SU956

Accuracy Assessment of DEMs in Different Topographic Complexity Based on an Optimum Number of GCP Formulation and Error Propagation Analysis

Saeed Nadi[1], Davood Shojaei[2], and Yusof Ghiasi[3]

[1] Assistant Professor, Dept. of Geomatics Engineering, Faculty of Civil Engineering and Transportation, Univ. of Isfahan, Isfahan, Iran (corresponding author). Snadi@eng.ui.ac.ir
[2] Centre for SDIs and Land Administration, Dept. of Infrastructure Engineering, Univ. of Melbourne, Victoria, 3010, Australia, shojaeid@unimelb.edu.au
[3] Dept. of Geography and Environmental Management, Faculty of Environment, Univ. of Waterloo, Ontario, Canada, syghiasi@uwterloo.ca

Abstract

One of the main concerns during DEM evaluation is the number of ground control points (GCP). Accordingly, in this paper, a new method is proposed for calculating the appropriate number of GCPs for DEM evaluation based on a Confidence Interval (CI) of RMSE. Then, the method is employed to determine the CI of the estimated vertical accuracy of AW3D30, SRTM and ASTER GDEM Free 30m resolution global DEMs in mountainous, hilly, flat and urban regions of two study areas. To provide a more reliable estimation of errors, robust statistical methods including Median, Normalized Median Absolute Deviation (NMAD) and Huber's μ and σ are also investigated. Furthermore, a new formulation is developed to analyse propagation of the errors in slope and aspect products of DEM. The results showed that to evaluate the accuracy of AW3D30, ASTER GDEM and SRTM with a CI of ±1m and the probability of 99%, in the study area, a minimum number of 2110, 1483 and 750 GCPs are required, respectively. The results also showed that in the flat, hilly and mountainous study areas AW3D30 is the most accurate DEM. However, SRTM fits better to the urban study area. Finally, the results of the error propagation analysis illustrates that the slope and aspect errors bear a striking relation to the surface gradient.

Keywords: Optimum Number of GCP, DEM, Accuracy Assessment, Robust Statistics, Error Propagation, Slope Error, Aspect Error.

Introduction

Since 1950, Digital Elevation Models (DEMs) are used as one of the main information resources in various spatial analyses (El-Sheimy et al. 2005). These models and their products such as slope and aspect maps are being widely used in various engineering disciplines, which require knowledge about the accuracy of these models. Characteristics of primary data including their vertical and horizontal accuracy, distribution pattern and density as well as properties of interpolation methods affect the accuracy of the produced model. Furthermore, filtering methods and algorithms for matching and geo-referencing in photogrammetry, laser scan and LiDAR can have significant effects in regions with different topographic characteristics (Liu et al. 2015; Muir et al. 2017). Accordingly, the accuracy of these DEMs must be evaluated regarding the topographic complexities (Bolkas et al. 2016; Shan et al. 2003).

Various global DEMs are freely available on the web. AW3D30, SRTM and ASTER GDEM are three popular examples of global DEMs (Yap et al. 2018). AW3D30 was first published in 2016 with global coverage by the Japan Aerospace Exploratory Agency (JAXA). AW3D30 is a product of Advance Land Observation Satellite (ALOS) which was launched on Jan 24, 2006, and was operational till 2011. ASTER GDEM is one of the products of ASTER sensor. The ministry of Economy, Trade and Technology of Japan (METI) in collaboration with the National Aeronautical Space Administration (NASA) mounted this sensor on the board of Terra Satellite in 2000 and produced the first version of ASTER GDEM and made it available to the public in 2009. NASA in collaboration with the National Geospatial Agency (NGA) produced SRTM through an eleven day mission of endeavor shuttle in 2000. Public availability of these models made them

very popular in various analyses in different disciplines such as transportation, hydrology and environmental science. These models are widely being used in many decision making processes where the availability of knowledge about the accuracy of them in regions with different topographic characteristics can affect the reliability of the decisions (Chappell et al. 2006; Dragut and Eisank 2011; Frey and Paul 2012; Ramlal and Baban 2008; Romanowicz et al. 2008; Zhu and Jekeli 2009).

Numerous researchers investigated the accuracy of DEMs from different aspects such as the effects of DEM accuracy on hydrological modeling (Czubski 2013), simplification of accuracy assessment of DEMs from end users' point of view (Darnell et al. 2008), The effect of low laying coastal plains (Du et al. 2016), mountainous terrain (Eckert et al. 2005; Mispan et al. 2015) and salt marsh area (Hladik and Alber 2012) on accuracy of DEM. Some researchers performed the accuracy assessment of GDEMs globally (Rodriguez et al. 2006) while some others investigated their accuracy in a local area (Rexer and Hirt 2014; Zhao et al. 2011). However, the optimum number of Ground Control Points (GCPs) has been always a concern for evaluating the accuracy of spatial data such as georeferenced images (Nguyen 2015; Oniga et al. 2018), and DEMs (Höhle and Höhle 2009). Accordingly, in this paper an innovative method is proposed to determine the required number of GCPs for assessing the accuracy of DEMs. This method is based on defining a confidence interval (CI) for Root Mean Square Error (RMSE). The method is employed to assess the accuracy of AW3D30, ASTER GDEM and SRTM free 30m global DEMs and determine the confidence interval for their estimated RMSE. Furthermore, the error distribution of these DEMs have been investigated to determine the appropriateness of using common statistical measures of error such as Mean Error (ME), Standard Deviation (SD) and RMSE as well as robust measures such as Median, Normalized Median Absolute Deviation (NMAD) and Huber's μ and σ for

accuracy assessment of these DEMs. Then, the accuracy of these models was compared in two case study areas in the center and west parts of Iran with different topographic characteristics using both common and robust statistical models. This paper provides a concrete guideline to assess these DEMs in mountainous, hilly, flat and urban regions. Finally, a new formulation is also developed to describe the propagation of errors on the slope and aspect products of these DEMs.

The reminder of the paper is structured as follows. In the following section, the previous studies are explored. Subsequent sections deal with a brief introduction about free global DEMs, the details of the proposed methods, accuracy assessment of DEMs and discussion of the results. Finally, the paper concludes with a summary and directions for future research.

Literature Review

DEMs are being widely used in various applications to analyze and visualize earth related phenomenon (Wolock and Price 1994). For example, Cai and Wang (2006) and Chappell et al. (2006) used DEM to study the behavior of water flow over the earth surface. Ramlal and Baban (2008) utilized DEM to simulate and manage fluid materials. Romanowicz et al. (2008) used DEM for planning roads and modeling their safety. In addition, one of the very important applications of DEM is the geoid determination which is considered by a number of researchers (Kiamehr and Sjöberg 2005; Merry 2003; Zhu and Jekeli 2009).

It is very important to control the accuracy of datasets including DEM in analyses. This helps interpreting the reliability of the results. The distribution of primary data of DEMs and their random, systematic and gross errors can significantly affect the reliability of the results of analyses (Elkhrachy 2017; Mukherjee et al. 2013). There are officially reported global validation of free global DEMs. For example, the RMSE of 5m is reported for AW3D30 (Tadono et al. 2014; Tadono et al. 2016; Takaku and Tadono 2017). For ASTER GDEM the RMSE is reported to be variable

in between 10m to 25m (Tachikawa et al. 2011) and it is also reported that RMSE of SRTM is less than 16m (Rodriguez et al. 2006; Rodriguez et al. 2005). Despite of the global validation of DEMs a number of researchers have evaluated these DEMs in local areas as well. Using control points to calculate RMSE is one of the most popular methods for investigating the accuracy of DEMs. Yap et al. (2018) used Global Navigation Satellite System (GNSS)/leveling ground control points to evaluate the horizontal and vertical accuracy of ASTER GDEM, SRTM and AW3D30 and reported the RMSE of 16.7m, 20.4m and 13.2m for AW3D30, ASTER GDEM and SRTM, respectively, in Cameroon. They also analyzed the effect of land cover and slope in DEM vertical accuracy. Eckert et al. (2005) used the same method to evaluate the accuracy of ASTER GDEM in a mountainous area of Switzerland. Patel et al. (2016) also used GNSS control points to evaluate different open source DEMs. The accuracy of ASTER GDEM and SRTM was also evaluated in India (Mukherjee et al. 2013). They reported RMS error of 12.62m and 17.76m for ASTER and SRTM DEM, respectively. There are also some other relevant studies which have confirmed that SRTM has higher accuracy than ASTER GDEM (Du et al. 2016; Zhao et al. 2011). Some studies investigated the accuracy of DEMs in different topography. Hu et al. (2017) have evaluated ASTER DEM, SRTM and AW3D30 over China and showed that the accuracy of all three DEMs in hilly study area is better than 11.7m. Rexer and Hirt (2014) also reported the RMSE of ASTER GDEM and SRTM as 9.4m and 6.8m, respectively, in the hilly area. In mountainous area they observed that the RMSE of ASTER GDEM and SRTM are 11.9m and 9.8m, respectively. Hladik and Alber (2012) also used GNSS control points to evaluate the accuracy of salt streaks map extracted from LiDAR data. Another error assessment method is to compare the reconstructed contour lines from both control points and DEM (Zhao et al. 2011). Constructing slope from control points and DEM and comparing them is another evaluation method which some researchers

have considered (Du et al. 2016; Mispan et al. 2015; Szabó et al. 2015). Furthermore, recently some researchers have used robust statistical models instead of using RMSE which provides more reliable results (Höhle and Höhle 2009; Wang et al. 2015). However, there is a lack of discussion about the optimum number of GCPs in DEM accuracy assessment, the confidence level of the estimated accuracies and the propagation of errors for DEM products which are addressed in this paper.

Free Global Digital Elevation Models

Various free global DEMs are introduced to the public with different accuracy, resolution and coverage. SRTM is one of the first global DEMs which was firstly released with spatial resolution of 3 arc-seconds. In 2014, its 1 arc-second global digital elevation model (~30 m) was released. Most parts of the world have been covered by this dataset ranging from 54°S to 60°N latitude except for the Middle East and North Africa, which was completed in August 2015. Another global model with almost 99% coverage of the globe is named ASTER GDEM. It was introduced to the public in 2009, which covers ±83 degree latitude. Spatial resolution of this model is 1 arc-second. Its third version is produced taking advantage of additional ASTER scenes (350,000) and further improvements in water body delineation (Gesch et al. 2016). AW3D30 is the most recent free global DEM with 1 arc-second spatial resolution. This model was first released in 2015 for Japan and its latest version (V2.1) was released in April 2018 with global coverage. ALOS had a Panchromatic Remote sensing Instrument for Stereo Mapping (PRISM) which was an optical sensor designed to generate worldwide elevation data. To utilize the global data, PRISM records triplet stereoscopic images in 2.5 meter ground resolution during the five year mission life of the sensor from 2006 to 2011. AW3D30 is a Digital Surface Model (DSM) resampled from the 5-meter mesh version of the "World 3D Topographic Data" which was generated utilizing the

original triplet stereo images of PRISM.

Moreover, there are a number of commercial DEMs. For example, World DEM is one of the sub-products of the TerraSAR-X satellite with 6-meter horizontal accuracy and 4-meter absolute and 2-meter relative vertical accuracy (Becek et al. 2016; Wessel et al. 2018). Despite of the better accuracy of the commercial models, due to their costs, they are not as popular as free DEMs.

This paper analyzes the vertical accuracy of the most recent versions of AW3D30, ASTER GDEM and SRTM as the most popular DEMs. The specifications of the versions of these models that are used in this paper are shown in table 1.

Methodology

Errors come from different sources in DEM production including data and processes. Some methods are available to reduce the effects of these errors. However, it is necessary to have a quantitative assessment of the remaining portion of errors in order to evaluate the suitability of a DEM for an application (Glennie 2018). The assessment is usually based on measuring some statistics for height difference between DEM and GCPs. As illustrated in Fig. 1, the proposed method in this paper follows 3 steps. In the first step, a new method based on the confidence interval is proposed to determine the optimum number of required GCPs to obtain a reliable estimation of RMSE. Then, different measures to evaluate the accuracy of DEMs are introduced. In the second step, the propagation of DEM errors to slope and aspect maps are formulated. Finally, in the third step, the proposed methods is employed to calculate the accuracy of AW3D30, ASTER GDEM and SRTM and the effects of these errors on slope and aspect maps are evaluated.

Required number of GCPs

The required number of GCPs, their spatial distribution and their accuracy are the most important

issues in selecting GCPs for accuracy assessment of DEMs (Höhle and Höhle 2009). General guideline for the number of GCPs is that they must be large enough and uniformly distributed all over the test area to obtain reliable accuracy measures (Nguyen 2015; Oniga et al. 2018). In this paper, we rely on this general guideline about the necessity of uniformly distributed GCPs and proposed and justified equation 1 to calculate required number of GCPs using a defined confidence interval for RMSE. This equation depends on the number of uniformly distributed GCPs and can be used to determine the required number of GCPs to reach an appropriate confidence interval. The proposition 1 gives confidence interval for RMSE of height difference between the GCPs and DEMs.

Proposition 1. Given a set of n uniformly distributed GCPs over a DEM to calculate its RMSE, the following probability is valid.

$$P\left(\frac{(n-1)\left(\overline{RMSE}^2-\overline{\Delta h}^2\right)}{\chi^2_{\left(n-2,1-\frac{\alpha}{2}\right)}}+\overline{\Delta h}^2<RMSE^2<\frac{(n-1)\left(\overline{RMSE}^2-\overline{\Delta h}^2\right)}{\chi^2_{\left(n-2,\frac{\alpha}{2}\right)}}+\overline{\Delta h}^2\right) \quad (1)$$
$$=1-\boldsymbol{\alpha}$$

where, $\Delta h = h_{reference\ point} - h_{DEM}$ and $\chi^2_{\left(n-2,1-\frac{\alpha}{2}\right)}$ is chi-square distribution with *n-2* degree of freedom for confidence level of α**.**

Proof. The ratio of sample and population variance has chi square distribution as illustrated in equation 2 (Montgomery and Runger 2010).

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \quad (2)$$

where, S and σ are the variances of sample and the whole population, respectively, and n is the degree of freedom. Knowing the actual heights for each cell of a DEM with *N* cell, equation 3 provides the population variance.

$$\sigma^2 = \frac{\sum_{i=1}^{N}(\Delta h_i - \mu_{\Delta h})^2}{N} = \sum_{i=1}^{N}\frac{{\Delta h_i}^2}{N} + \sum_{i=1}^{N}\frac{{\mu_{\Delta h}}^2}{N} - \sum_{i=1}^{N}\frac{2\Delta h_i \mu_{\Delta h}}{N} = RMSE^2 - {\mu_{\Delta h}}^2 \qquad (3)$$

where, $\mu_{\Delta h}$ and $RMSE$ are actual mean and root mean square error of height difference between actual height and DEM height of all cells. Similarly, for a sample of n GCPs estimated variance can be calculated using equation 4.

$$S^2 = \overline{RMSE}^2 - \overline{\Delta h}^2 \qquad (4)$$

where, $\overline{\Delta h}$ and $\overline{RMSE}$ are mean and root mean square error for height difference between GCPs and DEM. Substituting equation 3 and 4 in 2 results in equation 5.

$$\frac{(n-1)\left(\overline{RMSE}^2 - \overline{\Delta h}^2\right)}{\left(RMSE^2 - {\mu_{\Delta h}}^2\right)} \sim {\chi^2}_{(n-1)} \qquad (5)$$

In equation 5, actual mean of height differences ($\mu_{\Delta h}$) are unknown. By substituting it with its estimated value ($\overline{\Delta h}$) and decreasing the degree of freedom by one, equation 6 is obtained.

$$\frac{(n-1)\left(\overline{RMSE}^2 - \overline{\Delta h}^2\right)}{\left(RMSE^2 - \overline{\Delta h}^2\right)} \sim {\chi^2}_{(n-2)} \qquad (6)$$

Equation 7 shows the probability of this statistics.

$$P\left({\chi^2}_{\left(n-2,\frac{\alpha}{2}\right)} < \frac{(n-1)\left(\overline{RMSE}^2 - \overline{\Delta h}^2\right)}{\left(RMSE^2 - \overline{\Delta h}^2\right)} < {\chi^2}_{\left(n-2,1-\frac{\alpha}{2}\right)}\right) = 1 - \boldsymbol{\alpha} \qquad (7)$$

Equation 1 can be obtained directly from equation 7.

The confidence interval in equation 1 suggests that if the $\overline{RMSE}$ of a DEM is estimated using n uniformly distributed GCPs and repeated using different sample data sets, the actual value of RMSE of the DEM will be in this interval with the probability of $(1 - \boldsymbol{\alpha})$. Considering $\boldsymbol{\alpha}$ as a known significance level and n as an unknown parameter, this equation can be used to calculate the required number of control points (n) to reach the desired $\boldsymbol{\alpha}$. Furthermore, a rough estimation of $\overline{RMSE}$ and $\overline{\Delta h}$ is required to calculate n, which can be determined using a small sample of

uniformly distributed control points. These estimates represents the topographical complexity of the region. It can be seen from equation 1 that the higher values of $\overline{\Delta h}$ indicates the more topographical complexity of the region and the higher values of $\overline{RMSE}$ indicated that the region are more error-prone and therefore, more control points are required to reach a desired confidence interval. This means that in the regions with more topographical complexity, higher amount of control point must be used to have a reliable estimate of the calculated RMSE. The appropriate number of GCPs can then be calculated by using a computer program that check different values of n to reach this desired confidence level.

Accuracy measures

Most common statistical measures of accuracy such as RMSE are based on the assumption of normality. However, because of the non-normal distribution of errors in spatial data such as DEM, it is required to use robust statistical measures (Zandbergen 2008). Robust statistics is a convenient way of summarising data when they include a small proportion of outliers (Wilcox 2012). Most estimates of central tendency and dispersion such as the arithmetic mean and standard deviation depend on an implicit assumption that the data comprises a random sample from a normal distribution. However, analytical data often departs from that model. The robust measures are less sensitive to outliers in data and do not require the normality assumption (Amiri-Simkooei 2003).

Accuracy measures and their properties can formally be described using defined norms in Hilbert space. The *Lp* norm for each *k* dimensional vector *x* in Hilbert space of R^k is defined using equation 8 (Rudin 1964).

$$\|x\|_p = (\sum_{n=1}^{k} |x_n|^p)^{1/p} \qquad (8)$$

Let $p \geq 1$ be a real number, this equation provides a generic definition of different norms. For $p=1$,

it is the taxicab norm, for *p=2,* it is the Euclidean norm, and as p approaches ∞ the *p-norm* approaches the infinity norm or the maximum norm.

In Euclidean norm, the central tendency is an arithmetic mean. Equation 9 is the mean error (ME) of DEM in 2-norm.

$$\overline{\Delta h} = 1/n \sum_{i=1}^{n} \Delta h_i \quad (9)$$

where, Δh is the height difference between GCPs and their corresponding DEM cells. One of the other measures of errors in this norm is variance which shows the dispersion of the errors (Equation 10).

$$s^2 = \frac{\sum_{i=1}^{n} \left(\Delta h_i - \overline{\Delta h}\right)^2}{n} \quad (10)$$

where, $\overline{\Delta h}$ is ME obtained using equation 9. Furthermore, equation 11 is used to calculate RMSE which is the most common measure of errors in this norm.

$$RMSE = \frac{1}{n} \| \Delta h \|_2 \quad (11)$$

If the errors (Δh) follow a normal distribution, RMSE is the best measure showing the magnitude of errors. However, the existence of outliers deviates error's distribution from normal and makes the use of these measures a challenging issue. In the other word, the measures in L_2 norm are very sensitive to outliers (Amiri-Simkooei 2003).

The L_1-norm of a vector is defined as the sum of the absolute values of its components. Median is the central tendency in this norm and is found when the values of Δh are sorted from smallest to largest and then the value in the middle is selected. Median is one of the robust measures which is less sensitive to errors.

In this norm, Normalized Median Absolute Deviation (NMAD) is one of the robust measures for determining the dispersion of errors and can be calculated using equation 12 (Wilcox 2012).

$$NMAD = \frac{Median(|\Delta h_i - M|)}{0.6745} \qquad (12)$$

where, M is the median of height differences. Equation 13 is another robust measure for estimating dispersion of errors (Wilcox 2012).

$$\sigma_m = \frac{1}{2\sqrt{n} f(x_{0.5})} \qquad (13)$$

where n is the number of GCPs and $f(x_{0.5})$ can be calculated using equation 14.

$$f(x_{0.5}) = \frac{(A-B)\sqrt[5]{n}}{2.4n(q_{0.75} - q_{0.25})} \qquad (14)$$

where, *A* and *B* are the number of values less than or equal to *(M+h)* and *(M-h),* respectively, and $q_{0.75}$ and $q_{0.25}$ are the maximum values of the first and third quantiles.

Huber's method makes more use of the information provided by the data. This method progressively transforms the original data by a process called winsorization. In this process, the initial estimates of central tendency $(\hat{\mu}_0)$ and dispersion $(\hat{\sigma}_0)$ (median and NMAD in L_1 norm or mean and standard deviation in L_2 norm) are used to convert Δh_i using equation 15 (Huber 2011).

$$\widetilde{\Delta h}_i = \begin{cases} \hat{\mu}_0 + 1.5\hat{\sigma}_0 & \Delta h_i > \hat{\mu}_0 + 1.5\hat{\sigma}_0 \\ \hat{\mu}_0 - 1.5\hat{\sigma}_0 & \Delta h_i < \hat{\mu}_0 - 1.5\hat{\sigma}_0 \\ \Delta h_i & otherwise \end{cases} \qquad (15)$$

The new values of $\widetilde{\Delta h}_i$ will be used to estimate the new values of $\hat{\mu}_0$ and $\hat{\sigma}_0$. This procedure will be iterated by using the current improved estimates for the winsorization at each cycle. Eventually, the process converges to an acceptable degree of accuracy, and the resulting values are the robust estimates.

Error propagation for slope and aspect products of DEM

In this section, the propagation of DEM errors in its slope and aspect products is investigated.

The propagation of errors into the slope map is the effect of its variables. Equation 16 is used to calculate slope for any pixel *(i,j)* (DeMers 2002).

$$Slope = tan^{-1}\left(\sqrt{s_x^2 + s_y^2}\right) \qquad (16)$$

where, s_x and s_y are directional slopes calculated using equation 17.

$$\begin{aligned} s_x &= \frac{\left(z_{(i+1,j+1)} + 2z_{(i,j+1)} + z_{(i-1,j+1)}\right)}{8d} - \frac{\left(z_{(i+1,j-1)} + 2z_{(i,j-1)} + z_{(i-1,j-1)}\right)}{8d} \\ s_y &= \frac{\left(z_{(i+1,j-1)} + 2z_{(i+1,j)} + z_{(i+1,j+1)}\right)}{8d} - \frac{\left(z_{(i-1,j-1)} + 2z_{(i-1,j)} + z_{(i-1,j+1)}\right)}{8d} \end{aligned} \qquad (17)$$

where, d is spatial resolution of DEM and $z_{(i,j)}$ is the height of pixel *(i,j)*.

Proposition 2. The error of extracted slope from DEM can be calculated using equation 18.

$$\delta_{slope}^2 = \frac{3}{16d^2\left(1 + s_x^2 + s_y^2\right)^2}\delta_z^2 = \frac{3}{16d^2(1 + [\tan(slope)]^2)^2}\delta_z^2 \qquad (18)$$

where, δ_z^2 is the error of the corresponding DEM.

Proof. Using equation 16, the variance of slope is illustrated as equation 19.

$$\delta_{slope}^2 = \left(\frac{\partial slope}{\partial s_x}\right)^2 \delta_{s_x}^2 + \left(\frac{\partial slope}{\partial s_y}\right)^2 \delta_{s_y}^2 \qquad (19)$$

where, $\delta_{s_x}^2$ and $\delta_{s_y}^2$ are the variance of slope in *x* and *y* directions. Using equation 17, $\delta_{s_x}^2$ (variance of s_x) and $\delta_{s_y}^2$ (variance of s_y) can be calculated using equation 20 and 21.

$$\delta_{s_x}^2 = \left(\frac{\partial s_x}{\partial z_{(i+1,j+1)}}\right)^2 \delta_{z_{(i+1,j+1)}}^2 + \left(\frac{\partial s_x}{\partial z_{(i,j+1)}}\right)^2 \delta_{z_{(i,j+1)}}^2 + \left(\frac{\partial s_x}{\partial z_{(i-1,j+1)}}\right)^2 \delta_{z_{(i-1,j+1)}}^2$$
$$+ \left(\frac{\partial s_x}{\partial z_{(i+1,j-1)}}\right)^2 \delta_{z_{(i+1,j-1)}}^2 + \left(\frac{\partial s_x}{\partial z_{(i,j-1)}}\right)^2 \delta_{z_{(i,j-1)}}^2 \quad (20)$$
$$+ \left(\frac{\partial s_x}{\partial z_{(i-1,j-1)}}\right)^2 \delta_{z_{(i-1,j-1)}}^2$$

where, $\delta_{z_{(i,j)}}^2$ is vertical error of cell *(i,j)*. Suppose that vertical error of all cells is similar and substitute it with the error of DEM (δ_z^2), $\delta_{s_x}^2$ and $\delta_{s_y}^2$ can be calculated using equation 21.

$$\delta_{s_x}^2 = \delta_{s_y}^2 = \frac{3}{16d^2}\delta_z^2 \quad (21)$$

By substituting equation 21 and respective derivatives in equation 19, equation 18 is obtained. Equation 18 shows that slope error of each cell depends directly on DEM's error and has an inverse relation with the resolution and slope of the cell.

Aspect of each cell is calculated using equation 22 (DeMers 2002).

$$Aspect = \tan^{-1}\left|\frac{s_x}{s_y}\right| \quad (22)$$

where, s_x and s_y are directional slope calculated using equations 17.

Proposition 3. The error of extracted aspect map from DEM can be calculated using equation 23.

$$\delta_{aspect}^2 = \frac{3}{16d^2(s_x^2 + s_y^2)}\delta_z^2 = \frac{3}{16d^2[\tan(\text{slope})]^2}\delta_z^2 \quad (23)$$

where, δ_z^2 is the error of the corresponding DEM.

Proof. Using equation 22, the variance of aspect is calculated in equation 24.

$$\delta_{aspect}^2 = \left(\frac{\partial aspect}{\partial s_x}\right)^2 \delta_{s_x}^2 + \left(\frac{\partial aspect}{\partial s_y}\right)^2 \delta_{s_y}^2 \quad (24)$$

Substituting $\delta_{s_x}^2$ and $\delta_{s_y}^2$ from equations 21, equation 23 is obtained.

Equation 18 shows that aspect error of each cell depends directly on DEM's error and has an inverse relation with the resolution and slope of the cell.

AW3D30, ASTER GDEM and SRTM error evaluation

In this section, the study area is described and the confidence interval for RMSE estimation is calculated regarding the available number of GCPs. Then, the systematic error of the DEMs is investigated by comparing the trend of the elevation values extracted from DEMs and the elevation values of GCPs in all study regions. Furthermore, the histogram and the normal probability plot are used to check the normality of distribution of the DEMs' errors. If the errors do not follow a normal distribution, they may include outliers and robust statistical approaches are more appropriate to indicate the accuracy of DEMs. Finally, using the described methods in this paper, the accuracy of the DEMs are compared and the error maps of slope and aspect are produced.

Study area

In order to investigate the effect of urban land use and topography on DEM accuracy, two datasets were chosen from two different provinces in Iran. Table 2, shows the main specifications of these datasets. The first study area is located in Isfahan city, in the central part of Iran (Fig. 2a). The topography of Isfahan makes it suitable for testing the accuracy of DEM in urban areas. Another study area is located in Ilam province in the west part of Iran. Due to its topographic complexity, it is highly suitable for investigating the effect of different topographic characteristics on DEMs (Fig. 2b). In this study, three sample regions were selected from Ilam study area representing three typical topographic characteristics. The classification is done based on the slope of regions as shown in table 2 (Li et al. 2004).

In the first study area, 421 GNSS points were received from the Isfahan Municipality with the reported vertical and horizontal accuracy of ±20mm and ±5mm, respectively

(Isfahan municipality ICT Organization 2014). For the flat, hilly and mountainous regions, in the second study area, 59573 GNSS points were collected from the Iranian Ministry of Petroleum with the vertical and horizontal accuracy of ±15mm and ±7mm, respectively (National Iranian Oil Company 2016). GNSS points have ellipsoidal heights with respect to WGS84, and the DEMs have heights with reference to the EGM96 geoid. In order to compare with DEMs heights, the GNSS ellipsoidal heights were transferred from WGS84 to EGM96 datum using the EGM96 Geopotential Model which was jointly developed by National Imagery and Mapping Agency (NIMA) and National Aeronautics and Space Administration (NASA) (http://cddis.gsfc.nasa.gov/926/egm96/egm96.html). The Geopotential Model provides correction coefficient which are used in this study to transfer WGS84 ellipsoidal heights of GCPs to EGM96 surface. Table 2 shows the statistical characteristics of GCPs in each area. These statistics indicate the topographical complexity of each study area. For example in mountainous study area, the difference between minimum and maximum heights are around 1100 meters and the standard deviation of heights are around 300 meters which indicate a very complex mountainous area.

In order to investigate the sufficiency of the number of GCPs, equation 7 is used. Using the primary sample of 50 GCPs, $\overline{RMSE}^2$ and $\overline{\Delta h}^2$ for AW3D30, ASTER GDEM, and SRTM were calculated as 6.81m and 3.46m, 8.81m and 2.52m, and 9.67m and 2.61m, respectively. Table 3 shows the minimum number of required GCPs for AW3D30, ASTER GDEM and SRTM based on the different values of $\boldsymbol{\alpha}$.

As it is illustrated in this table for the available number of GCPs in this paper for flat, hilly and mountainous area (Table 2), the estimated RMSE will be in the confidence interval of ±1m with the probability of 99%, and for urban area it will be in the confidence interval of ±3m with the probability of 99%.

It must be cited that $\overline{\Delta h}^2$ can seriously affect the number of GCPs. As a result, it is recommended that its estimation should be based on the primary GCPs which are spread all over the study area.

Trend of elevations and error distribution

In order to investigate the correspondence of trends of each model with the natural ground, a scatter diagram is used for each DEM. Furthermore, the histogram of differences between the GCP values and each DEM values were produced to investigate the normality of the errors. Fig. 3 shows the linear trend of AW3D30, ASTER GDEM and SRTM with respect to corresponding GCP values and also the histogram of the differences between GCP values and AW3D30, ASTER GDEM and SRTM values in the mountainous study area.

As Fig. 3 shows, the trend of the models is similar to the GCPs. However, AW3D30 has less fluctuations compared to SRTM and ASTER GDEM. The Pearson product-moment correlation coefficients of 0.9999, 0.9995 and 0.9996 for AW3D30, ASTER GDEM and SRTM, respectively, also show that there is a very strong positive linear correlation between the elevation values of DEMs and the natural ground. The histogram shows that distribution of errors in SRTM and ASTER GDEM models are fairly similar but AW3D30 has less standard deviation. Moreover, it shows that there is a slight positive skewness in SRTM errors and a slight negative skewness in AW3D30 and ASTER GDEM errors. Therefore, it is reasonable to conclude that there are slight normal distribution violations in errors of these models. Fig. 4 illustrates the graphs for the hilly study area.

As it can be seen in this graph, the models have the same trend as the GCPs. The Pearson product-moment correlation coefficient values of 0.9990, 0.9950 and 0.9971 for AW3D30, ASTER GDEM and SRTM, respectively, show that AW3D30 values have the strongest correlation with GCP

values. The histogram also shows that unlike ASTER GDEM, the errors of AW3D30 and SRTM have less deviations from the central tendency. There is also a minor negative skewness in AW3D30 and ASTER GDEM errors and a small positive skewness in SRTM errors which slightly violate the normal distribution. Fig. 5 shows the graphs for the flat area.

The graph shows that all the models have similar trends as GCPs and Pearson product-moment correlation coefficients of 0.9936, 0.9610 and 0.9819 for AW3D30, ASTER GDEM and SRTM, respectively, show that AW3D30 has stronger correlation with GCPs than SRTM and ASTER GDEM. The histogram also shows that unlike ASTER GDEM, the errors of AW3D30 and SRTM show less dispersion. This means like hilly study area, the occurrence of big errors is less frequent in AW3D30 and SRTM in the flat study area. Fig. 6 shows the graphs for the urban area.

The linear trend in Fig. 6 shows the same trend for AW3D30, SRTM and ASTER GDEM with respect to the GCPs. However, the scatter diagram and Pearson product-moment correlation coefficients of 0.9952, 0.9805 and 0.9996 for AW3D30, ASTER GDEM and SRTM, respectively, show that unlike the other study areas, in urban areas, SRTM shows more correlation with GCPs than AW3D30 and GDEM ASTER. In these areas, ASTER GDEM histogram of errors shows more dispersion than AW3D30 and SRTM. AW3D30 and ASTER GDEM also show a slight positive skewness and SRTM shows a slight negative skewness which reflects some violation from the normal distribution.

As shown, the models in all areas follow a similar trend as the GCPs. Therefore, it is reasonable to conclude that there is no any major systematic error in the models. However, the slight skewness in the histograms of both models can be interpreted as the probability of existence of some outliers which slightly violate the normal distribution.

Statistical evaluation of errors and discussion

As discussed in section 4.2, while errors follow a normal distribution, conventional measures based on the L_2 norm provide an acceptable estimation for them. However, the existence of outliers violates the normal distribution. In this situation, robust statistical approaches must be used. In figures 3 to 6, a slight skewness is seen for the models in almost every area. Therefore, in this section, robust statistical measures in L_1 norm such as NMAD and Huber measures as well as conventional measures in L_2 norm such as RMSE were calculated. Table 4 shows the measures for the models in every four regions.

The results showed that, the overall accuracy of DEMs in all the study areas, in terms of RMSE of AW3D30, ASTRER GDEM and SRTM are less than 4.9m, 9.8m and 8.9m, respectively, which are slightly better than the results of officially reported global validation of these DEMs as 5m for AW3D30 (Tadono et al. 2014; Tadono et al. 2016; Takaku and Tadono 2017), 10m to 25m for ASTER GDEM (Tachikawa et al. 2011) and less than 16m for SRTM (Rodriguez et al. 2006; Rodriguez et al. 2005). For the whole study area, AW3D30 has shown a higher accuracy compared to SRTM and ASTER GDEM, and also SRTM has found to be better than ASTER GDEM. Jain et al. (2017) have reported the same results for lower Tapi Basin India study area, and also Yap et al. (2018) has demonstrated similar results in Cameroon study area. Some relevant studies also have confirmed that SRTM has higher accuracy than ASTER GDEM (Du et al. 2016; Zhao et al. 2011). However, Mukherjee et al. (2013) has demonstrated that ASTER GDEM is more accurate than SRTM.

In the urban study area, the RMSE of AW3D30, ASTER GDEM and SRTM are 4.2m, 5.9m and 3.1m, respectively, which indicates that SRTM has less RMSE than AW3D30 and ASTER GDEM; and also AW3D30 has less RMSE than ASTER GDEM. The same results have been reported by Jain et al. (2017) and Yap et al. (2018) which confirm SRTM to be more accurate

than ASTER GDEM and AW3D30. This means that in the urban study areas, unlike other study areas, the accuracy of the AW3D30 is less than the accuracy of SRTM. This might be due to a reason that while the GCPs are on the ground level, these DEMs are all Digital Surface Models (DSM) and the averaged version of AW3D30 is derived from resampling of 5-meter mesh version which can be highly influenced by urban construction near the control points. The negative ME of the DEMs (-3.3m for AW3D30, -2.6m for ASTER GDEM and -2.1m for SRTM) indicates that there are negative systematic shifts and these DEMs under-predict elevations in the urban study area. The same results for ASTER GDEM and SRTM were also observed by Mukherjee et al. (2013). Yap et al. (2018) also reported the same results for ASTER GDEM.

For the flat study area, considering RMSE, the accuracy of AW3D30, ASTER GDEM and SRTM are 3.7m, 8.4m and 5.3m, respectively, which indicates that AW3D30 has highest accuracy, and SRTM has better accuracy than ASTER GDEM. The same results were also observed by previous studies (Jain et al. 2017; Rexer and Hirt 2014; Varga and Bašić 2015; Yap et al. 2018). The positive ME of 2.3m, 3.8m and 1.7m for AW3D30, ASTER GDEM and SRTM, respectively, shows that there is a positive bias in the DEMs which indicates that these DEMs over-predict elevations in the flat study area. The same observations have been done by previous studies (Hu et al. 2017; Jain et al. 2017; Sun et al. 2003).

In the hilly study area, the RMSEs of AW3D30, ASTER GDEM and SRTM are 3.6m, 9.7m and 6.4m, respectively, which indicate that AW3D30 has the highest accuracy, and SRTM has also higher accuracy than ASTER GDEM. Hu et al. (2017) have evaluated these three DEMs over China and showed that the accuracy of all three DEMs in hilly study area is better than 11.7m. Rexer and Hirt (2014) also reported the same results for RMSE of ASTER GDEM and SRTM as 9.4m and 6.8m, respectively, in the hilly area. The positive ME values of 1.1m, 5.4m and 2.2m for

AW3D30, ASTER GDEM and SRTM, respectively, indicated that there is a positive bias in these DEMs in the hilly study area. However, the bias of ASTER GDEM is much more than the biases of AW3D30 and SRTM. These positive MEs indicate that all three DEMs over-predict elevations in the hilly study area. Rexer and Hirt (2014) have reported a negative ME of -3.6m for ASTER GDEM and a positive ME value of 3.5m for SRTM in the hilly area. Hu et al. (2017) believed that the different ME signs in different studies can be related to differences in terrain morphology or land cover of the study areas.

In the mountainous study area, the RMSE of AW3D30, ASTER GDEM and SRTM are 4.9m, 9.8m and 8.9m, respectively, which indicate that AW3D30 is better than ASTER GDEM and SRTM; and SRTM is slightly better than ASTER GDEM. Rexer and Hirt (2014) have confirmed this result by observing RMSE of 11.9m and 9.8m for ASTER GDEM and SRTM, respectively. Jain et al. (2017) also confirmed this result. They reported the RMSE of 2.5m, 3.6m and 2.64m for AW3D30, ASTER GDEM and SRTM, respectively. Yap et al. (2018) have also reported the RMSE of 16.7m, 20.4m and 13.2m for AW3D30, ASTER GDEM and SRTM, respectively, in Cameroon, which represent superiority of SRTM over the AW3D30 and ASTER GDEM. The positive ME values of 1.9m and 3.7m for AW3D30 and SRTM and the negative value of -0.6m for ASTER GDEM confirm the existence of biases in all three DEMs. This indicates that AW3D30 and SRTM over-predict elevations while ASTER GDEM under-predict elevations in the mountainous study area. Yap et al. (2018) and Rexer and Hirt (2014) have confirmed similar results for ME in the mountainous study area but Jain et al. (2017) have reported a positive ME for all three DEMs in mountainous study area .

The results showed that unlike the urban study area, in the mountainous, hilly and flat study areas, AW3D30 has better accuracy than SRTM; and SRTM has better accuracy than ASTER GDEM. The results also confirmed that SRTM is the best one in the urban study area.

The results also showed that considering RMSE, the accuracy of AW3D30 in the hilly and flat study areas (3.6m and 3.7m, respectively) is more than in the mountainous and urban study area (4.9m and 4.2m, respectively). But, the accuracy of SRTM and ASTER GDEM in urban study areas (3.1m and 5.9m, respectively) is more than in the flat (5.3m and 8.4m, respectively), hilly (6.4m and 9.7m, respectively) and mountainous (8.9m and 9.8m, respectively) study areas. Furthermore, the results of comparing different DEMs using conventional and robust measures showed similar results which confirms that the violation of errors from a normal distribution in the study areas is not significant.

Table 5 shows a 99% confidence interval for RMSE of each model in every region which is calculated using the equation 1.

The above-mentioned confidence intervals for each region mean that in an infinite number of independent experiments, the true value of RMSE is expected to fall within the interval for 99% of the times. Therefore, the length of each interval can be a measure of accuracy of the performed experiment. As it is shown in table 5 the length of confidence interval for estimated RMSE of AW3D30, SRTM and ASTER GDEM in the urban study area are 0.4m, 0.6m and 0.5m, in flat area 0.2m, 0.3m and 0.7m, in hilly area 0.1m, 0.1m and 0.7m and in mountainous area 0.3m, 0.4m and 0.7m, respectively. These values are all below one meter which shows the high accuracy of the estimated RMSE.

Error propagation to slope and aspect maps

Using equations 18 and 23 and substituting corresponding δ_z^2, from the table 4, for each dataset in

different study area, the maximum and minimum propagated error to slope and aspect maps are calculated and illustrated in table 6.

Table 6 shows that, the minimum and maximum slope errors for AW3D30 are 0.01×10^{-2} and 0.49×10^{-2} radian, respectively. These errors are 0.01×10^{-2} and 0.83×10^{-2} radian for SRTM and 0.02×10^{-2} and 2.22×10^{-2} radian for ASTER GDEM. The minimum and maximum values of propagated errors for aspect maps are 0.09×10^{-2} and 3.62×10^{-2} radian for AW3D30, 0.05×10^{-2} and 8.04×10^{-2} radian for SRTM and 0.23×10^{-2} and 6.60×10^{-2} radian for ASTER GDEM, respectively. It can be seen from equations 18 and 22 that the propagated errors for slope and aspect maps are directly related to the error of DEM and inversely related to the slope. Although the slope is higher in mountainous area, the more DEM error in mountainous areas results the maximum slope and aspect errors.

Conclusion

The increasing use of free global digital elevation models in various analysis and decisions has made it very necessary to control their error and its impact on output products derived from these models. In this regard, the number of ground control points is always one of the main concerns during DEM accuracy evaluation. By this, we mean that, what is the required number of GCPs to reach a desired confidence interval for RMSE estimation of DEMs? Accordingly, in this paper, a new method based on a confidence interval of RMSE is proposed for calculating the appropriate number of GCPs for DEM accuracy evaluation. Using the proposed method, the errors of AW3D30, ASTER GDEM and SRTM models in urban, flat, hilly and mountainous regions were evaluated and the confidence level of the estimated RMSE are calculated based on the available number of GCPs which determines the accuracy and completeness of the performed tests. Then, the effect of topographic changes on the error of these models was investigated. For this purpose,

different statistical measures in the L_1 and L_2 norms and also Huber robust statistical measures were used. The measures used in L_1 norm and also Huber method are robust and are less susceptible to outliers. Common measures in L_2 norm, along with robust measure make it possible to compare and accurately analyze the errors in these models. The results showed that in flat, hilly and mountainous study areas the accuracy of AW3D30 is better than SRTM and ASTER GDEM with the probability of 99% and CI of ±1m, but in the urban study area, SRTM has better accuracy than AW3D30 and ASTER GDEM with the probability of 99% and CI of ±3m. Another innovation in this paper is the investigation of error propagation for slope and aspect maps. The results of the error propagation analysis showed that the error in the aspect and slope is directly related to the slope changes. Therefore, the maximum error is observed in areas with a higher slope.

Data Availability Statement

- Some or all data, models, or code used during the study were provided by a third party (3D coordinates of GCPs). Direct requests for these materials may be made to the provider as indicated in the Acknowledgments.
- Some or all data, models, or code generated or used during the study are available from the corresponding author by request (the corresponding elevation of GCPs from SRTM, AW3D30 and ASTER GDEM).

Acknowledgement

The authors would like to thank the Municipality of Isfahan and the Exploration Management of the National Iranian Oil Company for providing the GCPs in urban area as well as the altimetry points in mountainous, hilly and the flat study areas in the west of Iran.

References

Amiri-Simkooei, A. (2003). "Formulation of l1 norm minimization in gauss-markov models." *J. Surv. Eng.-ASCE*, 129(1), 37-43.

Becek, K., Koppe, W., and Kutoğlu, Ş. H. (2016). "Evaluation of vertical accuracy of the worlddem™ using the runway method." *Remote Sens. Environ.*, 8(11), 934.
Bolkas, D., Fotopoulos, G., Braun, A., and Tziavos, I. N. (2016). "Assessing digital elevation model uncertainty using gps survey data." *J. Surv. Eng.-ASCE*, 142(3).
Cai, X., and Wang, D. (2006). "Spatial autocorrelation of topographic index in catchments." *J. Hydrol.*, 328(3), 581-591.
Chappell, N. A., Vongtanaboon, S., Jiang, Y., and Tangtham, N. (2006). "Return-flow prediction and buffer designation in two rainforest headwaters." *For. Ecol. Manage.*, 224(1), 131-146.
Czubski, K. (2013). "Accuracy of SRTM-X and ASTER elevation data and its influence on topographical and hydrological modeling: Case study of the pieniny mts. In poland." *Int. J. Geoinformatics*, 9(2), 7-14.
Darnell, A. R., Tate, N. J., and Brunsdon, C. (2008). "Improving user assessment of error implications in digital elevation models." *Comput. Environ. Urban Syst.*, 32(4), 268-277.
DeMers, M. N. (2002). *GIS modeling in raster*, Wiley, New York.
Dragut, L., and Eisank, C. (2011). "Object representations at multiple scales from digital elevation models." *Geomorphology*, 129(3), 183-189.
Du, X., Guo, H., Fan, X., Zhu, J., Yan, Z., and Zhan, Q. (2016). "Vertical accuracy assessment of freely aavailable digital elevation models over low-lying coastal plains." *Int. J. Digit. Earth*, 9(3), 1-20.
Eckert, S., Kellenberger, T., and Itten, K. (2005). "Accuracy assessment of automatically derived digital elevation models from aster data in mountainous terrain." *Int. J. Remote Sens.*, 26(9), 1943-1957.
El-Sheimy, N., Valeo, C., and Habib, A. (2005). *Digital terrain modeling: Acquisition, manipulation, and application*, Artech House, Massachusetts.
Elkhrachy, I. (2017). "Vertical accuracy assessment for SRTM and ASTER digital elevation models: A case study of Najran city, Saudi Arabia." *Ain Shams Eng. J.*, In Press.
Frey, H., and Paul, F. (2012). "On the suitability of the SRTM DEM and ASTER GDEM for the compilation of topographic parameters in glacier inventories." *Int. J. Appl. Earth Obs. Geoinf.*, 18, 480-490.
Gesch, D. B., Oimoen, M. J., Danielson, J. J., and Meyer, D. (2016). "Validation of the ASTER global digital elevation model version 3 over the conterminous United States." *XXIII ISPRS Congress*, International Society for Photogrammetry and Remote Sensing, Prague, Czech Republic, 143-148.
Glennie, C. (2018). "Arctic high-resolution elevation models: Accuracy in sloped and vegetated terrain." *J. Surv. Eng.-ASCE*, 144(1).
Hladik, C., and Alber, M. (2012). "Accuracy assessment and correction of a lidar-derived salt marsh digital elevation model." *Remote Sens. Environ.*, 121, 224-235.
Höhle, J., and Höhle, M. (2009). "Accuracy assessment of digital elevation models by means of robust statistical methods." *ISPRS-J. Photogramm. Remote Sens.*, 64(4), 398-406.
Hu, Z., Peng, J., Hou, Y., and Shan, J. (2017). "Evaluation of recently released open global digital elevation models of Hubei, China." *Remote Sens.*, 9(3), 262.
Huber, P. J. (2011). *Robust statistics*, Springer, Berlin Heidelberg.
Isfahan municipality ICT Organization (2014). "Design and construction of GPS-Leveling Network in Isfahan City."Isfahan, Iran (In persian).

Jain, A. O., Thaker, T., Chaurasia, A., Patel, P., and Singh, A. K. (2017). "Vertical accuracy evaluation of SRTM-GL1, GDEM-V2, AW3D30 and CartoDEM-V3. 1 of 30-m resolution with dual frequency GNSS for lower Tapi Basin India." *Geocarto Int.*, 1-20.

JAXA (n. d.). "ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)"." <https://www.eorc.jaxa.jp/ALOS/en/aw3d30/>. (November, 2018).

Kiamehr, R., and Sjöberg, L. (2005). "Effect of the SRTM global dem on the determination of a high-resolution geoid model: A case study in iran." *J. Geodesy*, 79(9), 540-551.

Li, Z., Zhu, C., and Gold, C. (2004). *Digital terrain modeling: Principles and methodology*, CRC press, Florida.

Liu, X. H., Hu, H., and Hu, P. (2015). "Accuracy assessment of lidar-derived digital elevation models based on approximation theory." *Remote Sens.*, 7(6), 7062-7079.

Merry, C. (2003). "DEM-induced errors in developing a quasi-geoid model for Africa." *J. Geodesy*, 77(9), 537-542.

Mispan, M. R., Rasid, M. Z. A., Rahman, N. F. A., Khalid, K., Haron, S. H., and Ahmad, N. (2015). "Assessment of ASTER and SRTM derived digital elevation model for highland areas of peninsular malaysia region." *Int. Res. J. Eng. Technol.*, 02(06), 316-320.

Montgomery, D. C., and Runger, G. C. (2010). *Applied statistics and probability for engineers*, John Wiley & Sons, New York.

Muir, J., Goodwin, N., Armston, J., Phinn, S., and Scarth, P. (2017). "An accuracy assessment of derived digital elevation models from terrestrial laser scanning in a sub-tropical forested environment." *Remote Sens.*, 9(8), 24.

Mukherjee, S., Joshi, P., Mukherjee, S., Ghosh, A., Garg, R., and Mukhopadhyay, A. (2013). "Evaluation of vertical accuracy of open source digital elevation model (DEM)." *Int. J. Appl. Earth Obs. Geoinf.*, 21, 205-217.

NASA (n. d.). "Earthdata Search Application: ASTER GDEM." <https://search.earthdata.nasa.gov/search/granules?p=C197265171-LPDAAC_ECS&m=31.078125!50.9765625!4!1!0!0%2C2>. (September, 2018).

NASA (n. d.). "Earthdata Search Application: SRTM." <https://search.earthdata.nasa.gov/search/granules?p=C1000000240-LPDAAC_ECS&g=G1004581359-LPDAAC_ECS&m=29.7421875!49.04296875!4!1!0!0%2C2>. (September, 2018).

National Iranian Oil Company (2016). "GPS-Leveling Point grid construction for exploration porposes."Teharn, Iran (In persian).

Nguyen, T. (2015). "Optimal ground control points for geometric correction using genetic algorithm with global accuracy." *Eur. J. Remote Sens.*, 48(1).

Oniga, V. E., Breaban, A. I., and Statescu, F. (2018). "Determining the optimum number of ground control points for obtaining high precision results based on uas images." *Proc., Multidisciplinary Digital Publishing Institute Proceedings*, 352.

Patel, A., Katiyar, S. K., and Prasad, V. (2016). "Performances evaluation of different open source DEM using Differential Global Positioning System (DGPS)." *Egypt. J. Remote Sens. Space Sci.*, 19(1), 7-16.

Ramlal, B., and Baban, S. M. (2008). "Developing a GIS based integrated approach to flood management in trinidad, west indies." *J. Environ. Manage.*, 88(4), 1131-1140.

Rexer, M., and Hirt, C. (2014). "Comparison of free high resolution digital elevation data sets (ASTER GDEM2, SRTM v2. 1/v4. 1) and validation against accurate heights from the australian national gravity database." *Aust. J. Earth Sci.*, 61(2), 213-226.

Rodriguez, E., Morris, C. S., and Belz, J. E. (2006). "A global assessment of the SRTM performance." *Photogramm. Eng. Remote Sens.*, 72(3), 249-260.

Rodriguez, E., Morris, C. S., Belz, J. E., Chapin, E. C., Martin, J. M., Daffer, W., and Hensley, S. (2005). "An assessment of the SRTM topographic product."NASA Jet Propulsion Laboratory D-31639.

Romanowicz, R. J., Young, P. C., Beven, K. J., and Pappenberger, F. (2008). "A data based mechanistic approach to nonlinear flood routing and adaptive flood level forecasting." *Adv. Water Resour.*, 31(8), 1048-1056.

Rudin, W. (1964). *Principles of mathematical analysis*, McGraw-Hill, New York.

Shan, J., Zaheer, M., and Hussain, E. (2003). "Study on accuracy of 1-degree DEM versus topographic complexity using GIS zonal analysis." *J. Surv. Eng.-ASCE*, 29(2).

Sun, G., Ranson, K. J., Kharuk, V. I., and Kovacs, K. (2003). "Validation of surface height from shuttle radar topography mission using shuttle laser altimeter." *Remote Sens. Environ.*, 88(4), 401-411.

Szabó, G., Singh, S. K., and Szabó, S. (2015). "Slope angle and aspect as influencing factors on the accuracy of the SRTM and the ASTER GDEM databases." *Phys. Chem. Earth*, 83–84, 137-145.

Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D., Oimoen, M., Zhang, Z., Danielson, J., Krieger, T., Curtis, B., and Haase, J. (2011). "ASTER global digital elevation model version 2–summary of validation results." M. Dave, ed., NASA Land Processes Distributed Active Archive Center and the Joint Japan-US ASTER Science Team, 27.

Tadono, T., Ishida, H., Oda, F., Naito, S., Minakawa, K., and Iwamoto, H. (2014). "Precise global DEM generation by ALOS PRISM." *Proc., ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 71-76.

Tadono, T., Nagai, H., Ishida, H., Oda, F., Naito, S., Minakawa, K., and Iwamoto, H. (2016). "Generation of the 30m mesh global digital surface model by ALOS PRISM." *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.* Prague, Czech Republic, 157-162.

Takaku, J., and Tadono, T. "Quality updates of 'AW3D'global DSM generated from ALOS PRISM." *Proc., Geoscience and Remote Sensing Symposium (IGARSS), IEEE International*, IEEE, 5666-5669.

Varga, M., and Bašić, T. (2015). "Accuracy validation and comparison of global digital elevation models over Croatia." *Int. J. Remote Sens.*, 36(1), 170-189.

Wang, B., Shi, W., and Liu, E. (2015). "Robust methods for assessing the accuracy of linear interpolated dem." *Int. J. Appl. Earth Obs. Geoinf.*, 34, 198-206.

Wessel, B., Huber, M., Wohlfart, C., Marschalk, U., Kosmann, D., and Roth, A. (2018). "Accuracy assessment of the global TanDEM-X digital elevation model with GPS data." *ISPRS-J. Photogramm. Remote Sens.*, 139, 171-182.

Wilcox, R. (2012). *Introduction to robust estimation and hypothesis testing (third edition)*, Academic Press, Boston.

Wolock, D. M., and Price, C. V. (1994). "Effects of digital elevation model map scale and data resolution." *Water Resour. Res.*, 30(11), 3041-3052.

Yap, L., Kandé, L. H., Nouayou, R., Kamguia, J., Ngouh, N. A., and Makuate, M. B. (2018). "Vertical accuracy evaluation of freely available latest high-resolution (30 m) global digital elevation models over Cameroon (Central Africa) with GPS/leveling ground control points." *Int. J. Digit. Earth*.

Zandbergen, P. A. (2008). "Positional accuracy of spatial data: Non-normal distributions and a critique of the national standard for spatial data accuracy." *Trans. GIS*, 12(1), 103-130.
Zhao, S., Cheng, W., Zhou, C., Chen, X., Zhang, S., Zhou, Z., Liu, H., and Chai, H. (2011). "Accuracy assessment of the ASTER GDEM and SRTM3 DEM: an example in the loess plateau and north china plain of china." *Int. J. Remote Sens.*, 32(23), 8081-8093.
Zhu, L., and Jekeli, C. (2009). "Gravity gradient modeling using gravity and DEM." *J. Geodesy*, 83(6), 557-567.

Figure Captions

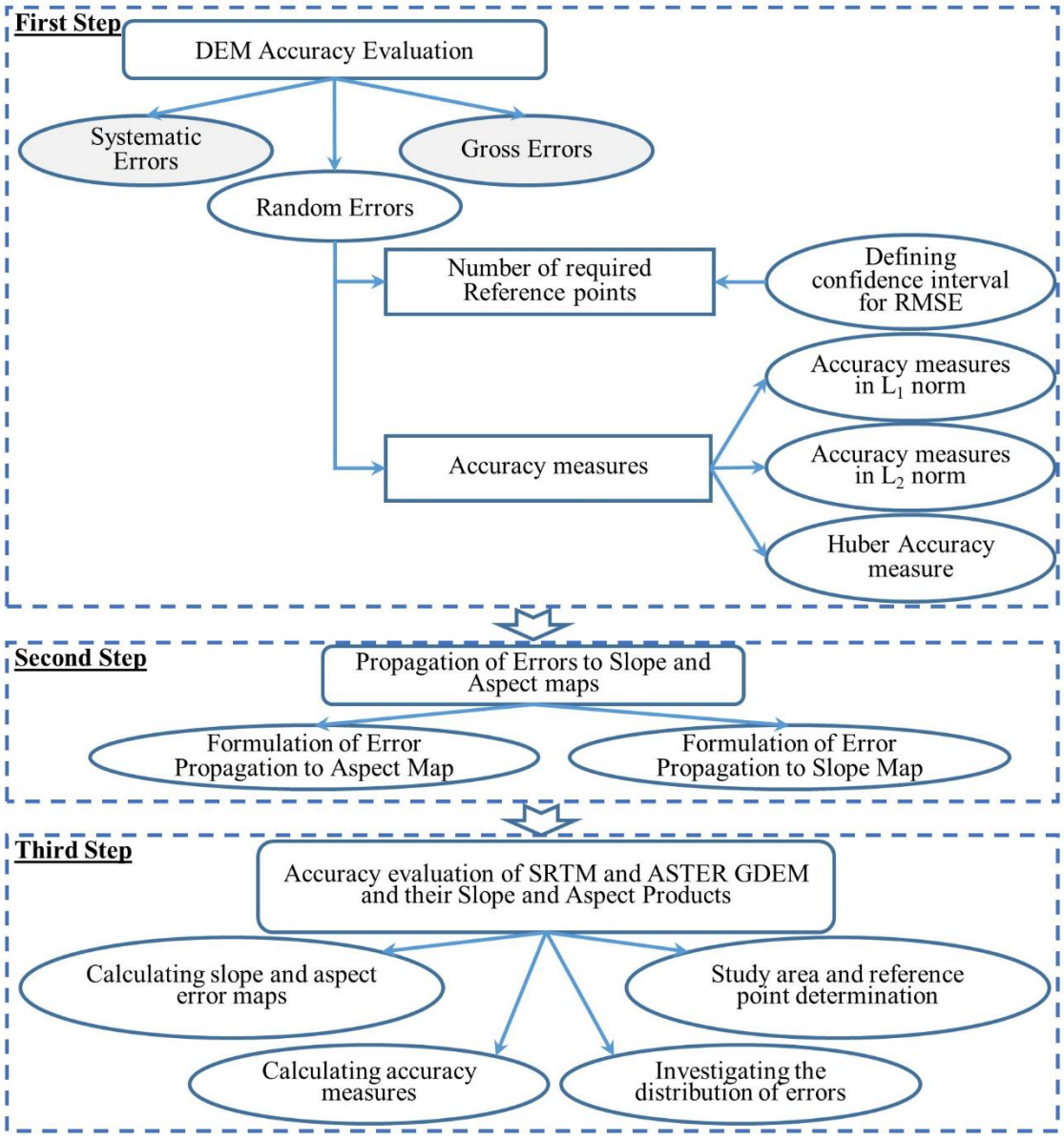

Fig. 1. Steps of proposed methodology.

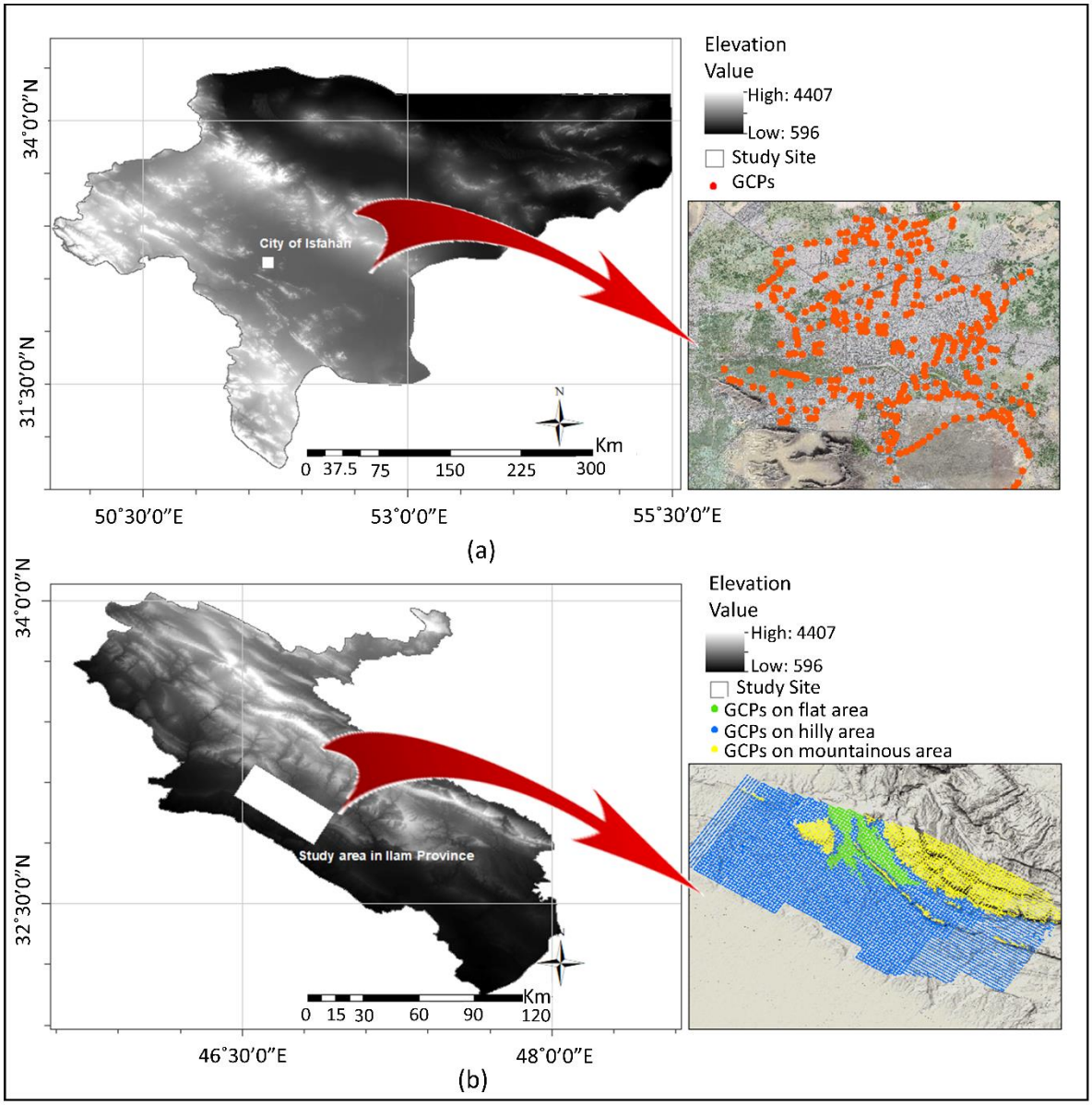

Fig. 2. Location of GCPs within the study area. (a) urban test site of Isfahan city. (b) flat, hilly and mountainous test site of Ilam province.

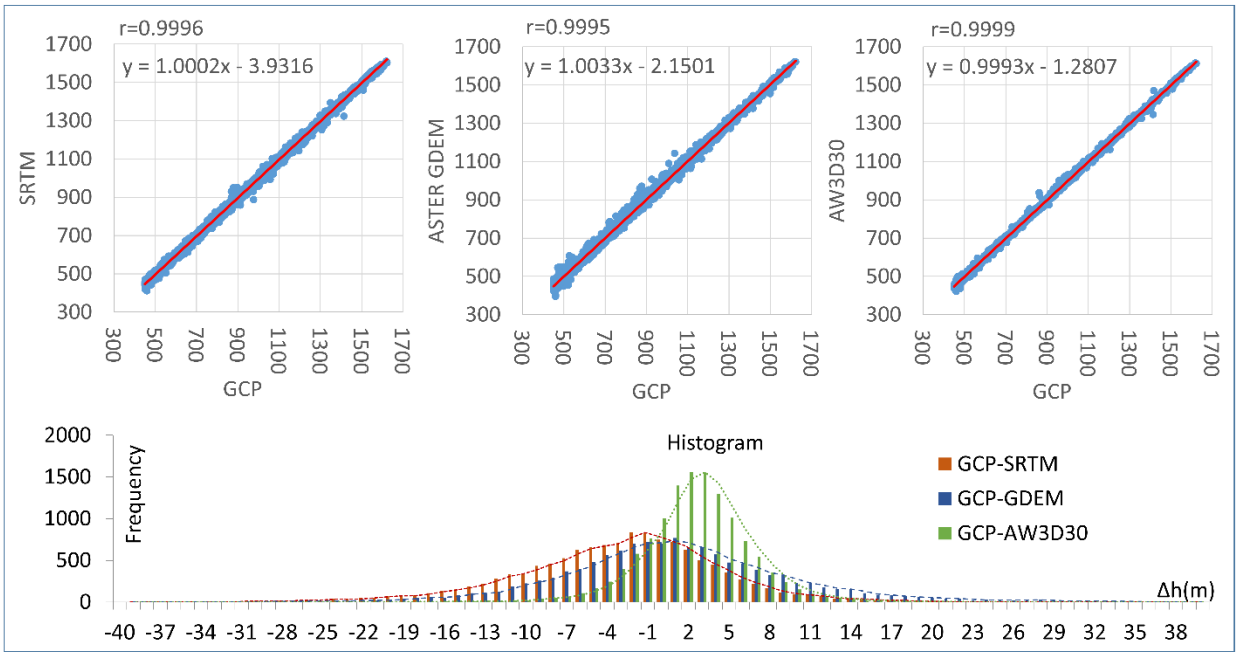

Fig. 3. Trends of AW3D30, ASTER GDEM, SRTM and GCP and the error histogram of GCP-AW3D30, GCP- ASTER GDEM and GCP-SRTM in the mountainous study area.

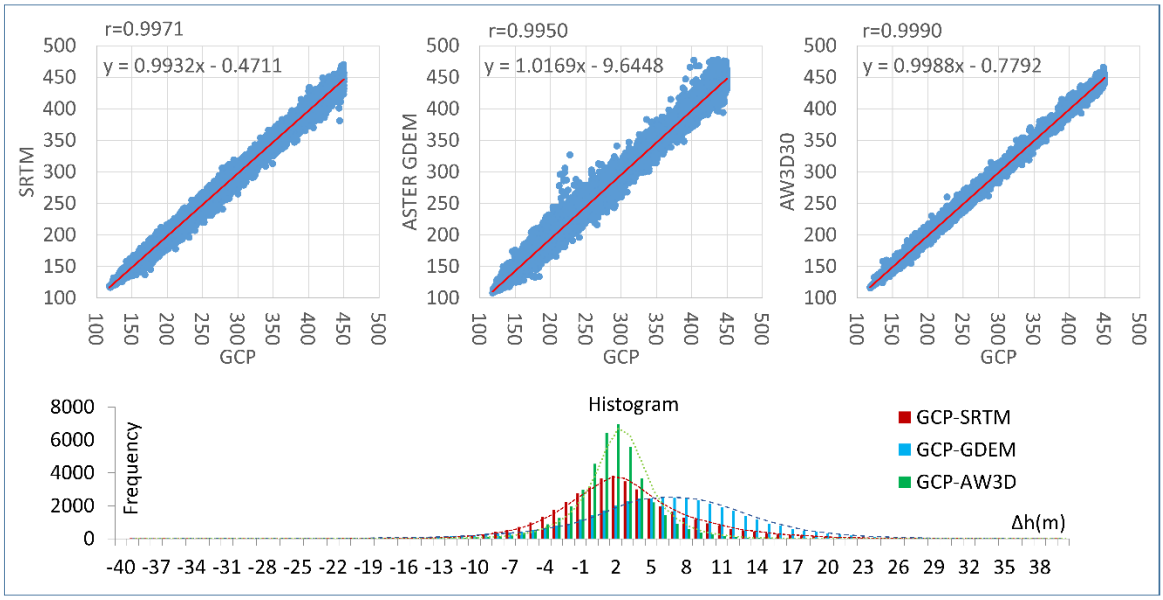

Fig. 4. Trends of AW3D30, ASTER GDEM, SRTM and GCP and the error histogram of AW3D30, ASTER GDEM and SRTM in the hilly study area.

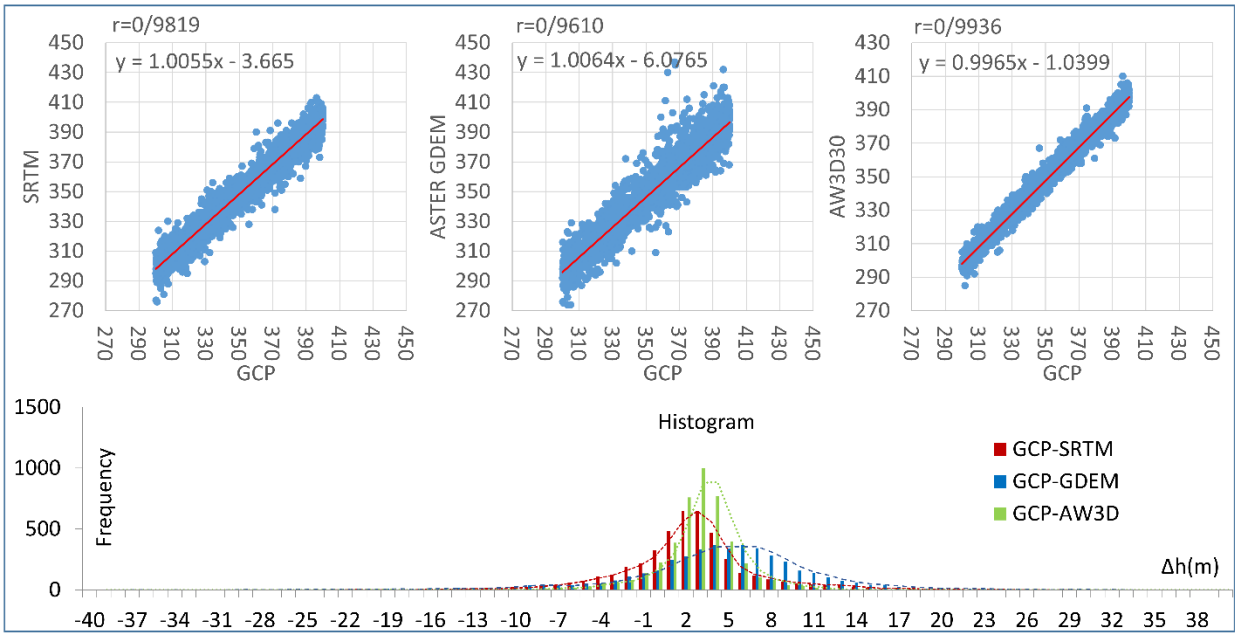

Fig. 5. Trends of AW3D30, ASTER GDEM, SRTM and GCP and the error histogram of AW3D30, ASTER GDEM and SRTM in the flat study area.

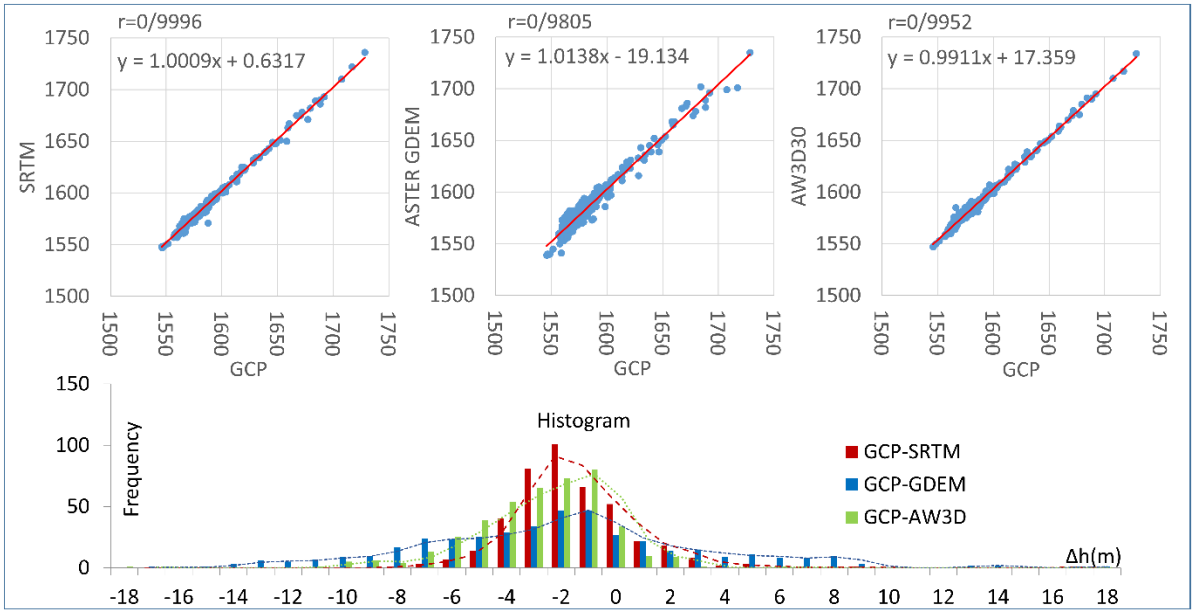

Fig. 6. Trends of AW3D30, ASTER GDEM, SRTM and GCP and the error histogram of AW3D30, ASTER GDEM and SRTM in the urban study area.

Table 1. Specifications of AW3D30, ASTER GDEM and SRTM

Specifications	Model		
	ASTER GDEM	*SRTM*	*AW3D30*
Data Source	ASTER sensor of Terra Satellite	Radar Image of Endeavour Shuttle	PRISM optical Sensor of ALOS
Version	Version 2	SRTMGL1 Version 3	Version 2.1
Spatial Resolution	1 arc-second	1 arc-second	1 arc-second
Coverage	±83º	56ºS to 60ºN	±82º
Horizontal Datum	WGS 84	WGS 84	WGS84
Vertical Datum	Geoid EGM96	Geoid EGM96	Geoid EGM96
Download Source	NASA Earthdata Search Application (NASA n. d.)	NASA Earthdata Search Application (NASA n. d.)	JAXA Web (JAXA n. d.)

Table 2. Specification of GCPs in each study area

Characteristics	*Isfahan Dataset*	*Ilam Dataset*		
	Type of study area			
	Urban	Flat (Slope < 2°)	Hilly (2°<Slope<25˚)	Mountainous (Slope>25˚)
Topography	Plain and gently sloping ground	Highly rugged and contains multiple and diverse topographic characteristics		
Minimum elevation (m)	1546	300	118	450
Maximum elevation (m)	1728	400	450	1623
Average of elevations (m)	1581	355	254	846
Standard deviation of elevations	26	26	79	304
Number of GCPs	421	4483	42355	12735

Table 3. Required number of GCPs regarding different $\boldsymbol{\alpha}$ and confidence intervals

Model	$\boldsymbol{\alpha}$	Confidence interval (m)	Number of required GCPs
ASTER GDEM 2	0.01	1	2110
		2	538
		3	247
	0.05	1	1224
		2	314
		3	145
	0.1	1	863
		2	223
		3	104
SRTM 1 arc-second	0.01	1	1483
		2	382
		3	178
	0.05	1	861
		2	223
		3	105
	0.1	1	608
		2	159
		3	75
AW3D30 V1	0.01	1	750
		2	180
		3	85
	0.05	1	437
		2	106
		3	51
	0.1	1	309
		2	76
		3	37

Table 4. Results of accuracy assessment of AW3D30, SRTM and ASTER GDEM using L_1 and L_2 statistical measures and Huber μ and σ

Type of study area	Accuracy Measure	GCP	AW3D30	ASTER GDEM	SRTM	GCP-AW3D30	GCP-ASTER GDEM	GCP-SRTM
Urban	Max (m)	1728.4	1734	1735.0	1736.0	5.0	18.0	16.6
	Min (m)	1545.9	1547	1539.0	1547.0	-18.7	-18.0	-8.9
	STD	25.8	25.7	26.7	26.0	2.5	5.3	2.3
	ME (m)					-3.3	-2.6	-2.1
	RMSE (m)					*4.2*	*5.9*	*3.1*
	Median (m)					-3.0	-2.8	-2.3
	NMAD					2.2	4.5	1.8
	$\hat{\mu}_{Huber}$ (m)					-3.3	-2.6	-2.2
	$\hat{\sigma}_{Huber}$					2.5	5.3	1.4
	Pairs with ΔH>20m (%)					0.00%	0.00%	0.00%
	Pairs with ΔH<-20m (%)					0.00%	0.00%	0.00%
	Skewness					-0.9	0.4	1.7
	Kurtosis					3.4	0.9	10.8
	Number of GCPs	421						
Flat	Max (m)	400.0	410	437.0	413.0	18.4	46.8	33.4
	Min (m)	300.0	285	274.0	276.0	-20.7	-70.0	-29.8
	STD	25.8	25.9	27.0	26.4	2.9	7.5	5.0
	ME (m)					2.3	3.8	1.7
	RMSE (m)					*3.7*	*8.4*	*5.3*
	Median					2.5	4.2	1.8
	NMAD					1.8	4.9	3.0
	$\hat{\mu}_{Huber}$ (m)					2.4	4.2	1.6
	$\hat{\sigma}_{Huber}$					1.6	4.6	2.6
	Pairs with ΔH>20m (%)					0.00%	1.60%	0.45%
	Pairs with ΔH<-20m (%)					0.02%	0.95%	0.18%
	Skewness					-0.6	-1.3	0.2
	Kurtosis					4.7	9.6	4.1
	Number of GCPs	4483						
Hilly	Max (m)	450.0	466	478.0	470.0	24.3	50.2	63.2
	Min (m)	118.0	116	108.0	117.0	-32.9	-99.5	-32.8
	STD	78.9	78.9	80.7	78.6	3.5	8.1	6.0
	ME (m)					1.1	5.4	2.2
	RMSE (m)					*3.6*	*9.7*	*6.4*
	Median					1.2	5.6	1.7
	NMAD					2.5	6.7	4.8
	$\hat{\mu}_{Huber}$ (m)					1.2	5.4	2.2
	$\hat{\sigma}_{Huber}$					2.1	8.1	6.0
	Pairs with ΔH>20m (%)					0.01%	2.80%	0.94%
	Pairs with ΔH<-20m (%)					0.03%	0.80%	0.06%
	Skewness					-0.4	-0.8	0.6
	Kurtosis					4.0	5.3	2.1
	Number of GCPs	42355						
Mountainous	Max (m)	1622.8	1615	1622.0	1611.0	66.9	66.3	88.9
	Min (m)	450.0	426	396.0	412.0	-79.3	-105.7	-72.0
	STD	304.3	304.1	305.5	304.5	4.6	9.8	8.1
	ME (m)					1.9	-0.6	3.7
	RMSE (m)					*4.9*	*9.8*	*8.9*
	Median					1.9	0.0	3.1
	NMAD					6.6	7.3	6.6
	$\hat{\mu}_{Huber}$ (m)					1.9	-0.6	3.7
	$\hat{\sigma}_{Huber}$					4.6	9.8	8.1
	Pairs with ΔH>20m (%)					0.23%	1.49%	3.20%
	Pairs with ΔH<-20m (%)					0.22%	3.08%	0.36%
	Skewness					-0.8	-1.0	0.5
	Kurtosis					21.1	8.0	5.4
	Number of GCPs	12735						

Table 5. The 99% Confidence interval for estimated RMSE of each model

Type of study area	CI of $RMSE_{AW3D30}$ (m)	CI of $RMSE_{ASTER\ GDEM}$ (m)	CI of $RMSE_{SRTM}$ (m)
Urban	[4.0 4.4]	[5.6 6.1]	[2.8 3.4]
Flat	[3.6 3.8]	[8.0 8.7]	[5.1 5.4]
Hilly	[3.6 3.7]	[9.4 10.1]	[6.3 6.4]
Mountainous	[4.8 5.1]	[9.5 10.2]	[8.7 9.1]

Table 6. Error propagation to slope and aspect in different study area

Type of study area	Aspect error *10-2 (rad)			Slope Error *10-2 (rad)		
	AW3D30	SRTM	ASTER GDEM	AW3D30	SRTM	ASTER GDEM
Urban	[0.10 1.09]	[0.05 1.71]	[0.39 1.94]	[0.02 0.14]	[0.01 0.12]	[0.09 0.65]
Flat	[0.66 1.46]	[1.70 1.04]	[5.90 3.91]	[0.11 0.19]	[0.32 0.57]	[0.87 1.30]
Hilly	[0.09 2.13]	[0.17 1.15]	[0.40 4.55]	[0.02 0.28]	[0.03 0.83]	[0.07 1.51]
Mountainous	[0.09 3.62]	[0.22 2.11]	[0.23 6.60]	[0.01 0.49]	[0.02 1.51]	[0.02 2.22]