89
Professional paper
Doi 10.61892/STP202601070P
ISSN 2566-4484

Filip Petrovski, Saints Cyril and Methodius University of Skopje, petrovski@gf.ukim.edu.mk

VALIDATION OF DIFFERENT DIGITAL ELEVATION MODELS OVER A REGION IN WESTERN MACEDONIA

Abstract:
This study investigates the performance of several digital elevation models (DEMs) over a selected area in Western Macedonia. The objective is to identify the most suitable DEM to serve as a reliable reference for the representation of topographic masses used in gravity anomaly determination, terrain correction computation, and geoid modeling. The evaluated DEMs are assessed for their capability to accurately represent topographic information needed for short-wavelength variations of the Earth's gravity field. Validation is performed through comparisons between DEM-derived orthometric heights and terrestrial measurements, as well as analyses involving corrector surfaces for removal of the systematic bias that exists between different vertical datums.
Keywords: digital elevation model, orthometric height, corrector surface

ВАЛИДАЦИЈА РАЗЛИЧИТИХ ДИГИТАЛНИХ МОДЕЛА ТЕРЕНА НА ПОДРУЧЈУ ЗАПАДНЕ МАКЕДОНИЈЕ

Сажетак:
Ова студија испитује перформансе више дигиталних модела терена (DEM) на одабраном подручју Западне Македоније. Циљ овог рада је идентификација најпогоднијег дигиталног елевационог модела (ДЕМ) који ће служити као поуздана референтна основа за репрезентацију података о топографским масама у процесима одређивања аномалија Земљине теже, прорачуна теренских корекција и моделирања геоида. Евалуирани ДЕМ-ови су оцењени у погледу њихове способности да тачно представе топографске информације неопходне за краткоталасне варијације Земљиног поља силе теже. Валидација је спроведена поређењем ортометријских висина добијених из DEM-а са терестричким мерењима, као и анализе које укључују примену поравнања у циљу уклањања систематских одступања која постоје између различитих вертикалних датума.
Кључне речи: дигитални модел терена, ортометријска висина, поравнања

1. INTRODUCTION

Digital elevation models (DEMs) constitute a fundamental dataset in geoid determination, as they provide the necessary topographic information for modeling short-wavelength variations in the Earth's gravity field. Their influence on the determination of geoid undulations is primarily local, corresponding to wavelengths shorter than 20 km ($n_{max} = 1000$).
The term "digital elevation model" is commonly used to describe models that provide topographic information for a given area. In the literature, the acronym DEM is most frequently employed when referring to topographic data obtained through satellite acquisition, airborne LIDAR, photogrammetry, digitized contours, as well as interpolation of terrestrial measurements. However, within this broad category, two additional related terms are distinguished: Digital terrain model (DTM) and Digital surface model (DSM). To clarify these distinctions, each is briefly described [1]:

- DEM - In most of the literature, DEM is used as a generic term encompassing both DSM and DTM. This term provides a general description of all digital datasets containing elevation information for terrain, topography, and bathymetry. The adjective "digital" reflects the fact that these datasets are primarily created, distributed, and analyzed in digital form. Conceptually, DEMs serve a dual purpose: as a pixel-based representation of the terrain and as a model that can be analyzed by software algorithms to extract detailed topographic information;
- DTM - This term is often used when the terrain represents the bare Earth surface, consistent with the broader framework of a DEM. DTMs are produced through additional software processing to provide improved representation of topographic features and terrain details. DTM products are widely used in applications such as flood mapping, urban planning, and geoid modeling, where they are indispensable components.
- DSM represent the outermost surfaces in a given location on Earth, i.e., the surfaces from which radar or lidar signals are reflected. In practice, DSMs capture the first surface encountered by the sensor, which may include dense vegetation, buildings, and other man-made structures. In contrast, DTMs retain only information corresponding to the Earth's bare surface, thereby providing models of the terrain itself without vegetation or artificial objects.

DEMs, together with global geopotential models (GGMs) and terrestrial gravity anomalies, constitute the essential datasets required for geoid modeling. This study outlines the theoretical framework and key characteristics of the DEMs employed, along with the validation methodology designed to identify the model exhibiting the lowest standard deviation.

1. DATA AND METHODOLOGY

1.1. DATA

The main datasets implemented in the study are four different DEMs:

- SRTM1" (Shuttle Radar Topography Mission) is a terrain model produced through a joint effort by NASA (National Aeronautics and Space Administration, USA), the German Aerospace Center (DLR), and the Italian Space Agency (ASI). The mission generated three-dimensional representations of the Earth's surface using radar interferometry, in which two overlapping radar images are combined to extract elevation data. These datasets were processed to create a global model covering latitudes from 60° North to 56° South [2]. The model achieves a vertical accuracy of approximately 16 meters at a 90% confidence level. SRTM products are publicly accessible through the United States Geological Survey [3] and are extensively applied in hydrological studies. The model uses the WGS84 horizontal datum and the EGM96 vertical datum.
- ASTER1" (Advanced Space-borne Thermal Emission and Reflection Radiometer) is an observational sensor installed on the Terra satellite in December 1999. The sensor represents a collaborative effort between NASA and the Japanese Ministry of Economy, Trade, and Industry (METI). ASTER-derived DEMs are generated from pairs of stereo images acquired in nadir and backward-looking angles over the same area, as performed in 2009. This approach allows for improved elevation accuracy due to the increased number of observations. The model provides near-global coverage, extending from 83° North to 83° South latitude, including remote and mountainous regions [2]. The data are distributed in georeferenced TIFF format with 1° × 1° tiles at a spatial resolution of 1 arc-second. Water bodies, such as seas or oceans, are represented with values of 0 or -9999 meters.

Since this research is conducted exclusively over land-based areas, such representations are not considered. Detailed specifications and data access are available through the Japanese Space System server [4]. The vertical accuracy of ASTER DEMs ranges from 10 to 20 meters. The model uses the WGS84 horizontal datum and the EGM96 vertical datum. ASTER DEMs are primarily applied in geomorphological studies.

- MERIT3" (Multi-Error-Removed Improved Terrain) is a high-resolution terrain model developed by a research team led by Professor Toshihiro Yamazaki at the University of Tokyo. This model, with a 3-arc-second resolution (~90 m), was generated by integrating existing SRTM and AW3D (Advanced World 3D) datasets. Systematic and random errors in these source datasets, including signal reflection errors, vegetation-induced errors, and other noise, were removed using multiple satellite data sources and advanced filtering techniques. This process effectively separates measurement errors from true topographic signals before analysis. After error correction, the proportion of land surfaces with an elevation error within ±2 m increased from 39% to 58%. Significant improvements are observed in lowland areas, particularly near flood-prone river regions. The vertical accuracy of MERIT ranges from 2 to 5 meters. The model is defined using the WGS84 horizontal datum and the EGM96 vertical datum. Data are provided in 5° × 5° tiles and are available in ESRI (Fortran Direct Access Format) or GeoTIFF formats. MERIT is widely used in environmental protection and hydrological applications [5].
- TANDEM-X1" (TerraSAR-X add-on for Digital Elevation Measurements) is a radar observation mission consisting of a SAR interferometer formed by two identical satellites flying in close formation. Cross-interferometry with baselines ranging from 120 to 500 meters enables highly precise interferograms that are largely unaffected by temporal decorrelation and atmospheric disturbances. The primary objective of the mission is to generate a global digital elevation model, which was achieved in 2016. The absolute vertical accuracy ranges from 1 to 2 meters. TANDEM-X DEMs are widely applied in geology, oceanography, meteorology, environmental studies, and cartography. The standard spatial resolution is 1 arc-second (~30 m), with higher-resolution versions available at 0.4 arc-seconds (~12 m) under special authorization. The model is provided with different vertical reference systems. The horizontal datum is WGS84, while the vertical datum can be either WGS84 (ellipsoidal heights) or EGM2008 (geoid undulations). Consequently, orthometric heights can be derived by combining both datasets to account for the difference between ellipsoidal heights and geoid undulations. TANDEM-X data products are accessible through the Earth Observation Center at the German Aerospace Center (DLR) [6].

The study area selected for this research is located in the northwestern part of the Republic of North Macedonia, extending from φ_{min} = 41°18′18″ (41.305°) to φ_{max} = 41°38′42″ (41.645°) latitude and λ_{min} = 20°39′18″ (20.655°) to λ_{max} = 20°59′42″ (20.995°) longitude. The boundary coordinates are also provided in decimal degrees to facilitate plotting on axes in graphical representations.

This region exhibits a highly variable topography, with elevations ranging from approximately 600 m to 2200 m above sea level, and an average elevation of around 1264 m. The western part of the study area is characterized by more rugged terrain, with elevations between approximately 1500 m and 2200 m, while the eastern part is relatively lower, ranging from roughly 600 m to 915 m. The total area spans approximately 20′24″ in both latitude and longitude, corresponding to an approximate surface area of 1072 km².

The validation of the models is done by using 165 points which are obtained from the AREC (Agency for Real Estate Cadastre) comparing their orthometric heights with those derived from the DEMs at the same locations. Positional coordinates are given in the European Terrestrial Reference System 1989 (ETRS89) and refer to the GRS80 ellipsoid. The orthometric heights of the points are obtained during the new high-precision levelling network (NVT3) campaign that was carried out in the period between 2010-2014. According to Federal Agency for Cartography and Geodesy (BKG), the offset between the national vertical datum of North Macedonia and the EVRF2019 vertical reference frame is approximately 55 cm. The datum of the NVT3 is defined by the fundamental benchmark – Skopje (FRSK) with its orthometric height from NVT2 campaign (epoch 1971.5). The accuracy of the levelled heights between the vertical benchmarks is approximatelly 0.3 mm, while the vertical accuracy of the adjusted heights is 6 mm. During these comparisons, corrector surfaces are implemented to account for any inconsistencies arising from differences in the vertical datum definitions.

Figure 1 shows the spatial distribution of the points used for the comparison which are marked with red squares, while Figure 2 presents the location of the study area (red line) within the Republic of North Macedonia.

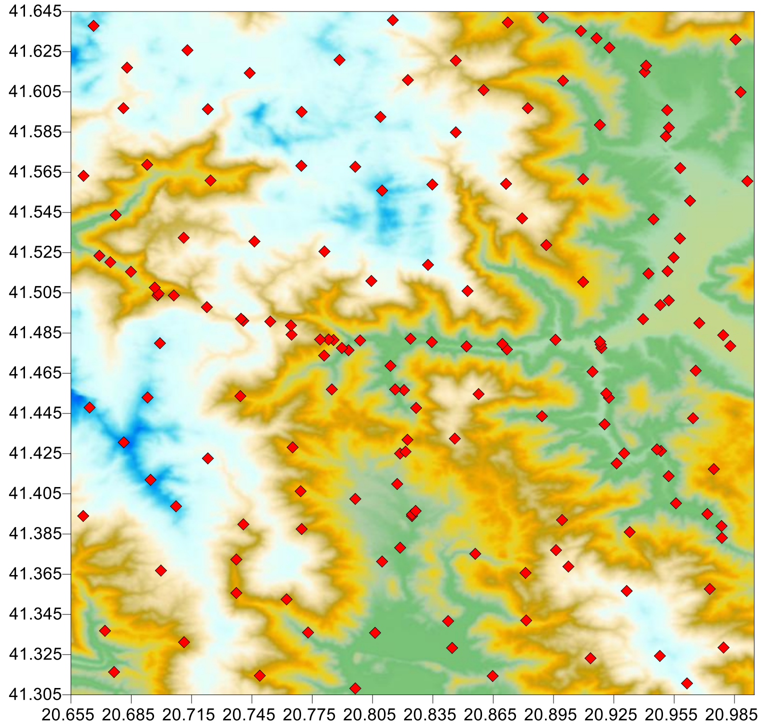

Figure 1. Distribution of the points for validation

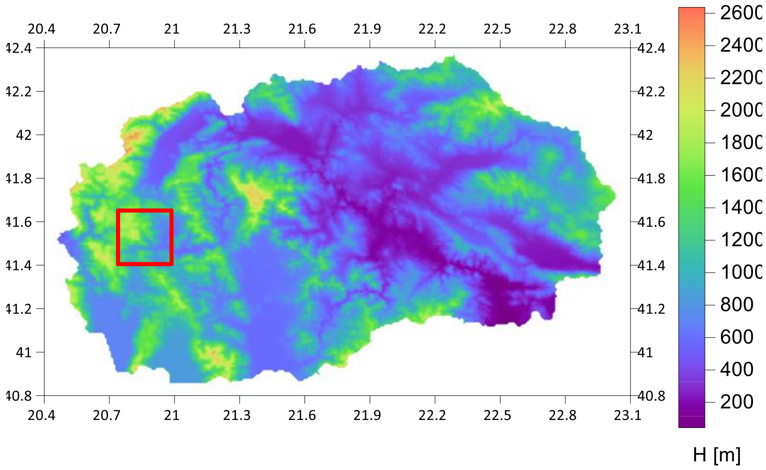

Figure 2. Study area within North Macedonia (red line)

1.2. METHODOLOGY

The methodology for comparison involves using measured orthometric heights and those extracted from the terrain models. Initially, differences are assessed between the terrain models themselves, followed by a comparison of each model with the corresponding terrestrial orthometric heights. In addition to direct comparison, correction surfaces are applied to account for systematic discrepancies.

Direct comparison (no-fit solution) between the models is done by using the following equation:

$$\delta H = H_{DEM1} - H_{DEM2} \tag{1}$$

while between the models and the terrestrial points is done with:

$$\delta H = H_{levelling} - H_{DEM} \quad (2)$$

On the other hand, the comparison is also performed with the implementation of a corrector surface (fitted solution) by the following equation:

$$\delta H_i = H_{levelling} - H_{DEM} = A_i^T x + v_i \quad (3)$$

where i is the computation point, x is vector of unknowns, and A_i is the vector of the known coefficients, while v_i is the residual. Equation (3) is also employed for comparisons between two DEMs.
Here, a four-parameter corrector surface is implemented with the following equation [7]:

$$A_i^T x = cos\varphi_i cos\lambda_i x_1 + cos\varphi_i sin\lambda_i x_2 + sin\varphi_i x_3 + x_4 \quad (4)$$

All comparisons are done by calculation of the mean and the standard deviation of the differences of the orthometric heights.

2. RESULTS

First comparison is done by validation between the DEMs at the discrete points. The results are shown in Table 1.

Table 1. Comparison of DEMs, no-fit, units [m]

Difference	Minimum	Maximum	Mean	STD
SRTM-ASTER	-9.30	20.09	0.40	4.64
SRTM-MERIT	-76.10	91.01	1.40	27.42
SRTM-TANDEM	-78.01	91.25	-3.73	28.21
ASTER-MERIT	-84.98	94.52	-2.60	28.29
ASTER-TANDEM	-86.90	94.76	-7.72	28.38
MERIT-TANDEM	-17.81	14.61	-5.12	5.20

The intercomparison of the analyzed DEMs highlights notable differences in their statistical characteristics, particularly with respect to the mean error and standard deviation. The strongest agreement is observed between the SRTM and ASTER models, as well as between the MERIT and TANDEM models, both pairs exhibiting comparatively small standard deviation. In contrast, the remaining pairwise combinations are characterized by substantially larger standard deviations, indicating increased dispersion and reduced consistency between those datasets. These discrepancies can be primarily attributed to differences in vertical accuracy and error structure. The MERIT and TANDEM models are known to incorporate advanced error correction and calibration procedures, resulting in improved vertical accuracy.
However, despite their low relative dispersion, the mean difference between MERIT and TANDEM is approximately −5 m, indicating the presence of a systematic vertical bias. This contrasts with the SRTM–ASTER comparison, where the mean error is close to zero, suggesting negligible systematic offset between these two datasets. Moreover, comparisons between the SRTM-ASTER and the MERIT-TANDEM consistently yield standard deviation values on the order of 27–28 m.
The consistency of these values suggests the presence of systematic inter-model discrepancies rather than random noise. Such behavior may be indicative of differences in vertical reference frameworks, datum realizations, or intrinsic model definitions, including spatial resolution, filtering approaches, and error propagation characteristics.
Next comparison is performed using a corrector surface. The results are presented in the Table 2.

Table 2. Comparison of DEMs, four-parameter fit, units [m]

Difference	Minimum	Maximum	Mean	STD
SRTM-ASTER	-14.26	12.91	0.00	4.55
SRTM-MERIT	-95.02	73.87	0.00	27.30
SRTM-TANDEM	-102.91	69.79	0.00	27.41
ASTER-MERIT	-101.45	78.90	0.00	27.91
ASTER-TANDEM	-109.34	74.82	0.00	27.73
MERIT-TANDEM	-16.64	11.16	0.00	4.71

Following the application of the fitting procedure, a slight reduction in the STD is observed, and the improvement is not significant. The SRTM–ASTER comparison consistently exhibits the smallest STD in both the fitted and no-fit cases. Since the four-parameter transformation eliminates the mean differences by definition, the evaluation of model agreement is based primarily on the STD. The results indicate that STD values remain largely unchanged after fitting, and the same grouping of models persists: SRTM–ASTER and MERIT–TANDEM comparisons show low dispersion, while all other combinations yield STD values of approximately 27–28 m. This indicates that the fitting procedure removes only systematic components, such as bias and large-scale trends, without significantly affecting the residual variability. The consistent STD values across inter-group comparisons suggest that the remaining discrepancies are primarily driven by systematic differences between the models, likely related to vertical reference surfaces, datum definitions, or processing methodologies, rather than random error alone. The third comparison is done by comparing the models to the terrestrial data. The results are shown in Table 3.

Table 3. Comparison of DEMs with levelling points, no-fit, units [m]

Difference	Minimum	Maximum	Mean	STD
SRTM	-44.02	205.62	67.14	59.10
ASTER	-48.13	192.81	62.35	58.45
MERIT	-56.05	178.18	52.14	52.85
TANDEM	-49.02	189.37	56.47	53.94

In this approach, when DEMs are compared to the levelling points, the model that yields smallest STD is MERIT, while the SRTM model exhibits the largest STD, as shown in Table 3.

Table 4. Comparison of DEMs with levelling points, four-parameter fit, units [m]

Difference	Minimum	Maximum	Mean	STD
SRTM	-133.89	130.13	0.30	58.84
ASTER	-124.99	130.66	-2.11	56.70
MERIT	-126.98	125.02	-0.26	52.44
TANDEM	-133.21	123.94	-0.64	53.19

After the removal of systematic bias through the application of the four-parameter transformation, the resulting statistics are presented in Table 4. The STD values remain comparable to those reported in the previous table, indicating that the fitting procedure does not significantly reduce the dispersion of the residuals. In contrast, the mean differences reported in Table 3 are relatively large and consistent, on the order of 50–70 m, suggesting the presence of a substantial systematic offset between DEM-derived heights and levelling-based heights. Furthermore, it is noted that the mean values in Table 4, obtained after the application of the four-parameter fit, are not exactly zero. Although systematic components are mitigated through a four-parameter fit, a non-zero residual mean may persist, particularly in models that include tilt terms or employ weighted estimation. Under ideal conditions, the fitting model would fully absorb the systematic bias, resulting in zero-mean residuals. This behavior contrasts with the results presented in Table 2, where the fitted comparisons yield mean values effectively equal to zero.

3. CONCLUSION

Based on the analyses presented in the previous chapters, it can be concluded that several widely used DEMs were implemented and evaluated in the context of geoid modeling. Since their assessment is essential prior to their use in geodetic applications, the validation process represents a critical step in ensuring the reliability of derived products. The intercomparison of the examined DEMs indicates a clear grouping behavior: older models tend to exhibit better mutual agreement, while newer models also demonstrate greater internal consistency within their respective group. In contrast, comparisons between older and newer DEMs reveal larger discrepancies, reflecting differences in data sources, processing strategies, and reference surfaces.

A key outcome of this study is the identification of significant systematic offsets between DEM-derived elevations and terrestrial levelling data prior to the application of any fitting procedure. These offsets, characterized by mean differences on the order of 50–70 m, highlight the presence of inconsistencies in vertical datums or reference surfaces and strongly motivate the use of a corrective surface. The four-parameter fitting model effectively mitigates these systematic components. Although discrepancies between vertical datums are typically expected at the decimeter- to meter-level, offsets of this magnitude suggest the presence of additional inconsistencies within the applied vertical reference systems, despite the fact that orthometric heights were used consistently in both the DEMs and the national levelling network. However, residual systematic effects may still persist, indicating that the fitting primarily harmonizes reference frames rather than eliminating all sources of bias. Following the fitting procedure, the STD values, which range approximately between 50 and 60 m for all evaluated DEMs, emerge as the principal quantitative result of the study. The similarity of these STD values across both no-fit and fitted solutions indicates that the fitting process does not significantly influence the dispersion of the residuals, but rather addresses systematic shifts. Consequently, these STD values can be interpreted as representing the achievable level of agreement between DEM-derived elevations and terrestrial heights, reflecting the inherent random errors and local terrain effects present in the datasets. From the perspective of physical geodesy, this level of agreement has important implications. The STD values of 50–60 m suggests that, while global DEMs are suitable for medium- to long-wavelength applications, such as gravity field modeling and regional geoid determination, their direct use in high-precision geodetic tasks is limited without additional refinement. For precise geoid modeling, these DEMs are best suited for terrain corrections and the representation of broader topographic features, whereas local elevation datasets remain necessary for achieving centimeter-level accuracy. Although occasional outliers were identified, the evaluated DEMs demonstrate consistent performance confirming their general suitability for geodetic and geophysical applications when appropriate preprocessing is applied. Future research should focus on the validation of more recent DEM products, such as FabDEM and FathomDEM, to further assess their potential advantages in terms of vertical accuracy and their applicability to gravity field and geoid modeling.

LITERATURE

[1] L. Croneborg, K. Saito, M. Matera, D. McKeown, and J. van Aardt, *Digital Elevation Models*. 2020.

[2] İ. Ö. Bildirici and R. A. Abbak, “Comparison of ASTER and SRTM digital elevation models at one-arc-second resolution over Turkey,” *J. Geod.*, vol. 75, no. 4, pp. 215–224, 2021.

[3] USGS, “United States Geological Survey,” earthexplorer.usgs.gov, 2026. Accessed May 19, 2026, [Online]. Available: https://earthexplorer.usgs.gov/

[4] ASTER, “Japanese Space System Server,” jspacesystems.or.jp, 2026. Accessed May 19, 2026, [Online]. Available: https://www.jspacesystems.or.jp

[5] D. Yamazaki et al., “A high-accuracy map of global terrain elevations,” *Geophys. Res. Lett.*, vol. 44, no. 11, pp. 5844–5853, 2017, doi: 10.1002/2017GL072874.

[6] EOC, “Earth Observation Center (EOC) of the German Aerospace Center (DLR),” geoservice.dlr.de, 2025. Accessed May 19, 2026, [Online]. Available: https://geoservice.dlr.de/web/

[7] C. Kotsakis and M. Sideris, “On the adjustment of combined GPS-Levelling-GeoidNetworks,” *J. Geod.*, vol. 73, pp. 412–421, 1999, doi: 10.1007/s001900050261.