Validation of the ASTER Global Digital Elevation Model (GDEM) Version 2 over the Conterminous United States

U.S. Geological Survey
Earth Resources Observation Science (EROS) Center
Sioux Falls, South Dakota, USA

Validation team: Dean Gesch, Michael Oimoen, Zheng Zhang, Jeffrey Danielson, David Meyer

August 12, 2011

Introduction

The ASTER Global Digital Elevation Model Version 2 (GDEM v2) was evaluated over the full extent of the conterminous United States (CONUS) in a manner similar to the validation conducted for the original GDEM Version 1 in 2009. The primary goal of the CONUS validation was to fully characterize the vertical accuracy of GDEM v2. The results reported herein contribute to the full validation of GDEM v2, which include results from testing conducted by colleagues at ERSDAC in Japan, the National Geospatial-Intelligence Agency (NGA), the Jet Propulsion Laboratory (JPL), and at NASA Goddard Space Flight Center.

Approach

All of the 934 1x1-degree tiles of GDEM v2 data covering CONUS were included in the validation effort. Absolute vertical accuracy of GDEM v2 was calculated by comparison with independent reference geodetic ground control points. GDEM v2 was also evaluated by pixel-to-pixel differencing with other 1-arc-second (30-meter) DEMs having complete coverage over CONUS, namely the National Elevation Dataset (NED) (Gesch, 2007) and the Shuttle Radar Topography Mission (SRTM) dataset (Farr *et al.*, 2007). Accuracy assessment results were segmented by land cover classes to look for relationships between vertical accuracy and cover type. One characteristic of GDEM v2, specifically the number of ASTER scenes (stereo pairs) used to derive an elevation for a pixel, was examined to see how it might affect vertical accuracy. The accuracy assessment results are presented here in summary statistics and charts.

Reference Data

The primary reference data were the “GPS on Bench Marks” dataset of geodetic control points (http://www.ngs.noaa.gov/GEOID/GPSonBM09/) from the National Geodetic Survey (NGS). These points represent NGS’s best x-y-z control point dataset for CONUS, and they are used by NGS for gravity and geoid determination (Roman *et al.*, 2004; Roman *et al.*, 2010). This set of control points is from NGS’s latest U.S. geoid model, GEOID09. The points have millimeter to centimeter-level accuracies, and as such are an excellent reference dataset against which to compare DEMs across CONUS. For the accuracy assessment presented here, 18,207 points (Figure 1) were intersected with GDEM v2. The elevations of the GPS benchmarks are provided in the NAVD88 vertical datum, whereas the elevations of GDEM v2 are referenced to the EGM96 geoid. Therefore, prior to comparison of the GDEM and the GPS points, the vertical referencing of the points was transformed to the EGM96 geoid. Over CONUS, the vertical offset between NAVD88 and the geoid averages about one-half meter (National Geodetic Survey, 2010).

The 2006 update of the National Land Cover Database (NLCD) (Homer *et al.*, 2004) was used to segment the accuracy assessment results by land cover class. NLCD includes land cover data in 19 classes derived from 30-meter Landsat data. The GPS benchmarks used for validation of GDEM v2 fall into 14 of the NLCD land cover classes.

Absolute Vertical Accuracy

The difference between the GPS benchmark elevation and the corresponding GDEM v2 elevation was recorded for each control point location. The recorded GDEM v2 elevation was derived through bilinear interpolation at the precise latitude/longitude location of the GPS point. At each point, the difference was calculated by subtracting the GPS benchmark elevation from the GDEM v2 elevation, and these differences are the measured errors in GDEM v2. Positive errors represent locations where the GDEM v2 elevation was above the control point elevation, and, conversely, negative errors occur at locations where the GDEM v2 elevation was below the control point elevation.

A plot of the GDEM v2 measured errors vs. elevations of the reference control points (Figure 2) indicates that there is no clear relationship of error with elevation. Also, it appears that there is no preference for positive or negative errors as the plotted GDEM v2 errors are uniformly distributed on both sides of the zero error axis.

Summary statistics of the measured GDEM v2 errors are presented in Figure 3 and Table 1. Note that the error distribution approximates a normal distribution (Figure 3). The Root Mean Square Error (RMSE) is an accuracy metric commonly used for elevation data, and the measured RMSE for GDEM v2 is 8.68 meters. This compares with the RMSE of 9.34 m for GDEM v1 (Table 1). Absolute vertical accuracy can also be expressed with a confidence level, in many cases 95%, or also referred to as "linear error at 95% confidence" (LE95). LE95 is derived directly from the measured RMSE (Maune *et al.*, 2007). GDEM v2 exhibits an LE95 of 17.01 meters, compared with an LE95 of 18.31 meters for GDEM v1 (Table 1). Note that the accuracy statistics for GDEM v1 were derived from a comparison with a previous smaller set of GPS benchmarks (13,305 points) from NGS, which was the most recent dataset available at the time of the GDEM v1 evaluation in 2009. However, most of these points are also included in the current GEOID09 GPS benchmark dataset used for GDEM v2 validation.

Another important descriptor of vertical accuracy is the mean error, or bias, which indicates if a DEM has an overall vertical offset (either positive or negative) from the true ground level. The GDEM v2 mean error of -0.20 meters is a significant improvement compared to the GDEM v1 mean error of -3.69 meters (Table 1).

The absolute vertical accuracy testing also included evaluation of the NED and SRTM datasets over CONUS. Because NED and SRTM are both supplied at the same 1-arc-second posting as GDEM v1, and they have been extensively tested with many results reported in the scientific literature, summary statistics are provided (Table 1) to help give context for the GDEM v2 results. Note that the number of GPS benchmarks used for evaluation of SRTM was reduced to 16,865 points due to the deletion of points that fell in SRTM void or fill areas.

<u>Land Cover Analysis</u>

The absolute vertical accuracy assessment results, both mean error (Figure 4) and RMSE (Figure 5), have been segmented by land cover to examine effects of cover types on measured errors. The land cover associated with each control point was determined by intersecting the GPS benchmark locations with the NLCD. While the RMSE varies little across cover types (Figure 5), the mean error (bias) does appear to be affected by land cover, ranging from +5.00 to -2.27 meters (Figure 4). Recall that ASTER images record the reflective surface, thus the derived elevations in GDEM v2 represent the height of those imaged surfaces. In areas with dense, taller vegetation or built structures, the derived ASTER elevation will represent the elevation of these features rather than ground level. The GDEM v2 mean errors by land cover class (Figure 4) verify that the presence of above ground features cause a positive elevation bias, as would be expected for an imaging system like ASTER (see particularly the mean error for the following classes: woody wetlands, mixed forest, deciduous forest, developed high intensity, evergreen forest, and developed medium intensity). The negative mean errors (on the right side of the chart in Figure 4) are all associated with land cover types that include little or no vegetation with significant above ground height, thus they could be considered open ground classes that would be expected to exert no vertical bias effects on elevations measured by remote sensing systems. However, in each case GDEM v2 exhibits a negative bias.

Figure 6 shows the results of aggregating into broad, generalized land cover classes. The GPS ground truth points were grouped into three broad land cover categories and the GDEM v2 mean error and RMSE were recalculated. The 14 NLCD classes were grouped into forest (deciduous, evergreen, mixed, woody wetlands), developed (open space, low intensity, medium intensity, high intensity), and open (barren land, shrub/scrub, grassland/herbaceous, pasture/hay, cultivated crops, emergent herbaceous wetlands). The chart in Figure 6 indicates the percentage of points that fell into each aggregated class. As with the individual classes, the RMSE varies little among the aggregated classes, but the mean error does appear to reflect the effects of land cover on measurement of elevations by ASTER. As expected, the generalized forest class exhibits a noteworthy positive bias of about 3 meters. However, the aggregated open ground class should have a mean error at or very close to zero, which is not the case. It appears that GDEM v2 may have a “true” overall negative bias on the order of 1 meter.

The mean errors and RMSE for NED and SRTM have also been plotted with the corresponding metric for GDEM v2 for comparative purposes, both by individual land cover classes (Figures 7 and 8) and by aggregated classes (Figures 9 and 10). The comparison of RMSE by land cover class among GDEM v2, NED, and SRTM (Figure 8) reflects the same overall conditions seen in the absolute vertical accuracy statistics (Table 1), with NED being the most accurate, SRTM having the second best accuracy, and then followed by GDEM v2. Comparison of the mean errors by land cover type across the three DEMs (Figure 7) reveals that in forested areas GDEM v2 is consistently registering higher elevations than SRTM, with the exception being evergreen forests. Like ASTER, SRTM is a “first return” system, and elevations above ground level would be expected for areas with trees and/or built structures. It is likely that ASTER is measuring elevations at or near the top of the forest canopy while SRTM is recording elevations part way down into the canopy. Such performance of SRTM in recording elevations within the vegetation canopy rather than at the top has been previously documented (Carabajal and Harding, 2006; Hofton, *et al.*, 2006).

Figures 11 and 12 show a comparison of GDEM v2 with GDEM v1 in terms of mean error and RMSE (by land cover type). The RMSE exhibited across land cover classes is roughly equivalent for the two versions (Figure 12), while the comparison of mean errors (Figure 11) shows the reduction in the magnitude of the overall bias in GDEM v2 with respect to GDEM v1. The five land cover classes on the right side of the chart in Figure 11 (grassland/herbaceous, developed open space, pasture/hay, shrub/scrub, cultivated crops) reveal the true elevation bias for both versions. These open terrain classes should have a mean error at or very close to zero, but GDEM v1 shows a negative elevation bias on the order of 5 meters for these classes, while GDEM v2 shows a negative elevation bias of about 1 meter for the same open ground cover types.

Scene Number Analysis

An ancillary data layer supplied with GDEM v2 indicates the number of individual ASTER scene (stereo pair) DEMs that were used to derive each elevation value. The individual ASTER DEMs are stacked and averaged to calculate the final elevation value for each pixel in GDEM v2. The “NUM” value (number of input individual DEMs) associated with each control point location was determined by intersecting the GPS benchmarks with the ancillary NUM layer. The reference points were then grouped into bins for each NUM value, and the measured GDEM v2 errors for the points in each bin were processed to calculate a mean error and average RMSE for each NUM bin. Figure 13 shows a plot of the mean error and RMSE associated with each NUM value. Note how both the mean error and RMSE improve rapidly as the NUM increases from 1 to 10 scenes. Beyond NUM values of about 15 scenes there appears to be little improvement in either mean error or RMSE values.

Comparison vs. Other Digital Elevation Models

GDEM v2 was differenced with NED and SRTM on a pixel-to-pixel basis across the full extent of their CONUS coverage. In the same manner as with the reference control points, the NED and SRTM were each subtracted from GDEM v2. Thus, positive differences represent locations where the GDEM v2 elevation was higher than the corresponding NED or SRTM elevation, and, conversely, negative differences occur at locations where the GDEM v2 elevation was lower than the NED or SRTM elevation. Prior to differencing, the NED elevations were converted from the NAVD88 vertical datum to the EGM96 geoid vertical reference frame. No such conversion was necessary for SRTM, as both GDEM v2 and SRTM are natively referenced to the EGM96 geoid. Difference statistics were calculated, and summary statistics (mean difference – Figure 14; RMSE – Figure 15) were segmented by NLCD land cover class.

The RMSE by land cover class (Figure 15) shows that in forested classes, GDEM v2 and SRTM generally agree better (as indicated by a smaller RMSE value) than GDEM v2 and NED. This is expected, as both ASTER and SRTM are first return systems that measure above ground elevations in tall vegetation canopies. As land cover becomes more open (for instance, the four classes on the right side of the chart in Figure 15), the GDEM v2-NED RMSE and GDEM v2-SRTM RMSE are nearly equivalent as all three DEMs are measuring near ground level elevations.

The chart of mean differences (Figure 14) supports previous observations from the absolute vertical accuracy assessment. In the forest classes (four classes on the left side of the chart in Figure 14), the GDEM v2-NED mean differences are large compared to the GDEM v2-SRTM mean differences. Again, this is the expected condition as NED by definition is a "bare earth" elevation model (Gesch, 2007), and ASTER is a first return system that measures canopy elevations in forested areas. Even though the GDEM v2-SRTM mean differences for three forest classes (mixed, deciduous, woody wetlands) are small compared to the corresponding GDEM v2-NED mean differences, the fact that they are positive supports the previously described observation that GDEM v2 has proportionally higher elevations than SRTM in many forested areas.

The negative mean differences for both GDEM v2-NED and GDEM v2-SRTM for the five open ground classes (shrub/scrub, pasture/hay, barren land, cultivated crops, grassland/herbaceous) on the right side of the chart in Figure 14 provide further evidence that GDEM v2 has an overall true negative elevation bias. Both NED and SRTM exhibit a mean error very close to zero for open ground land cover classes (Figures 7 and 9), so if GDEM v2 was performing in the same way over those open ground conditions the mean differences would be at or much closer to zero.

The GDEM v2-NED and GDEM v2-SRTM mean differences and RMSE were also segmented by NUM bins similar to the analysis described above for absolute vertical accuracy testing vs. GPS benchmarks. Although not presented here in chart form, the results show a very similar pattern in which mean difference and RMSE decrease quickly as NUM increases and then stabilize at a NUM value of about 15 scenes.

Conclusions

The validation testing described here has raised several important observations about the quality of elevation measurements contained in GDEM v2:

- There is an improvement in overall RMSE of nearly two-thirds of a meter (8.68 m vs. 9.34 m) when comparing the measured accuracies of GDEM v2 and GDEM v1. Likewise, there has also been an improvement in overall mean error (bias) in GDEM v2 when compared with GDEM v1 (-0.20 m vs. -3.69 m).

- It is clear that GDEM v2 includes non-ground level elevations for areas that have above ground features (tree canopies and built structures). Table 2 shows how the mean error increases in the developed land cover classes as the number and density of built structures increases. This condition is observed in both the comparison of GDEM v2 with GPS benchmarks, which represent ground level elevations, as well as in the GDEM v2-NED differencing, with NED representing ground level elevations.

- In many forested areas, GDEM v2 has elevations that are higher in the canopy than SRTM. This observation is based on both the comparison of GDEM v2 with GPS benchmarks, as well as the GDEM v2-SRTM differencing.

- An analysis of the number of ASTER individual scene DEMS that are stacked and averaged to derive the elevation value for every pixel in GDEM v2 shows that improvements to mean error and RMSE are minimal beyond about 15 scenes.

- GDEM v2 exhibits an apparent "true" negative elevation bias of about 1 meter, which was revealed through an analysis of mean error by land cover type. The overall mean error of -0.20 m (Figure 3 and Table 1) is certainly an improvement over the mean error of -3.69 for GDEM v1, but it somewhat masks the true performance of ASTER in measuring the elevation in open terrain conditions (non-vegetated, non-built-up). The overall mean error is dampened by the positive elevation biases contributed by forested and built-up land cover. While the true negative elevation bias of about 1 meter for GDEM v2 is a significant improvement over the true negative elevation bias of about 5 meters for GDEM v1, it is nonetheless a condition that users of GDEM v2 data should be aware of and factor into decisions regarding application of the product.

References

Carabajal, C.C., and Harding, D.J., 2006, SRTM C-band and ICESat laser altimetry elevation comparisons as a function of tree cover and relief: Photogrammetric Engineering & Remote Sensing, v. 72, no. 3, p. 287-298.

Crippen, R. E., 2009, Spatial resolution of the ASTER Global Elevation Model (GDEM). Presenation at the 35th ASTER Science Team Meeting, Kyoto, Japan.

Farr, T.G., Rosen, P.A., Caro, E.; Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D., 2007, The Shuttle Radar Topography Mission: Reviews of Geophysics, v. 45, RG2004, doi:10.1029/2005RG000183.

Gesch, D.B., 2007, The National Elevation Dataset, in Maune, D., ed., Digital Elevation Model Technologies and Applications: The DEM Users Manual, 2nd Edition: Bethesda, Maryland, American Society for Photogrammetry and Remote Sensing, p. 99-118.

Hofton, M., Dubayah, R., Blair, J.B., and Rabine, D., 2006, Validation of SRTM elevations over vegetated and non-vegetated terrain using medium footprint lidar: Photogrammetric Engineering & Remote Sensing, v. 72, no. 3, p. 279-285.

Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M., 2004, Development of a 2001 national land-cover database for the United States: Photogrammetric Engineering and Remote Sensing, v. 70, no. 7, p. 829-840.

Maune, D.F., Maitra, J.B., and McKay, E.J., 2007, Accuracy standards & guidelines, in Maune, D. (ed.), Digital Elevation Model Technologies and Applications: The DEM Users Manual, 2nd Edition, Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing, p. 65-97.

National Geodetic Survey, 2010, Proceedings of the 2010 Federal Geospatial Summit on Improving the National Spatial Reference System, Silver Spring, Maryland, May 11-12, 2010, 156 p. (http://www.ngs.noaa.gov/2010Summit/proceedings.shtml).

Roman, D.R., Wang, Y.M., Henning, W., and Hamilton, J., 2004. Assessment of the new national geoid height model – GEOID03: Surveying and Land Information Science, v. 64, no. 3, p. 153-162.

Roman, D.R., Wang, Y.M., Saleh, J., and Li, X., 2010, Geodesy, geoids, and vertical datums: a perspective from the U.S. National Geodetic Survey: Proceedings, FIG Congress 2010, Facing the Challenges – Building the Capacity, Sydney, Australia, April 11-16, 2010, 16 p.

Tachikawa, T., M. Kaku, and A. Iwasaki, 2009, ASTER GDEM validation. Presentation at the 35th ASTER Science Team Meeting, Kyoto, Japan.

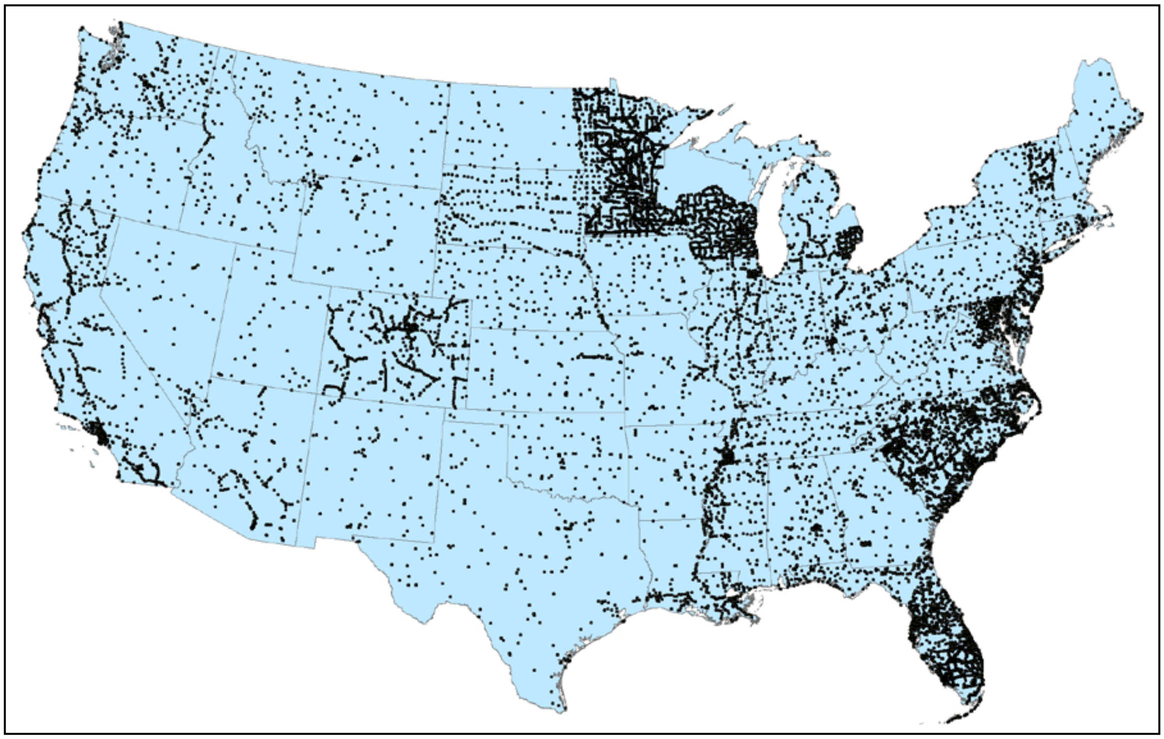

Figure 1. GPS benchmarks (18,207 points) used as GDEM v2 validation reference data.

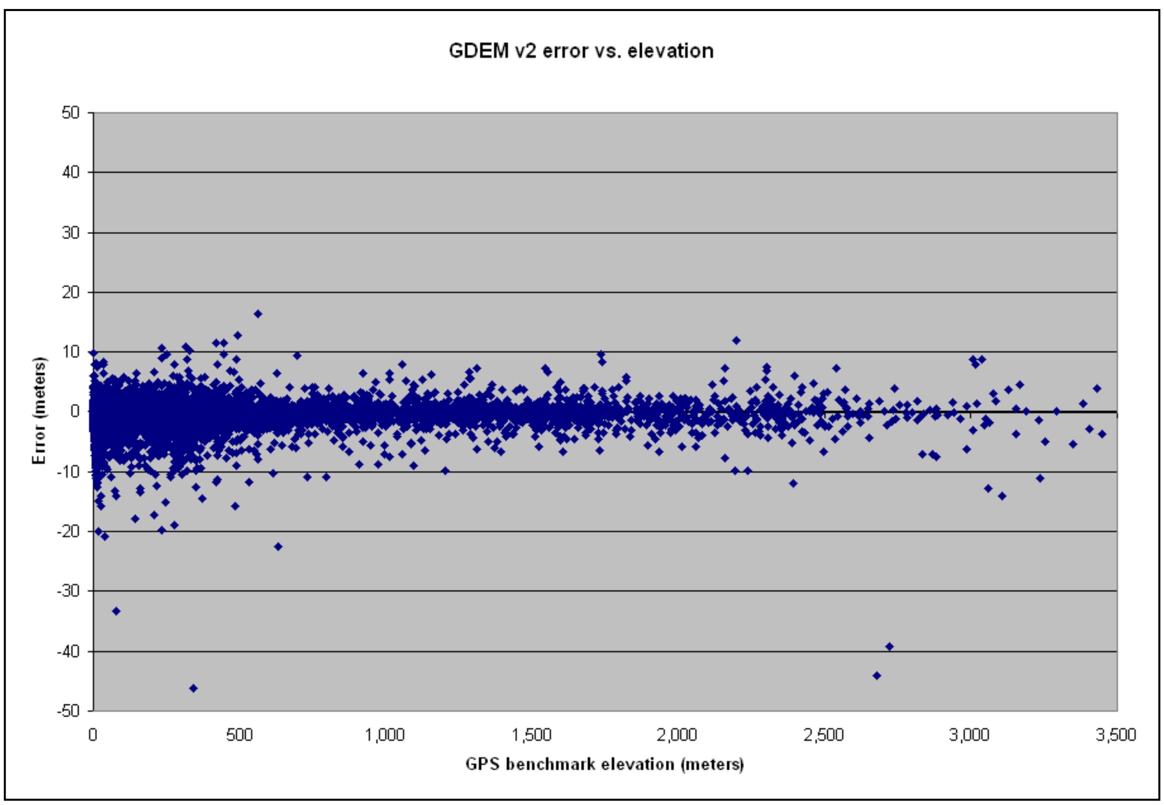

Figure 2. GDEM v2 measured errors plotted vs. elevation.

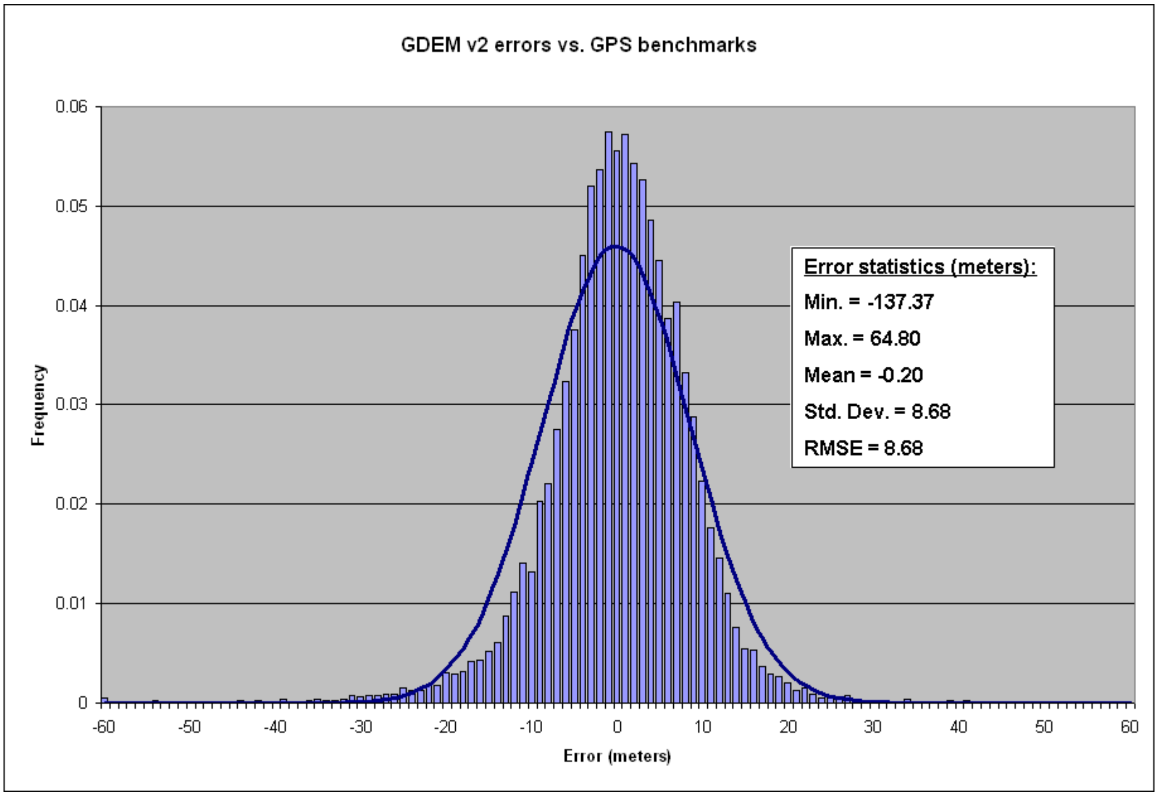

Figure 3. GDEM v2 absolute vertical accuracy.

DEM	Minimum	Maximum	Mean	Standard Deviation	RMSE	LE95
GDEM v2	-137.37	64.80	-0.20	8.68	8.68	17.01
NED	-46.21	16.42	-0.33	1.81	1.84	3.61
SRTM	-28.67	28.58	0.73	3.95	4.01	7.86
GDEM v1	-127.74	105.41	-3.69	8.58	9.34	18.31

Table 1. Error statistics from an accuracy assessment vs. NGS GPS benchmarks.

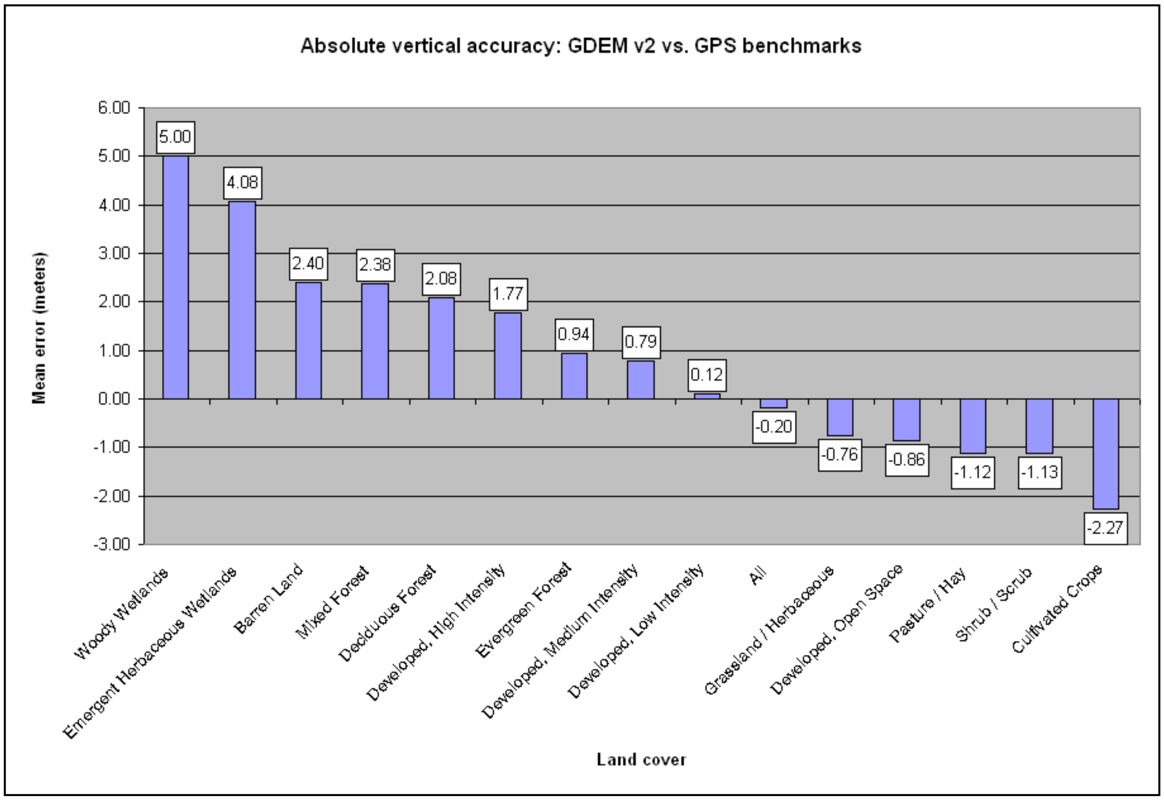

Figure 4. GDEM v2 mean error by land cover class.

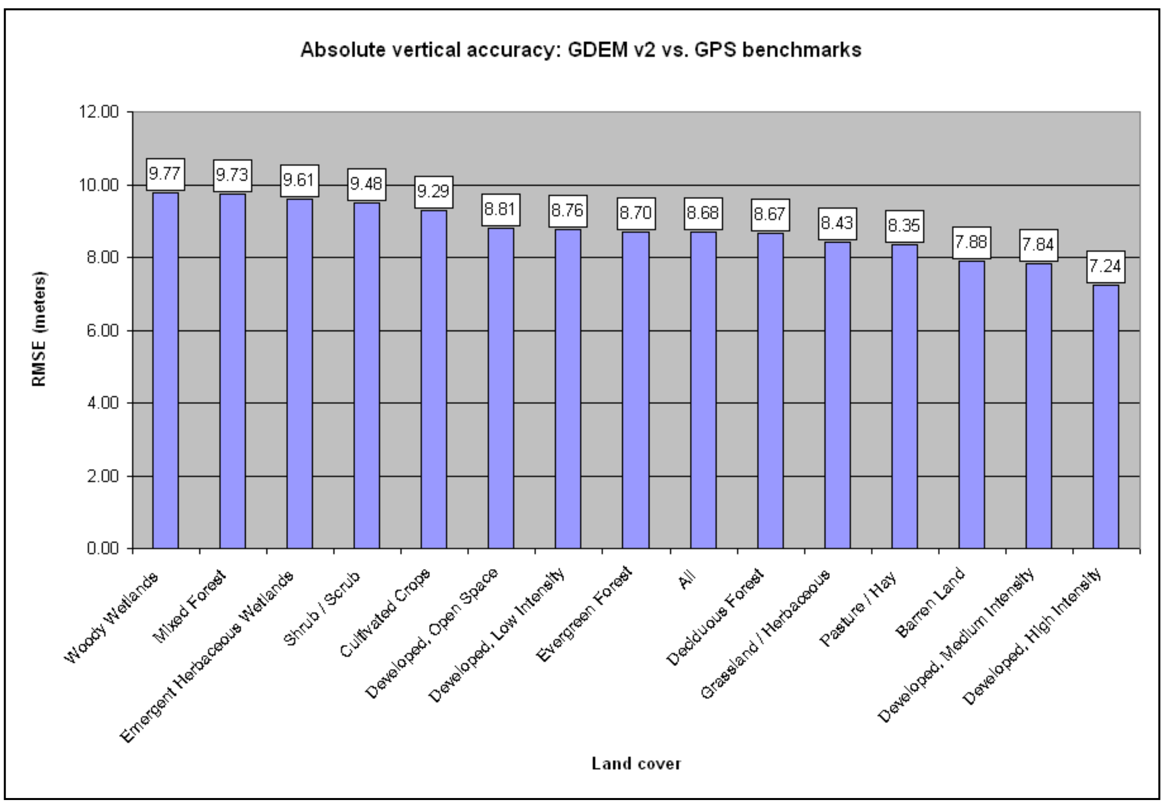

Figure 5. GDEM v2 RMSE by land cover class.

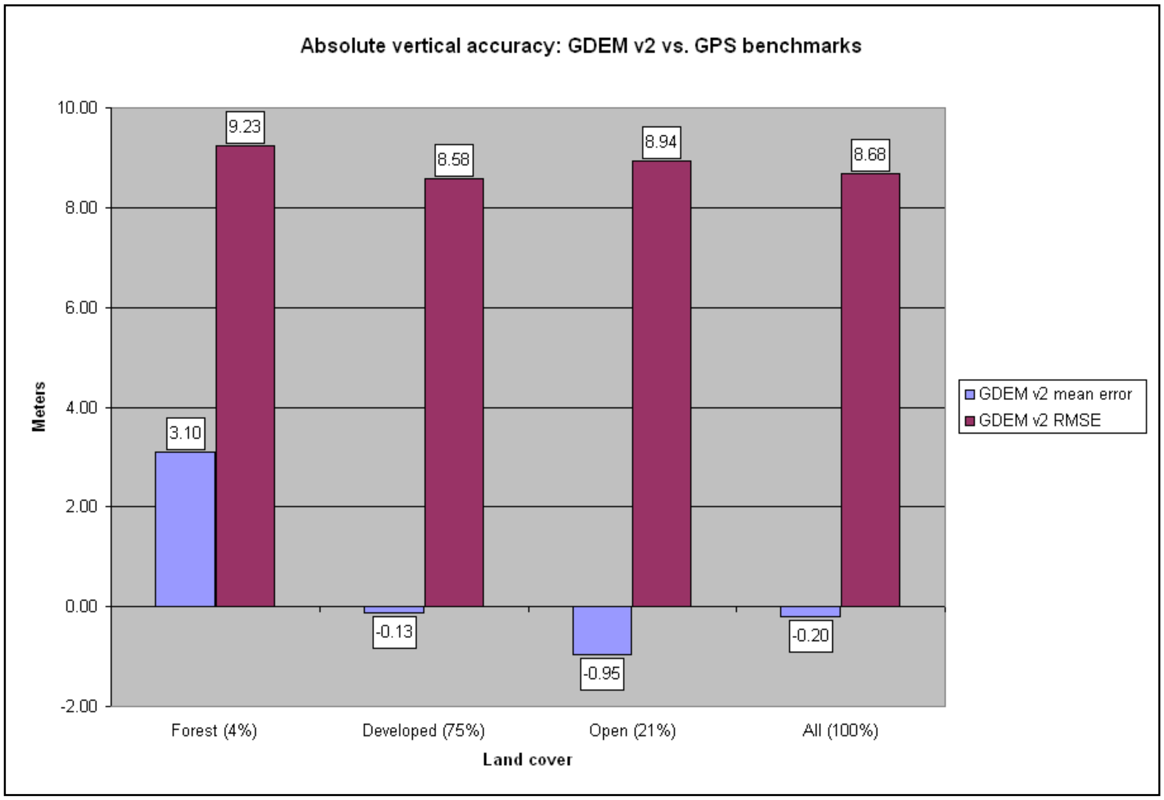

Figure 6. GDEM v2 mean error and RMSE by aggregated land cover class.

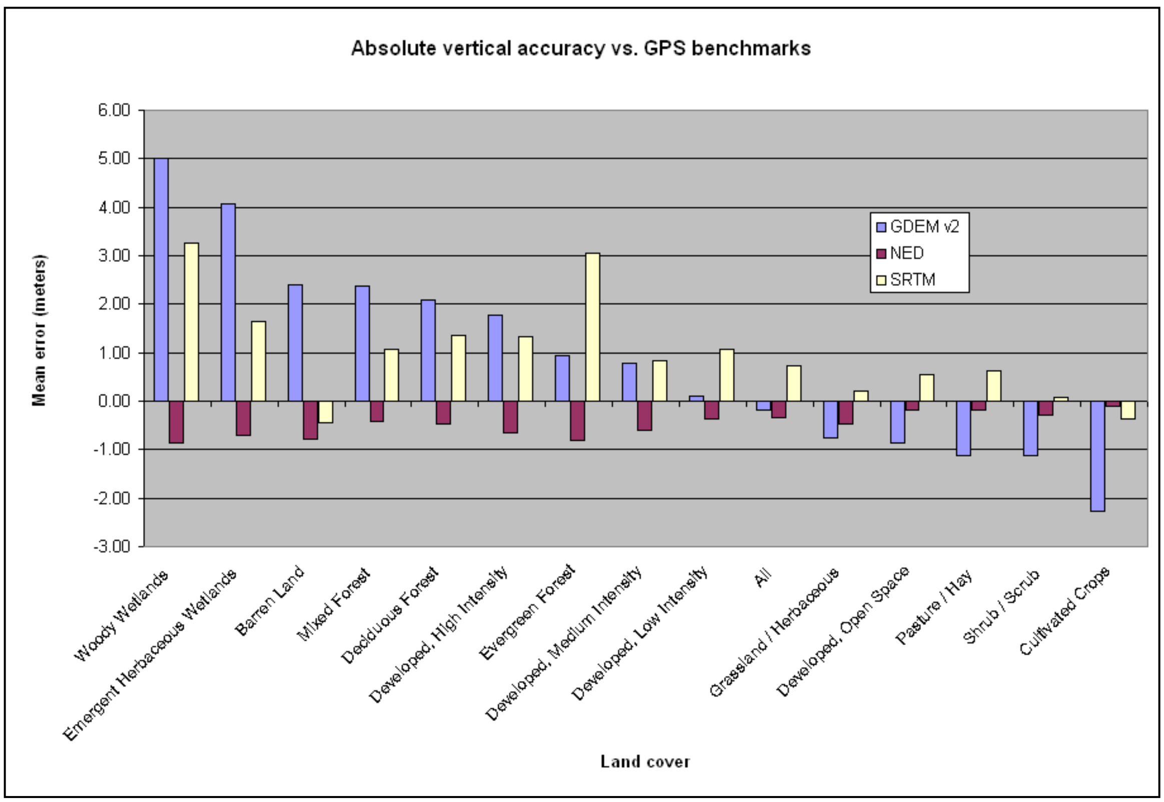

Figure 7. GDEM v2, NED, and SRTM mean errors by land cover class.

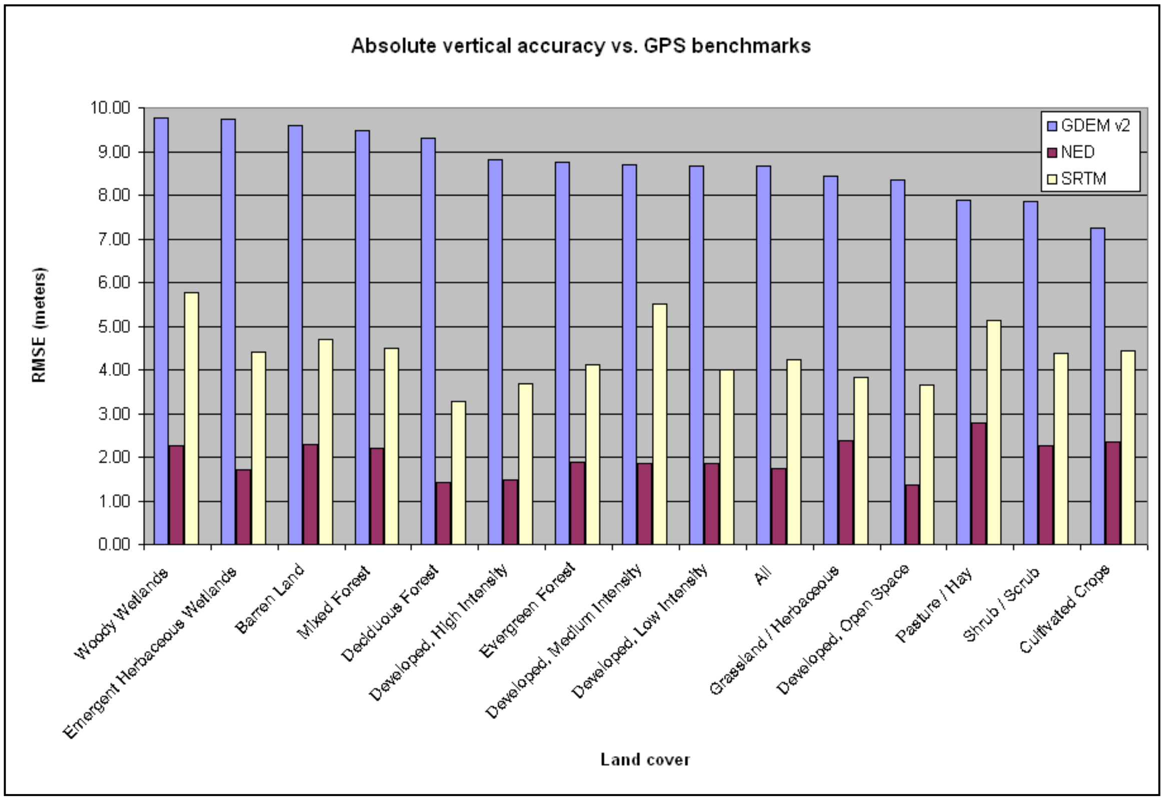

Figure 8. GDEM v2, NED, and SRTM RMSE by land cover class.

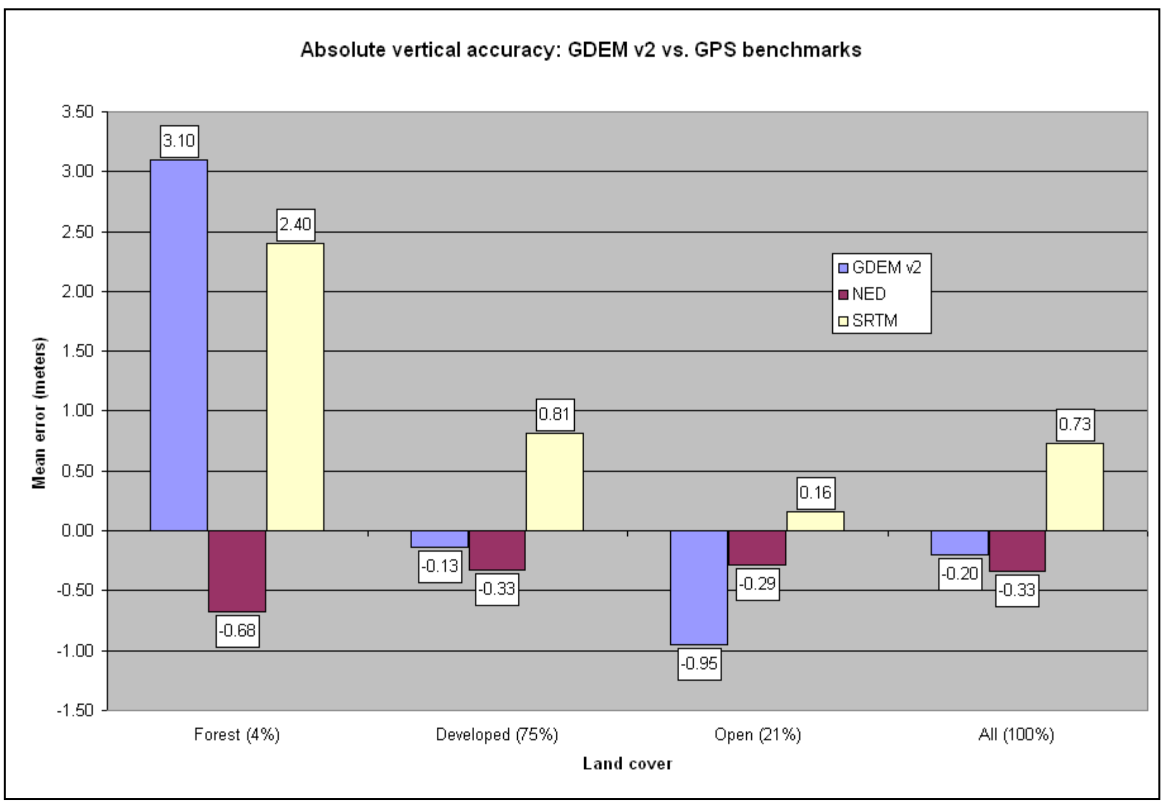

Figure 9. GDEM v2, NED, and SRTM mean errors by aggregated land cover class.

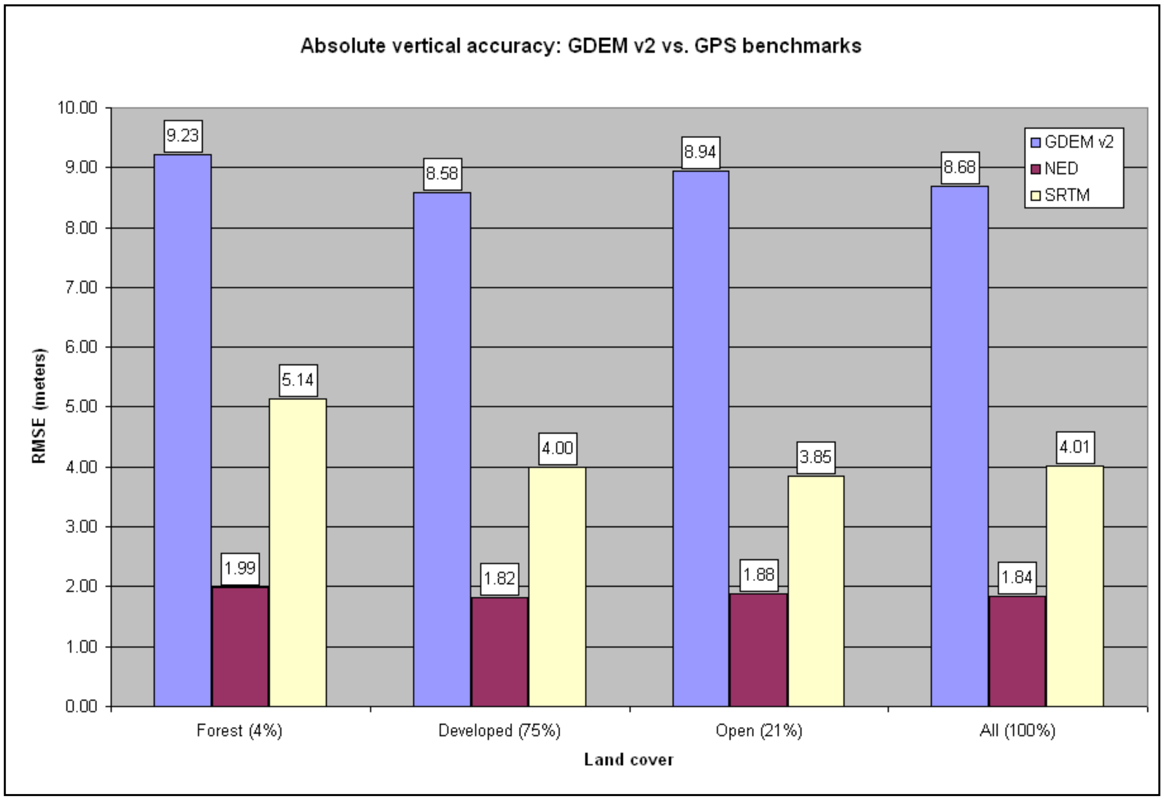

Figure 10. GDEM v2, NED, and SRTM RMSE by aggregated land cover class.

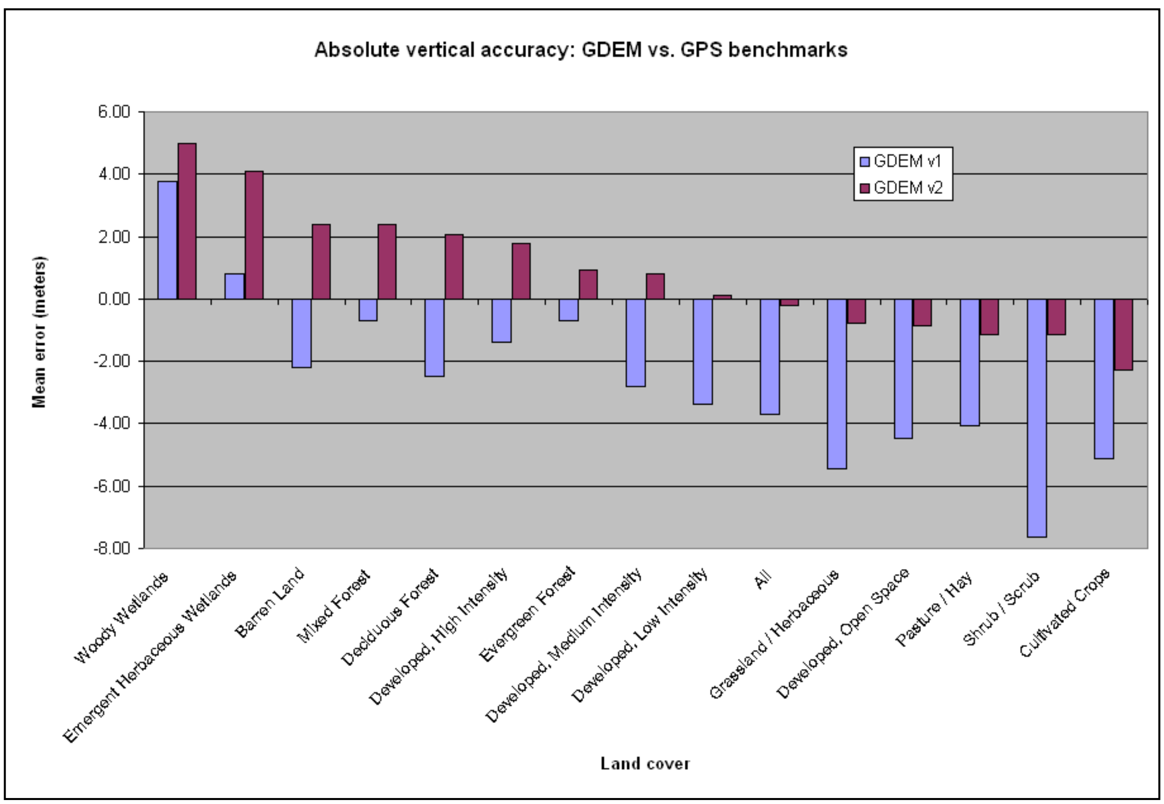

Figure 11. GDEM v2 and GDEM v1 mean errors by land cover class.

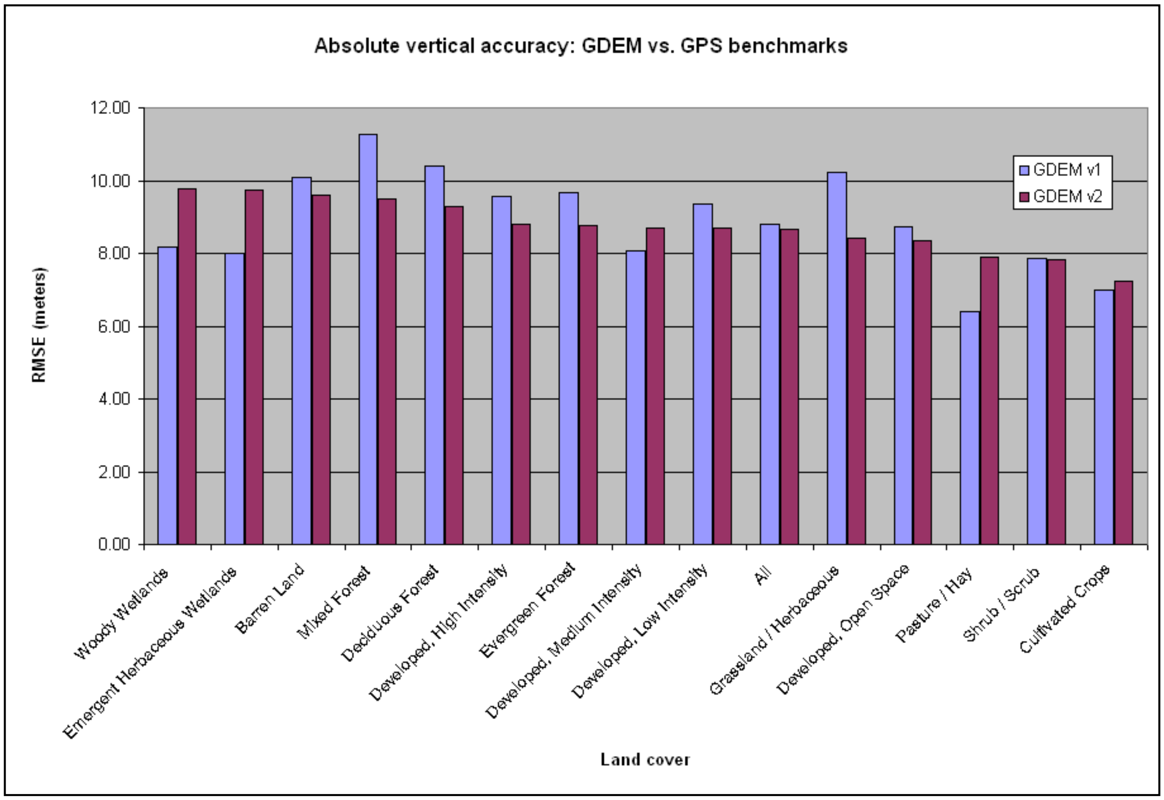

Figure 12. GDEM v2 and GDEM v1 RMSE by land cover class.

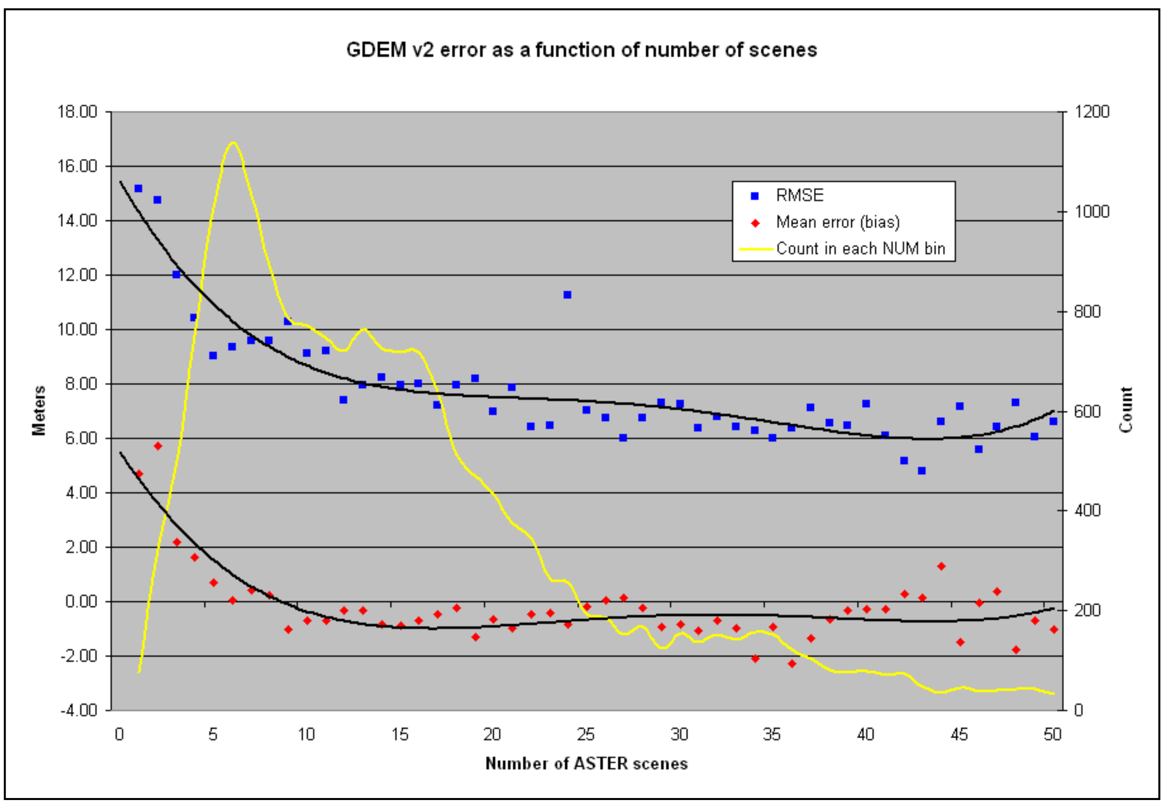

Figure 13. GDEM v2 mean error and RMSE vs. number of scenes used for elevation calculation.

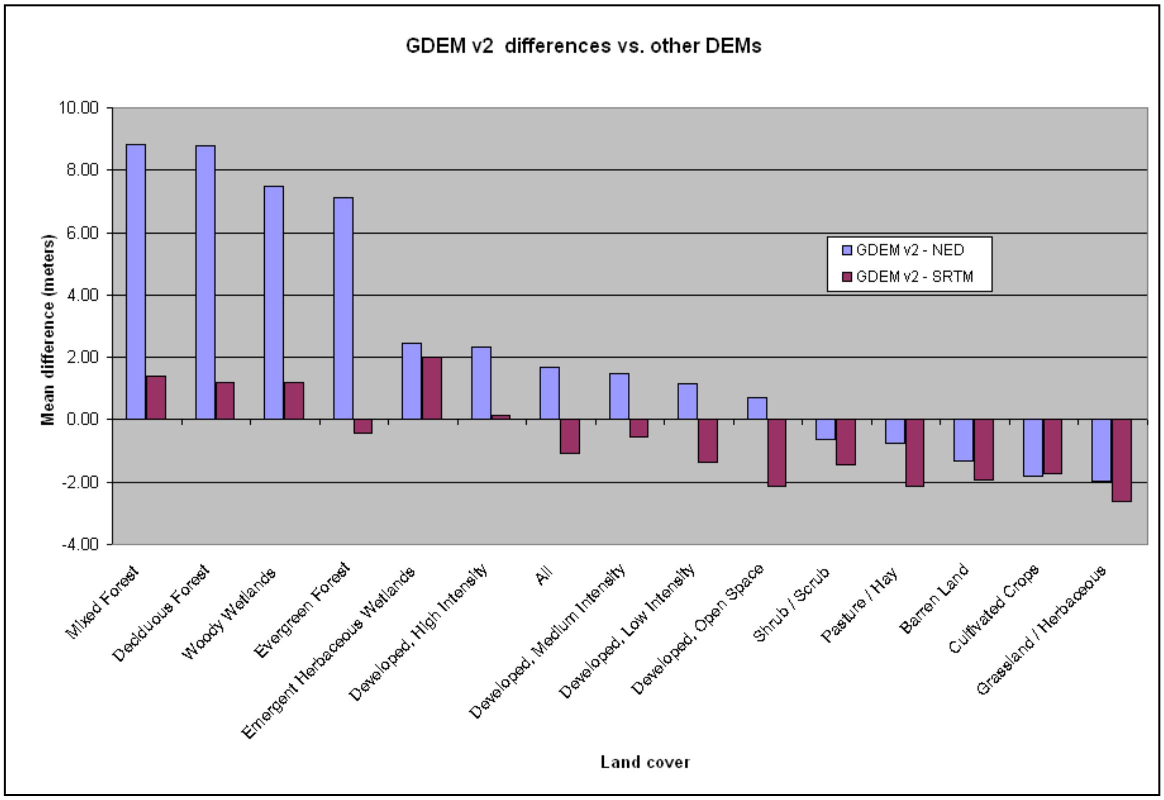

Figure 14. GDEM v2-NED and GDEM v2-SRTM mean differences by land cover class.

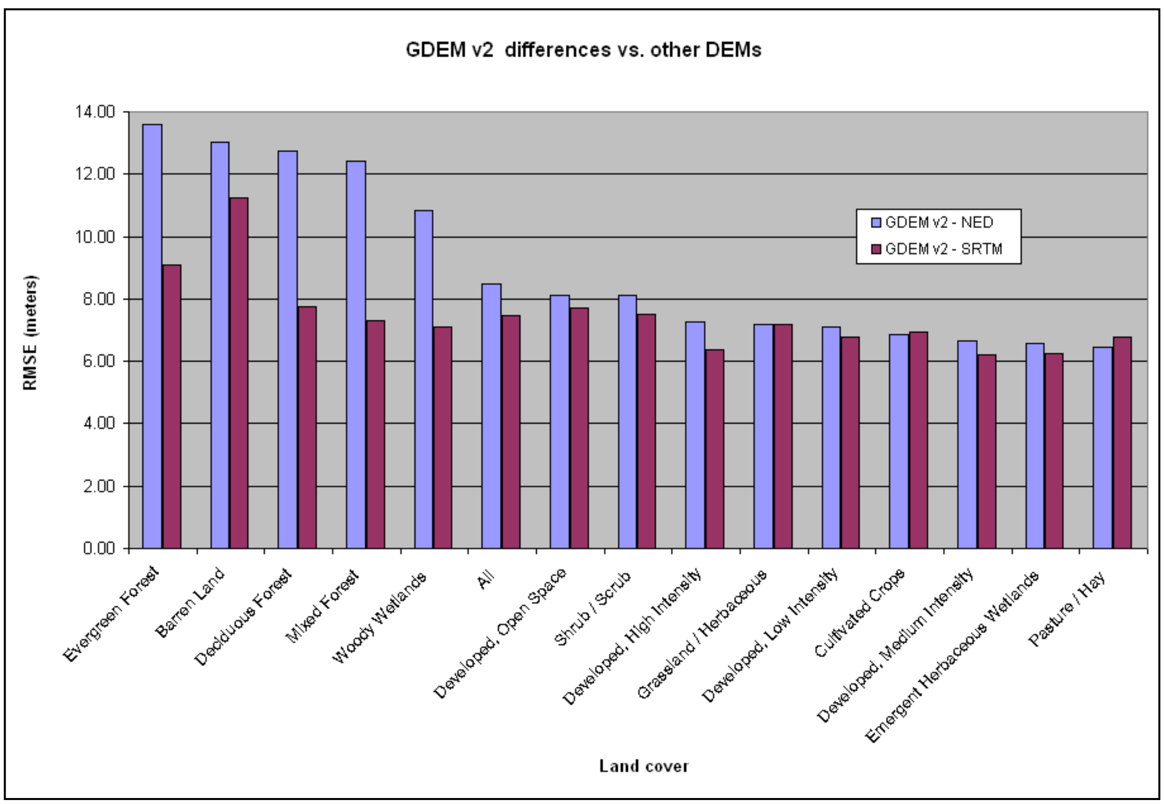

Figure 15. GDEM v2-NED and GDEM v2-SRTM RMSE by land cover class.

Land cover class	Description	GDEM v2 mean error vs. GPS benchmarks	GDEM v2 mean difference vs. NED
Developed, Open Space	• mostly lawn grasses, with some construction • <20% impervious surfaces • large-lot single-family housing units, parks, golf courses	-0.86 m	0.72 m
Developed, Low Intensity	• 20-49% impervious surfaces • single-family housing units	0.12 m	1.16 m
Developed, Medium Intensity	• 50-79% impervious surfaces • single-family housing units	0.79 m	1.48 m
Developed, High Intensity	• 80-100% impervious surfaces • apartment complexes, row houses, commercial/industrial	1.77 m	2.33 m

Table 2. Increasing GDEM v2 mean error with increasing density of developed land cover.