

NepalDTM

A guide to the national 8 m bare-earth terrain model

Achyut Paudel
achyut.pdl@gmail.com

October 2026

Abstract

This note explains the NepalDTM tiles: what they contain, how they were made, which numbers were measured, and how to rebuild them. The paper (Paudel, 2026) is the citable account. NepalDTM is a bare-earth elevation model on an 8 m grid. The High Mountain Asia stereo mosaic is the surface where it exists. Copernicus GLO-30 fills the gaps. A gradient-boosting model trained on ICESat-2 and GEDI predicts the height of vegetation and buildings and subtracts it. On 4,463,463 held-out ICESat-2 points inside Nepal the error is 6.10 m RMSE. The stored 8 m spacing is true terrain detail on 37% of the delivered area.

1 What the tiles contain

Each of the 324 tiles is 24 km by 24 km, $3,000 \times 3,000$ pixels, on UTM zone 45N (EPSG:32645). Heights are metres above the EGM2008 geoid (Pavlis et al., 2012). Three files are written for every tile:

- `*_dem.tif`, the surface, including trees, buildings, and snow;
- `*_dtm.tif`, the bare-earth NepalDTM, equal to the surface minus a predicted above-ground height;
- `*_aux.tif`, the source of the surface (1 = HMA 8 m mosaic, 2 = Copernicus fill) and the predicted absolute error.

The grid inside the national outline used for tiling covers 148,380 km². The code also writes three VRT mosaics, `nepal_dem_8m.vrt`, `nepal_dtm_8m.vrt`, and `nepal_aux_8m.vrt`. Those VRTs point at a local folder named `out/`. They find the tiles only when the GeoTIFFs sit in that folder next to the VRT.

2 Inputs

The mosaic is Shean (Shean, 2017). Copernicus GLO-30 is the European Space Agency product (European Space Agency, 2021). ICESat-2 is the ATL08 ground height (Markus et al., 2017; Neuenschwander and Pitts, 2019). GEDI is the L2A lowest mode (Dubayah et al., 2020). Canopy height is Tolan et al. (Tolan et al., 2024), and land cover is WorldCover 2021 (Zanaga et al., 2022). FABDEM (Hawker et al., 2022) and GEDTM30 (Ho et al., 2025) are comparison surfaces only. The Khumbu GPS points are an independent check (Quincey et al., 2016).

Table 1: Datasets used to build and to test NepalDTM.

Dataset	Spacing	Use
HMA DEM 8 m mosaics	8 m	Stereo surface where the mosaic is finite
Copernicus GLO-30	30 m	Fill, and a predictor
ICESat-2 ATL08	along track	Ground height for training and the head-line test
GEDI L2A	25 m footprint	Ground under canopy, at half weight
Canopy height	1 m	Mean, 75th percentile, and maximum in the 8 m cell
WorldCover 2021	10 m	Majority land-cover class
FABDEM V1-2	30 m	Comparison only
GEDTM30 v1.2	30 m	Comparison only
Khumbu dGPS	points	Independent check, not used in training

ICESat-2 points are kept when the height is valid, the segment is snow-free, the beam is strong (spots 1, 3, and 5), and `msw_flag` is at most 1. A cut on `h_te_uncertainty` is not applied. In steep terrain that uncertainty grows with slope and would drop the mountain points the test needs. GEDI points are lowest-mode elevations from SlideRule with the quality and degrade filters on, from a power beam, with sensitivity above 0.95 and at most 1. The retained set is 6,230,548 ICESat-2 points and 7,035,904 GEDI points, on 324 tiles, from 2,259 pass groups and 94 spatial blocks.

3 How the correction is fit

The mosaic heights arrive on the WGS84 ellipsoid. An EGM2008 offset is added before any comparison. Copernicus GLO-30 is already an orthometric surface. On a tile, mosaic pixels whose difference from Copernicus is more than 40 m from the median difference are removed. If more than a thousand mosaic pixels remain, Copernicus fills the holes. Otherwise the whole tile surface is Copernicus. The source flag is 1 where the mosaic remains and 2 where Copernicus is used.

The model predicts surface minus ground. Predictors are the three canopy summaries, the land-cover class, slope, the sine and cosine of aspect, terrain ruggedness, topographic position over 13 and 63 pixels (about 100 m and 500 m), surface minus Copernicus, the source flag, and the surface height. The learner is scikit-learn’s `HistGradientBoostingRegressor` (Pedregosa et al., 2011): 500 iterations, learning rate 0.05, at most 63 leaves, at least 40 samples per leaf, L2 regularisation 1, absolute-error loss, and the land-cover class treated as categorical. GEDI points have weight 0.5 and ICESat-2 points have weight 1. Each fit draws at most 3,000,000 points. The bare-earth height is the surface minus that prediction.

Two held-out tests are run before the final fit. One holds out whole satellite passes. The other holds out whole 48 km blocks, five folds. The block test is the one reported as the national result, because a pass can still cross a region the model has seen. The model written into the tiles is fit after those tests, again on at most 3,000,000 points. A second model, fit to ICESat-2 only, predicts the absolute out-of-fold error and is stored on the error tile.

4 What the tests count

The national number is not the error of the single final model rescored on the same points it was fit to. It is the spatial out-of-fold residual. The comparison with Copernicus, GEDTM30, and

FABDEM keeps a point only when every model has a value there, and the national and province rows keep only points inside the national outline. That set has 4,463,463 points.

Elevation-zone, forest, slope, and mosaic-versus-fill rows use the 6,227,829 points where all four models have a value, including points in the 1 km tile buffer outside the outline. Their counts sum to 6,227,829, not to 4,463,463. The spatial out-of-fold RMSE on all 6,230,548 ICESat-2 points, with no common-support cut, is 5.49 m. That number is not comparable with GEDTM30 or FABDEM. GEDTM30 is itself fit using ICESat-2 (Ho et al., 2025), so its column in these tables is an optimistic one.

5 Measured error

Table 2: Held-out ICESat-2 error inside Nepal, in metres. $n = 4,463,463$.

Model	RMSE	NMAD	Bias
NepalDTM, 48 km blocks	6.10	2.61	−0.15
NepalDTM, satellite passes	6.01	2.54	−0.13
Copernicus GLO-30	9.07	3.51	+3.87
GEDTM30 v1.2	8.47	3.96	+1.80
FABDEM V1-2	7.78	3.14	+1.90

Table 3: Spatial-holdout RMSE in metres by surface elevation, on the 6,227,829-point common set.

Zone	n	NepalDTM	Copernicus	GEDTM30	FABDEM
Below 300 m	1,251,581	1.96	4.53	2.12	2.46
300–1,000 m	487,932	5.18	8.53	5.87	6.16
1–2.5 km	1,076,095	6.24	8.69	7.57	7.33
2.5–4 km	776,214	8.27	12.88	10.45	10.27
Above 4 km	2,636,007	5.33	7.55	8.98	7.50

Table 4: Spatial-holdout RMSE in metres by province, on the 4,463,463 points inside Nepal.

Province	n	NepalDTM	Copernicus	GEDTM30	FABDEM
Bagmati	423,665	5.92	9.64	8.86	8.23
Gandaki	745,860	5.99	8.63	9.14	7.99
Karnali	1,325,510	6.86	9.88	9.39	8.55
Koshi	640,238	6.30	9.42	9.39	8.55
Lumbini	547,250	5.13	7.57	5.76	5.74
Madhesh	234,163	1.78	3.88	1.98	2.04
Sudurpashchim	546,777	6.28	9.67	7.68	7.54

On the common set, forest RMSE is 6.82 m ($n = 1,616,752$) against 11.20 m for Copernicus. Off forest it is 4.94 m ($n = 4,611,077$) against 6.89 m. Below 20 degrees of slope it is 3.26 m; from 20 to 40 degrees, 6.07 m; at 40 degrees and above, 10.81 m ($n = 516,657$). Where the stacked surface is the 8 m mosaic, RMSE is 4.00 m ($n = 3,199,580$). Where it is Copernicus fill, RMSE is 6.71 m ($n = 3,028,249$).

6 GPS, contours, and seams

The Khumbu control points were surveyed with differential GPS between October 2015 and October 2016 and were not used in training (Quincey et al., 2016). Table 5 scores the delivered surface tile and the delivered bare-earth tile separately. The surface agrees with the margins to 2.49 m RMSE. The bare-earth model is 4.85 m low on the glacier surface. Ice texture is being removed as if it were vegetation, and the survey is older than the satellite surface.

Table 5: Khumbu differential GPS minus the delivered tiles. RMSE and bias in metres.

Group	n	Surface RMSE	Surface bias	Bare-earth RMSE
Glacier margins	83	2.49	-1.08	3.89
Glacier surface	388	2.96	+0.86	6.01
Wind-tower grid	85	1.02	+0.46	6.59

Survey Department contour heights were compared with the delivered tiles at Bheemdatt and in the Bhote Koshi. At Bheemdatt the surface RMSE is 12.24 m and the bare-earth RMSE is 6.74 m ($n = 129,886$). In the Bhote Koshi both stay near 20 m (surface 21.99 m, bare earth 20.28 m, $n = 6,837,320$). A contour line is not a point survey, and this comparison does not rank NepalDTM against Copernicus. The files in `validation/` record the contour scores that were written.

The median absolute step from one tile to the next is 2.08 m. The median step between adjacent pixels inside a tile is 2.06 m. A search for a horizontal shift against ICESat-2, in each province, returns 0 m east and 0 m north at the best position. The predicted absolute error on the error tile has median 4.3 m and 90th percentile 8.8 m.

7 How to read the 8 m grid

Use the bare-earth tile for the ground, the surface tile for the top of trees and buildings, and the error tile as a per-cell warning rather than as a second elevation. The source band says which base surface the cell came from. A cell with source 2 can show a smooth 30 m hill even though the file is stored every 8 m. The correction changes the height. It does not invent valley or ridge detail the base surface never contained.

8 How to obtain and rebuild it

The Cloud Optimized GeoTIFFs are at <https://pub-7a51429102ce456f85c0aee664338f8f.r2.dev/cog/>. The browser map reads <https://pub-7a51429102ce456f85c0aee664338f8f.r2.dev/NepalDTM.pmtiles>. The repository is <https://github.com/achyut/NepalDTM>.

```
uv venv --python 3.12 .venv
uv pip install --python .venv/bin/python -r code/requirements-nw.txt
cd code
../.venv/bin/python -I run.py download
../.venv/bin/python -I run.py fuse
../.venv/bin/python -I run.py validate
```

`python run.py` all runs those three groups in order. `download` checks the EGM2008 grid and `~/earthdata_token`, builds the tile grid, downloads ICESat-2 and GEDI, and stacks the HMA mosaic, Copernicus, the canopy map, and WorldCover. `fuse` trains the correction and writes the tiles. `validate` writes `validation_report.json`. Cells and tiles already on disk are skipped. A national run is tens of gigabytes. The Earthdata token is not part of the repository.

9 Limits

The model adjusts for canopy. Terrain shape still comes from the base surface. Steep slopes and glacier ice are the weak places, and the tables above are the measurements of that weakness. Public airborne lidar was not an input to this build. A later lidar grid can replace tiles where it exists. The same tile layout and the same `run.py` steps are the path for that replacement. It is not a measured accuracy for a model that has not been built.

References

- Dubayah, R., et al. (2020). The Global Ecosystem Dynamics Investigation. *Science of Remote Sensing*, 1, 100002. <https://doi.org/10.1016/j.srs.2020.100002>
- European Space Agency (2021). Copernicus DEM. <https://doi.org/10.5270/ESA-c5d3d65>
- Hawker, L., et al. (2022). A 30 m global map of elevation with forests and buildings removed. *Environmental Research Letters*, 17, 024016. <https://doi.org/10.1088/1748-9326/ac4d4f>
- Ho, Y.-F., et al. (2025). GEDTM30. *PeerJ*, 13, e19673. <https://doi.org/10.7717/peerj.19673>
- Markus, T., et al. (2017). The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2). *Remote Sensing of Environment*, 190, 260–273. <https://doi.org/10.1016/j.rse.2016.12.029>
- Neuenschwander, A., and Pitts, K. (2019). The ATL08 land and vegetation product for the ICESat-2 Mission. *Remote Sensing of Environment*, 221, 247–259. <https://doi.org/10.1016/j.rse.2018.11.005>
- Paudel, A. (2026). NepalDTM: A national 8-metre bare-earth digital terrain model of Nepal from multi-source satellite data fusion. <https://github.com/achyut/NepalDTM>
- Pavlis, N. K., Holmes, S. A., Kenyon, S. C., and Factor, J. K. (2012). The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research: Solid Earth*, 117, B04406. <https://doi.org/10.1029/2011JB008916>
- Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Quincey, D., King, O., and Watson, C. S. (2016). Ground control points for investigating mass loss processes on Khumbu Glacier, Nepal. British Geological Survey, NGDC accession 75529. <https://webapps.bgs.ac.uk/services/ngdc/accessions/index.html#item75529>
- Shean, D. (2017). High Mountain Asia 8-meter DEM Mosaics Derived from Optical Imagery, Version 1. NASA NSIDC DAAC. <https://doi.org/10.5067/KX0VQ9L172S2>

- Tolan, J., et al. (2024). Very high resolution canopy height maps from RGB imagery. *Remote Sensing of Environment*, 300, 113888. <https://doi.org/10.1016/j.rse.2023.113888>
- Zanaga, D., et al. (2022). ESA WorldCover 10 m 2021 v200. <https://doi.org/10.5281/zenodo.7254221>