P. and Mehl, W. and Jarvis, A. 2009. “A first assessment of Aster GDEM tiles for absolute accuracy, relative accuracy and terrain parameters”. *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009, Vol 5.*, p. 240.
- [5] Guth, P.L., 2010. “Geomorphometric comparison of ASTER DEM and SRTM”. *Join Symposium of ISPRS Technical Commission IV, Orlando, FL*, p. 10.
- [6] Poggio, L., Soille, P., 2008. “Quality Assessment of Hydro-geomorphological Features Derived from Digital Terrain Models”. *EUR 23489 EN, Office for Official Publications of the European Communities, Luxemburg*, 87 pp.
- [7] Hengl, T., Reuter, H.I. (Eds), 2009. “Geomorphometry: Concepts, Software, Applications”. *Developments in Soil Science vol 33, Elsevier*, 796 p.
- [8] Hengl, T., Heuvelink, G. B. M., van Loon, E. E., 2010. “On the uncertainty of stream networks derived from elevation data: the error propagation approach”. *Hydrology and Earth System Sciences*, 14:1153-1165.
- [9] Toutin, T. 2008. “ASTER DEMs for geomatic and geoscientific applications: a review”. *Int. J. Remote Sens.*, 29(7), 1855–1875.