# Subjective Criterion for the DEMIX Wine Contest: Hillshade Maps

Peter L. Guth
Ocean & Atmospheric Sciences Dept
United States Naval Academy
Annapolis, MD, 21402 USA
pguth@usna.edu

Carlos H. Grohmann
Institute of Energy and Environment
Universidade de São Paulo
São Paulo 05508-010, Brazil
guano@usp.br

Sebastiano Trevisani
Culture del Progetto Dept
University IUAV of Venice
Venice 30123, Italy
strevisani@iuav.it

*Abstract*—**Evaluation of the hillshade map provides a significant tool for evaluating the quality of digital elevation models. The DEMIX wine contest provides a statistically rigorous way to compare and rank DEMs, and applies the method to evaluate 6 global 1 arc second DEMs. The wine contest used only quantitative criteria; we present an example using the qualitative hillshade map to verify that the wine contest works with either quantitative or qualitative criteria as DEMIX proposed. Our results verify the COPDEM and ALOS are much better than SRTM, NASADEM, and ASTER, and that those three should be retired with the advent of much better technology. We also highlight the challenges in getting enough judges to look at enough DEMs to approach the number of opinions possible with quantitative criteria. Qualitative test will probably remain a useful adjunct to much more numerous quantitative tests.**

## I. Introduction

Digital elevation models (DEMs) represent a fundamental building block for work in science, engineering, social science, government, and the military. DEMs at 1" (arc second, about 30 m) provide the best resolution freely available globally. The DEMIX group is working to compare and rank 6 of those DEMs, and created a database to support their work [1,2,3,4]. The DEMIX wine contest provides a framework for ranking DEMs and providing statistical significance for the results. An oenological wine contest frequently involves subjective assessments from experts, and the DEMIX group noted the ability to use subjective assessments for a DEM wine contest, but did not include any subjective criteria in their initial results. We will use a subjective, visual criterion, show the challenges in applying it to a large number of test areas, and demonstrate that our application of the subjective criterion validates the DEMIX group findings [3] that COPDEM, ALOS, and FABDEM are demonstrably much better than SRTM, NASADEM, and ASTER.

## II. Methods

Nothing in the wine contest precludes subjective criteria tests; for demonstration purposes, during spring 2022 we experimented with showing 16 "experts" hillshades of the DTM from DEMIX tile [5] N28VW018B covering part of La Palma in the Canary Islands (Figure 1). The DTM was created by aggregating a source DTM from the national mapping agency, using the 2 m DTM to create a 1 second DTM to match the global DEMs. Using a Google form [6], we asked the "experts" to rank the subjective visual quality of the maps. In addition to the images, they had an animation cycling though the hillshades which highlights differences. They were not allowed to have ties in their rankings.

During spring 2023 we repeated the contest with a larger number of "experts", several different test areas (two in the western US, and one in the Italian Alps), and improved methodology. Our initial assumption was the students who constituted the bulk of our "experts" did not know anything about the 6 DEMs, and the original test included the DEM names (as Figure 1). The revised test removed the animation and the DEM names, and presented the DEMs in a different random order for each test areas. We will also run the contest during Geomorphometry 2023 in Iasi, both to demonstrate the method and to collect additional data.

## III. Results

The Google Form [6] provides the test administrator a figure online (Figure 2) showing a quick visualization of the results, as well as individual results from each judge which we do not need. The Form program downloads the results in a CSV file for import into a spreadsheet. We rearranged the results to get the alternative graphic (Figure 1) which we feel more closely shows the results. We also ran statistics (Figure 3), using the wine contest Jupyter Notebook [7,8]. Table 1 summarizes the scoring for each iteration of the contest, and Figure 4 shows the overall evaluation of the overall results.

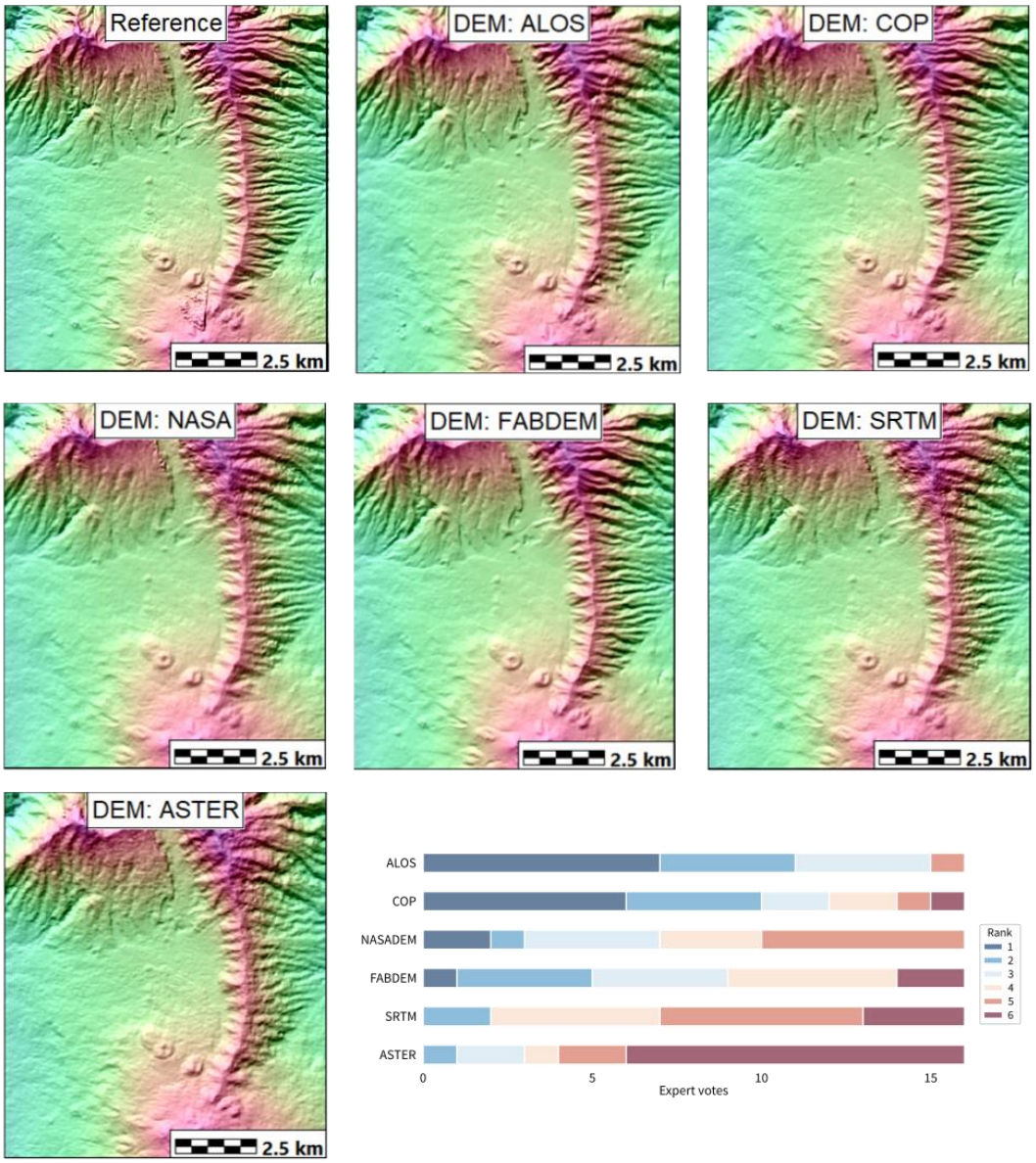


Figure 1. Hillshades of DEMIX tileTile N28VW018B, and the distribution of expert opinions for each of the 6 ranks. Low score in best.

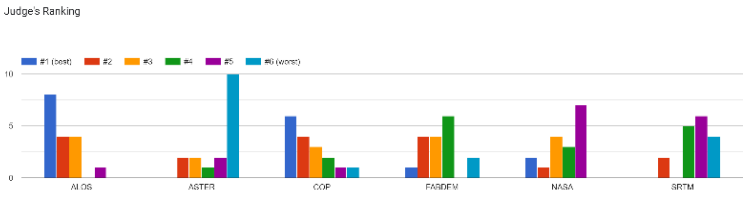


Figure 2. Google Forms automatic display of survey results.

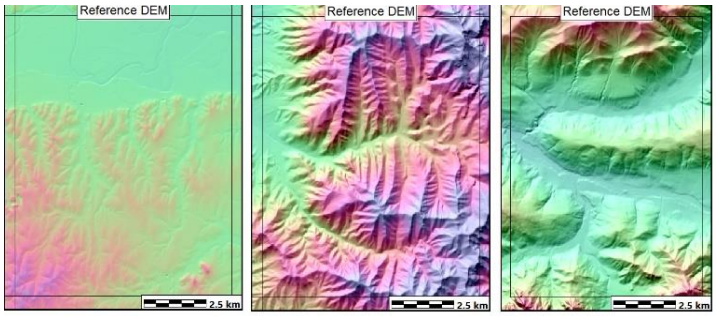


Figure 3. Hillshades for the three additional DEMs used for the second iteration of the contest.

Table 1. Wine contest average rankings for the 6 DEMs. Low score wins.

| Area | DEMIX_TILE | JUDGES | NUMBER JUDGES | COPDEM | ALOS | FABDEM | NASA | SRTM | ASTER |
|---|---|---|---|---|---|---|---|---|---|
| La Palma | N28VW018B | USNA 2022 | 17 | 2.47 | 1.94 | 3.35 | 3.71 | 4.59 | 4.94 |
| Canyon Range | N39RW113J | USNA 2023 | 24 | 2.13 | 2.71 | 1.92 | 4.67 | 4.63 | 4.96 |
| Canyon Range | N39RW113J | IUAV 2023 | 36 | 2.06 | 2.11 | 2.06 | 4.47 | 4.64 | 5.08 |
| Bolzano | N46XE012B | IUAV 2023 | 36 | 2.78 | 2.53 | 2.14 | 4.42 | 4.00 | 4.97 |
| Republican River | N39UW098F | São Paulo 2023 | 41 | 2.24 | 2.76 | 2.20 | 4.24 | 4.05 | 5.54 |
| Average all tests | | | | 2.27 | 2.39 | 2.51 | 4.27 | 4.61 | 4.95 |

```
Ranking without tolerance

No filters applied

Results of the DEMIX Wine Contest

For k=6, CL=0.05, and N=188, the critical value to compare is chi_crit=11.038
And since chi_r (430.675) is greater than chi_crit (11.038)...
Yay!! We can reject the null hipothesis and go to the Post-Hoc analysis!!
```

| | Rank | Sum of ranks | Sum of ranks divided by number of opinions | Ties with |
|---|---|---|---|---|
| **FABDEM** | 1.0 | 422.5 | 2.247 | ALOS, COP |
| **COP** | 2.0 | 457.5 | 2.434 | ALOS, FABDEM |
| **ALOS** | 3.0 | 465.0 | 2.473 | COP, FABDEM |
| **NASA** | 4.0 | 815.5 | 4.338 | SRTM |
| **SRTM** | 5.0 | 818.5 | 4.354 | NASA |
| **ASTER** | 6.0 | 969.0 | 5.154 | |

Figure 4. Wine contest ranking and statistical significance matrix. “Ties with” means the DEMs are not statistically different in this test.

## IV. Discussion

The results show a clear preference for COPDEM, FABDEM and ALOS; the results are quantitatively confirmed when using the wine contest statistics. The top three DEMs, and the bottom three significantly lower in the opinion of the judges, are the same as those from the DEMIX results [3] which relied on over 20,000 quantitative opinions for 15 criteria using 133 100 km² tiles from 19 areas spread over three continents.

While the hillshade maps show elevation with color, slope and surface roughness, derivatives of elevation, dominate the visual display. For many users these are more important characteristics of the DEM, but as emphasized by the DEMIX group, users must select the comparison criteria that most closely match their requirements.

For these areas, the difference between FABDEM and COPDEM are minimal and would be very hard to differentiate in a hillshade map, verified by the similarity in their contest scores. While NASADEM improved on SRTM for the elevation differences compared to the reference DEM, it generally has very little effect on the slope and roughness differences [3,4]. Since slope and roughness determine the hillshade, the judges did not clearly differentiate NASADEM and SRTM.

## V. Conclusions

Other potential subjective assessments for DEMs include topographic profiles [9,10] or elevation-slope plots [9,12,13]. DEM quality varies with land cover, land forms, and the slope of the terrain, so the test areas should cover a wide range of conditions.

The design and implementation of an expert-based approach to criteria evaluation is not a trivial task. The approach requires a considerable effort to collect this data and does not easily scale to multiple test regions. The demands on the judges to evaluate multiple DEMs mean that we could never reach the hundreds of test areas, and over a dozen criteria, which are possible with automated quantitative criteria. In our first iteration the DEMs were always in the same order; for the second iteration, we used multiple test areas which always had the DEMs in the same order, but varied in not showing the same DEM first or last in every test area. This was the best we could do using Google forms. It would require custom programming to make an ideal survey, and an effort to collect multiple experts willing to judge a number of tiles. Custom programming would also allow judges to give ties.

Despite the challenges, the test shows the power of the wine contest to evaluate DEMs, and that subjective criteria can be used. While the statistical validity of qualitative criteria may have caveats due to relatively small sample sizes, it provides another metric that users can evaluate in deciding which DEM they prefer to use, which in the end comes down to a value judgment. The mean of the differences to terrain parameters cannot be used in the wine contest, but means also provide information about where the candidate DEMs are low or high, too steep or too flat, and too rough or too smooth. SRTM, NASADEM, and ASTER should be retired, and users should choose among COP, ALOS, or FABDEM, all of which are very similar to the reference DTM.

## V. Acknowledgments

We thank the students serving as wine experts for this work, and our colleagues in DEMIX for many helpful discussions.

## References

[1] Guth, P.L.; Van Niekerk, A.; Grohmann, C.H.; Muller, J.-P.; Hawker, L.; Florinsky, I.V.; Gesch, D.; Reuter, H.I.; Herrera-Cruz, V.; Riazanoff, S.; López-Vázquez, C.; Carabajal, C.C.; Albinet, C.; Strobl, P. Digital Elevation Models: Terminology and Definitions. Remote Sens. 2021, 13, 3581. https://doi.org/10.3390/rs13183581out

[2] Strobl, P.A.; Bielski, C.; Guth, P.L.; Grohmann, C.H.; Muller, J.P.; López-Vázquez, C.; Gesch, D.B.; Amatulli, G.; Riazanoff, S.; Carabajal, C. The Digital Elevation Model Intercomparison eXperiment DEMIX, a community based approach at global DEM benchmarking. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2021, XLIII-B4-2021, 395–400. https://doi.org/10.5194/isprs-archives-XLIII-B4-2021-395-2021

[3] Bielski, C.; López-Vázquez, C.; Guth. P.L.; Grohmann, C.H. and the TMSG DEMIX Working Group, 2023. DEMIX Wine Contest Method Ranks ALOS AW3D30, COPDEM, and FABDEM as Top 1" Global DEMs: https://arxiv.org/pdf/2302.08425.pdf

[4] Guth, Peter L. (2022). DEMIX GIS Database (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7402618 Version 2 will soon be out.

[5] Guth, Peter L., Peter Strobl, Kevin Gross, & Serge Riazanoff. (2023). DEMIX 10k Tile Data Set (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7504791

[6] Google, 2023, Get insights quickly, with Google Forms: https://www.google.com/forms/about/ Accessed 12 June 2023.

[7] Grohmann, C.H., 2023. DEMIX Wine Contest Jupyter Notebook. URL: https://github.com/CarlosGrohmann/DEMIX_wine_contest last access 12 June 2023:

[8] Grohmann, C.H., . (2023). The DEMIX Wine Contest Jupyter notebook. Geomorphometry2023, Iasi, Romania. https://doi.org/10.5281/zenodo.77792562023-02-12 .

[9] Guth, P.L., 2010, Geomorphometric Comparison of ASTER GDEM and SRTM: ASPRS/CaGIS 2010 Fall Specialty Conference, Orlando, FL, November 15-19, 10 p. http://www.asprs.org/a/publications/proceedings/orlando2010/files/Guth.pdf

[10] 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, 212C:121-133

[11] Alganci, U.; Besol, B.; Sertel, E. Accuracy Assessment of Different Digital Surface Models. ISPRS Int. J. Geo-Inf. 2018, 7, 114. https://doi.org/10.3390/ijgi7030114

[12] Guth, P.L., 2006, Geomorphometry from SRTM: Comparison to NED: Photogrammetric Engineering & Remote Sensing, special issue based on Shuttle Radar Topography Mission—Data Validation and Applications Workshop, Reston, VA, 14 June 2005, vol.72, no.3, p.269-277.

[13] Grohmann, C.H., & Steiner, S.S. (2008) SRTM resample with short distance ‐ low nugget kriging, International Journal of Geographical Information Science, 22:8, 895-906, DOI: 10.1080/13658810701730152