ANALYSIS AND BIAS IMPROVEMENT OF HEIGHT MODELS BASED ON SATELLITE IMAGES

K. Jacobsen[1], R. Passini[2]

[1] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany - jacobsen@ipi.uni-hannover.de
[2]Mount Laurel, NJ, USA - ricardo.passini@outlook.com

Commission I, WG I/4

KEY WORDS: Height models, SRTM, AW3D30, TDM90, Analysis, Accuracy, Bias improvement, Morphology

ABSTRACT:

Height models are a fundamental part of the geo-information required for various applications. The determination of height models by aerial photogrammetry, LiDAR or space images is time-consuming and expensive. For height models with large area coverage, UAVs are not economic. The freely available height models ASTER GDEM-3, SRTM, AW3D30 and TDM90 can meet various requirements.
With the exception of ASTER-GDEM-3, which cannot compete with the other, the digital surface models SRTM, AW3D30 and TDM90 are analyzed in detail for accuracy and morphology in 4 test sites using LiDAR reference DTMs. The accuracy figures root mean square error, standard deviation, NMAD and LE90 are compared as well as the accuracy dependence on the terrain inclination. The analysis uses a layer for the open areas, excluding forest and settlement areas. Remaining elements that do not belong to a DTM are filtered. Particular attention is paid to systematic errors. The InSAR height models SRTM and TDM90 have some accuracy and morphological restrictions in mountain and settlement areas. Even so, the direct sensor orientation of TDM90 is better than for the other. Optimal results in terms of accuracy and morphology were achieved with AW3D30 corrected by TDM90 for the local absolute height level. This correction reduces the bias and also the tilt of the height models compared to the reference LiDAR DTM.

1. INTRODUCTION

Digital height models (DHM) are a basic requirement for geographic information systems (GIS) and various other applications. In addition to traditional mapping of survey administrations, large area covering DHM can be based on optical satellite images and on interferometric synthetic aperture radar (InSAR). With InSAR images from the Shuttle Radar Topographic Mission (SRTM) in 2000, the first almost global coverage height model with a homogenous and for several applications satisfactory accuracy was generated (Gesch et al. 2014). The SRTM DHM is used today as standard DHM for various purposes. Two pairs of interferometric antennas were used the Space Shuttle in 2000 - the US C-band and the German X-band. DHMs were created with both, but the X-band had the disadvantage of a smaller swath width with large gaps between the covered ones. For this reason the SRTM X-band DHM is not actually used.
The next notable large area covering DHM was in 2009 ASTER Global Digital Elevation (ASTER GDEM), based on the stereo models of the Japanese optical satellite TERRA ASTER with 15m ground sampling distance (GSD), later improved to ASTER GDEM2 (Tetsushi et al. 2011) and since August 2019 to ASTER GDEM3 (ASTER 2019), where more ASTER scenes are used and a water body tile is now included, especially improving the height model at shore lines.
The basic conditions for height models are significantly better with the Japanese tri-stereo satellite camera ALOS PRISM. ALOS PRISM, which was operated from 2006 to 2011, had a GSD of 2.5m and a better base-to-height-ratio up to 1.0 (Tadono et al. 2014, Takaku et al. 2014). The Japanese space organization JAXA has created the commercial DHM ALOS World 3D (AW3D) with 5m point spacing with all available ALOS PRISM images. AW3D with 30m point spacing (AW3D30) can be downloaded as free, reduced version.

The latest investigated height model is based on the German TanDEM-X InSAR satellite configuration. The German Aerospace Center (DLR) created with all from 2013 up to 2016 taken TanDEM-X InSAR image combinations, the global height model that is now commercially distributed by Airbus Defence and Space as WorldDEM (Wessel et al. 2018). It has a point spacing of 10m. A reduced version is available free of charge as TanDEM-X with (approximately) 90m point spacing (TDM90) (DLR 2018). This DHM has a point spacing of 3 arcsec (~92m at the equator).
All these height models are digital surface models (DSM) with the height of the visible surface. The X-band radar only slightly penetrates the vegetation. If digital terrain models (DTM) with the height of the bare ground are required, the DSM has to be filtered for elements that are not on the bare ground. This can only be successful if points are available on the bare ground, but it fails in areas such as forest where no points are available on the bare ground (Passini et al. 2002). (Smith and Berry, 2011) attempted to improve the filtering of SRTM 3 arcsec DSMs through a combination with ICESat height profiles. This was partially successful in forest areas where ICESat height profiles were available, but the generated ACE2 DTM shows major problems in areas between the height profiles (Aldosari and Jacobsen 2019), which means that the ACE2 DTM is not reliable.
Height models, which are determined by image matching of optical satellite images, have the advantage of good morphologic information in relation to InSAR DTMs, which have problems in built-up and in mountain areas due to foreshortening and radar shadows. In the TanDEM-X mission, this problem was reduced in mountain areas by a combination of flights with different viewing directions and different base length (Wessel et al. 2018). However, DSMs based on optical images show better morphologic details as InSAR DSMs (Aldosari and Jacobsen 2019). On the other hand, the absolute geo-referencing of radar images is better than for optical

images, which are limited to the accuracy of the satellite attitude. Both methods could be combined by combining the advantages of height models of both types. Radar imaging does not depend on cloud coverage as optical images. In areas that are permanently covered by clouds, gaps cannot be avoided by optical images.

The geometric accuracy of the height models is described by various accuracy specifications as root mean square error (RMSE), standard deviation of height (SZ), normalized medium absolute deviation (NMAD) (Höhle and Höhle 2009), linear error with 90% probability (LE90) or even the medium error. All of these accuracy descriptions have some justifications, but make it difficult to compare publications with different accuracy figures, also due to different threshold values for accepted height discrepancies between the investigated and the reference height model and the different character of the test fields.

Height models are required as reference information; ideally, these are LiDAR height models. Ground control points (GCPs) for reference are not optimal – GCPs are usually located in flat areas that do not correspond to the terrain being analyzed. An investigation with ICESat height profile points shows a trend for accuracy, but does not provide the correct accuracy for rough terrain due to the 66m diameter of the ICESat footprint.

Several publications with accuracy tests are available, particularly for the SRTM-DSM, but also for the other DSM included here. However, a direct comparison is often difficult due to different accuracy figures, different threshold values, different handling of systematic errors and often missing information on the influence of terrain inclination, separation of the land cover classes and filtering. There are no publications on the combination of the advantages of height models from InSAR and optical images. In addition, the influence of interpolation between the height points often is ignored or is not known. An attempt is made to make this situation clearer in the investigations described below.

2. ANALYZED HEIGHT MODELS

	SRTM (C-band)	ASTER GDEM 3	AW3D30	TDM90
organization	NGA+ NASA	METI, NASA	JAXA	Airbus / DLR
source	Shuttle InSAR	Optical ASTER	Optical ALOS PRISM	TanDEM-X InSAR
Latitude range	56° S to 60° N	+/- 83°	+/- 82°	global
Spacing	1 arcsec	1 arcsec	1 arcsec	3 arcsec

Table 1. Analyzed height models

The freely available height models listed in Table1, which cover almost or completely worldwide, were analyzed and the SRTM and AW3D30 were improved by the high absolute accuracy of TDM90. All of these height models are digital surface models (DSM) with the height of the visible surface. Two of them are based on optical space images and two of them are determined by InSAR. Both methods have their advantages and disadvantages.

The Shuttle Radar Topography Mission (SRTM) had InSAR data acquisition for 11 days in February 2000. A few years later, the height models based on it were freely available with a point spacing of 3 arcsec (~92m at the equator) and later with 1 arcsec (Gesch et al. 2014). Due to the 11 day observation, only one up to two covers was possible. In steep mountain areas, the radar layover caused larger gaps that have been filled with other

data such as height models from SPOT V HRS. The SRTM DSM got some geometric improvements over time.

The tri-stereo optical satellite images of ASTER had the disadvantage of only 15m GSD, which limited the accuracy of the height models generated. The standard deviation of the ASTER GDEM 2 and also of the GDEM 3 is the range of 6m to 7m even in flat terrain (Aldosari, Jacobsen 2019). For this reason, the results obtained in the test sites are not shown in this paper.

With the 2.5m GSD from ALOS PRISM images and the height-to-base-relation up to 1.0, significantly better conditions are provided for the generation of height models such as those from ASTER. During the operating period from 2006 up to 2011, multiple covers were achieved despite the problem of cloud cover. Only a few areas required a gap filling with other data. Together with the commercial version with 5m point spacing, the freely available AW3D30 (1 arcsec point spacing) has also been permanently improved. The AW3D30 comes with a mask-file showing the areas covered by clouds, snow and ice and a layer showing land and water. A quality assurance file contains areas of problem and the gap filling with other data. A stack file shows the number of covers. For the test sites, the covers were achieved on the average 4.0 times.

TDM90 is the latest of these height models, released in fall 2018. A very precise height model was generated based on the optimal and flexible TanDEM-X InSAR combination (Wessel at al. 2018). The reduced data with 3 arcsec point spacing is available free of charge. The height model comes together with merged radar images, a file with the number of covers, a height error map with estimated height accuracy, and a water mask. The average number of covers varies between 2.7 and 7.5 for the test sites used here. As it is common in mountain areas, the highest number of covers is available in test site Mountain in order to reduce the influence of radar foreshortening and shadows. The other height models are related to the earth geoid, only TDM90 has ellipsoid heights that require a transformation to the geoid. This was done with the EGM2008, which is available free of charge (Pavlis et al. 2012). EGM2008 is more accurate as EGM96 used for geoid height determination in earlier versions of SRTM and AW3D30, but this belongs to the accuracy of the height values determined by comparison with the LiDAR data which are related to the national geoids.

3. TEST DATA

Figure 1. Google image of test site "Flat"
Center: W95.91° N30.50°, 18km x 15km
Height range: 72m – 139m, DSM spacing 29m / 86m
Average multiple coverage for TDM90 = 2.7 and for
AW3D30=3.8; influence of interpolation over 29m= 0.25m,
over 86m= 0.53m (terrain roughness)

LiDAR digital terrain models (DTM) from the US surveying authorities could be used as reference. They had to be converted from the US state plane coordinate systems (Lambert and Transverse Mercator (Snyder 1987)) to UTM WGS84, as also the investigated height models from geographic to UTM WGS84, which was used for the investigation. The LiDAR DTMs have a point spacing of 2.5m and a standard deviation better than 20cm. Four test sites with different topography were used, called Flat, Rolling, Mountain and City (Figures 1 – 4).

The point spacing of the analyzed DSMs in east-west direction corresponds to the point spacing in north-south direction multiplied by the cosine of latitude. The spacing listed under Figure 1 is the average spacing in both directions for 1 arcsec and for 3 arcsec point spacing.

The test site Flat (Figure 1) is not completely flat, but the terrain inclination is limited. The area is partially covered by forest and includes single buildings and some individual trees.

With the reference LiDAR DTM with a point spacing of 2.5m, the root mean square influence of bilinear interpolation over 1 arcsec and 3 arcsec was computed – also listed in the details of the test site.

Figure 2, Google image of test site "Rolling"
Center: W98.94° N31.73°, 13km x 12km
Height range: 380m – 498m, DSM spacing 29m / 86m
Average multiple coverage for TDM90 = 2.5 and for AW3D30 = 5.1; influence of interpolation over 29m= 0.46m, over 86m= 0.77m (terrain roughness)

The test site Rolling (Figure 2) has a hilly character. It includes forest and some built-up areas.

Figure 3. Negative Google image of test site "Mountain"
Center: W104.78° N31.98°, 5.2km x 2.8km
Height range: 1520m – 2350m; average spacing 29m / 86m
Average multiple coverage for TDM90 = 7.5, AW3D30 = 7.4; influence of interpolation over 29m= 2.99m, over 86m= 7.33m (terrain roughness)

The "Mountain" test site (Figure 3) is located in New Mexico and has no significant vegetation. The Google image is converted into negative to give the correct human impression of hill shading on the northern hemisphere. The rough character is shown by the root mean square of the interpolation over 29m and 86m. For this purpose, the root mean square discrepancies of all LiDAR reference points in the mesh 29m x 29m, respectively 86m x 86m against the bilinear interpolation was computed.

A suburb of Houston is covered by the test site "City" (Figure 4). It is not densely built-up and also includes forest areas. The ground is visible between the buildings.

Figure 4. . Google image of test site "City"
Center: W95.55° N29.52°, 9km x 7km
Height range: 3m – 45m, average spacing 29m / 87m
Average multiple coverage for TDM90 = 2.5 and for AW3D30 = 4.2; influence of interpolation over 29m= 1.33m, over 87m= 1.80m (terrain roughness)

4. ANALYSIS RESULTS

4.1 Systematic height error (bias)

The direct sensor orientation of the optical and radar imaging satellites has some limitations. It is a little more complicated for optical images due to the limited attitude accuracy. In case of radar, the attitude is replaced by the inclined distance to the object, which can be more accurate. In addition, the technology has been improving since year 2000 when SRTM took place.

The geo-reference was checked in X, Y and Z by the Hannover program DEMSHIFT. The shifts of the investigated height models against the LiDAR-reference in X and Y are not only caused by the geo-reference of the DSMs, but typically also influenced by datum errors of the reference. The shifts in X and Y between the height models are between decimetres and up to 29m for SRTM. These horizontal shifts were used as pre-correction before the analysis.
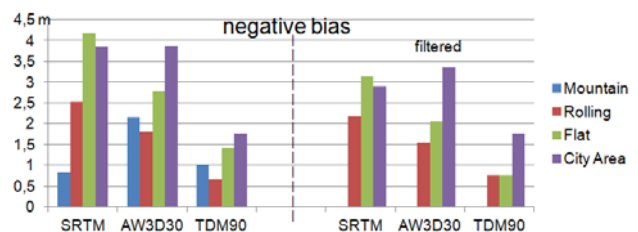

Figure 5. Negative Z-bias against LiDAR DTM for the open areas of the investigated test sites and all in detail shown DSMs

Datum problems of the reference are usually not a problem – systematic height errors are dominated by vertical shifts of the analyzed height models. The systematic height errors are somewhat more difficult, as will be shown later. Nevertheless, the average height shifts of a test site are an important indicator.

Due to the situation that the analyzed height models are DSMs and the references are DTMs, this analysis is only carried out for the open areas (excluding forests and built-up areas). But there are also individual houses and trees in the open areas that have to be filtered (Figure 5). No filtering is required for the test site Mountain, as there are no trees and buildings available. Of course, the individual results are random, but the four test sites show a clear trend – the bias of TDM90 is the least, followed by AW3D30 and then SRTM. This is confirmed by other test sites used in (Aldosari and Jacobsen 2019) and the intensive investigation of the TanDEM-X height models in (Wessel et al. 2018), which leads to an RMSE of 1.3m for height points on roads. Even so, the accuracy achieved on flat roads is not the same as with the entire height model including also rough terrain.

Figure 6. Test site Flat, color coded height differences against LiDAR, left: AW3D30, right: TDM90, color scale for both: +/-10.5m

Figure 7. Test site Flat, number of stacks, left: AW3D30 – up to 10 covers, right: TDM90 with 2, respectively 3 covers

Figure 6 shows the color coded height differences between AW3D30 and LiDAR as well as the height differences for TDM90. The red color indicates built-up areas and forest with significant height differences between the DSMs and the reference DTM. Most of these parts have been removed by a layer – the remaining parts are called "open area". In Figure 6, left side, in addition to the red areas on the left, there is a green color and on the right side there is a yellow color, which corresponds to the higher number of stacks on the right, shown by Figure 7 on the left. The yellow color corresponds to height differences of -4.5m up to -1.5m, while green corresponds to -1.5m up to +1.5m. This means that there are systematic height differences of the AW3D30 DSM especially for right hand side this stereo model. Similar effects are available for AW3D30 DSMs at other test sites. No files with number of covers are available for SRTM, but the systematic errors are also not the same for the whole test sites. Figure 7 also shows that the optical images are acquired with descending orbit while the InSAR images are from ascending orbit.

The systematic errors of the TDM90 scenes show no correlation to the number of covers, there is only a correlation to land cover classes in which the vegetation height is not taken into account by a layer or by filtering. This indicates a higher accuracy of TDM90 and smaller systematic errors (bias).

The significantly better absolute height orientation of TDM90 was used by the Hannover program ZFIT for an improvement of AW3D30 and SRTM. ZFIT uses a height correction that corresponds to average height differences between AW3D30, respectively SRTM against TDM90 for a moving average with a radius of 3000m or 6000m. Individual height values have almost no influence, but the general height level is adapted to TDM90. This is marked by "ZFIT" in the figures with the accuracy results.

4.2 Morphologic quality

The morphological quality can be recognized by the contour lines (Figures 8 – 10). Of course, the LiDAR reference shows the best of details, followed by AW3D30 and then TDM90 and finally the SRTM. TDM90 has the disadvantage of a point spacing of 3 arcsec compared to a point spacing of 1 arcsec point spacing for AW3D30 and SRTM. SRTM is not as precise as TDM90, but on the test site Mountain SRTM has some advantages over TDM90 due to the smaller point spacing.
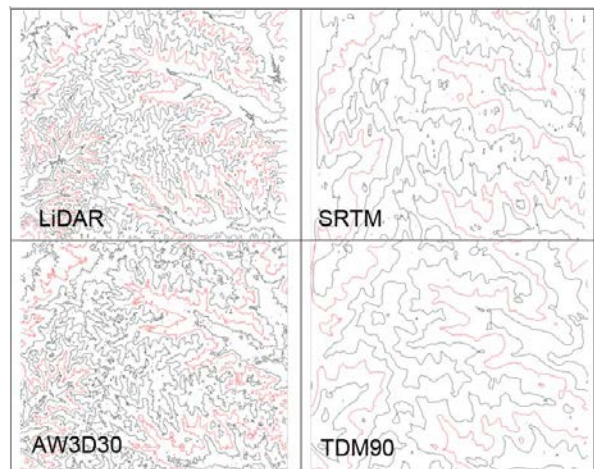

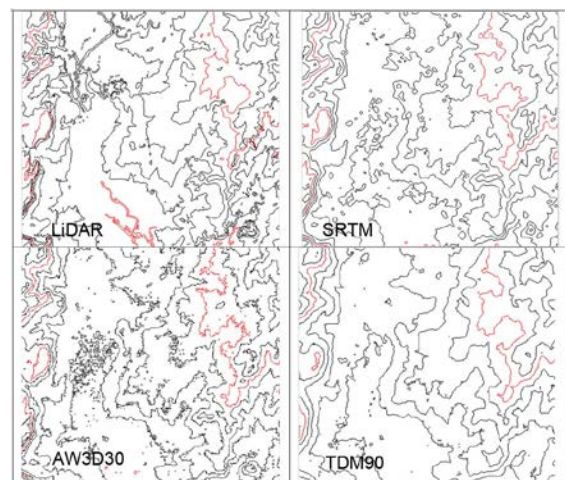Figure 8. Contour lines test site Flat, contour interval 10m

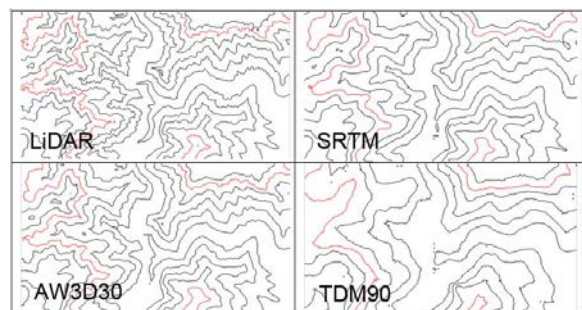Figure 9. Contour lines test site Rolling, contour interval 10m

Figure 10. Contour lines test site Mountain, contour interval 100m

A visual comparison of the contour lines based on SRTM, AW3D30 and TDM90 with the reference LiDAR contour lines shows the lowest details based on the 3 arcsec point spacing of TDM90. SRTM with the same point spacing as AW3D30 of 1 arcsec shows less details than AW3D30. AW3D30 shows the best match with the LiDAR contour lines. In flat areas, as partially in figure 9, the higher details of AW3D30 show noise caused by single trees and small forest parts – this is not the case with the smoothing of the SRTM InSAR and the lower point spacing of TDM90.

As a result of this pre-analysis, it can be mentioned that AW3D30, improved by ZFIT, combines the advantage of morphologic quality with the absolute height accuracy of TDM90.

4.3 Test site "Flat"

On the test side Flat there are some forest parts and a few built-up areas. These areas have been masked off for final accuracy validation. It is important to have the knowledge of the system accuracy in open areas with no forest and buildings. This result can be transferred to other areas. Forest and buildings always have an individual influence on the height model, depending on the individual object height.

With a height range of 67m, this test site is not completely flat - the average terrain inclination over 2.5m distance of the LiDAR DTM is only 0.06 (3°) and over 32m 0.09 (5°), but the result is strongly influenced by single trees and buildings that require final filtering for such objects not belonging to a DTM.
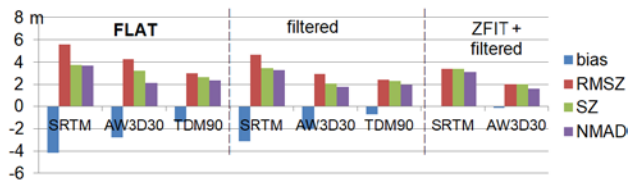

Figure 11. Accuracy results in open area of test site Flat

Due to the limited slope, the accuracy information for terrain inclination up to 0.1 (6°) is almost identical to the information including all terrain inclinations (Figure 11). The filtering improves the results, especially the bias and RMSZ, as larger differences are eliminated. The improvement of the absolute height level through TDM90 with the ZFIT program reduces RMSZ in particular, while SZ and NMAD are respecting the bias from the start. The filtered height values improved by ZFIT lead to an SZ of 2.02m and an NMAD of 1.62m for AW3D30 and 3.43m respectively 3.15m for SRTM. TDM90 alone still has a standard deviation of 2.33m and an NMAD of 1.97m. The filtering reduces the bias of all 3 data sets. The fit of SRTM and AW3D30 and the filtering nearly eliminates the bias totally (Figure 6, right hand side).

The RMSZ is influenced by the bias (Figure 5), while this is not the case with SZ and NMAD. The linear error with 90% probability LE90 depends on the 10% of largest height discrepancies. With normal distributed discrepancies, the relationship between RMSZ and LE90 is 1.65. In the test site Flat the relationship between RMSZ and LE90 is in the range between 1.51 and 1.76, which is not too far from the relation of the normal distribution. The normalized medium absolute deviation NMAD is not as strongly influenced by a higher number of larger discrepancies as corresponding to the normal distribution. Such a higher number may be caused by objects that do not belong to the DTM, which can be reduced by filtering.
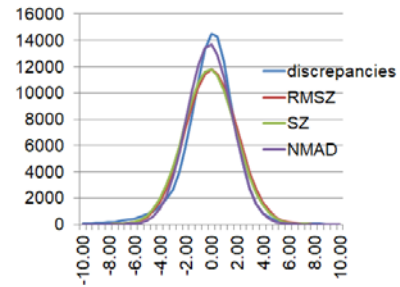

Figure 12. Frequency distribution of AW3D30 discrepancies improved by ZFIT and filtered against LiDAR

The justification and meaning of the accuracy numbers must be seen in relation to the frequency distribution of the height discrepancies. Figure 12 shows the frequency distribution according to "discrepancies" and normal distributions based on RMSZ, SZ and NMAD. The frequency distribution of AW3D30 improved by ZFIT and filtered against LiDAR (Figure 12), is no longer influenced by the bias, which was eliminated by adapting AW3D30 to TDM90 (ZFIT). As in most cases, the NMAD-based normal distribution fits the frequency distribution better than the other. Only in the range of the discrepancies from -4m up to -10m does the frequency distribution show higher numbers, but only limited total numbers than the normal distribution based on NMAD. For this reason, the square sum expressions RMSZ and SZ with 2.03m respectively 2.02m are larger than NMAD with 1.68m. The filtering cannot remove all vegetation influences above the DTM determined by LiDAR. Of course, LE90 with 3.15m has a justification, even if the relationship to RMSZ is close to the relationship of 1.65 for the normal distribution. LE90 only gives reliable information about the size of 90% of the observations that are smaller than this number, but it is not a correct description of the accuracy characteristic. NMAD best describes the frequency distribution; on average over the entire range, the discrepancies of the normal distribution based on NMAD compared to the frequency distribution are 4.03%, while this is 6.24% for SZ. RMSZ corresponds to SZ if there is no bias.

The RMSZ influence of interpolation over the distance of 29m spacing is limited to 0.25m. This means that in this test site the height values improved by ZFIT and filtered for AW3D30 have an NMAD of 1.66m instead of 1.68m without the influence of interpolation. Nevertheless, the influence of interpolation is part of the accuracy description, but the influence of interpolation changes from test site to test site.

4.4 Test site "Rolling"

The test site Rolling has a greater height range of 118m. Nevertheless, the average slope is smaller than for the test site Flat, which is caused by the less changing vegetation height. Test site Rolling has some built-up and forest areas that also require a layer for the open area. The greater influence of the rolling terrain can be seen from the influence of the interpolation over 29m, which is 0.46m instead 0.25m for test site Flat. For the spacing 86m it is 0.77m instead 0.53m for Flat.
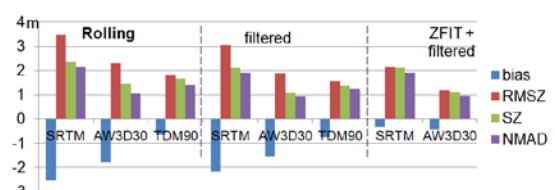

Figure 13. Accuracy results in open areas of test site Rolling

Also in the test site Rolling the influence of the terrain inclination to the accuracy figures is limited. For the open areas SZ for TDM90 is 1.68m and for the terrain with a slope below 0.1 (6°) with 1.53m it is only 9% better. It is similar for the other accuracy numbers. For AW3D30, improved by ZFIT and filtered, the SZ is 1.11m and for the terrain below an inclination of 0.1 it is 1.07m. Due to the larger size of the accuracy numbers for SRTM, the influence of the terrain inclination even is smaller.

In general, the trend of the accuracy relations is similar to that of the test site Flat, but the accuracy numbers are smaller for the open area in test site Rolling due to smaller influence of vegetation. Here too, without the influence of the bias, the accuracy figures for AW3D30 are smaller than for TDM90, which cannot only be explained by the spacing of 86m of TDM90.

In the test site Flat, the bias was reduced to -0.13m for SRTM and -0.06m for AW3D30 by adapting it to TDM90. On the test site Rolling the remaining bias is greater with -0.31m respectively -0.43m, but even this is a very good result.

The frequency distribution of the AW3D30 discrepancies improved by ZFIT and filtered, against LiDAR for the test site Rolling is very similar to the test site Flat.

4.5 Test site "Mountain"

The test site Mountain is very different from the other. There is no vegetation to be seen in the dry mountain region, and there is no built-up area. Because of this, no layer for open areas and no filtering for elements that do not belong to the DTM are required. The height range is 830m and the average terrain inclination is around 0.4 (22°) with 5% over 45°. As it is common in mountain areas, the number of TDM90 covers is higher with an average of 7.5. Even for AW3D30, the number of covers is 7.4, which is due to a low cloud cover in the dry area of New Mexico. The rough terrain increases the RMSE of the interpolation over 29m to 2.99m and for a spacing of 86m to 7.33m, which strongly influences the accuracy figures.
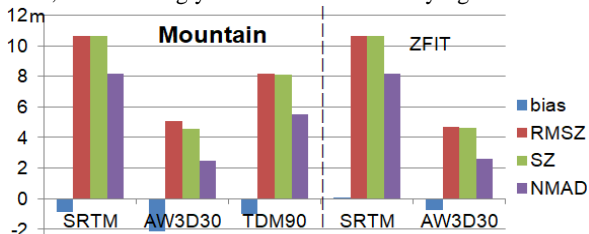

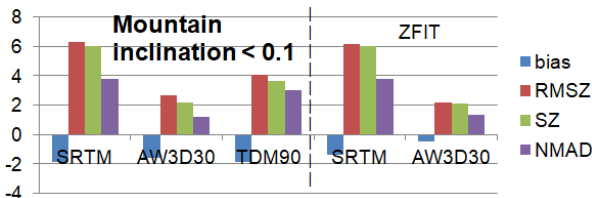Figure 14. Accuracy results in test site Mountain

Figure 15. Accuracy results in test site Mountain for inclination < 0.1

The accuracy figures are considerable larger than at the other test sites (Figure 14), which is caused by steep slope of the terrain. The accuracy is better for terrain inclination < 0.1 (6°) (Figure 15), but not as good as for the other test sites. As already mentioned, the interpolation has a strong influence especially for this steep terrain. InSAR has problems with radar lay over and shadows in such a mountain area. This is reduced by the number of covers for TDM90 with different base length

and viewing directions (Figure 16), but this cannot compensate everything. Multiple covers could not be performed for SRTM in the short imaging time of only 11 days. For SRTM, the RMSZ can be expressed by the relationship RMSZ=6.48m + 18.3m * tangent (slope).
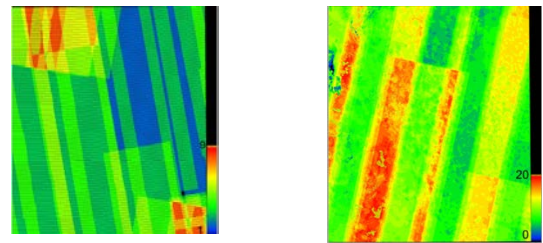

Figure 16: number of covers (colour scale on lower right), left: TDM90, right: AW3D30

ALOS has a sun-synchronous orbit in which the images are recorded in descending mode (Figure 16 right). The pair of radar satellites TanDEM-X has also a sun-synchronous orbit, but the imaging can also be done on the night side of the earth with an ascending orbit. The ascending trajectories dominate for TanDEM-X, but small parts are also recorded in descending orbits (Figure 16 left, upper left and lower right).
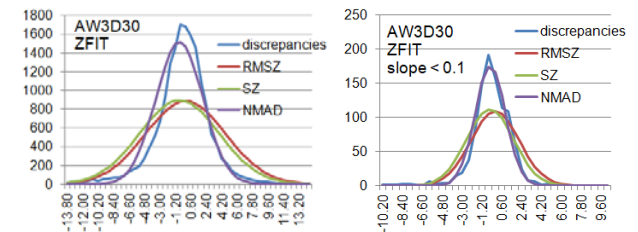

Figure 17. Test site Mountain,
Frequency distribution AW3D30 with ZFIT
left: whole data set   right: only for inclination < 0.1

The frequency distribution of the height discrepancies in test site Mountain for AW3D30 (Figure 17), improved by height shift to TDM90 (ZFIT) shows, as usual, a better fit of the normal distribution based on NMAD to the frequency distribution than the other accuracy numbers. This also applies to the height discrepancies for the terrain inclination below 0.1 (6°). SZ fits to the frequency distribution with deviations of 16.5%, while this applies for NMAD with 7.9%; limited to the discrepancies below an inclination of 0.1 it is for SZ 9.5% and for NMAD 4.0%.

4.6 Test site "City"

The test site City is a suburb of Houston. It is not densely built with green spaces between the buildings and approximately 50% of the area not belonging to built-up. The height range is 42m. The RMSE influence of interpolation over 29m with 1.33m and over 87m with 1.80m is greater than for an open area with a similar low undulation. This is caused by the influence of buildings and vegetation.
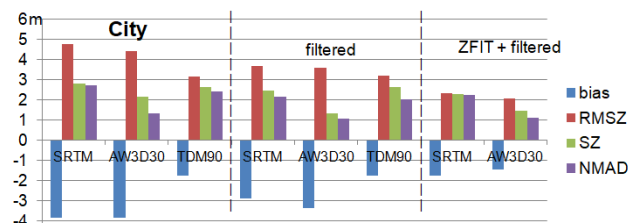

Figure 18. Accuracy results in open areas of test site City

InSAR has some problems in built-up areas with layover. The ground level in street canyons often cannot be determined (Soergel et al. 2013). Even the "open area" includes individual buildings and individual trees, which leads to discrepancies from the reference LiDAR DTM. Filtering can reduce this problem, but cannot eliminate it entirely. The large bias (Figure 18) is caused in part by objects remaining even in the filtered area. TDM90 also has a bias of -1.75m, which is apparently due to the InSAR problems in built-up areas. For this reason, SRTM and AW3D30 also have such a bias if they are adapted to TDM90 by program ZFIT. So the RMSZ of AW3D30, filtered and adapted to TDM90, is therefore significantly larger than at the test sites Flat and Rolling. Due to the flat surface, the terrain inclination has no influence to the result.
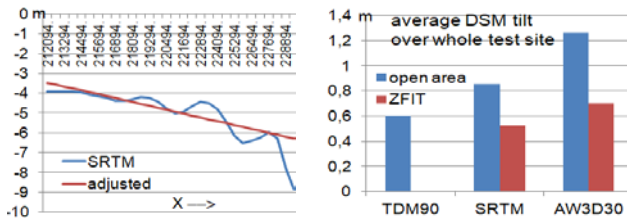
4.7   DSM tilt

Figure 19. Left: DSM-tilt in X-direction, SRTM in test site Flat
Right: average absolute DSM-tilt all test sites over a distance of 17.4 km (1m corresponds to 0.003°)

The height models are not levelled with respect to the reference LiDAR. As an example, the tilt in the X-direction of the SRTM-DSM in test site Flat against the reference is shown. The height discrepancies are averaged in X-groups over the entire DSM, shown in Figure 19, left, as blue line. The adjusted tilt, weighted by the number of observations in the groups, is shown as red line. In this case, the model tilt goes from -3.5m to -6.3m, which corresponds to a model tilt of 2.8m over the test site of 17.4 km. On the right side of Figure 19, the averaged absolute tilt in X- and Y-directions of all test sites for the three types of height models is shown. The TDM90 has a smaller tilt than the other. If the SRTM- and the AW3D30-DSMs are improved by fitting to TDM90, due to the averaged local fitting, the model tilts of SRTM and AW3D30 are also improved, as shown by the red bars in Figure 19, right.

5.   DISCUSSION

The worldwide or almost worldwide digital surface models ASTER GDEM, SRTM, AW3D0 and TDM90 are available free of charge. TDM90 has a point spacing of 3 arcsec (~92m at the equator), the other have a point spacing of 1 arcsec. For a comparison of the height models, the accuracy and the morphologic quality are important. The results obtained with ASTER GDEM are clearly not as good as for the others, which is why the results are not shown in detail here.
The individual test fields show varying results; still, there are some clear trends.
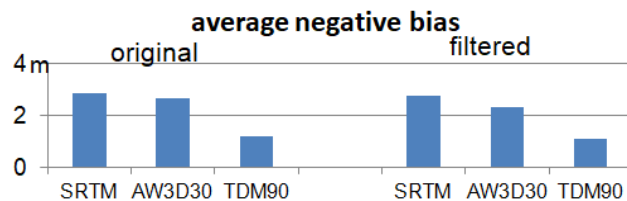

Figure 20. Averaged negative Z-bias of all test fields together

The averaged Z-bias of the test sites (Figure 20) shows a clear trend – the oldest DSM, SRTM, has the largest bias, followed by AW3D30. TDM90 with a bias of -1.21m and improved by filtering to -1,09m, clearly has the lowest bias and can be used to correct of AW3D30 and SRTM height models by a floating average of the height differences of SRTM respectively AW3D30 against TDM90. The average height differences with a radius of 3000m to 6000m were used for the individual height correction. For SRTM a radius of 6000m is optimal, for AW3D30 it depends on the covers by individual stereo models (Figures 7 and 16), but the size of the radius is not that important, it has only a minor influence.
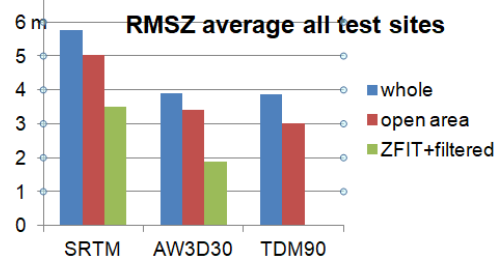

Figure 21. Average root mean square of all test sites and with test site Mountain for terrain inclination below 0.1
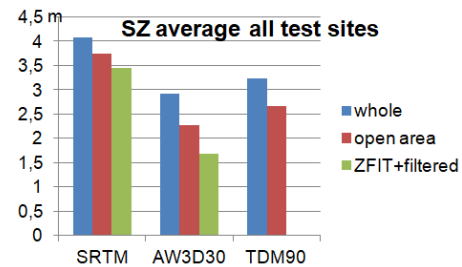

Figure 22. Average standard deviation of all test sites with test site Mountain for terrain inclination below 0.1
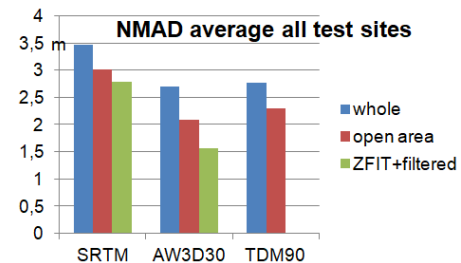

Figure 23. Average NMAD of all test sites with test site Mountain for terrain inclination below 0.1

The averaged accuracy figures (Figures 21 - 23) have only a limited individual meaning for the size, but they clearly show the accuracy relationships. SZ and NMAD have the same tendency; NMAD is generally slightly smaller than SZ, while RMSZ is influenced by the bias. As the individual results show, SRTM has the lowest accuracy. The RMSZ of the TDM90 is slightly smaller than that of the AW3D30, even if the TDM90 has three times the point spacing of the AW3D30. This is mainly caused by the lower bias of TDM90 (Figure 20). The relationship is different for the standard deviation, which is smaller for AW3D30 than for TDM90, especially when TDM90 is used for the improvement of AW3D30 (ZFIT).
The accuracy figures for the three height models (Figures 24 and 25), even if they are limited in Mountain to terrain with an inclination below 0.1 (6°), clearly show the lower accuracy of

InSAR in mountain areas. In test site City AW3D30 also has some advantages over the InSAR height models SRTM and TDM90. The test site rolling has less of an impact on the remaining vegetation and buildings, resulting in the smallest accuracy numbers of any test sites. Here, too, AW3D30 has some advantages over TDM90 due to smaller point spacing.
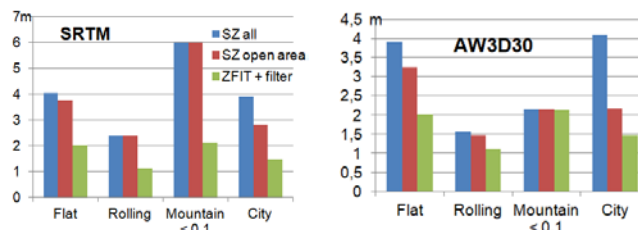

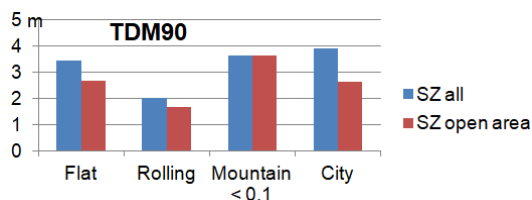Figure 24. Standard deviation of SRTM and AW3D30 DSM

Figure 25. Standard deviation of TDM90

6. CONCLUSION

The results can be summarized that the SRTM, AW3D30 and TDM90 have a level of accuracy that is satisfactory for various purposes. Nevertheless, SRTM is less accurate than the other, the morphological information from SRTM is not that good (Figures 8 – 10) and SRTM is based on radar images of February 2000 – older as the other. TDM90 has the best scene orientation, resulting in lowest bias, but InSAR has some disadvantages in mountain and urban areas and with a spacing of 3 arcsec, TDM90 has the greatest loss of accuracy by interpolation. The lower point spacing reduces the morphological quality (Figures 8 – 10). Ultimately, the best results were achieved with AW3D30, especially if it is filtered for elements that do not belong to a DTM and improved by the absolute orientation of TDM90 by adapting the AW3D30 to the averaged TDM90 height by program ZFIT. The morphological quality of the AW3D30 is better than that of the SRTM and TDM90 (Figures 8 – 10), and the matched optical image models have fewer problems in built-up areas and in steep mountains. The standard deviations of the improved AW3D30 at 1.1m to 2m in open areas with an inclination below 0.1 are better as expected. Even in the steep mountains the RMSZ and the SZ are only in the range of 4.5m or the NMAD in the range of 2.5m. In cities as well as in the forest, of course, the character of a DSM cannot be compared with a DTM.

Of course, the commercial height models AW3D with 5m point spacing and WorldDEM based on TanDEM-X with 10m point spacing have the particular advantage of better morphological information; however, except for the loss of accuracy by interpolation, they have the same accuracy as the AW3D30 and TDM90. It is therefore a question of requirement if the freely available DSMs should be used. In any case the use of SRTM should be replaced by the use of AW3D30 and AW3D30 should be fitted to the better absolute orientation of TDM90.

Of course, there are timing differences in data acquisition, but these changes in elevation model accuracy are usually not important. A reference radius of at least 3000m is used to fit the AW3D30 to the TDM90. Therefore, the height changes must cover large areas to have an important influence on improving the AW3D30.

REFERENCES

Aldosari, A., Jacobsen, K., 2019. Quality of Height Models Covering Large Areas, PFG Volume 87, Issue 4, pp 177–190, https://link.springer.com/article/10.1007/s41064-019-00072-11

ASTER Global Digital Elevation Model – Release of ASTER GDEM Version 3, 2019. https://ssl.jspacesystems.or.jp/ersdac/GDEM/E/4.html (April 2021)

DLR, 2018, The TanDEM-X 90m Digital Elevation Model, https://geoservice.dlr.de/web/dataguide/tdm90/ (January 2021)

Gesch, D.B., Oimoen, M.J.,Evans, G.A., 2014. Accuracy assessment of the US Geological Survey National Elevation Dataset, and comparison with other large-area elevation data sets: SRTM and ASTER, https://pubs.usgs.gov/of/2014/1008/pdf/ofr2014-1008.pdf (January 2021)

Höhle, J., M. Höhle. 2009. Accuracy assessment of digital elevation models by means of robust statistical methods, ISPRS JPRS, vol. 64, pp. 398-406

Passini, R., Betzner, D., Jacobsen, K.: Filtering of Digital Elevation Models, ASPRS annual convention, Washington 2002

Pavlis, N.K., Holmes, S.A., Kenyon,A.C., Factor, J.K. 2012. The Development and Evaluation of the Earth Gravitational Model 2008 (EGM2008), Journal of Geophysical Research, Vol 117, Issue B4, April 2012, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2011JB008916

Smith, R.G., P. Berry. 2011. Evaluation of the differences between the SRTM and satellite radar altimetry height measurements and the approach taken for the ACE2 GDEM in areas of large disagreement, J. Environ. Monit., 2011, 13, pp 1646-1652

Snyder, J.P. 1987. Map Projections – A Working Manual, U.S. Geological Survey Professional Paper 1395

Soergel, U., Jacobsen, K., Schack, L. 2013. TanDEM-X Mission: Overview and Evaluation of intermediate Results, Photogrammetric Weeks 2013

Tadono, T., H. Ishida, F. Oda, S. Naito, K. Minakawa, H. Iwamoto. 2014. Precise Global DEM Generation by ALOS PRISM, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-4, 2014

Takaku, J., T. Tadono, K. Tsutsui. 2014. Generation of High Resolution Global DSM from ALOS PRISM, ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-4, pp 157-162

Tetsushi, T., Manabu, K., Akira, I., Gesch, D., Oimoen, M., Zhang, Z., Danielson, J., Krieger, T., Curtis, B., Haase, J., Abrams, M., Crippen, R., Carabajal, C., 2011. ASTER Global Digital Elevation Model Version 2 – Summary of Validation Results, https://pubs.er.usgs.gov/publication/70005960 (April 2021)

Wessel, B., Huber, M., Wohlfahrt, C., Marschalk, U., Kosmann, D., Roth, A., 2018: Accuracy Assessment of the Global TanDEM-X Digital Elevation Model with GPS Data, ISPRS JPRS, vol. 139, pp 171-182