

The Digital Elevation Model Intercomparison eXperiment DEMIX, a community-based approach at global DEM benchmarking

Peter A. Strobl¹, Conrad Bielski², Peter L. Guth³, Carlos H. Grohmann⁴, Jan-Peter Muller⁵,
Carlos López-Vázquez⁶, Dean B. Gesch⁷, Giuseppe Amatulli⁸, Serge Riazanoff⁹, Claudia Carabajal¹⁰

¹ European Commission, DG Joint Research Centre, 21027 Ispra, Italy – peter.strobl@ec.europa.eu

²EOXPLORE UG, Germany – conrad.bielski@eoxplore.com

³Department of Oceanography, US Naval Academy, Annapolis MD 21402 USA – pguth@usna.edu

⁴Institute of Energy and Environment, University of São Paulo, São Paulo, Brazil – guano@usp.br

⁵Imaging group, Mullard Space Science Laboratory, Dept. of Space and Climate Physics, University College London,
Holmbury St Mary, Dorking, Surrey, UK – j.muller@ucl.ac.uk

⁶LatinGEO Lab IGM+ORT, Universidad ORT Uruguay, Uruguay – carloslopez@uni.ort.edu.uy

⁷Earth Resources Observation and Science Center, U.S. Geological Survey, Sioux Falls, SD 57198 USA – gesch@usgs.gov

⁸Yale University, School of the Environment, New Haven, CT, 06511, USA - giuseppe.amatulli@gmail.com

⁹VisioTerra, 77420 Champs-sur-Marne, France – serge.riazanoff@visioterra.fr

¹⁰SSAI, Inc. @ NASA Goddard Space Flight Center, Code 61A, Greenbelt, MD, USA – claudia.c.carabajal@nasa.gov

KEY WORDS: Digital Elevation Models, DEM, data quality, interoperability, CEOS

ABSTRACT:

This paper presents an initiative recently launched under the auspices of the Committee on Earth Observation Satellites (CEOS) aiming at providing harmonised terminology and methods, as well as practical guidelines and results allowing the intercomparison of continental or global Digital Elevation Models (DEM). As the work is still ongoing the main purpose of this article is not the dissemination of the outcome but rather to inform the wider community about the initiative, communicate the chosen approach to raise awareness, and attract possible further participants. Nevertheless, some preliminary results are included and an outlook on planned next steps is provided.

1. INTRODUCTION

A little more than a decade after the emergence of the first (quasi-)global Digital Elevation Model (DEM) based on the Shuttle Radar Topography Mission (SRTM) (Farr et al., 2007, Gesch et al., 2006), several similar products have entered the scene and through artefact removal, merging and re-processing, this multitude of DEMs has been even further increased. One of the latest additions came with the release of the Copernicus DEM which is based on the TanDEM-X mission (Rizzoli et al. 2017), making for the first time this innovative data set globally accessible on a free and open basis at resolutions down to 30m (ibid.). Today we find ourselves in a situation, in which it is often difficult, even for experts, to assess what the major strengths, weaknesses and differences are between the available data sets and to decide which DEM might be the most accurate or appropriate for a certain application or region. At the same time the dependency on global, accurate, and high-resolution DEM information is larger than ever, given their fundamental role in many domains dealing with high quality geoinformation products including the geometric correction of EO satellite data.

2. DEMIX INITIATIVE

2.1 Background

Whenever a new DEM or an update to an existing one is released, it raises a demand for information about the consistency and comparability with respect to other products and de-facto standards such as the aforementioned SRTM based suite of DEMs. The multitude of technologies used to produce DEMs and the differences in scope and scale often render this a difficult task, which is further complicated by the fact that terminology around DEMs (including the term itself) and their quality is neither exhaustively nor unambiguously defined and easily measurable.

While it is obvious that a systematic benchmarking of several global DEMs against each other will be of great use, the parameters, methods and algorithms to be applied in such an exercise remain unclear.

In January 2019, a workshop on Global Digital Elevation Model (DEM) Benchmarking was co-organised by the Joint Research Centre (JRC) of the European Commission and geomorphometry.org which took place in Ispra, Italy. The purpose of the workshop was to discuss terminology and methods for quality assessment with the aim of developing a framework for benchmarking current and future global DEMs. The workshop gathered about 30 scientists from around the globe representing DEM producers, analysts and various user communities. It recommended, among others, to reactivate the Terrain Mapping Sub-Group (TMSG) of the Working Group on Calibration and Validation within the Committee on Earth Observation Satellites (CEOS-WGCV) and work towards a comparison of state-of-art global DEMs based on a set of harmonized metrics. This initiative is called the ‘Digital Elevation Model Intercomparison eXperiment’ (DEMIX).

2.2 Set-up and organisation

In order to provide an added value with respect to the various existing but mostly regional or local DEM comparison studies (e.g., Grohmann 2018, Hayakawa et al., 2008, Hawker et al., 2019, Hirt et al., 2010), DEMIX is striving to have a global outreach to include all datasets which have at least continental coverage and are available under a free and open data policy (FAIR). These should include but are not limited to:

- NASADEM 1” (NASA, JPL, most recent release of the SRTM product line based on a full re-processing) (Crippen, 2016)

- SRTM, 3" and 1" to assess its adequacy and the need for a re-evaluation of the results from the many studies that have used it.
- ASTER-GDEM v3, (METI, NASA) (Abrams, 2020)
- AW3D30 1" (JAXA, free and open version of the Japanese ALOS based global DEM) (Tadono, 2016)
- TanDEM-X90, (DLR, free version of the TanDEM-X mission for scientific use) (Rizzoli, 2017)
- Copernicus DEM 3" and 1" (EC/ESA, free and open version of WorldDEM™, the commercial version of TanDEM-X produced by Airbus Inc.)
- MERIT, MERIT-hydro (Yamazaki et al, 2019)

All providers of these data have been invited through the CEOS-WGCV to participate in evaluations of the latest version of their respective data sets.

As a goal for DEMIX a number of expected outcomes have been identified and specified at either minimum threshold (t) or desirable goal (g) level.

- ✓ Consistent and comprehensive DEM definitions and terminology (t)
- ✓ Base (t) and extended (g) set of benchmarking metrics and respective algorithms (t) and open-source tools (g), able to provide accuracy assessments at pixel level.
- ✓ Detailed comparison results on test areas (t) and aggregated wall to wall benchmarking results (g).
- ✓ Recommendation of reference DEMs and consistent orthoimage (g)
- ✓ Final report (t) and peer-reviewed publications (g)

To pursue the DEMIX task, three subgroups were formed. Subgroup 1 is in charge of 'terminology and analytical basis' and is looking into reviewing the terminology and proposing a consistent set of terms which also matches with the analytical basis of geomorphology. Subgroup 2 is identifying and testing existing algorithms and toolkits to ensure their compatibility with the definitions and methods laid out by Subgroup 1. Finally, Subgroup 3 is preparing to run benchmarking tests first at smaller scale and later globally according to the tested and documented methods collated by the other groups.

Work in all three groups started in the second half of 2020 and is now well under way. By the end of 2020 about 40 experts from around the world have joined the exercise. All communications are facilitated by cloud-based collaboration tools enabling document sharing and all meetings are held in a common time-zone stretching from Tokyo to San Francisco.

3. FIRST RESULTS

3.1 Terminology and analytical basis

At the beginning of any collaboration bringing together experts from different backgrounds and user communities there should be an assessment of all key terminology which is needed in the discussion. Thus, this work had already started at the workshop in 2019. Under DEMIX, the group assembled a consistent, comprehensive and unambiguous framework to characterize and differentiate all relevant topographic phenomena which can be represented by a raster grid, occurring on planetary surfaces and in particular the Earth.

The base of this comprehensive definition of surface types is the separation of physical phenomena into six distinct categories which are known in geospatial sciences as "spheres", but are not always rigorously defined. It follows the definition of 'real surfaces' as presented by Florinsky (2016), while detailing and slightly extending it to develop a set of complete while mutually exclusive definitions of spheres based on observable properties of the contained matter. The defined bodies have surfaces (or boundary layers) along which they interface with other bodies in their surroundings. These surfaces can be categorized and differentiated according to the sphere on either side. A schematic view of how the identified six spheres and their respective interfaces on the Earth surface is given in Figure 1.

Unambiguously defining surfaces which can be referenced by elevation is necessary but insufficient as long as these 'real surfaces' are too complex for rigorous mathematical handling (Shary, 2008). To overcome this difficulty "real surfaces" are simplified and represented by a mathematically well understood model called the "topographic surface" (Florinsky, 2016). Topographic surfaces have been chosen to be the underlying concept of every DEM and its immediate derivatives.

Mathematical methods for the calculation of morphometric variables such as slope and aspect have been collected and scrutinized according to the definitions and translated into algorithms suitable for rasterized data such as DEMs including their extension to grids in latitude and longitude taking into account the differences in data spacing with non-rectangular pixels.

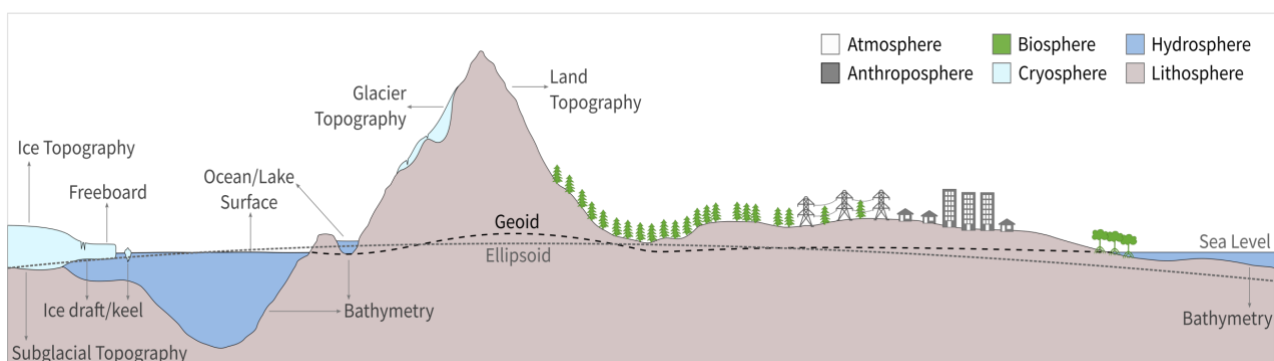


Figure 1: Schematic view of surfaces as interfaces between different spheres on the Earth Surface

3.2 Algorithms and Software

As a wide range of algorithms and software packages already exists, a systematic comparison of the results of different open-source and commercial tools was launched. Special attention is given to software/algorithm reproducibility, data reference frames and proper handling of distortions due to cartographic projections. Preliminary results show that by far, but not all, geospatial data processing tools commonly known and used can be applied blindly and compared directly. Another focus of the work is on identifying and correcting co-registration errors between DEMs as a direct comparison of different DEMs is only meaningful once horizontal and vertical shifts have been eliminated or at least minimised. Our goal is to ensure that DEM products are all on the same playing field before the actual comparisons are computed in order to eliminate or reduce the potential external biases from the benchmarking.

3.2.1 Computation of slope: Our first efforts looked at a slope algorithm (Sharpnack and Akin, 1969), for two reasons. First, slopes represent a major derived output from DEMs for many users in a variety of fields as diverse as landslides, cross country mobility, and ecology. Additionally, as the first derivative of elevation, slope amplifies any noise or errors in the DEM. Our focus was on the freely available DEMs listed in section 2.1 of this paper, which cover the world, or almost all of the world. These have arc second spacing, either one or three arc seconds, and are known to stress software which often recommends or requires reprojection of the DEM to UTM coordinates before computing slope. Reprojection modifies the DEM (and then the slopes computed from it) by computing interpolated elevations at new locations, and should not be the automatic response when alternative, correct solutions are available.

We used a test area in Alaska where we have a dense lidar point cloud. From the point cloud we created independent DEMs with 30 m and 1"x1" spacing. The location in Alaska tested the software, because the 1"x1" grid cells are about 15x30 m, decidedly not square, which many published slope algorithms require. Adapting the slope algorithms is relatively straightforward (Guth, 2010).

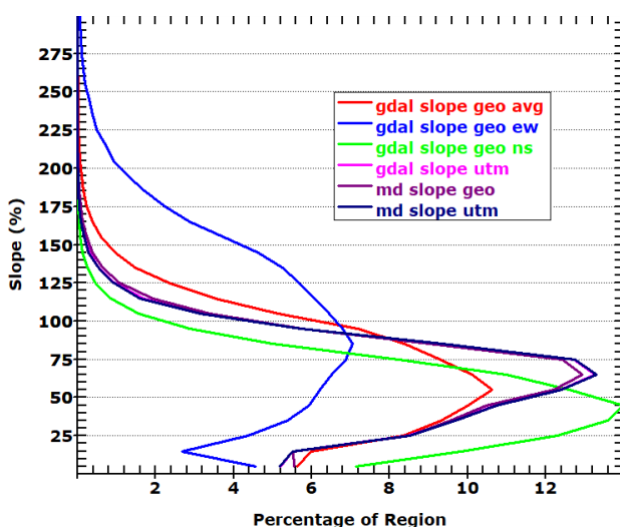


Figure 2. Slope distributions for UTM and geographic DEMs from a test area in Alaska from two GIS programs.

We tested 8 commercial and free software programs, calculating slope maps for both the UTM and geographic DEMs of the same area. Results from the two should be similar, but not identical because the pixels are of different sizes, and the overlap shifts throughout the area of the DEM. Figure 2 shows the results from two of the programs which are representative of the results. Table 1 summarizes the results for the 8 programs, and highlights which programs can be used with geographic (arc second) DEMs.

Software	Tested running	Windows	Linux	Works with Geographic DEM	Works with UTM DEM
GRASS	Command line or in QGIS	Yes	Yes	Yes	Yes
MICRODEM	GUI	Yes	No	Yes	Yes
SAGA	In QGIS	Yes	Yes	Yes	Yes
ArcGIS	GUI	Yes	No	Only Geodesic algorithm	Both algorithms
ERDAS	GUI	Yes	No	Only Spatial Model	Both algorithms
GDAL	Command line or in QGIS	Yes	Yes	No	Yes
ENVI	GUI	Yes	Yes	No	Yes
WhiteBox	GUI or command line	Yes	Yes	No	Yes

Table 1. Software Testing Results

All programs correctly handled the UTM DEMs. In Figure 2 the results from the two programs for the UTM DEM overlie almost perfectly with the results of the geographic DEM from one of the programs. Three programs correctly handled the geographic DEMs, by automatically computing the different cell spacings in the x and y directions from geodetic formulas, and modifying the slope formulas. They produce results almost identical to the UTM DEM, but there are systematic small differences between the slopes from the UTM and the geographic DEMs because they are sampling over different horizontal distances. Two programs offer two different algorithms to compute slope, only one of which correctly performs with the geographic DEMs; users should take care when using those programs with geographic DEMs. Three of the programs cannot correctly compute slope from the geographic (Lat/Lon) DEMs.

The software that cannot correctly compute slope from geographic DEMs typically asks the user to supply a correction factor, to convert from the arc second spacing to meters. This assumes that the pixels are square, and that one spacing will work in both the x and y directions. Figure 2 shows the attempt to use the north-south spacing, the east-west spacing, or the average spacing, but none of the three produce a slope distribution that matches the slope distribution from the UTM DEM, or the slope distribution from the programs that correctly handle the geographic DEMs. Users should take extreme care using these programs to compute slope from geographic DEMs, at least until they can adapt their algorithms.

3.2.2 Developing a test protocol and criteria: Let's assume we have under consideration k global DEMs, which are somewhat equivalent in coverage, attributes, etc. so on principle any of them can be used. The goal of an intercomparison is to be able to state a sentence like "... according to the criteria, the i -th DEM is statistically better than the others ..." or "... according to the criteria, the i -th DEM and the j -th DEM are better than the others, and no statistically significant difference between them has been found ...". In what follows we will discuss what can be meant by *criteria*, and how can we establish a statistically sound procedure to decide that one DEM is better than the others for a given confidence level.

We will define here as *criteria* any procedure that, applied to the k global DEM, can produce a ranking among them. If there are no ties, all the integers between 1 and k are used, attaching each one to one DEM. If there are ties, the equivalent DEMs will have each an averaged rank, computed in order that the sum of the ranks equals $k*(k+1)/2$. Thus, a valid ranking for $k=4$ will be any permutation of {1, 2, 3, 4}, or when ties are present it might be {1, 2.5, 2.5, 4} or {2, 2, 2, 4}, etc. In all cases the sum of the ranks should be $10=4*5/2$.

In the comparison we will use a number of independent Test sites. In each of them we have access to a reference dataset, deemed to be of higher accuracy. From the reference points we can compute an interpolated elevation at DEM grid points, and then we can use the discrepancies as a figure of merit. One familiar choice for the *criteria* is to compute the RMSE. In this case the *criteria* will be to rank the DEMs according to the value of the RMSE, being better ranking attached to smaller RMSE. Different test sites might render different rankings, and we will show later how to deal with this. Different *criteria* might require different reference datasets; for example, if we want to compare DEM based in their computed slope, the reference dataset might be derived from topographical local information without a high absolute elevation accuracy.

The use of the RMSE is not the single option, even considering just the elevation. In this protocol we will show that more than one *criteria* can be used under the modest requirement that it can provide a ranking among the available DEMs. The procedure might be based upon the familiar RMSE, or might be another rule which for example considers the discrepancy of the drainage network. We should require from the *criteria* that higher performing DEMs will have the better rankings (whereby "better" usually will be lower numbers), just in order to be able to compare. The *criteria* might be quantitative, and thus the rank arises after ordering the answers. Or conversely, the *criteria* might be qualitative, producing the ranking without intermediate results. We will not discuss here the various *criteria* under investigation as this is still work in progress and a final selection has not yet been made.

We will strive to find a number of Test Sites, spread over the world that are representative of varying landforms and landscapes. If we plan to declare that a particular global DEM is better than the others, there is no reason to expect that it will be better in all the Test Sites cases. We can perform a stratified sampling, separating for example plains from mountains cases, and ranking the DEMs separately. However, it will be difficult for an end user involved with global problems to select one DEM for mountain regions and another for plains. For such a user, we will assume that the DEM's choice will be just one, being a balanced option among the DEMs available.

However, it might be reasonable to perform a stratified sampling because there are users and applications which might work in a regional or local scale. For example, a study of a particular basin will be interested just in a portion of the global DEM. So the suggestion is to perform both studies, with and without stratified sampling. The conclusion about the best DEM might be the same or not, and the answer is unknown at present.

For each Test Site we can apply C different *criteria*, each able to rank the k DEMs under analysis. Common practice uses $C=1$, and is based only in the RMSE of the elevation against the reference data. If we have T Test sites, we can build a matrix holding the rankings composed of k columns, and $N=C*T$ rows. If we are interested in a stratified sampling, T will be smaller.

If the T test sites can be described as independent, as well as the C *criteria* used, we can create a matrix of rankings organized in k columns and N rows. Under some mild assumptions this is a fairly standard situation in statistics. The problem is equivalent to a wine contest, where k wines are to be appreciated by N judges. Another example is the evaluation of k medical treatments by N patients, etc. The statistical analysis of the wines or the medical treatments is a two-step one. First, we should check whether or not there exists a statistical difference between the k wines. The null hypothesis is that there is no difference, and the Friedman's test (Pereira et al., 2015) is used to reject or accept it for a given confidence level. If the data shows that the null hypothesis should be rejected, the second step is to perform a Post-hoc analysis. It is a pair-wise comparison between wines, which will show if one is better than the other for a given confidence level. The Post-hoc analysis is a low power statistical method, so it can be applied only after the Friedman's test rejects the null hypothesis.

To illustrate the method, we assembled a preliminary ranking table for a test region in Alaska using three available DEMs. As mainly explanatory, this example is not meant to provide more than preliminary benchmarking results.

Criterion	DEM1	DEM2	DEM3
Visual, subjective assessment of the hillshade	1	2	3
Topographic profiles	1	2.5	2.5
Compare Elev vs slope to reference DSM	2	2	2
Visual comparison to slope histogram	1	2	3
Uniform aspect distribution without spikes	1	3	2
Slope skewness	1.5	1.5	3
Slope kurtosis	1.5	1.5	3
SSO diagram	1.5	3	1.5
RMSE	1	3	2
L90	1	2	3
IQR95	1	2	3

Table 2. exemplification of ranking results for five qualitative (green) and six quantitative *criteria* (blue) assessed over a sample area in Alaska

In the example shown the null hypothesis can be clearly rejected using Friedman's test. This means there is a significant difference between the test candidates. It can be further established that summarising all criteria 'DEM1' is significantly better than both DEM2 and DEM3, while the difference between the latter is rather insignificant.

The described statistical procedure is fairly standard in empirical sciences. The use of the non-parametric Friedman's test circumvents the requirements of specific distributions (normal, uniform, etc. that in practice might be difficult to confirm) while still being powerful enough. It does not require large k or N . In our problem, k is moderate (maybe less than 5) while there is no bound on the number of criteria C or the number of Test Sites T . Being non-parametric the Friedman's test does not pose any fundamental requirement on the size of k or N . To accept or reject the null hypothesis some statistical tables are used for mid to low k and N values, and for other case there exist an asymptotic analytical estimate using the chi-square distribution. In particular the problem has been analyzed in detail recently, substantially extending the statistical tables for mid values of (k, N) , both for the case with (López-Vázquez and Hochstain, 2020) and without ties (López-Vázquez and Hochstain, 2019).

3.3 Platforms and processing

Subgroup 3 concentrates on applying the methodologies developed within the other subgroups. Specifically, this relates to making available to a wider audience tools for the visualisation of the results for the inter-comparison of DEMs. This will be achieved by providing preferably open-source tools and performance tests on a variety of geographical regions around the world in order to demonstrate and verify the applicability of the DEM inter-comparison exercise.

Furthermore, subgroup 3 will help to define the minimum standards and requirements regarding user access, data availability, programming interfaces and processing power, to which open candidate platforms should comply in order to participate in DEM benchmarking. Thereafter this subgroup will also be in charge of implementing and monitoring DEM benchmarking and comparison at global scale. Organisations willing to provide access to their computing infrastructure should get in touch because their help would be very much appreciated in working to process as many DEM products around the world.

4. OUTLOOK

While work in the various subgroups is steadily advancing first results are under way to be published in the second half of 2021. The plan is to then finish a first round of global benchmarking by the end of 2021. The expected legacy of DEMIX, however, will extend way beyond static comparison results of existing DEMs. The flexibility of the chosen approach allows the collation and interpretation of results to be adapted to specific application profiles and user's needs such as regionalization. To achieve this, benchmarking can be tailored by domain such as hydrology, geomorphometry or orthorectification or stratified to specific geographic areas or environmental zones. This includes adding additional quality criteria and tests which might also help in improving existing DEMs and specifying future ones. In addition, the collected reference data and methods will stay available and allow a wide range of users to benchmark their own products and also to add their own reference data. DEMIX, as a community-based approach remains open to interested individuals who wish to contribute with own methods, data, or infrastructure.

REFERENCES

- Abrams, Michael; Crippen, Robert; Fujisada, Hiroyuki. 2020. "ASTER Global Digital Elevation Model (GDEM) and ASTER Global Water Body Dataset (ASTWBD)" *Remote Sens.* 12, no. 7: 1156. <https://doi.org/10.3390/rs12071156>
- Crippen, R., Buckley, S., Agram, P., Belz, E., Gurrola, E., Hensley, S., Kobrick, M., Lavalley, M., Martin, J., Neumann, M., Nguyen, Q., Rosen, P., Shimada, J., Simard, M., and Tung, W.: NASADEM GLOBAL ELEVATION MODEL: METHODS AND PROGRESS, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B4, 125–128, <https://doi.org/10.5194/isprs-archives-XLI-B4-125-2016>, 2016.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S. Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007, The Shuttle Radar Topography Mission, *Rev. Geophys.*, 45, RG2004.
- Florinsky, I.V., 2016, Digital terrain analysis in soil science and geology: Elsevier, 2d Edition, 432 p.
- Gesch, D., Muller, J-P., Farr T.G., 2006. The shuttle radar topography mission - Data validation and applications. *Photogrammetric Engineering and Remote Sensing*. 72:233-235. (Special Issue on SRTM)
- Grohmann, C. H. 2018. Evaluation of TanDEM-X DEMs on selected Brazilian sites: Comparison with SRTM, ASTER GDEM and ALOS AW3D30. *Remote Sensing of Environment*, 212: 121–133.
- Guth, P.L. (2010). Slope, Reflectance, and Viewshed Algorithms for Arc-Second Digital Elevation Models: ASPRS 2010 Conference, San Diego, CA, April 26-30, 2010, 11 page paper on Conference CD-ROM.
- Hayakawa, Y. S., Oguchi, T., Lin, Z., 2008, Comparison of new and existing global digital elevation models: ASTER G-DEM and SRTM-3, *Geophys. Res. Lett.*, 35, L17404,

Hawker, L., Neal, J., Bates, P., 2019. Accuracy assessment of the TanDEM-X 90 Digital Elevation Model for selected floodplain sites, *Remote Sensing of Environment*, 232:111319.

Hirt, C., Filmer, M.S., Featherstone, W.E., 2010. Comparison and validation of the recent freely available ASTER-GDEM ver1, SRTM ver4.1 and GEODATA DEM-9S ver3 digital elevation models over Australia, *Australian Journal of Earth Sciences*, 57:3, 337-347

López-Vázquez, C. & Hochsztain, E. 2020. Tables for the Friedman's Test with ties [files] <http://hdl.handle.net/20.500.11968/4208>

López-Vázquez, C. & Hochsztain, E., 2019. Extended and updated tables for the Friedman rank test, *Communications in Statistics - Theory and Methods*, 48:2, 268-281

Pereira, Dulce G.; Afonso, Anabela & Melo Medeiros, Fátima (2015) Overview of Friedman's Test and Post-hoc Analysis, *Communications in Statistics - Simulation and Computation*, 44:10, 2636-2653

Rizzoli, P., Martone, M., Gonzalez, C., Wecklich, C., Borla Tridon, D., Bräutigam, B., Bachmann, M., Schulze, D., Fritz, T., Huber, M., Wessel, B., Krieger, G., Zink, M., and Moreira, A., 2017. Generation and performance assessment of the global TanDEM-X digital elevation model. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol 132, pp. 119-139.

Sharpnack, D.A., & Akin, G. (1969). An algorithm for computing slope and aspect from elevations: *Photogrammetric Engineering*, vol.35, no.3, p.247-248.

Shary, P.A., 2008, *Models of Topography*. In: Zhou Q., Lees B., Tang G. (eds) *Advances in Digital Terrain Analysis*. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg, p.29- 57.

Tadono, T., Nagai, H., Ishida, H., Oda, F., Naito, S., Minakawa, K., Iwamoto, H., 2016. Initial Validation of the 30 m-mesh Global Digital Surface Model Generated by ALOS PRISM. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B4, pp.157-162.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., Pavelsky, T. M. (2019). MERIT Hydro: a high-resolution global hydrography map based on latest topography dataset. *Water Resources Research*, 55, 5053–5073. <https://doi.org/10.1029/2019WR024873>