3D Evaluation of Fine-Scale Normalised DSMs in Urban Settings

Andrè Breytenbach

Smart Places, CSIR

DOI: http://dx.doi.org/10.4314/sajg.v9i2.26

Abstract

Humankind often needs to accurately model, identify and spatially quantify aboveground phenomena on the Earth's surface for informed decision-making. Height data derived from digital elevation models (DEMs) is often used to achieve this. This study conducted a deterministic assessment of three normalised digital surface models (nDSMs) of different spatial resolutions, namely 2m, 4m and 12m, derived from VHR digital stereo aerial photography, tri-stereo Pléiades imagery and Tandem-X InSAR data, respectively. Covering a predominantly built-up area within a city landscape, the nDSMs were vertically and volumetrically compared to assess their quality and fit-for-use. In each case a consistent systematic evaluation was accomplished against a lidar derived reference surface at matching spatial resolutions (co-registered) using a semi-automated GIS routine. The relative height and volumetric errors were statistically analysed and described, including those computed individually over nine urban land cover/land use (LCLU) classes and several selected large buildings. Higher vertical accuracies were reported across single storey structures and areas with no to little or short vegetation, as apposed to substantially lower accuracies obtained over multi-levelled buildings and tall (dense) woody vegetation. Here significant underestimations of volumes exacerbated by lower spatial resolutions were also observed across each nDSM. Conversely, notable volume overestimations were found over predominantly grass-covered areas in especially the finer-scaled nDSMs. VHR elevation data is recommended to model and quantify aboveground elements spatially in 3D (e.g. buildings, earthworks and woody vegetation) in urban landscapes, but a sensitivity test beforehand remains critical to ensure more reliable outcomes for users and stakeholders alike.

1. Introduction and Background

State-of-the-art remote sensing (RS) technology nowadays has enabled mankind to model and map the Earth's surface and features of interest thereon in increasingly higher detail and accuracy. Whether relying on radar, laser or various other ortho-optic data acquisitions from near-space high-resolution satellite image (HRSI) systems or by manned and unmanned aerial surveys, the captured spatial data enables users to generate a DEM for their specific needs. A quality, consistent (seamless) DEM offers large potential benefits not only to those involved in geomorphometry (Sofia *et al.*, 2016), but also to a variety of other disciplines (Hajnsek *et al.*, 2014). Often the need

exists to reliably model, identify, demarcate (or extract) and quantify existing aboveground anthropogenic or natural phenomena. Provided quality geo-referenced height data is available over the target area can the linear dimensions (*e.g.* size, structure or form) and volume of these features or objects be estimated using the appropriate RS/GIS software and specialised algorithms (Beumier and Idrissa, 2016). For example, it may involve creating complex geo-spatial geometries of built-up structures to assimilate building types (Krauss and Reinartz, 2010) or detect urban changes (Beumier and Idrissa, 2012) in 2.5 or virtually in 3D. Lacking in-field observations, the constructed building information might serve as proxies for their uses and possible number of inhabitants (Alahmadi *et al*., 2016) for the pricing of insurance policies or risk management, or to report on informal settlement growth trends (Kuffer and Sliuzas, 2017), as well to investigate cityscape designs (Biljecki *et al*., 2015) and urban heat island formation (Voelkel and Shandas, 2017). The elevation data can also be applied to monitor civil construction progress or aboveground mining activities (Yoshida *et al.*, 2019). In addition, research may want to use DEMs to quantify aboveground biomass or develop an accurate canopy height model (CHM), often in context of CO_2 storage and emission calculations (He *et al*., 2018; Raciti *et al.*, 2014) or natural resource management (Urbazaev *et al.*, 2016; Ginzler and Hobi, 2015; Deng *et al.*, 2014), respectively. Nonetheless, to achieve the above a fit-for-use seamless DEM with the optimal spatial resolution – normally strongly related to the user requirements, application and scope of the assessment – is vital.

A DEM constructed from topographic data always retain a physical support size that equals the original scanning resolution. This is the fixed area or volume of the land surface sampled and subsequently represented by the elevation value at the grid cell node (Hengl and Evans, 2009). In other words, it is the average value of all possible elevations in that pixel. For example, laser sensor measurements would have a smaller support size (in millimetres) and irregular spatial distribution in comparison to a passive air- or spaceborne sensor (in metres) with regular postings. Gridded DEM construction methods consisting of different interpolation and filtering techniques routinely produce seamless DEM from the sampled elevation data that correspond to terrain relief (Kramm and Hoffmeister, 2019; Reuter *et al*., 2009). The initial DSM represents a continuous land surface referenced to mean sea level (MSL) that includes the orthometric height of first return off-terrain objects such as buildings and vegetation in metres. Thus, a gridded DEM is a digital representation of continuous elevation values over a topographic surface by a regular array of x,y, and z values (Nelson *et al.*, 2009). The vertical and horizontal components are each referenced to their particular datum. A DSM may be further converted to a digital terrain model (DTM) – a digital representation of variables relating to the Earth's topographic surface (Beumier and Idrissa, 2016). The 'bare Earth' DTM is still referenced to MSL, but now technically devoid of any man-made features and vegetation after being subjected to a choice of various editing, filtering and interpolation processes (Barbarella *et al.*, 2019). Some of the main production challenges faced in these processes is to preserve the scale characteristics of different DEM (Poli *et al.*, 2009) and to preserve terrain continuity (Doytsher *et al.,* 2009). The nDSM or digital height model (Longbotham *et al*., 2012)

can then be generated by computing the vertical offset (in metres) between the overlapping DSM and DTM cell node values (Peeroo *et al.*, 2017). If subtracting an earlier DEM from a later one to detect and measure any spatial changes (Hsieh *et al.*, 2016), it is called DEM differencing. Fundamentally representing a height model, the nDSM values would commonly range from zero metres at ground level, *i.e.* the DTM, to the vertical height of the tallest feature or elevation point observed in the DSM (Figure 1).

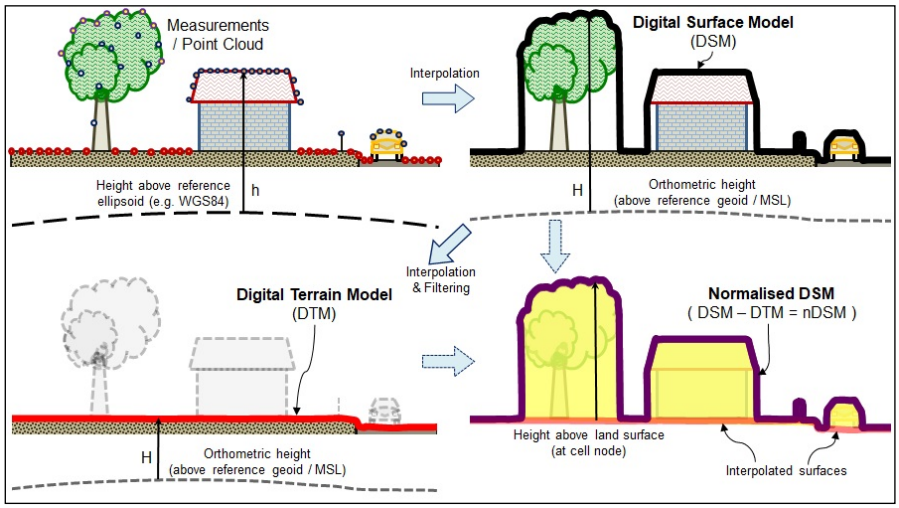

Figure 1. Various seamless DEM types generated from air- or spaceborne sensor data

If the DEM ground sampling distance (GSD) equates to exactly one metre, would each cell value theoretically also represent the precise volume (m^3) of that pixel in 3D space, as referenced from the land surface (Gröhmann *et al.*, 2011). By definition though, since only a single elevation value can be stored per grid cell or pixel, this dimension will only resonate true for solid materials or city fabrics at a particular scale and much less so where storied vegetation, tunnels, overhangs and other typical occlusions occur in urban landscapes.

The need to analyse the quality and accuracy of a DEM with a sensitivity test prior to its use has been recommended by many (Santos *et al.*, 2020; Chudý *et al.*, 2013; Temme *et al.*, 2009; Höhle and Höhle, 2009). Through the broad literature review conducted, very few studies performed a consistent fine-scale nDSM accuracy evaluation in a predominantly urban setting (Peeroo *et al.*, 2017; Beumier and Idrissa, 2016), whilst also taking in consideration the influence of the different urban LCLU types on DEM accuracy (Breytenbach and Van Niekerk, 2019; Alganci *et al.*, 2018; Balenović *et al.*, 2015). Together with a locally produced photogrammetric DEM data set, is the performance of other contemporary commercial wide-area DEM editions, e.g. the GEO Elevation Suite (Airbus, 2019[1]) and near global WorldDEM™ products (Airbus, 2019[2]), also largely untested hitherto domestically. The research therefore questions the 3D performance of fine-scale nDSM when applied in a characteristically heterogeneous urban environment along with abundant woody vegetation. This study aimed to determine the quality (*i.e.* vertical and volumetric accuracy) of these three different fine-scaled nDSMs over a well-established urbanized area. It also addresses uncertainties regarding the influence that spatial resolution and prevailing LCLU elements may have on DEM accuracy at the anticipated work

scale. It includes the systematic computation of descriptive error statistics across nine primary urban LCLU types and a substantial number of dissimilar large buildings.

2. Materials and Methods

2.1. Study Area

The study area is located in the City of Tshwane, South Africa. The 3.2 km^2 site displays relatively gentle topography with mixed natural vegetation that occurs mainly along the low hills running through the central part of the study area (Figure 2).

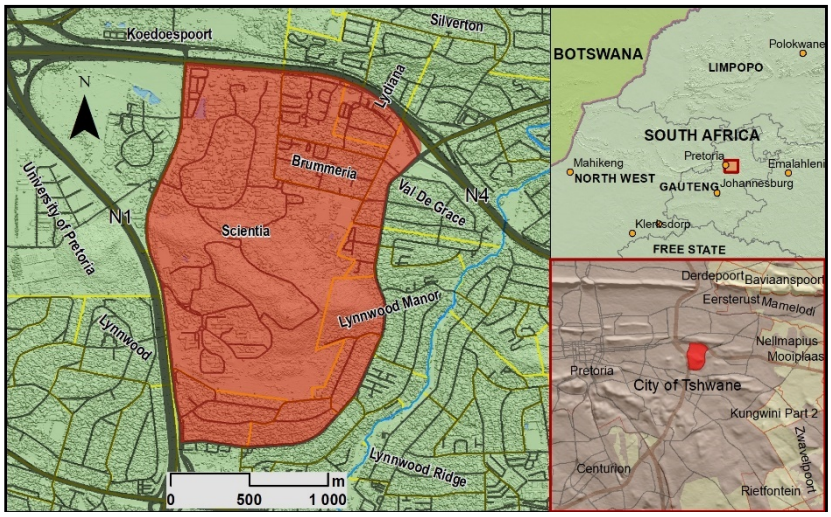

Figure 2. Location of the study area within the City of Tshwane

The most dominant land cover consists of various sized patches of natural or cultivated grass (24%) interspersed with larger woody vegetation types such as shrubs, bushes (11%) and trees (32%). Single and multi-levelled buildings (15%) and the associated heterogeneous urban impervious surfaces and transport infrastructure (14%) occur in the landscape, together with some natural bare soils and rocks (3%) and a solitary retention dam (1%).

2.2. Data Collection and Pre-Processing

The DEMs obtained, prepared and evaluated in terms of their vertical accuracy and volumetric correctness over the area of interest were overlapping samples of i) an experimental photogrammetric 2m DSM and DTM, ii) the Elevation4™ DSM and DTM products at 4m postings (Airbus, 2019[1]), and iii) the 0.4 arc-second (12m) WorldDEM™ DSM and DTM products (Airbus, 2019[2]). The technical specifications and origins of the source data, including the RS technology used to realise the respective DEM products above are listed in Table 1. All the elevation grids were hydrologically corrected 32-bit precision DEMs obtained as GeoTIFFs. The commercial data was (re-)projected to Lo29-WGS84 using a bi-linear interpolator (Wise, 2011) and by implementing the 'Snap Raster' setting in the ArcGIS 10.7 'Project Raster' tool environment. This

ensured precise pixel-to-pixel co-registration with the relevant matching reference grid at the same spatial resolution in each instance. Since all the DSMs are ultimately normalised by subtracting its corresponding DTM, any possible absolute vertical bias due to different vertical datums associated with some data pairs were thus negated. Three normalised DSMs or height models were constructed from these subsets and their quality systematically evaluated against a base nDSM derived from a lidar point cloud. Henceforth these nDSMs were (mostly) referred to in this paper as the APnDSM2, E4nDSM4 and WDnDSM12, respectively.

Table 1. The three elevation grids and reference data, their specifications and sources, the enabling RS configuration and relevancy, respectively

Sample grid	Grid size	Source	Supplier / Owner	Data origin (platform)	Product coordinate system	Year
APnDSM2	2m	50cm Aerial photographs	CSIR Pretoria	Digital mapping camera	TM (Lo29), Hartebeesthoek1994	2012
E4nDSM4	4m	Elevation4™ DEMs	Airbus (France)	Tri-stereo HRSI (Pleiades1-B)	UTM-35S, WGS84	2013
WDnDSM12	12m	WorldDEM™ DEMs	Airbus (France)	Single-pass InSAR (Tandem-X)	Geographic, WGS84-G1150	2014
Reference nDSM	1m	Airborne laser scanning	City of Tshwane	ALS50-II LiDAR	TM (Lo29), Hartebeesthoek1994	2013

Lidar have gained in popularity in those spheres of industries where highly detailed and accurate 3D data is essential (ASPRS, 2013). Routinely mounted on a variety of airborne platforms, airborne laser scanning (ALS) systems can capture accurate topographic data related to the Earth's surface. By measuring the location and attitude of the cruising aircraft, the Euclidean distance to ground and scan angle (with respect to the base of the laser scanner housing), a precise 3D ground position for the impact point of each laser pulse can be determined (Tinkham *et al.*, 2012). This yields direct, 3D measurements of the ground surface, vegetation, buildings and various infrastructures as required (Peeroo *et al.*, 2017; Basgall *et al.*, 2014). The ability to digitize either the signal strength or the range to the reflecting surface is dependent on that surface having adequate reflectivity (Demir *et al.*, 2009). Provided each target results in adequate signal strength for detection, a lidar system is normally capable of detecting up to four targets for each outbound laser pulse (i.e., first, second, third and last return). The base DSM, DTM and derived nDSM reference grids used in this case study originated from ALS data, which fundamentally ascribed to the recommendation that the reference data should be three times better than the tested data (ASPRS, 2013). Surveyed from a manned aircraft flying at an altitude of around 1,500m, it covered the entire Tshwane metropole. The required vertical accuracy for the ALS survey was set to 0.08m (*RMSE*) whilst premarks were utilized as ground control points. The ALS post-survey validation results, sampled across 30 points, revealed a mean vertical difference of +0.078m, standard deviation of 0.059m, and $RMSE_z$ measuring 0.062m. Acquired simultaneously with this mission was optical data captured with a Kodak KAI-11002 dual charge coupled device to produce VHR digital RGB aerial ortho-photos (10cm GSD). With multi-pulse mode enabled using an ALS50-II lidar instrument and with 30% strip overlap achieved, this resulted in (on average) eight

observations per square metre on ground level. The return-signal intensity data was processed into a dense point cloud and ultimately classified into discreet ground and non-ground points using a single algorithm across the entire dataset. These classified *x,y,z* measurements in turn formed the primary input when generating both a seamless 1m DSM and DTM, also hydrologically-corrected with the ANUDEM algorithm using the drainage enforcement option (Hutchinson *et al.*, 2011). These 32-bit baseline products came in the preferred geodetic reference framework (*i.e.* Lo29-WGS84). The primary nDSM used to generate the upscaled (bi-linearly) reference surfaces matching the postings of the test nDSMs over the study area was produced from these two base elevation grids.

DEM quality is inherently influenced by surface morphology and vegetation structure as shown earlier. Tinkham *et al.* (2012) however found that vegetation structure have almost no influence on lidar derived DEM errors, whereas increased variability in the vertical error metrics was observed on steeper slopes (>30°). It thus illustrates that lidar classification algorithms are not limited by high-biomass forests, but rather that slope and sensor accuracy both play important roles. The lidar data was thus well suited to reliably model vegetation and man-made structures in the derived base DSM. The resulting base 1m nDSM was also employed to enhance the required LCLU classification layer over the study area. A VHR (1.24m) multi-spectral WorldView-3 image collected over the study area in 2014 was first used to perform a supervised LCLU classification that initially comprised of twenty-three land cover classes. After significantly improving the classification result with the incorporation of the classified height data (Gxumisa and Breytenbach, 2017), further extensive manual edits and corrections were performed (mainly concentrated on the built-up zones) to ultimately deliver a highly accurate LCLU dataset at 2m GSD. Similar to the multi-DSM evaluation approach followed by Alganci *et al.* (2018) where accuracy was assessed across seven primary urban land cover categories within the Istanbul metropolitan area in Turkey, this study proceeded to derive nine primary classes from the initial 23 LCLU classes in the study area for evaluation purposes.

2.3. Computing Height Errors and Accuracy Statistics

Several established statistical descriptors were computed as listed in Table 2 to evaluate the *relative* (point-to-point) accuracy of each nDSM as measured against the corresponding co-registered reference surface. Other than mean (*ME*) and standard deviation (*SD*) of the height errors, Euclidean distance based mean absolute error (*MAE*), quadratic mean (*RMSE*) and normalised mean absolute deviation (*NMAD*) measures were also calculated because they are regarded as less sensitive to resolution differences and equally useful when variates have positive and negative signs (Höhle and Höhle, 2009; Kramm and Hoffmeister, 2019). Using various ArcGIS 10.7 tools in the 'Model-Builder' environment, the above error metrics were computed for each height raster in an automated manner across its entire surface and within each of the nine LCLU classes that were prepared beforehand (Figure 3, right). Along with the generally smaller

residential dwellings and complexes too concealed by cultivated (urban) tree canopy cover, the built-up LCLU mix varied between large institutional and government buildings (including an international conference centre facility), together with small zones related to commercial and light industrial activities or educational facilities.

Table 2. The applied vertical error statistics, formulas and their respective variables

Description	Unit	Formula	Variables
Height: total offset (Δh)	m	$h_{sample} - h_{ref}$	Sampled nDSM height (h_{sample}) and reference nDSM height (h_{ref})
Height: mean error (*ME*)	m	$\frac{1}{n}\sum_{i=1}^{n}\Delta h_i$	Number of grid nodes (n) sampled
Height: standard deviation of errors (*SD*)	m	$\sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(\Delta h_i - ME)}$	
Height: mean absolute error (*MAE*)	m	$\frac{1}{n}\sum_{i=1}^{n}\lvert\Delta h_i\rvert$	
Height: root mean square error (*RMSE*)	m	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\Delta h_i^2}$	
Height: normalised mean absolute deviation (*NMAD*)	m	$1.4826 \cdot median_i\left(\lvert\Delta h_i - m\Delta h\rvert\right)$	Median of the individual errors $i = 1 \ldots n$ ($m\Delta h$)
Aboveground volume (V_a)	m^3	$\sum_{i=1}^{n} h \cdot (cell\ size)^2$	Qualified cells (per class)

Figure 3. The spatial distribution of the nine primary LCLU types (on the left) and the sampled 72 large buildings along with the NDVI mask in green shades (on the right)

Since all the different RS technology involved in this study have appreciably different vegetation penetration capabilities (Deng *et al.*, 2014) and because a DEM grid can by definition only have a single elevation/height value per cell node (Reuter *et al.*, 2009), the next accuracy test determined the height errors only across several solid, large individual buildings – the overall majority having level rooftops. Seventy-two large buildings were selected in total and their exact footprints were digitally captured in GIS (Figure 3, left) using the 10cm RGB orthophotos and verified by several field inspections. The minimum footprint size was set at three pixels of the coarsest subset being tested, *i.e.* 432m^2 (3×12m×12m). To account for possible occlusions by tree canopies over any of these 72 built-up structures, a distinct vegetation layer was produced. This was achieved by first classifying the VHR multi-spectral WorldView3 and RGB images with a modified NDVI algorithm before combining the outputs and temporally adjusting it. This layer (also shown in Figure 3) was finally converted to a simple 1/0 Boolean mask (zero being vegetation free) at a 2m grid resolution. Thus, all non-zero cells traversed during sampling were summed and expressed as a percentage of the total number of intersecting grid cells per building footprint. This ratio could thus later be correlated against individual results exhibiting excessive height errors to consider whether the presence of any tree cover occlusions could possibly have been a major contributing factor to the erroneous offset.

All height offsets computed thus far described the average errors over the complete nDSM, individual LCLU classes or buildings in question and the total sum thereof would be nonsensical. Yet, this is different when measured in 3D space as volume per pixel. Aboveground volume (V_a) at each cell node was equated as height multiplied by cell area (m^3) for each nDSM before comparing it with the corresponding pre-processed reference volumes. The total summed volume represented by the extracted reference grid cells over a particular class or area can then be subtracted from the summed volume produced by the corresponding cells in the sample data (ΔV_a) in order to estimate its 3D dimensions and magnitude (Gröhmann *et al.*, 2011).

3. Evaluation Results

3.1. Relative Height Accuracies

Statistical analysis of the relative vertical offsets measured between the three nDSMs and their corresponding reference grid over the study area produced different height accuracies for each. All the computed error statistics are listed in Table 3, including the respective sample sizes. The APnDSM2 performed best in terms of vertical accuracy (*i.e.*, 1.67m *NMAD*), followed by the E4nDSM4 (*i.e.*, 2.44m *NMAD*), and the WDnDSM12 slightly worse (*i.e.*, 2.79m *NMAD*). With negative mean errors for all nDSMs overall, it suggested a general underestimation of the actual heights. In all three cases, these were mostly contributable to the height inaccuracies in the source data over the vegetative component of the land cover, particularly larger trees and dense bushes. The significant magnitude and spatial distribution of these relative vertical errors along with others, here classified in natural breaks (Jenks), can be viewed in Figure 4. The topographic height

errors of the APnDSM2 were the most uniformly distributed, whereas the E4nDSM4 did have some other localised areas that produced noticeable errors – with both positive and negative offsets discernible.

Table 3. Point-to-point height error statistics computed across the entire extent

		Relative vertical error statistics (m)				
Subset	***n***	***ME***	***SD***	***MAE***	***RMSE***	***NMAD***
APnDSM2	795,616	-1.02	2.67	1.78	2.32	1.67
E4nDSM4	200,530	-0.81	3.03	2.11	2.69	2.44
WDnDSM12	22,287	-1.77	3.35	2.69	3.67	2.79

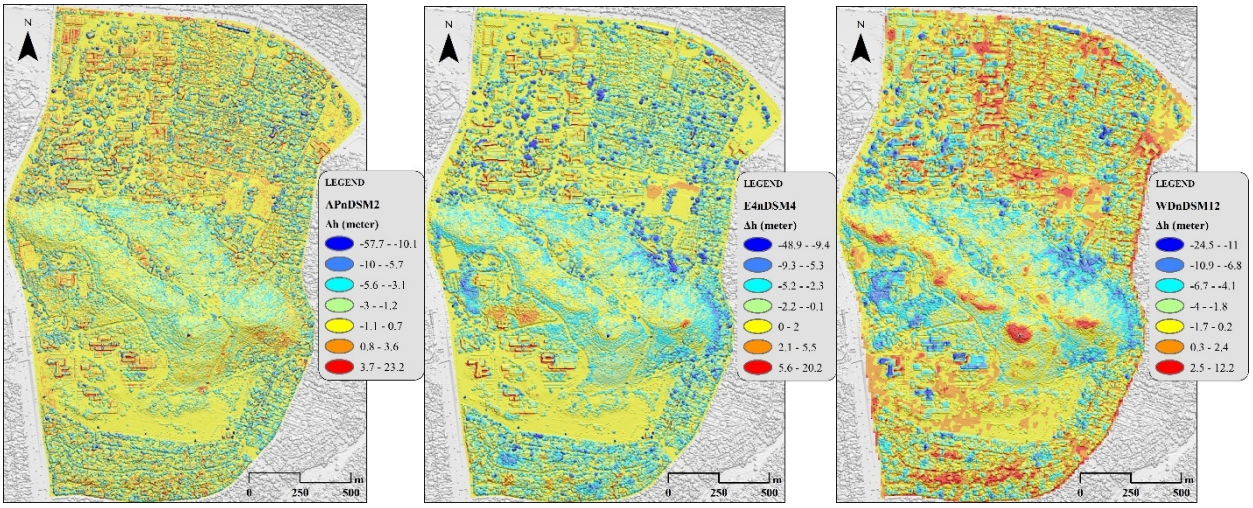

Figure 4. Magnitudes and spatial distributions of the relative height errors for the (left) APnDSM2, (middle) E4nDSM4, and (right) WDnDSM12.

So did the WDnDSM12, but more pronounced as evident with the underestimations in areas of dense and/or high vegetation in general. Thus, regardless of the sensor involved, here the blue shades observed over the dense patches of vegetation substantiated the general under measurement of height. Localised height overestimations (red shades) were also more prominent with the WDnDSM12, particularly on the extreme hills tops and ridges (in the central part of the study area) and in parts of the formal residential areas. These relatively larger underestimations across all the three nDSMs in tall vegetation, together with the prominent positive errors found along the vertical break lines (footprint perimeter) of many large buildings were also clearly expressed in the detailed accuracy statistics computed across each of the nine LCLU classes (Table 4). Conversely, rocks, bare soils or with areas with scarce or short vegetation (*e.g.*, cultivated grass) generally recorded the best vertical accuracies. When the built-up component were scrutinized, the same reasonable accuracies applied for single storey houses, but – as with tall trees – rapidly deteriorated when most multi-storey buildings were encountered. Similarly, the lower than expected accuracies in the 'Built-up–Artificial/Bare/Paved' class could possibly also be attributed to extensive tree canopy cover along paved roads, urban streets and courtyards, including within paved parking areas, here also often covered with translucent carport structures and numerous stationary vehicles during daytime. Although artificially levelled, the accuracies over the waterbody class differed by 0.3m at most between the three nDSMs.

Table 4. Relative height error statistics within each of the nine primary LCLU classes

APnDSM2		Relative height error statistic (m)			
LCLU	***n***	***ME***	***SD***	**MAE**	**RMSE**
VEG – Trees	255,543	-3.17	2.71	3.39	4.18
VEG – Tall shrub/Grass	93,502	-0.95	1.69	1.61	1.94
VEG – Grass	145,627	0.82	1.62	0.96	1.82
VEG – Grass & Bare soil	44,950	0.15	1.42	0.72	1.43
BLT – Artificial/Bare/Paved	110,876	0.23	1.85	0.96	1.86
BLT – Single storey	63,102	-0.05	1.81	1.28	1.81
BLT – Multi storey	54,580	-0.67	2.80	1.72	2.88
NAT – Bare rock/Gravel/Soil	25,655	-0.91	1.40	1.17	1.67
WTB – Artificial dam	1,781	0.29	1.58	0.78	1.60
E4nDSM4					
VEG – Trees	63,979	-3.47	3.28	3.81	4.77
VEG – Tall shrub/Grass	23,194	-0.75	1.61	1.35	1.78
VEG – Grass	36,523	1.05	1.64	1.57	1.95
VEG – Grass & Bare soil	11,276	1.19	1.45	1.42	1.88
BLT – Artificial/Bare/Paved	29,273	1.12	1.68	1.57	2.02
BLT – Single storey	15,760	-0.18	1.58	1.15	1.59
BLT – Multi storey	13,649	-0.43	2.58	1.63	2.62
NAT – Bare rock/Gravel/Soil	6,427	0.19	1.24	1.02	1.25
WTB – Artificial dam	449	0.54	1.64	1.08	1.72
WDnDSM12					
VEG – Trees	8,191	-4.16	3.19	4.36	5.24
VEG – Tall Shrub/Grass	2,178	-1.13	1.98	1.91	2.28
VEG – Grass	3,725	0.62	1.67	1.23	1.78
VEG – Grass & Bare soil	933	0.00	1.24	0.84	1.24
BLT – Artificial/Bare/Paved	3,378	0.77	2.09	1.61	2.20
BLT – Single storey	1,731	-0.61	1.91	1.57	2.00
BLT – Multi storey	1,504	-4.15	3.54	4.33	5.46
NAT – Bare rock/Gravel/Soil	606	-0.92	1.50	1.42	1.76
WTB – Artificial dam	41	1.07	1.57	1.43	1.88

Examining the relative mean height errors computed across each of the 72 large building footprints, the APnDSM2 and E4nDSM4 performed notably well (Table 5).

Table 5. Relative height errors and statistics measured across seventy-two large buildings

Subset	Relative vertical height errors and statistics (m)					
	MIN	**MAX**	**ME**	**SD**	**MAE**	**RMSE**
APnDSM2	-2.43	2.54	-0.15	0.69	0.50	0.69
E4nDSM4	-2.72	1.31	-0.21	0.73	0.55	0.76
WDnDSM12	-12.10	0.95	-4.76	2.73	4.80	5.48

Calculated at around -15cm and *MAE* at 50cm, the APnDSM2 demonstrated high accuracy with the E4nDSM4 around 5cm less accurate according to the same descriptors. This opposed to the coarser WDnDSM12 that underestimated the mean height at -4.8m, and a *MAE* of the same magnitude. The consistent underestimations evident in all the assessed grids nonetheless confirms the considerable influence that the collection of incorrectly classified land cover and/or height pixels commonly found along the perimeter of these built structures (due to the original view-angles and the distinct presence of shadows) have on the overall accuracy of the data when included within their footprints. Plotting the vertical mean height errors per nDSM across all 72

buildings arranged from lowest (~3.9m) to tallest (~16.3m) mean reference structure height (Figure 5(a)), the accuracies of the two higher resolution datasets were clearly superior, seemingly irrespective of building height overall.

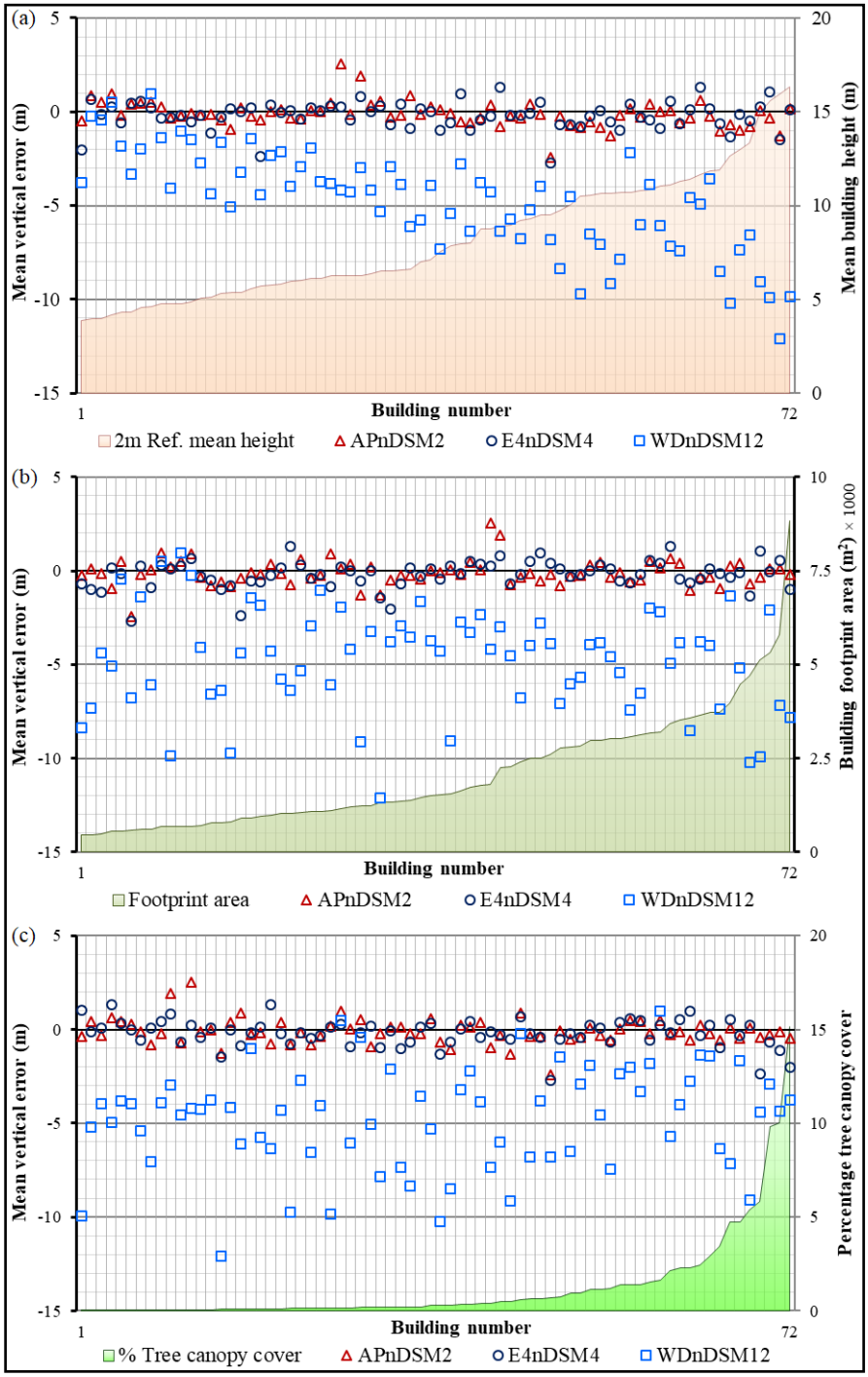

Figure 5. Mean height errors measured across 72 large buildings when arranged by the reference (a) mean building height, (b) footprint area and (c) possible tree canopy cover percentage

Their error values were clustered closely around the *x*-axis and significantly diverging from zero and each other in only a few instances, yet not necessarily with the taller building structures. In contrast, the WDnDSM12 displayed a significantly larger and divergent decrease in accuracy (mean heights underestimations) as the built-up structures became taller. When arranged from

smallest (~463m^2) to largest (~8,834m^2) reference building footprint area (Figure 5(b)) or from none (zero) to largest (~15%) proportion of each building that might potentially be covered by tree canopies (Figure 5(c)), it was apparent that neither had any discernible influence or correlation on the vertical mean accuracies obtained. Thus, the physical aerial extent of the 72 large buildings and (possible) vegetation cover played no significant role in the relative height accuracy results on these structures for all three nDSMs.

3.2. Aboveground Volumetric Accuracies

The accuracy results from the aboveground volumetric assessment over the nine land cover types can be inspected in Figure 6.

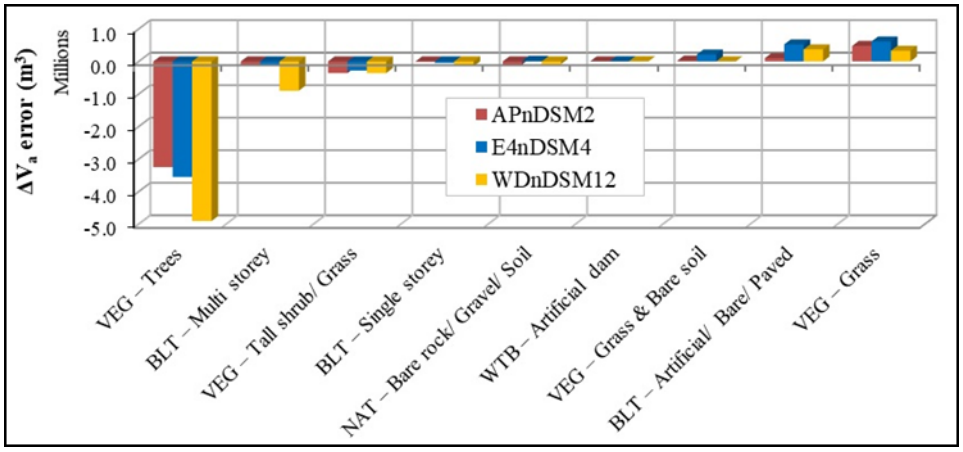

Figure 6. The aboveground volumetric differences computed over nine land cover types

The VEG–Trees class produced the largest total V_a discrepancies, with substantial underestimations by all three nDSMs assessed (*i.e.*, the WDnDSM12 the most at -4,905,460m^3, the E4DSM4 at -3,550,440m^3 and the APnDSM2 at -3,245,100m^3). Under-measurements were also returned over multi-storey buildings, but notably smaller in magnitude *(i.e.*, the APnDSM2 at -146,180m^3 and E4nDSM4 at -94,980m^3). Conversely, overestimations of V_a were observed where cultivated grass species were more abundant and over built-up impervious and artificial surfaces (*e.g.*, across the paved transport network and tennis courts). Here the E4nDSM4 was less accurate and to a lesser extent the APnDSM2, whereas the WDnDSM12 was at times comparatively more accurate than the E4nDSM4 in these land cover types. The four remaining LCLU classes returned relatively smaller volumetric differences from the three nDSMs. Another way to comprehend the volumetric errors is to express it as percentage over- or underestimation per LCLU class as measured against their respective reference volume totals (Figure 7). In the VEG–Trees class the APnDSM2 V_a underestimation was around 39%, the E4nDSM4 about 46% and the WDnDSM12 just over 62%, whereas the finer-scaled data sets underestimated the multi storey class by less than 8% and the WDnDSM12 around 52%.

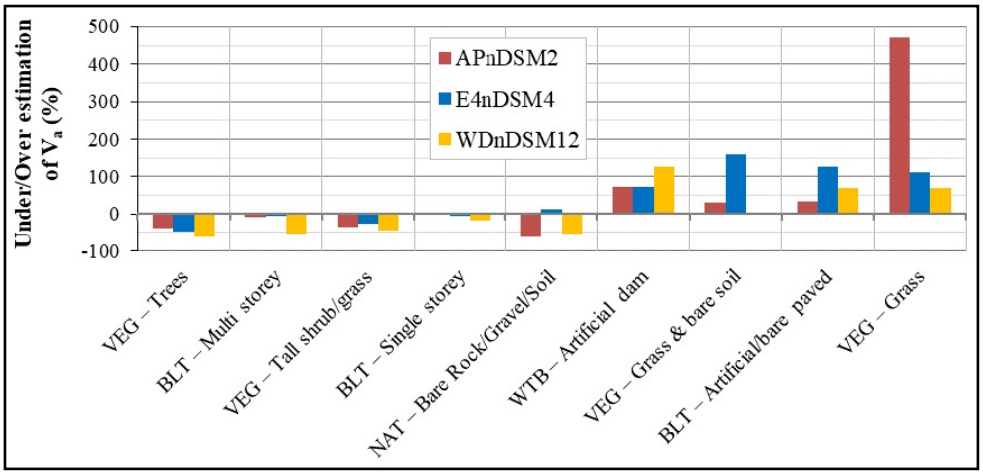

Figure 7. Percentage aboveground volumetric under/over estimations over nine land cover types

The BLT–single storey class had the least deviation, where the APnDSM2 underestimated V_a by almost 2%, the E4nDSM4 at around 5% and the WDnDSM12 by almost 18%. Areas with tall shrubs and bush recorded volume underestimations from around 26% (E4nDSM4) to almost 43% (WDnDSM12). On the other hand, the E4nDSM4 overestimated the (mixed) grass and bare soil class volumes by almost 160%, the artificial and bare built-up areas by almost 127% and the VEG–grass class by 112%. In last-mentioned class the APnDSM2 notably overestimated V_a at around 470% compared to the WDnDSM12 at about 70%. With the retention dam being relatively small and the water surface/level artificially smoothed in all the data sets, were the differences regarded as inconclusive here.

Observation of the V_a accuracies across the 72 large buildings, this time sorted from smallest to largest reference volume (Figure 8(a)), largely echoed the error trends observed earlier when mean height was considered. The APnDSM2 performed best in estimating aboveground building volume, followed closely by the E4nDSM4, but now only faltered significantly with the (last) two largest sized structures. The WDnDSM12 volume errors displayed a similar, but relatively smoother, more gradual correlation with total building volume than with building height, and again with diminishing accuracy (underestimations) as the total volume increased per building. Once more, it was distinguished that the building footprint area did not have any significant influence in the volumetric differences found here (Figure 8(b)) with any of the nDSMs tested.

4. Discussions

In terms of relative height accuracy the fine-scale APnDSM2 results demonstrated its superiority, with the E4nDSM4 following closely. The *RMSE* and *NMAD* values concurred with those reported by Kramm and Hoffmeister (2019) and demonstrated their usefulness for a variety of applications that employs 3D data in an urban context. The WDnDSM12 consistently produced the lowest vertical accuracy results, other than the relatively better outcomes produced within certain land cover types, particularly where cultivated grass and areas with less to no vegetation were encountered.

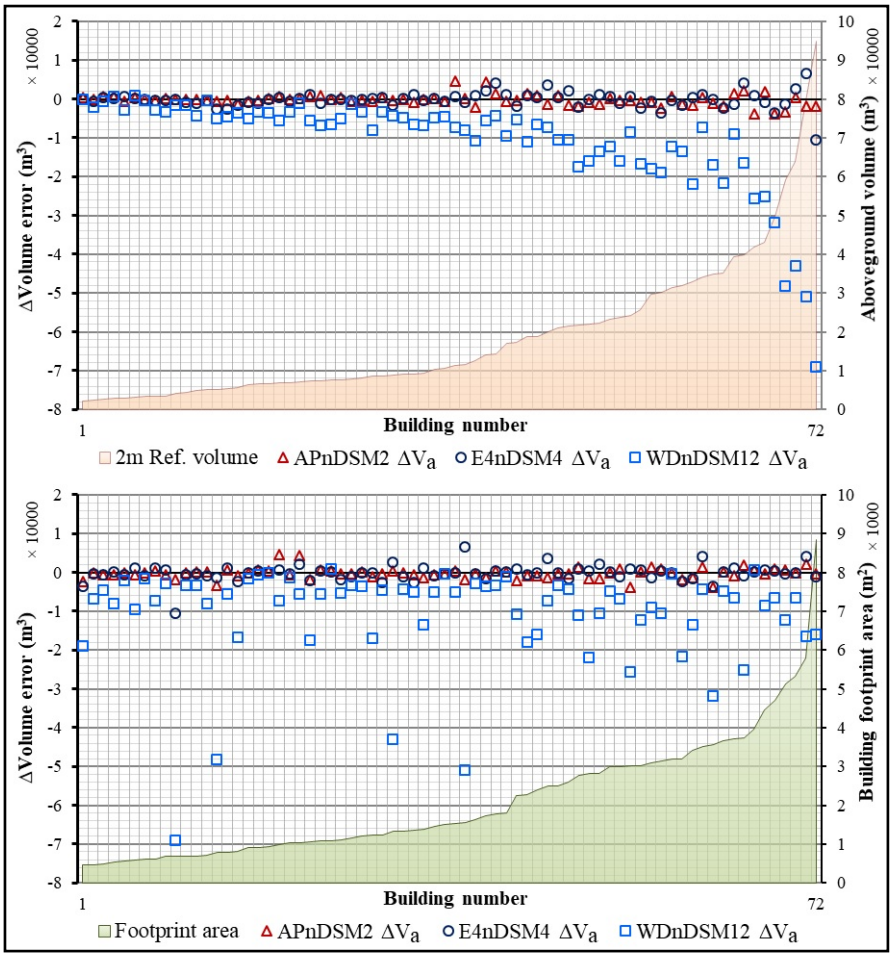

Figure 8. Volumetric differences measured across 72 large buildings when arranged by the reference building volume (top) and footprint area (bottom)

However, all three nDSMs produced relatively larger erroneous offsets where multi-levelled (dense) woody vegetation or buildings were encountered in comparison to the other LCLU classes. Here significant underestimations were returned in both height and volume as these features increased in the vertical dimension – the WDnDSM12 in particular. Taking into consideration the WorldDEM™ product's global coverage at 0.4 arc-sec degree however, it would imply some over-filtering and/or excessive interpolation to realistically achieve the same standard and fall within the mission specifications at all relevant locations worldwide. Conversely, significant overestimations of V_a may occur over natural and cultivated grass surfaces as demonstrated by even the fine-scaled ortho-optic products, with the APnDSM2 producing a notably overestimation of around 470%. Seeing that the aerial proportion of VEG–Trees (32%) and VEG–Grass (24%) classes combined covers more than half the study area, may such under- and overestimations be viewed as problematic for some decision-makers. This might lead to substantial under- and overestimations of aboveground biomass, building volumes or other process-based changes if not factored into the calculations.

Research by Gröhmann *et al.* (2011) attested that the larger the cell size, the less accurate DEM volume calculations became. This was clearly demonstrated by the APnDSM2 and E4nDSM4 being highly accurate when measured over the exact extent of (mainly) level-roofed large

buildings, whereas the coarser WDnDSM12 showed notable deterioration in accuracy, especially when the building height or volume progressively increased. In concurrence with other past research, the VHR ortho-optic models would in practice thus be most useful in the more localised and fine-scaled urban geometric applications such as change detection, cityscape designs or construction progress monitoring that depend on accurate topographic height data (Beumier and Idrissa, 2016; Krauss and Reinartz, 2010). The wide-area WorldDEM™ products would effectively be more suitable for geo-spatial investigations on a larger metropolitan or regional work scale (Breytenbach and Van Niekerk, 2019) than localised studies, particularly when processing time is also taken into account. The total 2D footprint size and possible tree canopy cover percentage nonetheless had no discernible influence in or correlation with the vertical accuracy results obtained across the 72 large buildings assessed. The last-mentioned factor may however only apply to the larger multi-levelled buildings and not the smaller single storey built-up structures, such as residential dwellings and individual small business set-ups, which are here substantially more obscured by cultivated tree canopies.

5. Conclusion

When considering specialised DEM applications in a predominantly urban setting, VHR elevation data (*e.g.* DEMs derived from dense point clouds) will provide better outputs than the available high to medium resolution (near global) topographic products. Comparatively this was clearly demonstrated in this study by the 12m WDnDSM12 that (due to relatively high spatial autocorrelation) produced the lowest vertical height and volumetric accuracy overall as opposed to the significantly higher accuracies produced by the 2m APnDSM2 and 4m E4nDSM4. Nevertheless, all three nDSMs underestimated the height and volume of dense woody vegetation and large buildings, especially the coarser dataset. In addition, significant overestimations over predominantly grass-covered areas by all three data sets were also demonstrated – the APnDSM2 in particular. Thus, some caution should be appropriated if calculating critical aboveground building, earthworks and vegetation volumes in relation to urban change detection or monitoring applications and subsequent decision-making at a local scale. A sensitivity test and evaluation of user parameters in line with the project goals beforehand is therefore recommendable. Despite the abundant fine-scale topographic data and seamless DEM products available nowadays, more research towards the most effective and efficient methodology to construct a quality DTM from the first stage DSM (Beumier and Idrissa, 2016) is needed to in turn arrive at a fit-for-use nDSM. Concerning fine-scale HRSI and aerial ortho-optic data, the correct choice of ground filtering algorithm remains pertinent to effectively remove all man-made structures and correctly process areas with high density vegetation (Barbarella *et al.*, 2019), particularly in comparison to the advantageous multi-returns and small support size offered by lidar data (Basgall *et al*., 2014).

6. Acknowledgement

The author would like to thank: The Council for Scientific and Industrial Research (CSIR) for the hard- and software environment and time to produce the photogrammetric DEM data set and conduct the research, as well as procuring the Elevation4™ products; The City of Tshwane and AAM Geomatics for permitting the CSIR to freely utilise the lidar data, VHR aerial photography and metadata; Airbus France for freely supplying the WorldDEM product samples conditional to a Technical Evaluation License Agreement with the CSIR.

7. References

Airbus 2019[1], ElevationProductSuite_web_201906.pdf, viewed 1 November 2019, <https://www.intelligence-airbusds.com/elevation-models/#elevation>.

Airbus 2019[2], WorldDEM™ Technical Product Specification, viewed 8 July 2019, <https://www.intelligence-airbusds.com/files/pmedia/public/r51492_9_2019-04_worlddem_technicalspecs_version2.5_i1.0.pdf>.

Alahmadi, M, Atkinson, PM, & Martin, D 2016, 'A comparison of small-area population estimation techniques using built-area and height data, Riyadh, Saudi Arabia', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1959-1969.

Alganci, U, Besol, B & Sertel, E 2018, 'Accuracy assessment of different digital surface models', ISPRS *International Journal of Geo-Information*, vol. 7, no. 114, 1-16.

ASPRS 2013, 'American Society of Photogrammetry and Remote Sensing Accuracy Standards for Digital Geospatial Data', *Photogrammetric Engineering & Remote Sensing*, December 2013, pp. 1073-1085.

Balenović, I, Marjanović, H, Vuletić, D, Paladinić, E, Sever, MZS & Indir, K 2015, 'Quality assessment of high density digital surface model over different land cover classes', *Periodicum Biologorum*, vol. 117, no. 4, pp. 459-470.

Barbarella, M, Cuomo, A, Di Benedetto, A, Fiani, M & Guida, D 2019, 'Topographic base maps from remote sensing data for engineering geomorphological modelling: An application on coastal Mediterranean landscape', *Geosciences*, vol. 9,500, pp. 1-29.

Basgall, PL, Kruse, FA & Olsen, RC 2014, 'Comparison of lidar and stereo photogrammetric point clouds for change detection', in MD Turner, GW Kamerman, LM Wasiczko-Thomas and EJ Spillar (eds.), *Laser Radar Technology and Applications XIX; and Atmospheric Propagation XI*, vol. 9080 (14 pp.), Baltimore, May 2014, SPIE, Washington, 2014.

Beumier, C, Idrissa, M 2012, 'Building change detection from uniform regions', in L Alvarez, M Mejail, L Gomez, J Jacobo (eds), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Lecture Notes in Computer Science, vol. 7441 (pp 648-655), Springer-Verlag, Berlin.

Beumier, C & Idrissa, M 2016, 'Digital terrain models derived from digital surface model uniform regions in urban areas', *International Journal of Remote Sensing*, vol. 37, no. 15, 3477-3493.

Biljecki, F, Stoter, J, Ledoux, H, Zlatanova, S & Çöltekin, A 2015, 'Applications of 3D city models: State of the art review', *ISPRS International Journal of Geo-Information*, 2015, vol. 4, pp. 2842-2889.

Breytenbach, A & Van Niekerk, A 2019, 'Analysing DEM errors over an urban region across various scales with different elevation sources', *South African Geographical Journal*, vol. 102, no. 2, 133-169.

Chudý, R, Iring, M & Feciskanin, R 2013, 'Evaluation of the data quality of digital elevation models in the context of INSPIRE', *GeoScience Engineering*, vol. 2, 9-24.

Demir, N, Poli, D & Baltsavias, E 2009, 'Combination of image and LiDAR data for building extraction' *Optical 3D Measurement Techniques*, Vienna, July 2009, pp. 1-10.

Deng, S, Katoh, M, Guan, Q, Yin, N & Li, M 2014, 'Estimating forest aboveground biomass by combining ALOS PALSAR and WorldView-2 data: A case study at Purple Mountain National Park, Nanjing, China', *Remote Sensing*, vol. 6, pp. 7878-7910.

Doytsher, Y, Dalyot, S & Katzil, Y 2009, 'Digital terrain models: a tool for establishing reliable and qualitative environmental control processes' in R De Amicis *et al.* (eds.), *Geospatial visual analytics: Geographical information processing and visual analytics for environmental security* (pp. 215-234), Springer-Verlag, Berlin.

Ginzler, C & Hobi, ML 2015, 'Countrywide stereo-image matching for updating digital surface models in the framework of the Swiss National Forest Inventory', *Remote Sensing*, vol. 7, no. 4, pp. 4343-4370.

Gröhmann, CH, Sawakuchi, AO & Mendes, VR 2011, 'Cell size influence on DEM volume calculation', in T Hengel, IS Evans, JP Wilson and M Gould (eds.), *Geomorphometry 2011*, Redlands, September 2011, pp. 63-66, International Society for Geomorphometry.

Gxumisa, A & Breytenbach, A 2017, 'Evaluating pixel vs. segmentation based classifiers with height differentiation on SPOT 6 imagery for urban land cover mapping', *South African Journal of Geomatics*, vol. 6, no. 3, pp. 436-448.

Hajnsek, I, Busche, T, Krieger, G, Zink, M, Schulze, D & Moreira, A 2014, 'Announcement of Opportunity: TanDEM-X Science Phase', Doc.: TD-PD-PL-0032, Issue 1.0, 19.05.2014, TanDEM-X Ground Segment, Microwaves and Radar Institute (DLR-HR), Germany, pp. 1-27.

He, Q, Zeng, C, Xie, P, Liu, Y & Zhang, M 2018, 'An assessment of forest biomass carbon storage and ecological compensation based on surface area: A case study of Hubei Province, China', *Ecological Indicators*, vol. 90, pp. 392-400.

Hengl, T & Evans, IS 2009, 'Mathematical and digital models of the land surface', in T Hengl and HI Reuter (eds.), *Developments in Soil Science 33* (pp. 31-63), Elsevier, Amsterdam.

Höhle, J & Höhle, M 2009, 'Accuracy assessment of digital elevation models by means of robust statistical methods', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, pp. 398-406.

Hsieh, Y, Chan, Y & Hu, JC 2016, 'Digital elevation model differencing and error estimation from multiple sources: A case study from the Meiyuan Shan Landslide in Taiwan', *Remote Sensing*, vol. 8, no. 199, pp 1-20.

Hutchinson, MF, Xu, T & Stein, JA 2011, 'Recent progress in the ANUDEM elevation gridding procedure', in T Hengel, IS Evans, JP Wilson and M Gould (eds.), *Geomorphometry 2011*, Redlands, September 2011, pp. 19-22, International Society for Geomorphometry, Wageningen, 2011.

Kramm, T & Hoffmeister, D 2019, 'A relief dependent evaluation of digital elevation models on different scales for northern Chile', *International Journal of Geo-Information*, vol. 8, no. 430, 1-25.

Krauss, T & Reinartz, P 2010, 'Enhancement of dense urban digital surface models from VHR optical satellite stereo data by pre-segmentation and object detection' *The International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences,* vol. XXXVIII, Part 1, pp. 1-6.

Kuffer, M & Sliuzas, R 2017, 'Global slum analysis achievement and challenges', Presentation, *Human Planet Forum*, Twente, September 2017, viewed 20 November 2019, <https://www.itc.nl/hpi-forum/forum-programme/documents/kuffer-sliuzas-lobalslumchallenges.pdf>.

Longbotham, N, Chaapel, C, Bleiler, L, Padwick, C, Emery, WJ & Pacifici, F 2012, 'Very high resolution multiangle urban classification analysis', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1155-1170.

Nelson, A, Reuter, HI & Gessler, P 2009, 'DEM production methods and sources', in T Hengl and HI Reuter (eds.), *Developments in Soil Science* (pp. 65-85), Elsevier, Amsterdam.

Peeroo U, Idrees MO & Saeidi V 2017, 'Building extraction for 3D city modelling using airborne laser scanning data and high-resolution aerial photo', *South African Journal of Geomatics*, vol. 6, no. 3, pp. 363-376.

Poli, D, Wolff, K & Gruen, A 2009, '3D geodata recovery from high resolution satellite imagery', *Remote Sensing of Environment*, Ispra, May 2009, vol. 1, Part 2, pp. 1081-1085.

Raciti, SM, Hutyra, LR & Newell, JD 2014, 'Mapping carbon storage in urban trees with multi-source remote sensing data: Relationships between biomass, land use, and demographics in Boston neighborhoods', *Science of the Total Environment*, vol. 500–501, pp 72-83.

Reuter, HI. Hengl, T, Gessler, P & Soille, P 2009, 'Preparation of DEMs for geomorphometric analysis', in T Hengl and HI Reuter (eds.), *Developments in Soil Science 33* (pp. 87-120), Elsevier, Amsterdam.

Santos, AP, Medeiros, NG, Dal Poz, AP, Santos, GR, Rodrigues, DD & Emiliano, PC 2020, 'Methodology for the extraction of homologous points from a DEM/DSM to evaluate the relative positional accuracy', *Bulletin of Geodetic Sciences*, vol. 26, no. 2, 1-16.

Sofia, G, Hillier, JK & Conway, SJ 2016, 'Frontiers in Geomorphometry and Earth Surface Dynamics: possibilities, limitations and perspectives', *Earth Surface Dynamics* vol. 4: 721-725.

Temme, AJAM, Heuvelink, GBM, Schoorl, JM & Claessens, L 2009, 'Geostatistical simulation and error propagation in geomorphometry' in T Hengl and HI Reuter (eds.), *Developments in Soil Science* 33 (pp. 121-140), Elsevier, Amsterdam.

Tinkham, WT, Smith, AMS, Hoffman, C, Hudak, AT, Falkowski, MJ, Swanson, ME & Gessler, PE 2012, 'Investigating the influence of LiDAR ground surface errors on the utility of derived forest inventories', Technical Report 182, pp. 1-12, USDA Forest Service, Lincoln, Lincoln Faculty Publications.

Urbazaev, M, Thiel, C, Migliavacca, M, Reichstein, M, Rodriguez-Veiga, P & Schmullius, C 2016, 'Improved multi-sensor satellite-based aboveground biomass estimation by selecting temporally stable forest inventory plots using NDVI time series', *Forests*, vol. 7, no. 169, pp. 1-16.

Voelkel, J & Shandas, V 2017, 'Towards systematic prediction of urban heat islands: grounding measurements, assessing modeling techniques', *Climate*, vol. 5, no. 41, pp. 1-17.

Wise, S 2011, 'Cross-validation as a means of investigating DEM interpolation error.' *Computers and Geosciences* vol. 37, no. 8, pp. 978–991.

Yoshida, K, Okuoka, K, Miatto, A, Schebek, L & Tanikawa H 2019, 'Estimation of mining and landfilling activities with associated overburden through satellite data: Germany 2000–2010', *Resources*, vol. 8,126, pp. 1-17.