Urban correction of global DEMs using building density for Nairobi, Kenya

Victor Olajubu 1 · Mark A. Trigg 1 · Christian Berretta 1 · Andrew Sleigh 1 · Marco Chini 2 · Patrick Matgen 2 · Stephen Mojere 3 · Joe Mulligan 4

Received: 25 September 2020 / Accepted: 9 June 2021 / Published online: 30 June 2021
© The Author(s) 2021

Abstract
Urban flood models that use Digital Elevation Models (DEMs) to simulate extent and depth of flood inundation rely on the accuracy of DEMs for predicting flood events. Despite recent advances in developing vegetation corrected DEMs, the effect of building height and density errors in global DEMs in urban areas are still poorly understood, and their correction remains a challenge. In this research we developed a methodology for building error correction that can be applied to any other case study, where building density data and a local reference DEM data are available. This methodology was applied to Nairobi, Kenya using six global DEMs (SRTM, MERIT, ALOS, NASADEM, TanDEM-X 12 m, and TanDEM-X 90 m DEM). Our results show building error at highest building density varying between 1.25 m and 5.07 m for the DEMs used, with the MERIT DEM showing the smallest vertical height error from the reference DEM. The six DEMs were corrected by deriving a linear relationship between building density and DEM error. Our findings show that the removal of building density error resulted in the improvement of the vertical height accuracy of the global DEMs of up to 45% for MERIT and 40% for ALOS. This methodology was also applied to the Central Business District (CBD) area of Nairobi, characterised by taller buildings and high building density. The error parameters in the CBD area resulted to be between 15 to 45% higher than those of the Nairobi city wide area for the six global DEMs, thus providing further insights into the contribution of building heights to errors in global DEMs. Building height data is still unavailable on a global scale and our results show that global DEMs can be usefully corrected for building density errors in urban areas, even where specific building height data are not available.

Keywords Global DEMs · Urban correction · Building density · Corrected DEMs. Urban flooding

Introduction
Rapid urban growth puts strain on existing infrastructure and discourages the preservation of natural habitat in favour of new housing developments, shopping malls, urban infrastructure, etc. that can exacerbate the problem of urban flooding. Flooding is the most prevalent natural disaster, often characterised as a high intensity event that requires rapid emergency service response in order to minimise substantial human and economic losses (Apel et al. 2009). Climate change and urbanisation have been reported as the major contributors to the increasing damaging effects of flooding to lives and livelihoods worldwide (Aerts et al. 2014). Topography has been identified as a key method of estimating flood extent (Horrrit and Bates 2001) and many models of flood extent rely on DEMs in order to simulate paths of water flow, flood extent and depth. Errors in DEMs (DEMs) can substantially affect the results of flood models (Stephens et al. 2012; Hawker et al. 2018).
Global DEMs used in flood models are representations of physical ground surface and the spatial resolution of a DEM refers to the area of land being represented by single regular or irregular grid, with the value of each grid element representing the height of the ground at the corresponding datum (Vaze et al. 2010). There are many open access global scale DEMs such as the Shuttle Radar Topography Mission (SRTM), and its derivatives, the Multi-Error-Removed Improved-Terrain DEM (MERIT DEM) and NASA DEM (NASADEM), as well as Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) DEM and TerraSAR-X add-on for Digital Elevation Measurement (TanDEM-X 90 m) etc. The global coverage of these DEMs makes them highly suitable for use in scientific applications where they are used extensively in flood models and have been critical in facilitating important flood studies, particularly in data-sparse areas, where local data is often difficult to access or unavailable (Hawker et al. 2018).

Chen and Hill (2007) investigated the influence of DEM resolution on flood hazard modelling in urban areas and found that both vertical height error and spatial resolution of DEMs can impact on flood inundation depth and extent in urban flood modelling. Although, spaceborne DEMs provide fundamental input to many geoscience studies, they suffer from non-negligible height errors (Yamazaki et al. 2017). Sources of error in spaceborne DEMs include: (i) incomplete spatial sampling; (ii) measurement errors, such as positional inaccuracy, data entry errors; and (iii) processing errors such as computational numerical errors, interpolation errors, and classification and generalisation errors (Burrough 1986). Global DEMs suffer from many different types of errors, some of which are significant at local scales; for example, (Rodriguez et al. 2006) reported a global mean and standard vertical height error of 8.2 ± 0.7 and 6.9 ± 0.5 m for SRTM X- and C-band data, respectively. There is a number of published work on the correction of errors in global DEMs, especially vegetation errors. (Falomi et al. 2005; bhang et al. 2007; Dong et al. 2015; Gallant et al. 2012; Baugh et al. 2013; O’Loughlin et al. 2016; Chen et al. 2018). Also, there are many previous studies focused on the assessment of the vertical height accuracy of DEMs by comparing elevation values of DEMs to that of a reference local DEM having a higher vertical accuracy. A more accurate reference DEM such as the Light Detection and Ranging (LiDAR) is required in order to make an assessment of the vertical accuracy of global DEMs (Dong et al. 2015; Wessel et al. 2018; Acharya et al. 2018).

Although many studies (Robinson et al. 2014; Yamazaki et al. 2012; Yamazaki et al. 2017) have developed new vegetation-corrected DEMs, by either editing or adjusting existing global DEMs. However, despite significant advances in developing vegetation-corrected DEMs, there is limited understanding of DEM errors that can be attributed to building heights and building density in urban areas. Local DEMs that are based on airborne light detection and ranging (LiDAR) are preferential over open access, global DEMs due of their superior vertical accuracy, horizontal resolution, and ability to distinguish between ‘bare earth’ from built structures and vegetation (Yamazaki et al. 2017). However, (LiDAR) DEMs (<10 m horizontal resolution) are only available for a very small percentage of Earth’s land surface (~0.005%), and data acquisition is often expensive (Hawker et al. 2018).

Building heights and building density inhibit the ability of radar signals to penetrate land surfaces, especially in densely populated urban areas where higher DEM resolution does not necessarily ensure accurate mapping (Rossi et al. 2012). Gridded elevation datasets, such as the radar-measurement-derived SRTM, exhibit signal reflection from built structures and vegetation so that further data processing may be required to enable accurate flood modelling (Sanders 2007). (Kim et al. 2020) selected the SRTM and Sentinel 2 multispectral imagery to train the artificial neutral network in order to improve the quality of SRTM DEM and then evaluated the performance of the resulting SRTM DEM over two dense urban cities. The ‘new’ DEM (iSRTM) showed better results than the original SRTM, achieving 38% reduction in the root mean square error (RMSE). Similarly, (Klonner et al. 2015) leveraged on the advantages of the Airborne Laser Scanning (ALS) and the Open Street Map (OSM) to create an up-to-date Digital Surface Model (DSM) combining 2D OSM and ALS data.

Digital surface models (DSMs) can provide a good source of high-quality data for the extraction of building height maps in urban areas and (Alganci et al. 2018) explored the feasibility of using open access DSMs, such as the ALOS (AW3D30), ASTER, and SRTM datasets, for extracting digital building height models and compared their accuracy. The potential for DSMs as a rich data source for the extraction of building height data has been highlighted as a significant challenge in their use at the same time as representations of DTM in urban flood modelling (Alganci et al. 2018). Despite efforts made in the processing of global DEM data prior to making the data publicly accessible, DEMs frequently contain artefacts such as spikes, holes and line errors. (Hirt 2018) recommended that all DEM datasets undergo a complete global screening for artefacts prior to public release. Further advising users to check quality before using global DEMs. Despite recent advances in removing error components from DEMs, such as tree height bias, speckle noise, stripe noise and absolute bias, much work remains in the urban correction of building biases in global DEMs.

According to (Hawker et al. 2018), there is no forthcoming high-accuracy open-access global DEM, therefore, for the
foreseeable future, the primary means of improving flood simulation will be to use editing or stochastic simulation using existing DEM data. Urban correction of existing global DEMs remains a key research challenge. In this context, the present paper develops a methodology for the urban correction of six global DEMs, tested using building density data from the city of Nairobi, Kenya. Although the scope of this study is currently limited to the use of building density data, however, we anticipate that once building height data becomes globally available, our methodology can be extended to urban correction of DEMs using building height data.

**Study site**

Nairobi is the capital and largest city of Kenya and chosen as the study area for this research due to its rapid urban expansion within the last two decades, Fig. 1. Nairobi has witnessed a population growth from 0.51 to 4,397,073 million people at a growth rate of 3 to 4% per year in past 50 years leading up to the 2019 national census (KNBS 2019). The lies within an administrative area of 696 km² (269 sq. mi), whilst the metropolitan area has a population of 9,354,580. The city lies on the River Athi in the southern part of the country and has an elevation of 1795 m (5,889 ft) above sea level (Nippon 2014). Approximately 2 million people that make up nearly half the population of Nairobi live in the informal settlement area (5%) of the city occupying meagre 1% of the total 696 km² land area (Amnesty-International 2019).

In their recent work (Henderson et al. 2016), developed a dynamic model of a growing city that shows the urban expansion of Nairobi, Kenya. The study highlighted the nature of the intensified land use within Nairobi and its increasing building heights, with a key distinction between formal and informal, or slum sectors. The study painted a picture of the built environment of Nairobi, both in the spatial cross-section and its evolution through time between 2003 and 2015. The built volume of the whole city increased at 3.9% p.a., expanding by 59% between 2003 and 2015. The growth and expansion within the central business district and formal sector redevelopment increased building volume by 35%. The expansion in the city was achieved by the demolition of over one third of buildings and redevelopments that saw three times increase in building heights. The study painted a picture of a monocentric city with tall but variable building height at the centre and then diminishing moving away from the centre of the city.

Fig. 1 Map of the study area showing Nairobi
Dataset description

Global DEMs derived from spaceborne and remote sensing data, and which are used in many global flood studies, are important data sources of ground surface height information (Hawker et al. 2018). DEMs are a type of raster, regular or irregular grids of spot heights that provide a three-dimensional (3D) model of the earth surface that can be categorised into two groups: (i) digital terrain models DTMs, which are free of trees, buildings, and all other types of object; and (ii) digital surface models DSMs (Fig. 2), which reflect the earth’s surface, including all man-made, natural objects and other features elevated above the ‘Bare Earth’ (Martha et al. 2010; Maune and Nayegandhi 2017).

DTMs are obtained by different methods such as the interpolation of contour lines that include not only heights and elevations, but also other geographical elements and natural features such as rivers, ridge lines, and so on (Moore et al. 1991). Whilst DSMs are mostly used for landscape modelling and applications for projection of cities in 3D, etc., DTMs have applications for global flood modelling, geoscience studies, drainage modelling, land use studies etc. (Rayburg et al. 2009; Alganci et al. 2018). In this study, we focused on six of the most widely used global DEMs as fundamental input for many geoscience studies: SRTM; MERIT; ALOS; NASADEM; TanDEM-X 12 m; and TanDEM-X 90 m. Fig. 3 provides an illustration of the visual comparison of the six global DEMs over the study area of Nairobi, Kenya whilst Table 1 shows a summary of the characteristics of the global DEMs used in this study.

Global DEMs

The Shuttle Radar Topography Mission (SRTM) was a joint endeavour of NASA, the National Geospatial-Intelligence Agency, and the German and Italian Space Agencies that flew in February 2000. It used dual radar antennas to acquire interferometric radar data, processed to digital topographic data at 1 arc sec resolution (Farr et al. 2007). Three official versions of SRTM have been released. Version 1.0 is the (almost) raw data obtained during the mission and its quality is considered research-grade. The Non-Void Filled version 2.1 is the data from Version 1 cleaned-up to correct processing errors and to clip data to water boundaries. This version still contains “void” areas for which there is no elevation data. These void areas are due to problems obtaining data using the radar methodology, such as in areas with steep terrain, and areas of low reflectivity such as flat deserts. The last official version of the SRTM (V3 or “SRTM Plus”) data with 01” resolution (~30 m at the Equator) removes all of the void areas by incorporating data from other sources such as the ASTER GDEM and was publicly released in 2014 (Kolecka and Kozak 2014).

The Multi Error removed Improved Terrain (MERIT) DEM was developed by removing multiple error components (absolute bias, stripe noise, speckle noise, and tree height bias) from the existing spaceborne DEMs (SRTM3 v2.1 and AW3D-30 m v1) using multiple satellite data sets and filtering techniques (Yamazaki et al. 2017). MERIT represents the terrain elevations at a 3 s resolution (~90 m at the equator), and covers land areas between 90 N-60S, referenced to EGM96 geoid. Following the removal of the various error components, land areas mapped with ±2 m or better vertical accuracy were increased from 39% to 58%. Significant improvements were found in flat regions where height errors larger than topography variability, and landscapes such as river networks and hill-valley structures, became clearly represented (Yamazaki et al. 2017).

SRTM produced an unprecedented near-global DEM of the world (Farr et al. 2007). Since its release, the SRTM DEM is widely used in many research studies, commercial, and military applications. The objective of the NASADEM project was to improve the SRTM DEM vertical height accuracy and data coverage. The improvements were achieved by reprocessing the original SRTM radar echoes and telemetry data with updated algorithms and auxiliary data not available at the time of the original SRTM production (Crippen et al. 2016; Vaka et al. 2019). One known issue of the SRTM DEM is the observed height ripples caused by uncompensated SRTM antenna boom motion. The NASADEM compensates for these elevation ripples based on a high-resolution correction of strip data in radar geometry (Crippen et al. 2016).
NASADEM data is available for download via https://lpdaac.usgs.gov.

The TanDEM-X (TerraSAR-X add-on for Digital Elevation Measurements) is a spaceborne radar interferometer that is based on two TerraSAR-X radar satellites flying in close formation since 2010 to map all land surfaces at least twice and difficult terrain mapped even up to four times. Krieger et al. (2007). The TanDEM-X Digital Elevation Model (DEM) is a DEM with a complete global coverage at 3 arc-seconds resolution (~90 m at the equator) of the earth’s surface (Zink et al. 2016). The TanDEM-X 90 m resolution DEM product is open and free for download from https://download.geoservice.dlr.de/TDM90/ but the 0.4 arc sec (12 m resolution) version is only free to the science community for educational purpose. An application was made to the German Aerospace Centre for the release of TanDEM-X 12 m DEM used in this study.

Accuracy assessment of the TanDEM-X 90 m DEM used in this study has been undertaken by different studies to determine their suitability for global flood modelling and found the TanDEM-X DEM has improved flood inundation predictive capacity when compared to other DEMs, but not MERIT (Wang et al. 2012; Yan et al. 2015; Mason et al. 2016). (Hawker et al. 2019), carried out error accuracy assessment of the TanDEM-X DEM on the freely available TanDEM-X 90 for selected floodplain sites in comparison to other popular global DEMs with results indicating that the average vertical accuracy of TanDEM-X 90 and MERIT are similar and are both a significant improvement on SRTM. Also, results suggested that TanDEM-X 90 is the most accurate global DEM in all land cover categories tested except short vegetation and tree-covered areas where MERIT is demonstrably more accurate.

The ALOS World 3D – AW3D30 (ALOS) global DEM data were produced using the data acquired by the Panchromatic Remote Sensing Instrument for Stereo Mapping (PRISM) operated on the ALOS from 2006 to 2011 (Takaku et al. 2016). The operator of the satellite is the

| DEM         | Resolution (m) | Vertical accuracy (m) | Reference                  |
|-------------|----------------|-----------------------|----------------------------|
| SRTM        | 90             | 6 m (MAE)a            | (Farr et al. 2007)         |
| MERIT       | 90             | 5 m (LE90)b           | (Yamazaki et al. 2017)     |
| ALOS        | 30             | 4.4 m                 | (Tadono et al. 2016)       |
| NASADEM     | 30             | < 6 m ((MAE) a)       | (Buckley et al. 2020)      |
| TanDEM-X    | 12 & 90        | <10 m                 | (Mason et al. 2016)        |

a Mean error
b 90th percentile linear error

**Fig. 3** Visual comparison of the six global DEMs, applied in Nairobi
Japan Aerospace Exploration Agency (JAXA) and the mission led to the production of the global ALOS DEM using approximately 3 million images. The free version of the DEM has a 1" resolution, which is equivalent to approximately 30 m at the Equator and model is downloadable in 1° × 1° tiles. The grid elevations (m) are referenced to the EGM96 geoid and the geographic coordinates are referenced to the GRS80 ellipsoid (Caglar et al. 2018). The dataset is downloadable from: www.eorc.jaxa.jp/ALOS/en/aw3d30.

**Facebook high resolution settlement layer (HRSL) data**

Facebook, in partnership with the Centre for International Earth Science Information Network (CIESEN) at Columbia University developed population grids dataset for 140 countries by using machine learning applied to high resolution satellite imagery (Tiecke et al. 2017). The high resolution settlement layer (HRSL) provides estimates of human population distribution at a resolution of 1 arc-second (approximately 30 m) using population estimates assigned to settlements delineated by machine learning algorithm in both urban and rural areas. Each 30 m grid has a population value assigned to an identified structure. For building density purposes, the data assumes grids with no population have no buildings and those with population have a building covering the whole grid. The Data is accessible via https://data.humdata.org/dataset/highresolutionpopulationdensitymaps

**Sentinel-1 SAR data derived global building map**

(Chini et al. 2018), introduced a technique for automatically mapping built-up areas using synthetic aperture radar (SAR) backscattering intensity and interferometric multi-temporal coherence generated from Sentinel-1 data. The data represents global building maps in 20 m resolution and derived from multi-temporal InSAR coherence, a systematic and consistent feature that allows for a better characterization of urban areas. The urban footprint data are on average in 92% agreement with the Global Urban Footprint (GUF) map derived from the TerraSAR-X mission data (Chini et al. 2018).

**Reference topography relief map data**

As a reference raster, we used data from an interpolated contour map of Nairobi, which has an estimated vertical error of ±2 m and is derived from aerial photogrammetry. The detailed contour map was produced in 2003 by the Japanese International Co-operation Agency (JICA) for the government of Kenya (Nippon 2014). In 2003, JICA performed an aerial triangulation that mapped the entire city of Nairobi (595 km²), excluding the Nairobi National Park (~107 m²); the mission required 15 aerial photography flight strips over Nairobi, including 20 GPS validation photo points. The standard deviation (SD) of the final coordinates of all the newly installed photo control points was within an acceptable limit (within 30 cm vertical height error). JICA released this data for use in the present study, providing the original topography contour map as a vector file which we converted to a raster elevation file using the TIN Interpolation plugin conversion tool in QGIS.

**Methodology**

**Procedure for urban correction of global DEMs**

We developed a five-step method for removing building density error from the six global DEMs namely: (i) NASADEM, (ii) SRTM, (iii) MERIT, (iv) ALOS, (v) TanDEM-X 12 m, and (vi) TanDEM-X 90 m. presents a flowchart of the datasets and methodology used in our five-step method, and our each step is also described in detail.

This five-step method is first applied to the whole of greater Nairobi area, followed by a separate application to just the Central Business District. The CBD area is the commercial hub of Nairobi and is composed of tall buildings, skyscrapers, government offices etc. and it should provide some insights into the effect of building heights on DEM errors (Fig. 4).

**Step 1: Pre-process raster data** To allow for consistent geospatial analysis, all six global DEMs data were transformed to the EGM96 Geoid, if not already, and resampled to a 90 m raster resolution. This was carried out for all six global DEMs and for the reference DEM. The resampling was carried out using the bilinear method using QGIS (v3.12) raster resampling tools. The SRTM, MERIT, ALOS, NASADEM elevation data are orthometric heights referenced to the EGM96 Geoid, whilst the TanDEM-X 12 & 90 m elevations are referenced to the WGS84 (G1150) ellipsoid. Therefore, in order to compare elevations, the TanDEM-X 12 m & TanDEM-X 90 m elevations were transformed to the EGM96 Geoid using the NOAA’s VDatum transformation tool, version 4.0.1 accessible via (https://vdatum.noaa.gov/). Using QGIS Triangulated Irregular Network (TIN) interpolation plugin tool, we created a DEM raster map of the study area from the original topography contour map of the study area to serve as the reference DEM, resampled to 90 × 90 m grids to match the horizontal resolution of the global DEMs.

**Step 2: Calculate global DEM error** Using the GIS raster algebra tool, we calculated the vertical accuracy for the six global DEMs by creating error maps as shown in Fig. 5. We produced the error rasters by subtracting elevations of the JICA reference DEM, which has a higher vertical accuracy, from the six global DEMs of the study area, (Eq. 1). We then...
calculate the root-mean square error (RMSE), mean error (ME), standard deviation (SD) and median (M) of each global DEM. We analysed the differences in the elevations of global DEMs by using error metrics, density distribution plots and the DEM error maps.

$$\text{Error of DEM} = Y_{GD} - Y_{ref} \quad \text{(1)}$$

$$\text{ME} = \frac{1}{n} \sum Y_{GD} - Y_{ref} \quad \text{(2)}$$

$$\text{RMSE} = \sqrt{\frac{\sum(Y_{GD} - Y_{ref})^2}{n}} \quad \text{(3)}$$

Fig. 4 Flowchart of the datasets and methodology used to correct building density errors in global DEMs in our five-step method. This can be applied to any spatial extent.

Fig. 5 Error map of global DEMs at 90 m resolution.
Fig. 6 Building density raster of Nairobi derived from (a) the Facebook HRSL population density map, and (b) global urban building map

$$SD = \sqrt{\frac{\sum_n^2 ((Y_{GD} - Y_{ref}) - ME)^2}{n}}$$  \hspace{1cm} (4)

where $Y$ is elevation in metres, $GD$ refers to the global DEM, $ME$ refers to the mean error and $ref$ refers to the reference JICA elevation.

**Step 3: Calculate building density** We calculate building density rasters by processing the Facebook high-resolution settlement layer data (HRSL) and the Sentinel-1 SAR urban footprint map developed by LIST to generate building density maps. We resampled the HRSL and LIST building maps to

Fig. 7 Map of the Central Business District (CBD), Nairobi
a coarser grid size and using GIS raster aggregation tools, we computed an aggregate over all of the input raster grids whose centres lie within the output grid of a coarser resolution (270 × 270 m) urban footprint map of the study area. We used QGIS tools (qgis/grass/r.resamp.stats) to resample the building maps to a coarser grid using aggregation to generate building density maps. An aggregate is computed over all of the input raster grids whose centres lie within the output cell. The aggregate uses the values from all input raster grid cells of 20 m resolution of the LIST building map which intersect the coarser resolution (270 m) output cell, weighted according to the proportion of the source cell which lies inside the output cell to generate building density maps for the study area. The aggregate uses the weighted values to create a building density raster with a building to land area fraction within the study area of between 0 and 1 Fig. 6.

A visual comparison of the Facebook HRSL and the LIST data with google earth image of the study area show that the LIST urban footprint map is of better agreement with building footprint of the study area. Therefore, we found the building density map derived from the LIST urban footprint map to be of higher accuracy in comparison to the output building density map derived from the HRSL population density map, Fig. 6. The higher resolution nature of the LIST data i.e. 20 m compared to the 30 m for the Facebook HRSL data is a plausible explanation for the differences in the results and accuracy of the two output maps. Consequently, we progressed this study based on the use of the building density map derived from the LIST urban footprint map for the study area.

### Table 2 Statistical error parameters for the global DEMs, in metres, before (and after) urban correction

| DEM          | SRTM       | MERIT      | NASADEM    | ALOS       | TanDEM-X 12 | TanDEM-X 90 |
|--------------|------------|------------|------------|------------|-------------|-------------|
| Mean         | −0.87 (−0.72) | 0.77 (0.35) | 1.99 (1.76) | −1.92 (−1.25) | 1.83 (1.68) | 1.87 (1.72) |
| RMSE         | 0.99 (0.85)  | 0.88 (0.48) | 1.41 (1.33) | 1.54 (0.92)  | 1.35 (1.22) | 1.37 (1.28) |
| SD           | 5.92 (5.32)  | 2.97 (2.53) | 3.46 (3.28) | 4.34 (3.53)  | 3.03 (2.89) | 3.29 (2.60) |
| Median       | −0.41 (−0.36) | 1.02 (0.70) | 1.45 (1.19) | −1.39 (−1.26) | 1.28 (0.98) | 1.42 (1.17) |

### Step 4: Determine DEM error relationship with building density
In order to calculate the building density error for each tile of the global DEMs, we established a relationship between DEM error and building density. Using the gdalqxyz plugin in QGIS tools, we exported raster values for the error maps and building density maps from the GIS platform and converted the exported data to csv format for further processing. We created plots of DEM error versus building density for all six global DEMs represented by a linear regression fit and $R^2$ values as illustrated in Fig. 9. As is common in other DEM correction studies, we have used a linear relationship due to the noisiness of the data (Baugh et al. 2013) and (OLoughlin et al. 2016). The resulting DEM error coefficient for each DEM represented increases in mean error measured in meters for every increase in building density and set between 0 and 1, with zero representing areas of no buildings at all and value of 1 for very dense areas respectively.

### Step 5: Apply error relationship to correct global DEM
The next step is to remove the fraction of vertical error component.
that is associated with building density. We calculated this for each DEM grid, grid-by-grid from the linear regression functions by using the DEM error coefficient, building density predictor, and a constant value, (Fig. 9). We repeated the procedure for all six global DEMs. Subsequently, using the raster calculator tool in QGIS, we created a building density error map for each of the global DEMs with an example for the SRTM DEM shown in Fig. 10a. The building density error

Fig. 9 Scatter plots of building density with DEM error, with superimposed linear regression lines of best fit for the tested global DEMs, applied to Nairobi: (a) ALOS; (b) SRTM; (c) MERIT; (d) NASADEM; (e) TanDEM-X 12 m and (f) TanDEM-X 90
maps are created using the linear regression function for each of the global DEMs and the building density map derived from the LIST urban building map. To create the new urban corrected DEM for all six global DEMs, we subtracted the building density error map from the original global DEM to arrive at the final product illustrated with the SRTM DEM in Fig. 10b. The new product is the urban corrected NASADEM, SRTM, MERIT, ALOS, TanDEM-X 12 m, TanDEM-X 90 m DEM for Nairobi, Kenya.

The Central Business District (CBD) area (Fig. 7) of Nairobi features many tall buildings, government offices, skyscrapers etc. and we wanted to understand if taller buildings will provide some further insights into the nature of the error. Therefore, we extended the analysis of the urban correction of the global DEMs to the CBD area by repeating the 5 steps described above for the CBD area.

Results and discussion

Distribution of vertical errors

The results show the MERIT DEM with the smallest vertical height deviation from the reference DEM, with an SD of 2.97 m, followed by TanDEM-X 12 and TanDEM-X 90, which had similar SDs of 3.03 m and 3.29 m, respectively. Figure 8 provides an illustration of comparison of density distribution plots for the six global DEMs. The error statistics for the six global DEMs are shown in Table 2. The SRTM, NASADEM & ALOS DEMs show a standard deviation of 5.92 m, 3.46 m, and 4.34 m respectively.

The results show that the MERIT and TanDEM-X 12 m & 90 m global DEMs have lower vertical height errors in comparison to the NASADEM, SRTM & ALOS DEMs if the SD metric only is considered. In addition, if the RMSE metric of the errors is considered alongside mean and median values, MERIT still provides lowest overall values and highest accuracy of all six global DEMs. The MERIT DEM is a multiple error-reduced improved version of SRTM (Chen et al. 2018) with tree height bias, stripe noise, absolute bias, and speckle noise removed from the original SRTM. MERIT is a corrected version of the SRTM, therefore, providing a plausible explanation for its higher accuracy.

(Hawker et al. 2019) investigated the vertical height accuracy of the TanDEM-X DEM 90 m, in comparison to other popular global DEMs by using high resolution (<10 m) LiDAR (Light Detection and Ranging) DEMs as a reference dataset. Their results show mean error values of 1.09 m, 1.30 m and 1.06 m for the MERIT, SRTM and TanDEM-X 90 DEMs respectively. The results correspond well with our own mean error magnitudes of 0.77 m, 0.87 m, and 1.72 m for the MERIT, SRTM and TanDEM-X 90 m DEMs.

The density distribution plot for all six global DEMs shown in Fig. 8 demonstrates that all six global DEMs have a unimodal distribution, except for SRTM which shows a weak bi-modal distribution. The kurtosis of the error distribution for all six global DEMs are generally positive for MERIT, NASADEM, ALOS, TanDEM-X 12 and TanDEM-X 90 m DEMs, but is less positive for the SRTM DEM by showing a less acute peak around the mean than the other DEMs. SRTM, MERIT, and NASADEM DEMs show a nearly symmetric error distribution with a near zero skewness whilst ALOS, TanDEM-X 12 and TanDEM-X 90 m DEMs all have positive skewness and show more extreme positive outliers than negative ones.

The while the SRTM DEM shows a relatively low positive mean error (+0.87 m), this is only as a result of averaging cancelling out a large positive and negative spread of errors, evidenced by the highest Standard Deviation (5.32 m) of all the DEMs. The MERIT and NASADEM are derivatives of the SRTM DEM, with improvements made to reduce errors and in MERIT’s case also remove vegetation bias. Unsurprisingly therefore these have better error characteristics than SRTM. However, neither MERIT nor NASADEM have been corrected for urban bias and this perhaps explains some of the remaining positive error bias (overprediction of elevation), which is larger for the NASADEM (+1.99 m) compared to the MERIT (+0.77 m). The higher mean error of NASADEM is likely explained by the lack of vegetation correction compared to MERIT. TanDEM-X, another radar instrument derived dataset also suffers from a similar positive error bias (for 12 and 90 m, 1.83 and 1.87 m respectively) presumably also for the lack of vegetation and urban error correction.

There are only minor differences between the 12 m and 90 m TanDEM-X DEMs in our analysis, unsurprising as the

| Table 3 | Statistical error parameters for SRTM & TanDEM-X 90 m DEMs in Nairobi city wide and the CBD area, in metres, before (and after) urban correction |
|---------|-------------------------------------------------------------------------------------------------|
| **DEM** | **SRTM (Nairobi)** | **SRTM (CBD)** | **TanDEM-X 90 (Nairobi)** | **TanDEM-X 90 (CBD)** |
| Mean | −0.87 (−0.72) | −1.25 (−1.04) | 1.87 (1.72) | 2.49 (2.01) |
| RMSE | 0.99 (0.85) | 1.39 (1.02) | 1.37 (1.28) | 2.18 (1.85) |
| SD | 5.92 (5.32) | 6.84 (6.24) | 3.29 (2.60) | 4.84 (4.04) |
| Median | −0.41 (−0.36) | −0.88 (−0.49) | 1.42 (1.17) | 1.94 (1.52) |
90 m DEM is derived from the 12 m DEM in the first place. However, we might expect if we were resampling from 12 m to 90 m this may reduce random noise error due to the averaging process. This does not seem to be the case here, indicating that the positive bias is indeed related to a systematic bias, likely vegetation and urban artefacts (and possibly other errors). The most unusual error characteristics are observed in the ALOS DEM.

The error statistics for our urban corrected global DEMs are also shown in Table 2 along with the original DEM error statistics to allow direct comparison.

### DEM error and building density relationships

We found that there is a linear and positive, but noisy relationship between DEM error and building density Fig. 9. All the DEMs show a noisy relationship; with SRTM having the noisiest Fig. 9(b)) and TanDEM-X DEMs the least noisy (Fig. 9(e) & (f)). At zero building density the DEM error is not necessarily zero. For each DEM, the highest error in all DEMs is found at the highest building density of 1 as shown in Table 4.

The CBD features many of Nairobi’s important buildings, government offices, headquarters of business and corporations – both national & international, skyscrapers etc. The scatter plots of building density of the CBD area against DEM error for the six global DEMs is shown in the analysis of the urban correction of the global DEM for the Central Business District area (Figure 7) of Nairobi consisting of taller buildings provided some further insights into the nature of the errors. Similar to the results of the analysis undertaken at a city scale for Nairobi, we found that there is a linear and positive, but noisy relationship between DEM error and building density. All the DEMs show a noisy relationships and the statistical error parameters for the six global DEMs both before (and after) the urban correction for the CBD area is shown in Table 2.

The relationships for both CBD area and Nairobi appear to be weak when the values of the $R^2$ are considered. However, the very sensitive nature of the impacts of vertical height accuracy on DEMs means that these results are real and can be significant. We noticed a higher error for the CBD area across all error metrics of ME, RMSE, and SD for all six global DEMs. For example, Table 3 shows a comparison of the error parameters for the SRTM and the TanDEM-X 90 m DEMs for Nairobi city wide and for the CBD area both before and after the urban correction. The error parameters in the CBD area is between 15 to 45% higher than those of the Nairobi city wide area for the DEMs. The very tall nature of the buildings in the CBD area appears to have contributed to the percentage increase in the errors. The focus of this study is on building density error and the results obtained for the CBD area show building heights can be an important contributor to DEM errors in urban areas and is worthy of further study.

| DEM        | Regression equation | Correction error (m) at highest BD | Correction error (m) at lowest BD |
|------------|---------------------|------------------------------------|----------------------------------|
| SRTM       | 4.73x+0.34          | 5.07                               | 0.34                             |
| MERIT      | 0.73x+0.52          | 1.25                               | 0.54                             |
| NASADEM    | 2.21x+2.26          | 4.47                               | 2.26                             |
| ALOS       | 0.53x+1.74          | 2.27                               | 1.74                             |
| TanDEM-X 12| 1.49x+1.56          | 3.05                               | 1.56                             |
| TanDEM-X 90| 1.53x+1.81          | 3.34                               | 1.81                             |

Fig. 10 (a) Building density error raster for SRTM DEM, (b) urban-corrected SRTM DEM
We corrected the DEMs by applying a correction based on the linear error relationship fitted to each DEM (Table 4). For example, to correct the SRTM DEM we calculated the vertical error for the given building density of a corresponding DEM grid using the regression equation in Table 4 to create a building density error raster Fig. 10a. Subsequently, the building density error map is subtracted from the original DEM to create the urban corrected DEM in Fig. 10b.

It should be noted that there will be the requirement for a local reference DEM data of vertical height accuracy higher

![Scatter plots of building density with DEM error](image)

**Table 5** Statistical error parameters for the global DEMs in the CBD area, in metres, before (and after) urban correction

| DEM        | SRTM  | MERIT  | NASADEM | ALOS  | TanDEM-X 12 | TanDEM-X 90 |
|------------|-------|--------|----------|-------|--------------|--------------|
| Mean       | −1.25 (−1.04) | 1.05 (0.68) | 2.47 (1.96) | −2.32 (−1.46) | 2.15 (1.87) | 2.49 (2.01) |
| RMSE       | 1.39 (1.02)  | 1.28 (0.88) | 2.06 (1.73) | 2.19 (1.43)  | 2.09 (1.72) | 2.18 (1.85) |
| SD         | 6.84 (6.24)  | 3.02 (2.87) | 3.98 (3.60) | 4.95 (3.99)  | 4.36 (3.73) | 4.84 (4.04) |
| Median     | −0.88 (−0.49) | 1.33 (0.91) | 1.98 (1.64) | −1.74 (−1.51) | 1.61 (1.15) | 1.94 (1.52) |
than the global DEMs before the methodology described in this paper can be adapted for similar urban case study areas. Accuracy assessment of the global DEMs will involve the subtraction of elevation values belonging to each grid cell of the reference DEM from the corresponding cells of the global DEMs. Also, the six global DEMs and the reference DEM datasets used in the study were acquired over different periods spanning decades and could be a possible factor influencing the higher accuracy of the most recent DEMs (Fig. 11).

Building error DEM artefacts in urban areas have two major components: building density and building height. Ideally, both of these components should be removed from DEM data; however, building height data is unavailable on a global scale. Therefore, this paper only addresses errors due to building density biases. Our results show that global DEMs can be usefully corrected for building density errors in urban areas, even where specific building height data are not available (Table 5).

Conclusions

Open-access global DEMs are not only useful tools for estimating flood risks, but they also provide baseline data for flood studies. Despite significant advances in developing vegetation-corrected DEMs, there is limited understanding of DEM errors that can be attributed to building heights and building density in urban areas. Current global DEMs are not corrected for building errors. Because building height data is unavailable on a global scale, this paper addresses errors due to building density biases.

In this research we developed a methodology for building error correction that can be applied to any other case study, where building density data and a local reference DEM data of vertical height accuracy higher than the global DEMs are available. In this study, we have quantified the building error for the city of Nairobi, Kenya for six of the most widely used global DEMs: SRTM; MERIT; ALOS; NADAM; TanDEM-X 12 m; and TanDEM-X 90 m. Our results show building error at highest building density varying between 1.25 m and 5.07 m for the DEMs used. The results show the MERIT DEM with the smallest vertical height deviation from the reference DEM, with an SD of 2.97 m, followed by TanDEM-X 12 and TanDEM-X 90 (3.03 m and 3.29 m respectively). In addition, if the RMSE metric of the errors is considered alongside mean and median values, MERIT still provides the lowest overall values and highest accuracy. A plausible explanation for its higher accuracy is that the MERIT DEM is a multiple error-reduced improved version of SRTM with tree height bias, stripe noise, absolute bias, and speckle noise removed.

By deriving a relationship between DEM error and building density we were able to correct the evaluated building error. We found that there is a linear and positive, but noisy relationship between DEM error and building density. All the DEMs show a noisy relationship; with SRTM having the noisiest and TanDEM-X 12 m & 90 m DEMs the least noisy. Our findings show that the removal of building density error from global DEMs resulted in the improvement of the vertical height accuracy of the global DEMs of up to 45% for MERIT and 40% for ALOS. Thus, our results show that global DEMs can be usefully corrected for building density errors in urban areas, even where specific building height data are not available.

In this paper we also show the results of the presented methodology for the Central Business District (CBD) area of Nairobi which is characterized by taller buildings and high building density. Results show the error parameters in the CBD area is between 15 to 45% higher than those of the Nairobi city wide area for the six global DEMs. These results provided some further insights into significance of building heights contributing to errors in global DEMs. Therefore, future work is required to understand the nature of building height errors in global DEMs and how these errors can be corrected.

Acknowledgements The field work for this study was supported by the Climate research bursary fund, the Priestly International Centre for Climate, University of Leeds, UK. The authors would like to offer their special thanks to the Japanese International Cooperation Agency (JICA) for sharing local GIS data used in this study, especially the topography contour map Nairobi. Also, special thanks to Marco Chini and Patrick Matgen for the contribution of their urban building work funded by the National Research Fund of Luxembourg (FNR) through the MOSQUITO (grant number: C15/SR/10380137) and the CASCADE (grant number: C17/SR/11682050) projects. Finally, the authors would also like to thank Stanley Kimani of the Kenya National Survey Authority, Nairobi for their contribution.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Acharya T, Yang I, Lee D (2018) Comparative analysis of digital elevation models between AW3D30, SRTM30 and airborne LiDAR: A case of chuncheon, South Korea. J Korean Soc Surveying Geodesy Photogrammetry Cartography 36:17–24
Apel H, Aronica GT, Kreibich H, Thieken AH (2009) Flood risk
Baugh CA, Bates PD, Schumann G, Trigg MA (2013) SRTM vegetation
Amnesty-International (2019) Kenya: the unseen majority: Nairobi’s two
million slum-dwellers. Amnesty International Publications: London, 3
Ape H, Aronica GT, Kreibich H, Thieken AH (2009) Flood risk
analyses—how detailed do we need to be? Nat Hazards 49:79–98
Baugh CA, Bates PD, Schumann G, Trigg MA (2013) SRTM vegetation
removal and hydrodynamic modeling accuracy. Water Resour Res 49:5276–5289
Bhang JK, Schwartz FW, Braun A (2007) Verification of the vertical error
in C-band SRTM DEM using ICESat and Landsat-7, otter
Tail County, MN. IEEE Trans Geosci Remote Sens 45:36–44
Buckley, S. M., Agram, P.S., J. E. Belz, R. E. Crippen, E. M. Gurrola,
Hensley, M. K., M. Lavalle, J. M. Martin, M. Neumann, Q. D.
Nguyen, & P. A. Rosen, J. G. S., M. Simard, W. W. Tung 2020.
Nasadem Userguide Version 1
Burrough PA (1986) Principles of geographical information systems for
land resources assessment. Clarendon, Oxford
Caglar b, Becak K, Mekik C, Ozendli M (2018) On the vertical accuracy
of the ALOS world 3D-30m digital elevation model. Remote Sens
Lett 9:607–619
Chen H, Liang Q, Liu Y, Xie S (2018) Hydraulic correction method
(HCM) to enhance the efficiency of SRTM DEM in flood modeling.
J Hydrol 559:56–70
Chen J, Hill A (2007) Modeling urban flood hazard: just how much does
dem resolution matter? Appl Geogr 30:8
Chini M, Pelich R, Hostache R, Matgen P, Lopez-Martinez C (2018)
Towards a 20 m global building map from Sentinel-1 SAR data.
Remote Sens 10:1833
Crippen R, Buckley S, Agrarn P, Belz E, Gurrola E, Hensley S, Kobrick
M, Lavalle M, Martin J, Neumann M, Nguyen Q, Rosen P, Shimada
J, Simard M, Tung W (2016) Nasadem global elevation model:
methods and progress. ISPRS - International Archives of the
Photogrammetry, Remote Sensing and Spatial Information
Sciences, XLI-B4, 125–128
Dong Y, Chang H-C, Chen W, Zhang K, Feng R (2015) Accuracy as-
alysis of the ALOS world 3D-30m digital elevation model. Remote Sens
Environ 173:15–28
Maune DF, Nayegandhi A (2017) Digital elevation model technologies
and applications: the DEM users manual. 3rd edition, 655
Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: A
review of hydrological, geomorphological, and biological applications.
Hydrof Process 5:3–30
Nippon K (2014) The project on integrated urban development master
plan for the City of Nairobi in the Republic of Kenya. Final Report,
Part 1, 692
O’Loughlin FE, Paiva RCD, Durand M, Alsford DE, Bates PD (2016)
A multi-sensor approach towards a global vegetation corrected SRTM
DEM product. Remote Sens Environ 182:49–59
Rayburg S, Thorns M, Neave M (2009) A comparison of digital elevation
modeling methods from different data sources. Geomorphology 106:
261–279
Robinson N, Regetz J, Guralnick RP (2014) EarthEnv-DEM90: A nearly-
global, void-free, multi-scale smoothed, 90m digital elevation model
from fused ASTER and SRTM data. ISPRS J Photogramm Remote
Sens 87:57–67
Rodriguez E, Morris CS, Belz JE (2006) A global assessment of the
SRTM performance. Photogramm Eng Remote Sens 72:249–260
Rossi C, Fritz T, Eineder M, Erten E, Zhu X, Germann M (2012)
Towards AN urban dem generation with satellite SAR interferome-
try. ISPRS - International Archives of the Photogrammetry, Remote
Sensing and Spatial Information Sciences, 39B7, 73
Sanders BF (2007) Evaluation of on-line DEMs for flood inundation
modeling. Adv Water Resour 30(8):1831–1843
Stephens EM, Bates PD, Freer JE, Mason DC (2012) The impact of
uncertainty in satellite data on the assessment of flood inundation
models. J Hydrol 414–415:162–173
Tadono T, Nagai H, Ishida H, Oda F, Naito S, Minakawa K, Iwamoto H
(2016) generation of the 30 m-mesh global digital surface model by
Alos Prism. ISPRS - International Archives of the Photogrammetry, Remote
Sensing and Spatial Information Sciences, XLI-B4, 157–162
Takaku J, Nakano K, Ichikawa M (2016) Validation of the
"AW3D" global DSM generated from Alos Prisma. ISPRS Annals
of Photogrammetry, Remote Sensing and Spatial Information Sciences, III, 25
Tiecke, T., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E. & Dang, H.-A. 2017. Mapping the world population one building at a time
Vaka DS, Kumar V, Rao YS, DeSo R (2019) Comparison of Various DEMs for Height Accuracy Assessment Over Different Terrains of India. IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, 28 July-2 Aug. 2019. 1998-2001
Vaze J, Teng J, Spencer G (2010) Impact of DEM accuracy and resolution on topographic indices. Environ Model Softw 25:1086–1098
Wang W, Yang X, Yao T (2012) Evaluation of ASTER GDEM and SRTM and their suitability in hydraulic modelling of a glacial lake outburst flood in Southeast Tibet. Hydrol Process 26:213–225
Wessel B, Huber M, Wohlfart C, Marschalk U, Kosmann D, Roth A (2018) Accuracy assessment of the global TanDEM-X digital elevation model with GPS data. ISPRS J Photogramm Remote Sens 139:171–182
Yamazaki D, Baugh CA, Bates PD, Kanae S, Alsdorf DE, Oki T (2012) Adjustment of a spaceborne DEM for use in floodplain hydrodynamic modeling. J Hydrol 436:437:81–91
Yamazaki D, Ikeshima D, Tawatari R, Yamaguchi T, O’loughlin F, Neal JC, Sampson CC, Kanae S, Bates PD (2017) A high-accuracy map of global terrain elevations. Geophys Res Lett 44:5844–5853
Yan K, Di Baldassarre G, Solomatine DP, Schumann GJ-P (2015) A review of low-cost space-borne data for flood modelling: topography, flood extent and water level. Hydrol Process 29:3368–3387
Zink M, Moreira A, Bachmann M, Bräutigam B, Fritz T, Hajnsek I, Krieger G, Wessel B (2016) TanDEM-X mission status: The complete new topography of the Earth. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 10–15 July 2016. 317–320

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.