ISPRS Journal of Photogrammetry and Remote Sensing 139 (2018) 171–182

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Accuracy assessment of the global TanDEM-X Digital Elevation Model with GPS data

Birgit Wessel *, Martin Huber, Christian Wohlfart, Ursula Marschalk, Detlev Kosmann, Achim Roth

German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany

ARTICLE INFO

Article history:
Received 13 June 2017
Received in revised form 21 December 2017
Accepted 20 February 2018

Keywords:
TanDEM-X
Digital Elevation Model
Kinematic GPS
Vertical accuracy
Validation

ABSTRACT

The primary goal of the German TanDEM-X mission is the generation of a highly accurate and global Digital Elevation Model (DEM) with global accuracies of at least 10 m absolute height error (linear 90% error). The global TanDEM-X DEM acquired with single-pass SAR interferometry was finished in September 2016. This paper provides a unique accuracy assessment of the final TanDEM-X global DEM using two different GPS point reference data sets, which are distributed across all continents, to fully characterize the absolute height error. Firstly, the absolute vertical accuracy is examined by about three million globally distributed kinematic GPS (KGPS) points derived from 19 KGPS tracks covering a total length of about 66,000 km. Secondly, a comparison is performed with more than 23,000 "GPS on Bench Marks" (GPS-on-BM) points provided by the US National Geodetic Survey (NGS) scattered across 14 different land cover types of the US National Land Cover Data base (NLCD). Both GPS comparisons prove an absolute vertical mean error of TanDEM-X DEM smaller than ±0.20 m, a Root Means Square Error (RMSE) smaller than 1.4 m and an excellent absolute 90% linear height error below 2 m. The RMSE values are sensitive to land cover types. For low vegetation the RMSE is ±1.1 m, whereas it is slightly higher for developed areas (±1.4 m) and for forests (±1.8 m). This validation confirms an outstanding absolute height error at 90% confidence level of the global TanDEM-X DEM outperforming the requirement by a factor of five. Due to its extensive and globally distributed reference data sets, this study is of considerable interests for scientific and commercial applications.

1. Introduction

Since September 2016 the new TanDEM-X Digital Elevation Model (DEM) can be seen as one of the most consistent, highly accurate and completest global DEM data sets of the Earth surface. This novel product will play a major role in a wide range of various regional and global applications analyzing physical and biological processes of the Earth surface (Zink et al., 2014). The height information was derived by applying single pass Synthetic Aperture Radar (SAR) interferometry. The corresponding pairs of images were acquired by the twin satellites TerraSAR-X and TanDEM-X, which fly in a close helix formation with distances between 300 and 500 m of each other (Zink et al., 2014). Bi-static interferometry is applied by transmitting pulses from the antenna of only one of the satellites and by receiving the backscattered signals simultaneously with both. Although SAR interferometry is well suited to globally map the Earth's surface in a short period of time, due to its 'day and night' and its 'all-weather' observation capability, the measured height corresponds to the reflective surface of the X-Band signal. In general, the TanDEM-X height model can be regarded mostly as a Digital Surface Model (DSM) rather than a Digital Terrain Model (DTM). However, there exist some exceptions for areas where the SAR signal penetrates the surface by some meters – e.g. in cases of ice, snow or vegetation. Consequently, the umbrella term DEM, comprising any kinds of elevation models, is the best suited for the TanDEM-X DEM. This product is available for scientific users at the German Aerospace Center (DLR, March 2018), commercial users can get the DEM from Airbus Defense and Space as so-called WorldDEM in different versions, e.g. with geoid elevations or further value-additions (WorldDEM, March 2018). The used SAR data for the global DEM production were acquired between December 2010 and January 2015 in StripMap mode with horizontal transmit and receive polarization (Krieger et al., 2007, Wessel, 2016). All land masses are covered at least twice (Borla Tridon, et al., 2013) to facilitate dual-baseline phase unwrapping (Lachaise et al., 2018) and to reach the random height accuracies by averaging individual DEM scenes with an

* Corresponding author.
E-mail address: Birgit.Wessel@dlr.de (B. Wessel).

https://doi.org/10.1016/j.isprsjprs.2018.02.017

interferometric SAR-specific mosaicking approach (Gonzalez and Bräutigam, 2015, Gruber et al., 2016). The generated TanDEM-X DEM has a 0.4 arc second posting and resolution with a specified 10 m absolute vertical accuracy (90% linear error, LE90) and 2 m relative accuracy resp. 4 m for areas with slopes larger than 20% (90% linear point-to-point error). The absolute height error of the final DEM was established by calibrating the individual data takes based on ground control points from ICESat (Ice, Cloud and land Elevation Satellite) GLA14 products and image control points (Gruber et al., 2012).

The next step is the validation of the specified quality of the final TanDEM-X DEM at a larger scale, which would be of considerable interests for the scientific and commercial users. Due to the limited availability of the TanDEM-X DEM so far, only a few works on regional scale report about the validation of the DEM (Baade and Schmullius, 2016; Rexer and Hirt, 2016). The internal validation effort regarding the absolute vertical accuracy of the global TanDEM-X DEM is based on ICESat data (Rizzoli et al., 2017). ICESat points are also integrated into other global DEM validation work, e.g. for AW3D (ALOS World 3D) from the optical PRISM senor onboard the ALOS satellite (Takaku et al., 2016). The study of Rizzoli et al. (2017) states that the TanDEM-X DEM reaches with 3.5 m the global absolute accuracy goal of 10 m (90% linear error, LE90) by far. In our study, the absolute height accuracy of the TanDEM-X DEM should be validated globally as well, but with independent, higher accuracy data sets that have not been used for the generation of the DEM. Global Positioning System (GPS) points, which were chosen for this task, are a common measure to assess the accuracy of global DEMs (Rodríguez et al., 2006; Jacobsen and Passini, 2010; Mouratidis et al., 2010; Gesch et al., 2012; Baade and Schmullius, 2016; Bolkas et al., 2016; Gesch et al., 2016; Rexer and Hirt, 2016). Gesch et al. (2012) and Gesch et al. (2016) use GPS data to describe the absolute height error for 14 land cover classes in the United States separately for the global Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 (GDEM2). Bolkas et al. (2016) express the importance of selecting ground control points according to different terrain characteristics like slope and land cover types to get a representative Root Mean Square Error (RMSE) for the estimation of the DEM uncertainty. However, the availability of GPS points suited for world-wide validation is very limited and of high uncertainty. Therefore, one unique GPS data set used in this work was acquired in special campaigns in 2008 and 2009 for exactly this purpose (Kosmann et al., 2010).

The primary goal of the validation work performed in this paper is the characterization of the vertical accuracy of the TanDEM-X DEM by two independent, highly accurate and extensive GPS data sets. All reference data sets and the used DEM validation approach are introduced in the second chapter. Chapter three describes and discusses the results of the GPS analyses. In addition, comparisons with two high-resolution DSMs and a DTM are conducted. Chapter four completes the paper by a summary of our results.

2. Materials and methods

Absolute vertical accuracy assessment requires highly accurate and independent reference data. The accuracy of such data sets should be at least three times more accurate than the evaluated data set (Maune, 2007, p. 407). Here, we want to validate the 10 m LE90 absolute height error requirement, which results in a maximum error of 3.3 m LE90 for a reference data set. This leads to an accuracy requirement of a Standard Deviation (STD) being less than 2 m, which in turn is a very ambitious figure for a global validation data set. A potential existing data set that fulfils the demanded requirements is the ICESat GLA14 elevation product (Zwally, 2002; Carabajal and Harding, 2005). A subset of around 10% of the GLA14 data was used within the DEM generation process for the block adjustment of the individual DEM acquisitions (Wessel et al., 2008; Huber et al., 2009; Gruber et al., 2012). The remaining ICESat points built the base for a DEM accuracy analysis as described in Rizzoli et al. (2017). In this study the following independent GPS data sets were used.

2.1. Kinematic GPS tracks

An extensive ground truth data campaign was set up for the TanDEM-X mission in 2008 and 2009 to gain global reference data with an accuracy of STD <2 m (Kosmann et al., 2010). Transects on every continent (except Antarctica) were measured with the kinematic Global Positioning System (KGPS) method. The basic concept of this collection was to mount a GPS antenna on top of a car and drive along roads across the continents.

The TanDEM-X data was basically acquired in strips along the north-south direction. In order to enable the detection of systematic errors smaller than the absolute accuracy across the track direction, the GPS transects were driven mainly in east-west direction. This world-wide acquisition was realized in close cooperation with the FIG (International Federation of Surveyors). Interested scientists and organizations were invited to participate in the DEM verification efforts with KGPS tracks. The vertical error of the kinematic GPS tracks should not exceed 0.5 m. The use of precise differential GPS (PDGPS) using local reference stations is very time-consuming and cost-intensive and Continuously Operating Reference Station (CORS) networks are not available world-wide. Therefore, an extensive GPS post-processing approach called Precise Point Positioning (PPP) had been used (Ramm and Schwieger, 2007; Schweitzer et al., 2010). This approach did not require a GPS reference station network, but precise orbit and time information. Additionally, satellite antenna offsets and variations as well as phase wind up corrections had to be considered.

The GPS data were processed independently using two different software packages: the GIPSY software (GIPSY, 2017) and the PPP service provided by the Natural Resources of Canada (CSRS-PPP, 2017). In order to ensure a high accuracy, the outcomes of both tools were averaged in a way that points with height differences above 1.0 m were eliminated from the data (Schwieger et al., 2009). The final RMSE of the combined results reached 0.48 m and the availability rate was 53.5%. This can be explained by signal outages that occurred very often, especially in urban and forest areas. To start with an optimal accuracy below 0.1 m, an initialization phase of approximately 30 min was mandatory for every single track.

A total number of 14 million KGPS points were acquired along tracks encompassing a length of 66,000 km distributed across six continents. An overview of the collected GPS Ground Control Points (GCPs) for each continent is shown in Table 1. The GPS point collection attempted to acquire at least three GPS points per TanDEM-X pixel (approximately $12\,m \times 12\,m$ resolution on ground). GPS points within one pixel were averaged for further analyses.

2.2. GPS benchmark data

The reference data set "GPS on Bench Marks" is a highly accurate (millimeter to centimeter) set of GCPs which is measured and provided by the US National Geodetic Survey (NGS) for Northern America (GPS-on-BM, Mai 2017). These elevations are primarily used for geoid modelling. For our analysis, a total number of 23,961 points which were used (Fig. 1) were distributed across the United States. The coordinates of the points were provided in

Table 1
Set of measured world-wide kinematic GPS track data.

Continent	Length in km	No. of GPS in samples/1000	No. of averaged GPS in samples/1000	Avail-ability rate
Africa	9402	2098	406	61.7%
North America	22,290	4362	938	58.3%
South America	9689	1709	353	41.4%
Asia	11,642	2455	538	39.9%
Australia	4602	616	149	49.2%
Europe	8637	3208	736	60.0%
ALL	**66,262**	**14,424**	**3120**	**53.5%**

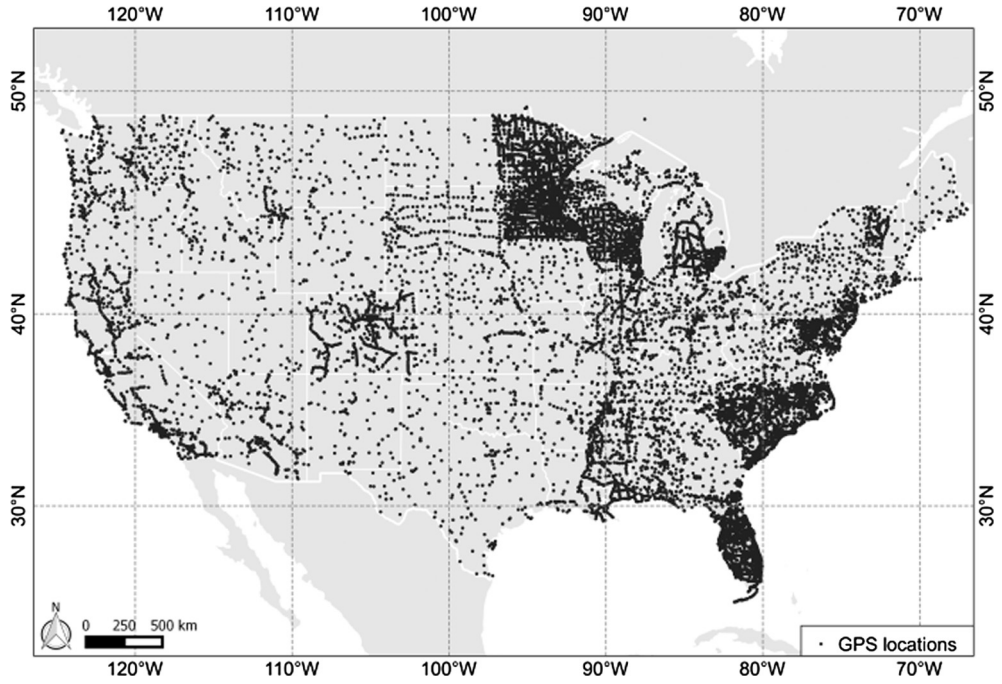

Fig. 1. GPS benchmarks (23,728 points) used for TanDEM-X DEM validation.

the North American Datum (NAD 83) with elevations annotated in the North American Vertical Datum of 1988 (NAV88). The locations and heights were transformed into geographic coordinates and ellipsoidal heights with the NGS software "Horizontal Time-Dependent Positioning" (HTDP, 2017). TanDEM-X DEM elevations are referenced to ITRF 2005 (Wessel, 2016). Therefore, the target datum for the GPS benchmarks was also ITRF 2005, Epoch 2010. For each point, the height difference (Δh) was calculated by subtracting the GPS elevation from the corresponding TanDEM-X DEM height. Outliers outside mean height difference plus/minus three times standard deviation ($\Delta h > \Delta h_{mean} + 3*STD$, $\Delta h < \Delta h_{mean} - 3*STD$) were eliminated resulting in 23,728 points used for validating the TanDEM-X DEM. The GPS benchmark data set has the drawback that the GPS points are often located in exposed and better accessible sites where the points are mainly representing the ground (Gesch et al., 2012). Steep slopes and high altitude points are underrepresented whereas developed areas are highly overrepresented.

To assess the accuracy of the TanDEM-X DEM by land cover type, the 2011 National Land Cover Database (NLCD) was used (Homer et al., 2015; NLCD, 2011). This dataset defines various land types of natural vegetation, cultivated classes and artificial structures, resulting in a total of 20 land cover classes, out of which the GPS benchmark points are positioned within 14 of these classes. The NLCD classification scheme is based on a decision-tree classification of Landsat data and has a spatial resolution of 30 m.

2.3. Higher resolution DEM test sites

To complement the characterization of the height error with GPS we compared exemplarily TanDEM-X DEM with existing high-resolution DEMs. For this analysis two DSMs and one DTM with an extent of up to 100 km^2 were used located in Thuringia (Germany), Kumamoto (Japan) and Cape Town (South Africa). Cape Town represents a DTM refering to the bare ground and was derived from Laserscanning like Thuringia (DSM), while the third one located in Japan (DSM) was from optical satellite imagery.

Table 2 gives the specifications of the three reference DEMs. The DTM, covering the entire municipal area of the city of Cape Town with an extent of 2,460 km^2 (Cape Town DEM, Mai 2017), is characterized mainly by bare, flat terrain with large wine-growing districts. The prominent Table Mountain is located in the southwestern part. The second test site covers a smaller area of 10 km $\times$ 10 km in a hilly, mainly dediduous forest covered region of Thuringia in Germany. A small village is located in the southeastern part. The open access DSM was provided by the Thuringia land surveying office (Geoportal Thuringia, Mai 2017). From the ALOS PRISM mission (Takaku et al., 2016) a 1° $\times$ 1° DSM containing the city of Kumamoto was obtained. The landscape is dominated by dense forest covered mountains. The flat coastal regions are urban regions characterized by agricultural fields in the surroundings.

In the first step, all high resolution reference DEMs were resampled to the best available TanDEM-X resolution, which was

Table 2
High resolution reference DEMs used for accuracy assessment.

Area	Data source	Exposure	Pixel-spacing	Height range (WGS84)	Elevation type	Terrain characteristics
Cape Town, South Africa	LiDAR	2011–2015	10 m	−4 m to 1621 m	DTM	open, flat, Table Mountain
Thuringia, Germany	LiDAR	2012	2 m	235 m to 588 m	DSM	forested, hilly
Kumamoto, Japan	AW3D (optic)	2006–2011	0.15″	−5 m to 1674 m	DSM	forested mountains, coasts

0.4″ × 0.6″ for Thuringia and 0.4″ × 0.4″ for Cape Town and Kumamoto. The heights for Thuringia had to be transformed from the german height reference (DHHN2016) into ellipsoidal WGS84 heights via the German Combined Quasigeoid 2016 (GCG2016).

2.4. Accuracy assessment method

In this assessment, accuracy is understood as a description of systematic and random errors. The systematic error is estimated by a statistical bias and the random error by the deviation of the height difference, both measured to a higher accurate reference. For each point, the difference was calculated by subtracting the GPS data from the corresponding TanDEM-X DEM pixel $\Delta h = h_i - h_{GPS}$. Usually the sampling of the kinematic GPS points was higher than the DEM pixel spacing. Therefore, the mean was taken from all points within a TanDEM-X pixel before differencing. In average 4.6 points were fused (ratio of the total number of GPS points before and after averaging in Table 1). For the GPS benchmark data set the nearest neighbor approach was chosen. Calculating the difference leads to positive difference height values Δh where TanDEM-X DEM is above the reference elevation and to negative differences where TanDEM-X DEM lies below the reference data. Initially, we removed outliers by applying the 3-sigma-rule (Three-sigma rule, Mai 2017). A first impression of the distribution of the height differences can be obtained from a histogram. Assuming a normal distribution, the accuracy measures mean error (ME)

$$ME = \frac{1}{n}\sum_{i=1}^{n} h_i - h_{ref} = \frac{1}{n}\sum_{i=1}^{n} \Delta h_i, \quad (1)$$

root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \Delta h_i^2}, \quad (2)$$

and standard deviation (STD)

$$STD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (\Delta h_i - ME)^2} \quad (3)$$

were applied to assess the error. Furthermore, robust accuracy measures for non-normal error distributions had to be considered (Höhle and Höhle, 2009). We used median (50% quantile)

$$\widehat{Q}_{\Delta h}(0.5) = m_{\Delta h}, \quad (4)$$

the median absolute deviation (MAD)

$$MAD = median_j(|\Delta h_j - m_{\Delta h}|), \quad (5)$$

the normalized median (NMAD)

$$NMAD = 1.4826 \cdot median_j(|\Delta h_j - m_{\Delta h}|) \quad (6)$$

and the absolute deviation at the 90% quantile (LE90)

$$LE90 = \widehat{Q}_{|\Delta h|}(0.9). \quad (7)$$

The NMAD is proportional to MAD and can be regarded as an estimate for the standard deviation for heavy tail distributions (Höhle and Höhle, 2009). In case of normally distributed errors, STD is identical to NMAD. In case of larger discrepancies STD will be larger than NMAD. For normally distributed observations, the linear error at 90% confidence level is LE90 = STD * 1.65. As height differences tend to follow a non-normal distribution, here the LE90 is directly equated to the 90th percentile of the sorted absolute differences calculated by the minimum rank method, i.e. the smallest value in the list. In other words 90% of the data is less than or equal to that value. Systematic positive height differences are expected for urban and forest areas by comparing ground level heights from GPS with TanDEM-X heights. Therefore, the accuracy assessment was further cleared by separating urban and forest classes from open ground areas. Subsequently, the error measures were separately calculated for individual land cover classes.

An estimate of the random height error, directly from the interferometric SAR measurements, can be obtained by the difference in the interferometric phase φ and its standard deviation. The standard deviation of the interferometric phase σ_φ is strongly related to geometrical considerations, the number of looks and the measured coherence (Krieger et al., 2007; Gonzalez and Bräutigam, 2015). The standard deviation of the interferometric phase can be regarded as the theoretical random height error and its value is annotated for each pixel within the Height Error Map (HEM) of the TanDEM-X DEM product. The final annotated values have been calculated by the mean of all input DEM acquisitions. Less coherent objects, like forest, ice, snow or water, incorporate a high random height error and therefore result in a less accurate absolute height. Also steeper areas tend to have higher height errors, especially where SAR specific effects like foreshortening, layover and shadow influence the heights. The influence of slope and low coherence areas annotated in the HEM was also evaluated.

3. Results and discussion

3.1. Absolute vertical accuracy

In this section, we examined the height accuracy of TanDEM-X DEM by comparison with two unique GPS data sets. The accuracy numbers of the measured TanDEM-X height differnces against KGPS are summarized in Table 3, while Fig. 2a shows the histogram of the height differences which indicates an approximatively uniform distribution around zero. The distribution is obvious narrower than a normal distribution, but it is more tailored. The RMSE is comparable to the STD. By inspecting all kinematic GPS data, the STD (1.28 m) is larger than NMAD (0.94 m) because of the tailored distribution. The absolute vertical accuracy in terms of the linear error at 90% confidence is 1.93 m for KGPS data. Note that kinematic GPS data was limited to roads and will selectively represent paved and relatively flat terrain. Thus, the values presented in Table 3 are probably optimistic for other terrain types.

Table 3 summarizes the accuracy numbers for each continent which are consistent across all continents. The overall low KGPS mean error of −0.17 m indicates almost no vertical offset. This corresponds well with the mean error of 0.04 m calculated from the ICESat validation work conducted by Rizzoli et al. (2017). This validation study used more than 144 million points of ICESat data. There, the absolute height error for 90% of the validation points is below 3.49 m. The NMAD of the KGPS height differences, as a more robust measure for the 68% probability level than the RMSE or the standard deviation, is even below 1 m; the LE90 is below 2 m.

Table 3
Accuracy numbers for height differences TanDEM-X minus kinematic GPS.

KGPS Track	ME (m)	STD (m)	RMSE (m)	MAD (m)	NMAD (m)	LE90 (m)
Africa	−0.29	1.34	1.37	0.73	1.08	2.11
North America	−0.28	1.27	1.30	0.63	0.93	1.95
South America	−0.08	1.33	1.33	0.66	0.97	2.09
Asia	−0.16	1.04	1.05	0.52	0.77	1.52
Australia	−0.48	1.16	1.26	0.58	0.86	1.77
Europe	0.06	1.35	1.35	0.65	0.97	2.02
All KGPS	**−0.17**	**1.28**	**1.29**	**0.63**	**0.94**	**1.93**

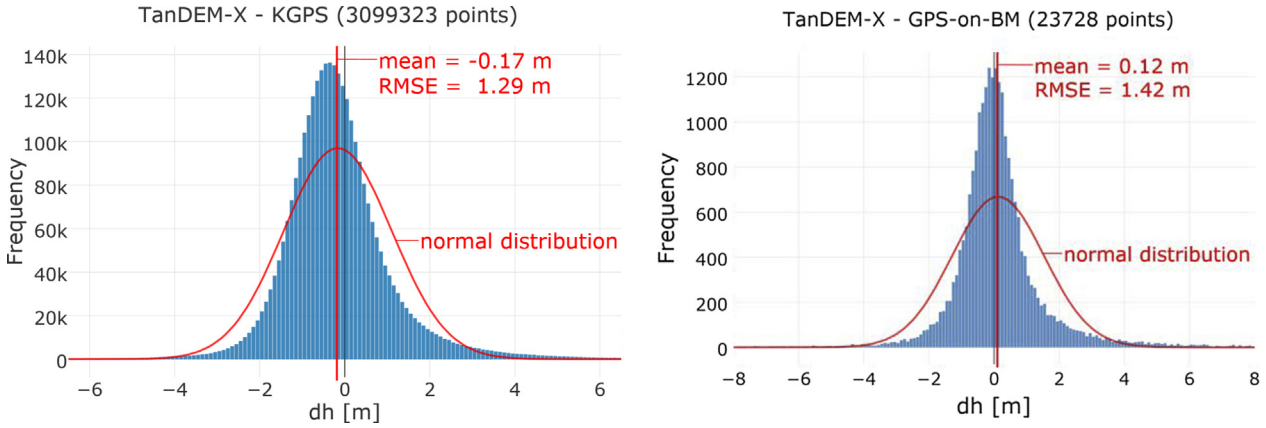

Fig. 2. Histogram plot of height differences for TanDEM-X DEM minus kinematic GPS for roads.

The height differences of TanDEM-X heights minus KGPS, are shown in Fig. 3, where the mean height difference within one geocell is depicted. No long wavelength error across continents can be identified when examining Fig. 3. The magnitude of the error is mainly within a ±1 m range. The calculated height differences in this study can be seen as an estimate for the present height error, because of the huge amount of points.

The results for all GPS-on-BM points are listed in Table 4 and show a consistent behavior to the KGPS results. The results for mean error (0.12 m), STD (±1.42 m) and NMAD (0.82 m) are in a similar range, as well as the distribution of errors (Fig. 5a).

3.2. Land cover analysis

The accuracy results discussed above represent the validation results for paved roads and their surroundings. Land cover can effect both the interferometric height measurement as well as the GPS signal of the reference point. To examine the influence of

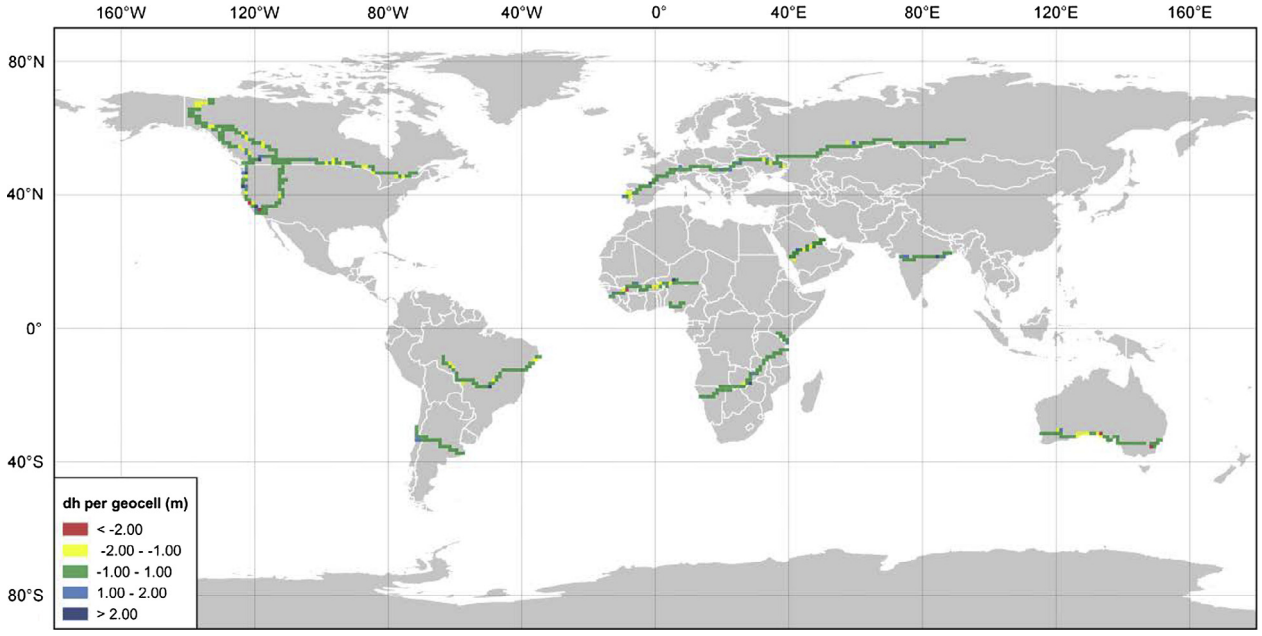

Fig. 3. Mean height difference, TanDEM-X DEM minus kinematic GPS heights, averaged over one geocell (mostly 1° × 1° of latitude and longitude). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Land cover specific accuracy numbers for height differences TanDEM-X minus GPS-on-BM points.

Land cover class	No. of Samples	ME (m)	STD (m)	RMSE (m)	MAD (m)	NMAD (m)	LE90 (m)
Developed							
open	7403	0.15	1.28	1.29	0.48	0.71	1.81
low	6491	0.13	1.39	1.40	0.56	0.84	2.03
medium	3311	0.12	1.70	1.71	0.68	1.01	2.59
dense	817	0.07	1.80	1.80	0.80	1.18	2.76
Low vegetation							
Barren land	138	−0.09	1.56	1.57	0.69	1.02	2.44
Grassland	758	−0.12	1.20	1.20	0.50	0.74	1.67
Pasture	1107	0.10	1.18	1.18	0.46	0.68	1.71
Cropland	1757	−0.05	0.95	0.95	0.42	0.63	1.27
Herb. wetlands	756	0.06	1.43	1.43	0.47	0.70	2.03
Forest							
Shrub/scrub	342	−0.05	1.27	1.28	0.61	0.91	2.15
Woody wetlands	336	0.34	1.78	1.81	0.72	1.07	2.55
Mixed forests	81	0.54	1.88	1.96	1.08	1.60	2.92
Deciduous forest	284	0.79	2.18	2.32	0.92	1.36	3.84
Evergreen forest	147	1.21	2.27	2.57	1.01	1.50	4.37
ALL	**23,728**	**0.12**	**1.42**	**1.42**	**0.55**	**0.82**	**2.03**

land cover on the vertical accuracy of the TanDEM-X DEM, all GPS-on-BM points were grouped according to the corresponding land cover class. The accuracy measures were calculated for each of the 14 land cover types separately and were summarized in Table 4. This part of the validation is similar to the validation conducted by Gesch et al. (2012) and Gesch et al. (2016) for the ASTER GDEM v2. The mean error (ME) over all classes resulted in 0.12 m, but showed rather high variation across the land cover types ranging from −0.12 to 1.21 m (Fig. 4). Higher biases can be attributed particularly to denser vegetated classes, such as forests and woody wetlands, whereas open and developed areas are determined by lower ME or even slightly below zero (barren land, grassland, cropland, shrub/scrub). Standard deviation and absolute errors show a similar pattern. RMSE values range from 0.95 to 2.57 m, being highest on developed and forest areas (Fig. 4). Less divergence across land cover classes hold the accuracy numbers of MAD (all classes: 0.55 m) ranging from 0.42 (cropland) to 1.01 m (evergreen forest) and for NMAD (all classes: 0.82 m) 0.63 (cropland) to 1.60 m (mixed forest). The LE90 accuracy values range from 1.27 m (cropland) to 2.76 m (dense developed), whereas for the forest classes they are mostly above 3 m.

To facilitate the land cover analysis, we fused the 14 land cover types into three main land cover classes: developed, low vegetation and forest (Table 5). Fig. 5b–d gives the corresponding error distributions. Developed and forest areas show a small positive bias of 0.13 m and 0.37 m, respectively. However, the height differences of these classes, i.e. comparing TanDEM-X surface heights with heights on the ground, seem negligibly small. Having a closer look at the location of the reference points, we see that the GPS-on-BM points are often located in open places. This confirms the low standard deviations. Although, the 90% absolute error rises from 1.53 m for low vegetation, up to 2.08 m for developed and 2.74 m for forested areas.

3.3. Random height error

In Sections 3.1 and 3.2 a minor negligible vertical offset was confirmed. Therefore, in this section the relationship between the measured height differences and the theoretical error was investigated. Fig. 6 shows the relationship between the theoretical height errors annotated in the HEM (Section 2.4) and the measured height differences at the GPS benchmark points. A clear linear relationship

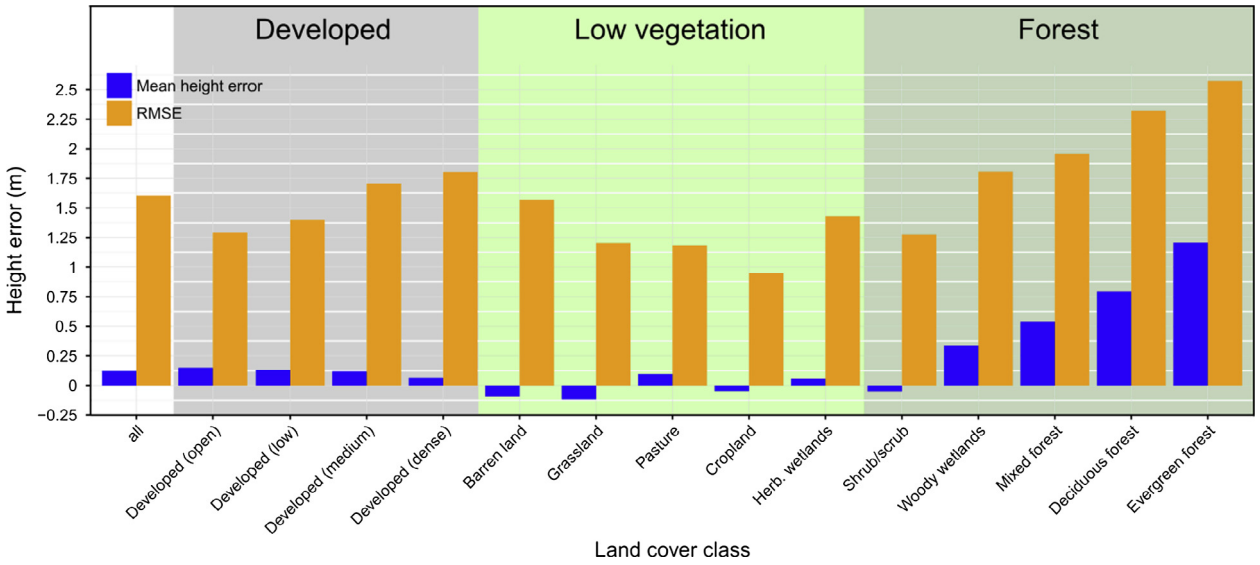

Fig. 4. Mean height error and RMSE by land cover class (TanDEM-X minus GPS-on-BM).

Table 5
Land cover specific accuracy results for main land cover classes (height differences TanDEM-X minus GPS-on-BM).

Class	ME (m)	STD (m)	RMSE (m)	MAD (m)	NMAD (m)	LE90 (m)
Developed	0.13	1.43	1.44	0.56	0.83	2.08
Low vegetation	−0.02	1.12	1.12	0.47	0.70	1.53
Forest	0.38	1.80	1.84	0.66	0.98	2.74

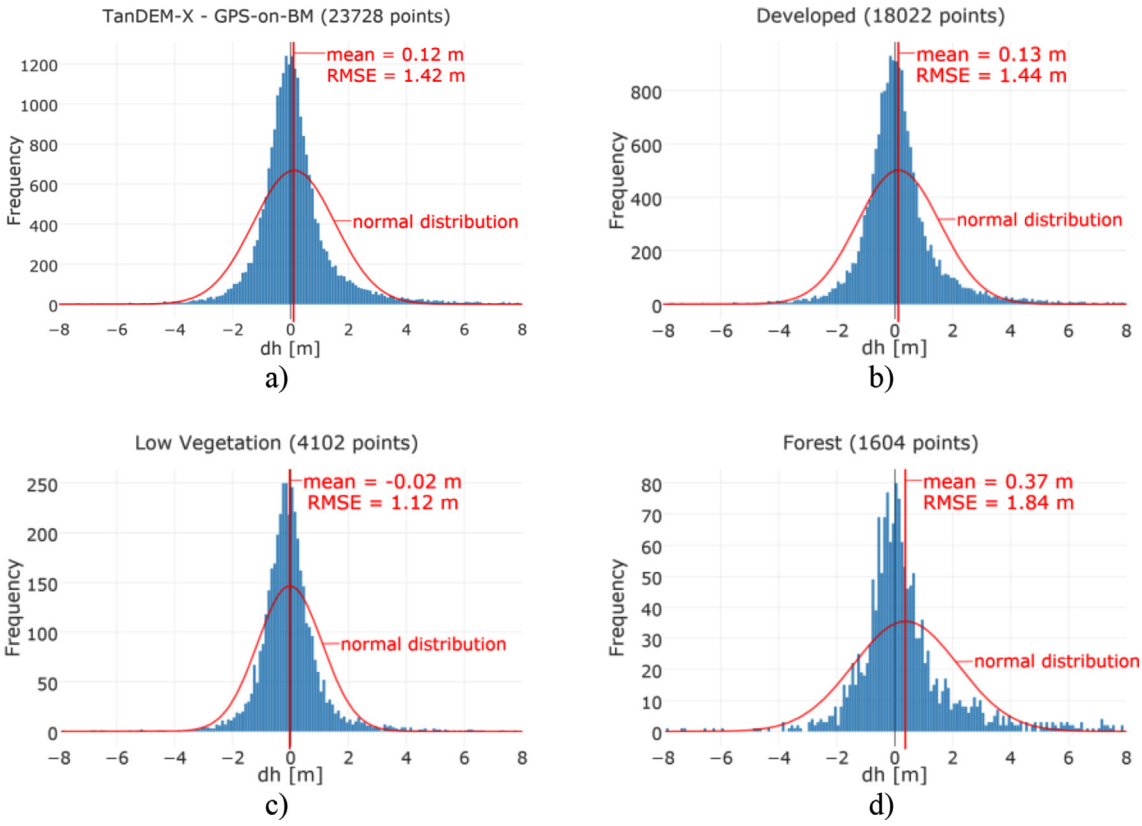

Fig. 5. Absolute vertical accuracy by histogram plots for GPS benchmarks for all points a) and divided into three main land cover classes: developed b), low vegetation c) and forest d).

from the Box-Whisker-Plot could be observed. It showed the higher the HEM values, the higher the measured height differences were. It is remarkable that the 75th percentile boxes did not increase significantly for higher theoretical HEM values. The linear fit of the theoretical HEM values and the measured absolute height error proved that the HEM can be regarded as a reliable estimate for the random height error.

Fig. 7 illustrates the mean HEM value for each geocell globally. Global variations of the random error are obvious for tiles with large forest, ice or snow areas, i.e. for those areas with a larger amount of volume decorrelation the random error increases. Also, greater errors became present for regions with distinct relief. The majority of the TanDEM-X DEM data revealed height errors below 2 m, even for forest. These results are analogous to the results obtained for the three main land cover types (Fig. 5 and Table 5), which also indicate standard deviations below 2 m including forest.

3.4. Elevation and slope

The TanDEM-X height differences to GPS-on-BM are plotted against the absolute elevation of the GPS benchmark data (Fig. 8) indicating a small bias for low elevations towards positive height differences. This might be caused by the huge amount of developed areas in the GPS benchmark data that have a slightly positive bias. In contrast, higher elevations are distributed around zero with lower variation, though only few points are present for higher elevations.

In the following, the reference points were grouped into bins depending on slope. The slope was calculated on the TanDEM-X DEM in one arc second resolution. The measured errors and the number of points are displayed for each slope bin in Fig. 9. Furthermore, the accuracy measures associated with slope values smaller than 10° and greater or equal than 10° are calculated (Table 6). The mean value is almost constant for all slopes with ME of 0.12 m below and 0.24 m above 10°. But for slopes above 10° the RMSE and NMAD are about 1 m higher and LE90 raises up to 5.49 m.

3.5. Model-to-model accuracy assessment

Punctual reference data like GPS points have the disadvantage of not fully describing the terrain details. Therefore, we additionally conducted a model-based accuracy analysis of the TanDEM-X DEM with higher resolution DEMs. For this purpose three test sites were selected around the globe with diverse and heterogeneous landscapes, a DTM for the case of Cape Town, South Africa and DSMs for the case of both Thuringia, Germany and Kumamoto, Japan.

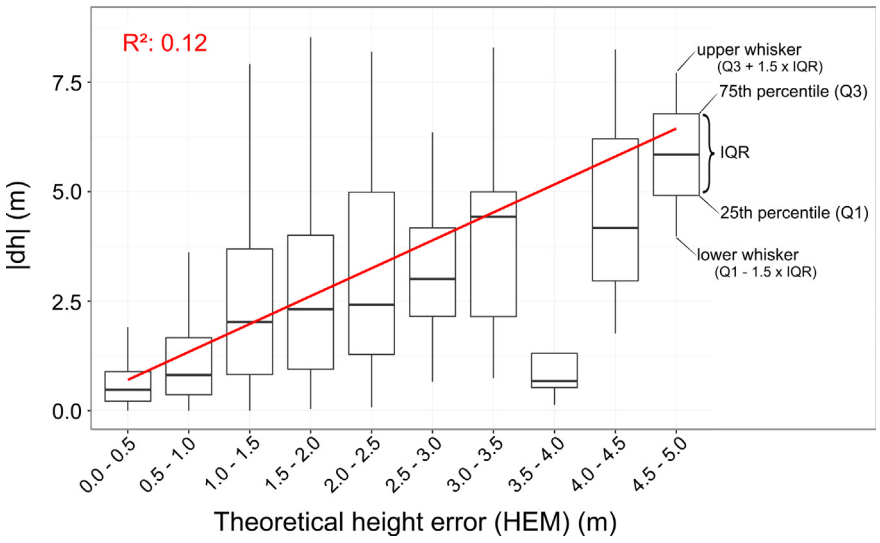

Fig. 6. Annotated theoretical TanDEM-X HEM values versus measured absolute height errors (|Δh|) [m] by GPS benchmark showing an approximately linear trend. The boxes represent lower and upper whiskers, 25th percentile (Q25), median, 75th percentile (Q75).

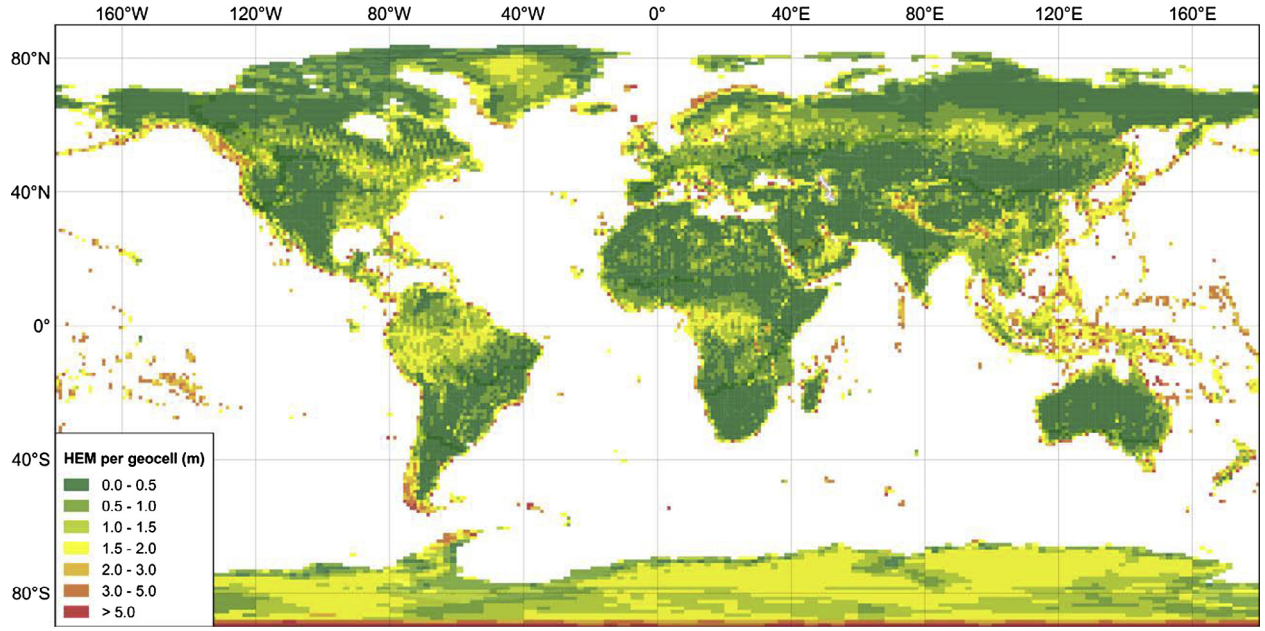

Fig. 7. Mean theoretical height error (HEM) per geocell in meters of the TanDEM-X DEM.

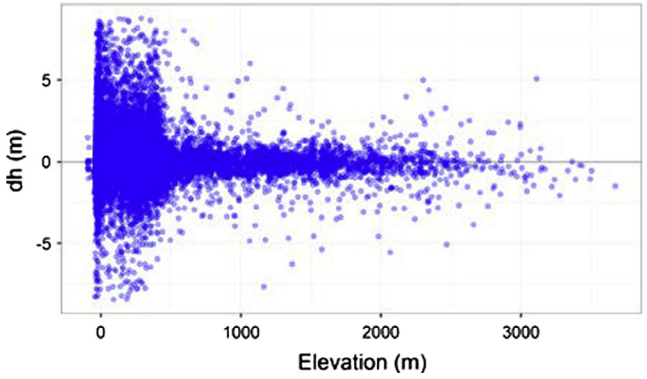

Fig. 8. Height differences TanDEM-X DEM minus versus GPS benchmarks plotted versus reference elevation.

Table 7 shows the accuracy numbers of the height differences between the TanDEM-X DEMs and the three high resolution reference DEMs. The calculations were carried out for the whole area and on different slope angle ranges (0°–10°, 10°–20°, 20°–30° and >30°).

The height differences over Cape Town showed very good results. The landscape is characterized by grassland, vine yards and agriculture fields having low vegetation cover in rather flat terrain with slopes less than 10°. Here, the values for ME, NMAD and LE90 are similar to those from the previous evaluations of the GPS data. Fig. 10a and b shows the TanDEM-X DEM of Cape Town and the height differences between TanDEM-X DEM and the DTM, scaled from −10 m to +10 m. Due to the fact that the 10 m-resolution DTM reflects the bare earth surface, mainly vegetation and urban areas show positive offsets in the height difference image. The yellow areas in the northern and central regions

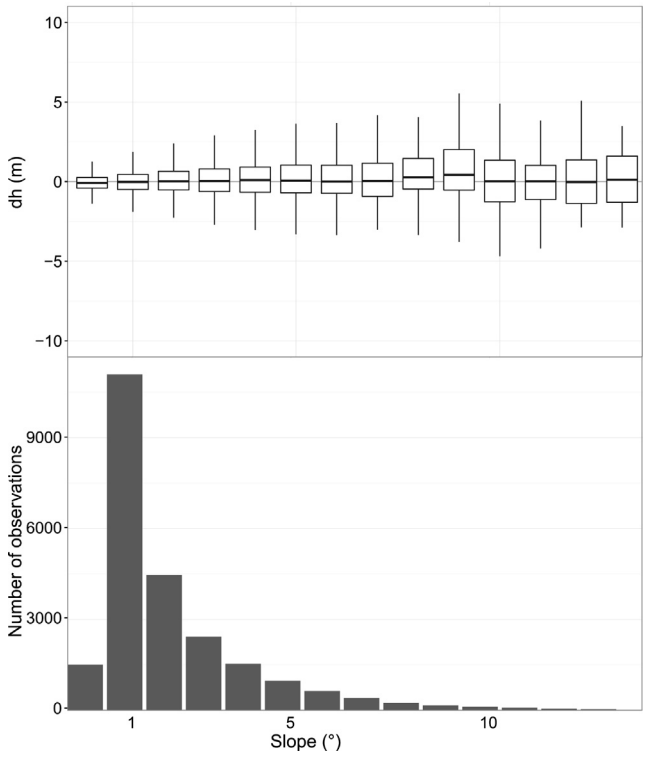

Fig. 9. Upper panel: Relationship between slope and height differences (TanDEM-X DEM minus GPS benchmark) in a Box-Whisker-Plot, the boxes show lower and upper whiskers, 25th percentile (Q25), median, 75th percentile (Q75); lower panel: frequency distribution of GPS benchmarks over slope.

indicate bushland as well as lower, sparse forest. The red linear features represent trees along roads and rivers, as shown in the detailed subset of Fig. 10b. The wooden eastern slopes of the Table Mountain appear in yellow and red colors, i.e. representing positive offsets. The flat-top level of the Table Mountain has no offset, whereas the top level of flat-top Mountains often show offsets in InSAR DEMs due to phase unwrapping errors induced by the extreme steep slopes. The slopes of the western Table Mountain are affected by phase unwrapping errors as seen in the saturated values in Fig. 10b, i.e. offsets above 10 m are present in this part. The significant high error values for slopes larger than 30°can be explained by some broad outliers, which raise these values above-average for LE90 (43.86 m), STD (36.98 m) and RSME (37.00 m). Striking is that the MAD and NMAD are less affected by this outlier region because the absolute median representing a median height offset is less sensitive to outliers.

The results of the Thuringia test site correspond well to the results obtained in Cape Town, aside from a slightly systematic negative bias with an overall mean error of −0.79 m. By analyzing the height difference image (Fig. 10d) it is apparent that there are higher negative biases especially over the forested areas even though the Thuringia reference model is a DSM like the TanDEM-X DEM. Although DSMs, InSAR and LIDAR, should generally fit over forests, the LIDAR DSM heights are somewhat below the TanDEM-X DEM in terms of forest areas, yet taking into account that the LIDAR data was captured in early spring before foliation. In a deeper evaluation of the LIDAR DSM it showed that some artefacts were present at steeper slopes, however requiring more attention to the quality assessment of these regions.

The Kumamoto site obtained slightly worse results with an overall mean of 3.79 m, RMSE of 6.18 m and NMAD of 2.90 m. This reflects the challenging topography and dense forest coverage in this region. With increasing degree of slope the errors increase as well. Regarding Fig. 10f, the quality of the reference DSM has also to be considered. For DSMs from optical imagery, the height accuracy is related to the resolution (0.15″ for AW3D) and decreases with slope, i.e. the accuracy might not be sufficient to evaluate TanDEM-X DEM.

In this model-to-model comparison study we found the best vertical agreement between TanDEM-X DEM and the LIDAR DTM

Table 6
Accuracy numbers (m) for TanDEM-X versus GPS benchmarks height differences for terrain with slope values less and greater than 10°.

Slope < 10° (24,889 points)						Slope ≥ 10° (336 points)					
ME	STD	RMSE	MAD	NMAD	LE90	ME	STD	RMSE	MAD	NMAD	LE90
0.12	1.39	1.39	0.55	0.81	2.59	0.24	2.46	2.47	1.27	1.89	5.49

Table 7
Accuracy numbers for high-resolution DEM test sites depending on slope (height differences TanDEM-X minus high-resolution reference DEM).

Slope (degree)	No. of points	ME (m)	STD (m)	RMSE (m)	MAD (m)	NMAD (m)	LE90 (m)
Cape Town (LIDAR, DTM)							
0–10	18,289,693	0.30	3.88	3.89	0.60	0.89	2.19
10–20	1,678,456	0.22	6.37	6.37	0.62	0.92	2.40
20–30	645,451	0.40	13.21	13.22	0.84	1.24	4.54
>30	370,296	−1.16	36.98	37.00	3.21	4.76	43.86
Whole area	**20,983,896**	**0.27**	**6.78**	**6.78**	**0.62**	**0.92**	**2.39**
Thuringia (LIDAR, DSM)							
0–10	422,941	−0.62	2.33	2.41	0.71	1.05	3.48
10–20	183,157	−1.22	4.25	4.42	2.15	3.19	6.96
20–30	62,015	−1.06	6.19	6.28	3.59	5.32	10.26
30 < u	20,907	0.31	8.45	8.45	5.15	7.64	14.02
Whole area	**689,020**	**−0.79**	**3.72**	**3.80**	**1.18**	**1.75**	**5.90**
Kumamoto (AW3D, DSM)							
0–10	24,015,923	2.70	3.39	4.33	1.33	1.97	5.88
10–20	15,566,979	3.88	4.23	5.74	1.98	2.94	8.41
20–30	11,124,794	4.49	5.15	6.83	2.56	3.80	10.38
>30	6,730,087	6.35	8.09	10.29	3.83	5.68	15.54
Whole area	**57,437,783**	**3.79**	**4.88**	**6.18**	**1.96**	**2.90**	**8.98**

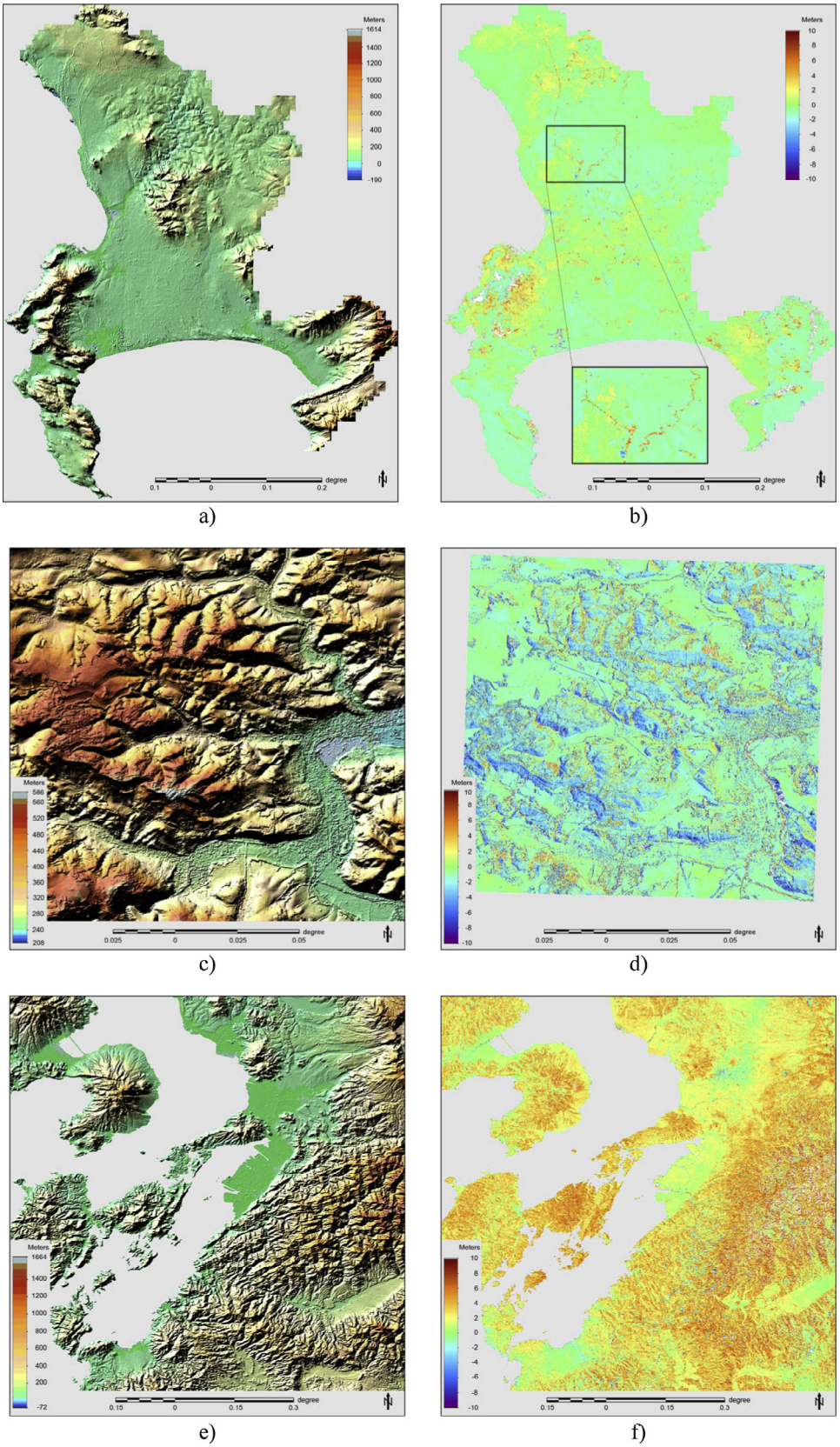

Fig. 10. TanDEM-X DEM of Cape Town (a), Thuringia (c) and Kumamoto (e) and corresponding height differences to TanDEM-X DEM against Cape Town DTM (b), Thuringia DSM (d) and Kumamoto DSM (f), scaled to ±10 m.

from Cape Town. These findings stem from the predominantly flat terrain and low vegetation. Although the results for the two DSMs show a good quality in general, the results are uncertain for forested areas. In addition, the comparison of the TanDEM-X DEM with AW3D is influenced by different penetration of X-Band SAR and optical measurements for forested areas. Future model-to-model validation work should be carried out on a larger variety of landscapes with an adequate accuracy of the reference data. In consequence to our results, we would suggest for further validation data sets accuracies below 1 m standard deviation considering the good quality of the TanDEM-X DEM.

4. Conclusions

In this study we analyzed the vertical accuracy of the TanDEM-X DEM (0.4" resolution, approx. 12 m) product globally using (a) kinematic GPS data with an accuracy of <0.5 m distributed across all continents (except the Antarctic), (b) punctual, stationary GPS data covering the whole of the US ("GPS on Bench Marks") provided by NGS for land cover-specific analysis and (c) two high resolution DTMs and a DSM. Several important observations about the accuracy of the TanDEM-X elevations can be drawn from our analyses.

The TanDEM-X DEM shows a high agreement between all the ground truth data sources. In terms of an absolute vertical bias, TanDEM-X DEM shows very low mean errors, when comparing to both GPS data sets (−0.17 m for KGPS and +0.12 m for GPS-on-BM). The determined LE90 of 1.9 m (kinematic GPS) and 2.0 m (GPS benchmarks) confirm an excellent absolute error at 90% confidence level below 2 m. The NMAD - as a more robust measure for the 68% probability level than the RMSE or the standard deviation - is even below 1 m.

The TanDEM-X DEM represents a surface height model and includes non-ground-level elevations for natural and artificial aboveground features (tree canopies and built structures). However, the positive mean errors for aboveground land cover classes are lower than expected for an interferometric SAR system such as TanDEM-X. The mean errors of 0.13 m for ground level in developed classes and 0.38 m for ground level in forest classes are probably underestimated as the majority of the measured GPS points are located in open areas. The RMSE values for ground level in developed and forest areas show slightly higher deviations of ±1.4 m and ±1.8 m, respectively, whereas the low vegetation class remains with a RMSE of ±1.1 m.

The height error annotated in the TanDEM-X HEM layer is a good estimate of the height error. The measured height differences correlate linearly with the HEM values, i.e. increased annotated height errors in the height error map indicate increased measured height differences.

For all examined data sets, the TanDEM-X data meets and even exceeds the 10 m (90 percent) performance goal, often by a factor of 5. Nevertheless, at local scale, a case-by-case validation of the vertical accuracy of TanDEM-X DEM is essential for the understanding of the potential and limitations of using this global dataset for a specific area.

Acknowledgments

The TanDEM-X project is partly funded by the German Federal Ministry for Economics and Technology (Förderkennzeichen 50 EE 0601). For providing high-resolution reference DEMs we thank the NTTDATA Corporation and the Remote Sensing Technology Center of Japan for the AW3D, the open data portal of the City of Cape Town and geoportal Thuringia. Further open source data were provided by the Multi-Resolution Land Characteristics (MRLC) consortium by the NLCD data and the NGS by the GPS on benchmarks data set. For acquiring the GPS data our special thanks go to the teams on the roads and to Prof. Volker Schwieger and team from University of Stuttgart for processing and consulting. Finally, the author would like to thank the anonymous reviewer for their valuable comments to improve this manuscript.

References

Baade, J., Schmullius, C., 2016. TanDEM-X IDEM precision and accuracy assessment based on a large assembly of differential GNSS measurements in Kruger National Park, South Africa. ISPRS J. Photogramm. Remote Sens. 119, 496–508. https://doi.org/10.1016/j.isprsjprs.2016.05.005.

Bolkas, D., Fotopoulos, G., Braun, A., Tziavos, I.N., 2016. Assessing digital elevation model uncertainty using GPS survey data. J. Surv. Eng. 142 (3).

Borla Tridon, D., Bachmann, M., Schulze, D., Ortega-Miguez, C., Martone, M., Böer, J., Zink, M., 2013. TanDEM-X: DEM acquisition in the third year era. Int. J. Space Sci. Eng. 1 (4), 367–381.

Carabajal, C.C., Harding, D.J., 2005. ICESat validation of SRTM C-band digital elevation models. Geophys. Res. Lett. 32, L22S01.

Cape Town DEM, Mai 2017. Open data portal of the City of Cape Town. URL http://web1.capetown.gov.za/web1/opendataportal/DatasetDetail?DatasetName=Digital%20elevation%20model.

CSRC-PPP, Mai 2017. Online PPP service of Natural Resources Canada. URL http://www.nrcan.gc.ca/earth-sciences/geomatics/geodetic-reference-systems/tools-applications/10925#ppp.

DLR, March 2018. TanDEM-X science service system. https://tandemx-science.dlr.de.

Geoportal Thuringia, Mai 2017. Geoportal Thuringia. URL http://www.geoportal-th.de/de-de/downloadbereiche/downloadoffenegeodatenthüringen/downloadhöhendaten.aspx.

Gesch, D.B., Oimoen, M.J., Zhang, Z., Meyer, D., Danielson, J.J., 2012. Validation of the ASTER global digital elevation model version 2 over the conterminous United States. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XXXIX-B4, 281–286.

Gesch, D.B., Oimoen, M.J., Danielson, J.J., Meyer, D., 2016. Validation of the ASTER Global Digital Elevation Model version 3 over the conterminous United States. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLI-B4, 143–148.

GIPSY, September 2017. GIPSY software. URL https://gipsy-oasis.jpl.nasa.gov.

Gonzalez, C., Bräutigam, B., 2015. Relative height accuracy estimation method for InSAR-based DEMs. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (11), 5352–5360.

GPS-on-BM, Mai 2017. GPS On Bench Marks used to make GEOID12B. URL https://www.ngs.noaa.gov/GEOID/GEOID12B/GPSonBM12B.shtml.

Gruber, A., Wessel, B., Huber, M., 2012. Operational TanDEM-X DEM calibration and first validation results. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 73, 39–49.

Gruber, A., Wessel, B., Martone, M., Roth, A., 2016. The TanDEM-X DEM mosaicking: Fusion of multiple acquisitions using InSAR quality parameters. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (3), 1047–1057.

Höhle, J., Höhle, M., 2009. Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS J. Photogramm. Remote Sens. 64 (4), 398–406.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogramm. Eng. Remote Sens. 81 (5), 345–354.

HTDP, Mai 2017. NGS Horizontal Time-Dependent Positioning software. URL https://www.ngs.noaa.gov/TOOLS/Htdp/Htdp.shtml.

Huber, M., Wessel, B., Kosmann, D., Felbier, A., Schwieger, V., Habermeyer, M., Wendleder, A., Roth, A., 2009. Ensuring globally the TanDEM-X height accuracy: analysis of the reference data sets ICESat, SRTM, and KGPS-Tracks. Proceedings of IGARSS 2009, Cape Town, South Africa.

Jacobsen, K., Passini, R., 2010. Analysis of ASTER GDEM elevation models. Int. Arch. Photogramm. Remote Sens. 38 (1), 1–6.

Kosmann, D., Wessel, B., Schwieger, V., 2010. Global Digital Elevation Model from TanDEM-X and the Calibration/Validation with worldwide kinematic GPS-Tracks. XXIV FIG International Congress 2010, Sydney, Australia.

Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M., Zink, M., 2007. TanDEM-X: a satellite formation for high resolution SAR interferometry. IEEE Trans. Geosci. Remote Sens. 45 (11), 3317–3341.

Lachaise, M., Fritz, T., Bamler, R., 2018. The dual-baseline phase unwrapping correction framework for the TanDEM-X mission part1: Theoretical description and algorithms. IEEE Trans. Geosci. and Remote Sens. 56 (2), 780–798.

Maune, D.F. (Ed.), 2007. Digital Elevation Model Technologies and Applications: The DEM User Manual, second ed. ISBN: 1-57083-082-7.

Mouratidis, A., Briole, P., Katsambalos, K., 2010. SRTM 3″ DEM (versions 1, 2, 3, 4) validation by means of extensive kinematic GPS measurements: a case study from North Greece. Int. J. Remote Sens. 31, 6205–6222.

NLCD 2011. National Land Cover Database 2011. URL https://www.mrlc.gov/nlcd2011.php.

Ramm, K., Schwieger, V., 2007. Requirements on Kinematic GPS-Measurements for the Evaluation of Height Accuracy within the TanDEM-X Project", DLR TanDEM-X Ground Segment Document TD-PGS-TN-3061.

Rexer, M., Hirt, C., 2016. Evaluation of intermediate TanDEM-X digital elevation data products over Tasmania using other digital elevation models and accurate heights from the Australian National Gravity Database. Aust. J. Earth Sci. 63 (5), 599–609. https://doi.org/10.1080/08120099.2016.1238440.

Rizzoli, P. et al., 2017. Generation and performance assessment of the global TanDEM-X Digital Elevation Model. ISPRS J. Photogramm. Remote Sens. 132, 119–139 http://www.sciencedirect.com/science/article/pii/S092427161730093X.

Rodríguez, E., Morris, C.S., Belz, J.E., 2006. A global assessment of the SRTM performance. Photogramm. Eng. Remote Sens. 72 (3), 249–260.

Schweitzer, J., Zheng, B, Schwieger, V., Kosmann, D., 2010. Evaluation of the TanDEM-X Digital Elevation Model by PPP GPS – Analysis and Intermediate Results. XXIV FIG International Congress, Sydney, Australia.

Schwieger, V., Schweitzer, J., Kosmann, D., 2009. GPS precise point positioning as a method to evaluate global TanDEM-X digital elevation model. VII FIG Regional Conference, Hanoi, Vietnam.

Takaku, J., Tadono, T., Tsutsui, K., Ichikawa, M., 2016. Validation of 'AW3D' Global DSM generated from ALOS PRISM. Int. Ann. Photogramm. Remote Sens. Spatial Inf. Sci. III-4, 25–31.

Three-sigma rule, Mai 2017. URL https://www.encyclopediaofmath.org/index.php/Three-sigma_rule.

Wessel, B., Marschalk, U., Gruber, A., Huber, M., Hahmann, T., Roth, A., Habermeyer, M., Kosmann, D., 2008. Design of the DEM Mosaicking and Calibration Processor for TanDEM-X. Proceedings of EUSAR 2008, Friedrichshafen, Germany, 4, 111–114.

Wessel, B., 2016. TanDEM-X Ground Segment – DEM Products Specification Document. EOC, DLR, Oberpfaffenhofen, Germany, Public Document TD-GS-PS-0021, Issue 3.1, 2016. [Online]. Available: https://tandemx-science.dlr.de/.

WorldDEM, March 2018. WorldDEM™: The New Standard of Global Elevation Models URL http://www.intelligence-airbusds.com/files/pmedia/public/r5434_9_int_022_worlddem_en_low.pdf.

Zwally, J., 2002. ICESat's laser measurements of polar, ice, atmosphere, ocean, and land. J. Geodyn. 34, 405–445.

Zink, M. et al., 2014. TanDEM-X: the new global DEM takes shape. IEEE Geosci. Remote Sens. Mag. 2 (2), 8–23.