Supplementary Materials for Regional 'Bare-Earth' Digital Terrain Model for Costa Rica Based on NASADEM Corrected for Vegetation Bias

Olga Pimenova, Craig Roberts and Chris Rizos

S1 Comparison of the validation datasets

To evaluate how the two proposed validation methods for DEMs compare, first, Laser Vegetation Imaging Sensor (LVIS) bare-earth ZG elevations were compared to the points from the levelling dataset. LVIS footprint is about 20 m in diameter. Therefore, we set several thresholds (20 m, as well as 5, 10 and 30 m) to search for point pairs between levelling and LVIS data. For comparison in terms of LVIS ellipsoidal heights, elevations of the levelling points were transformed from orthometric heights to ellipsoidal heights by the addition of EGM2008 geoid undulations computed with the spherical harmonics synthesis software Geopot7 and WGS84 ellipsoid parameters. The results are listed in Table S1. For the 5 m threshold for distances, only 30 point pairs were identified. For these, the mean height difference is 0.1 m, STD of 3.6 m and the difference range spans 17.3 m. For higher thresholds, the mean difference is about 0.7 m and STD of around 10 m. Furthermore, a point with a significant difference of 65.1 m appears to be an outlier, most probably in the levelling dataset. Both methods have uncertainties from different error sources. Nevertheless, each of them also provides different advantages: the levelling dataset is relatively sparse but has much better coverage. While LVIS data provides millions of points, but its spatial distribution is limited, covering only some parts of the study area.

Table S1: Statistics of the differences between LVIS bare-ground ZG elevations and ellipsoidal heights of levelling stations.

Distance threshold	**# Pairs**	**mean**	**STD**	**min**	**max**	**range**
5 m	30	0.1	3.7	-9.5	7.8	17.3
10 m	114	0.7	9.5	-30.3	64.1	95.4
20 m	168	0.7	10.4	-47.9	65.1	113.0
30 m	172	0.7	10.3	-47.9	65.1	113.0

S2 Validation of GFCH2019 model with LVIS datasets

For the comparison of the Global Forest Canopy Height 2019 (GFCH2019) model to LVIS2019 data, we evaluated several options for LVIS canopy heights. We expect that the elevation of the highest detected signal within the echo return, ZT, is not comparable to the GFCH2019 canopy heights. The other parameter ZH, the mean elevation of the highest detected mode within the waveform, is more appropriate for comparison. Furthermore, comparison to relative heights (RH) at which a certain percentage of the waveform energy occurs is also done. From Figure S1 it can be observed that the RH90 values have the best coincidence with GFCH2019, followed by RH85 and ZH-ZG values. However, for comparison of the prepared Canopy Height Model 2000 (CHM2000), none of these values is provided in LVIS1998 and LVIS2005 data products. We confirmed that GFCH2019 coincides primarily with the 90th percentile of LVIS-based canopy height (RH90).

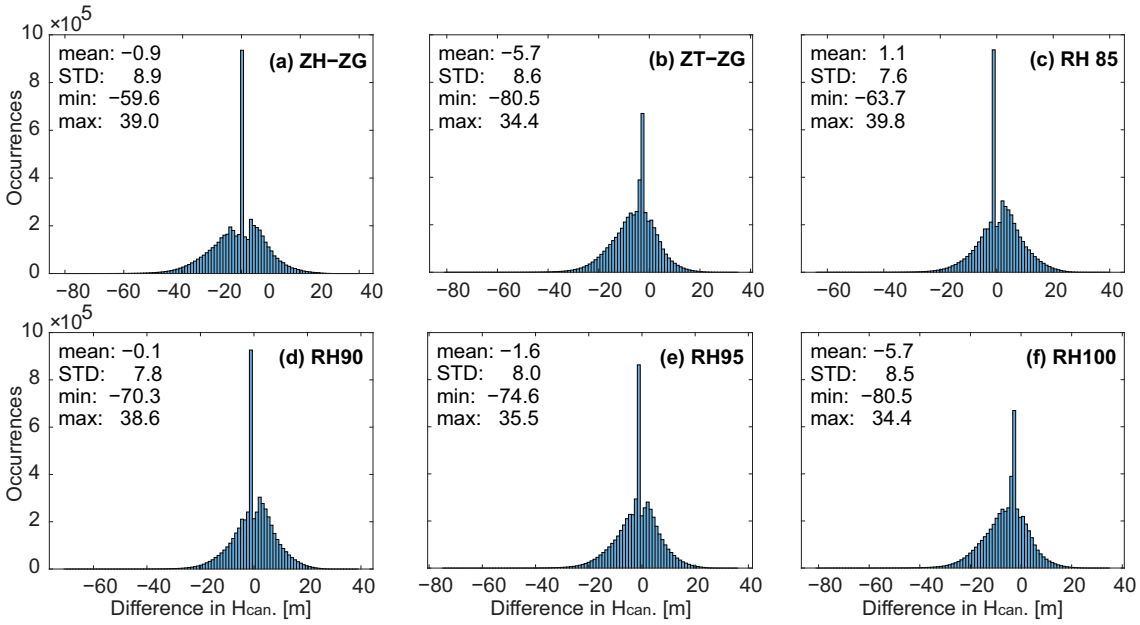

Figure S1: Histograms and statistics of differences in canopy heights between Global Forest Canopy Height 2019 model and Laser Vegetation Imaging Sensor (LVIS) canopy heights estimated as (**a**) ZH-ZG. (**b**) ZT-ZG. (**c**) RH85. (**d**) RH90. (**e**) RH95. (**f**) RH100. The statistics is provided in terms of the mean difference, Standard Deviation (STD), and minimum (min) and maximum (max) differences.

S3 Canopy change computed for the extended area

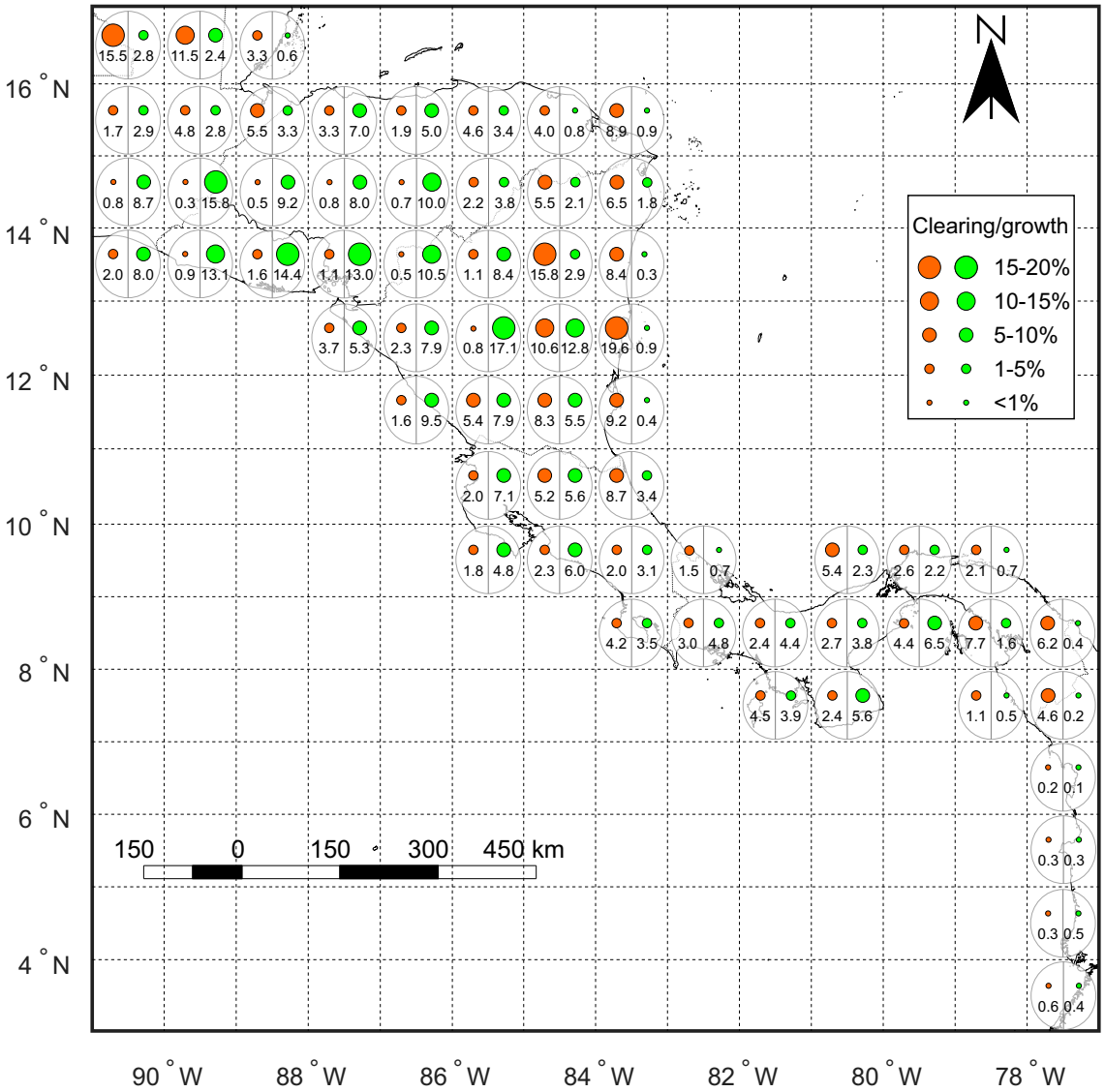

Figure S2: Tile-wise percentage of growth and clearing cases for an area around Costa Rica between latitudes 3°N and 17°N and longitudes 91°W and 77°W. The threshold for the GFCH2019 canopy heights to define 'growth' cases was set to 5 m. Therefore, the vegetation with heights between 3 and 5 m high were not considered. The percentages were computed as the number of cells identified as clearing/growth in each tile, divided by the total number of terrestrial cells in the tile.

S4 Evaluation of ten DEMs with levelling dataset

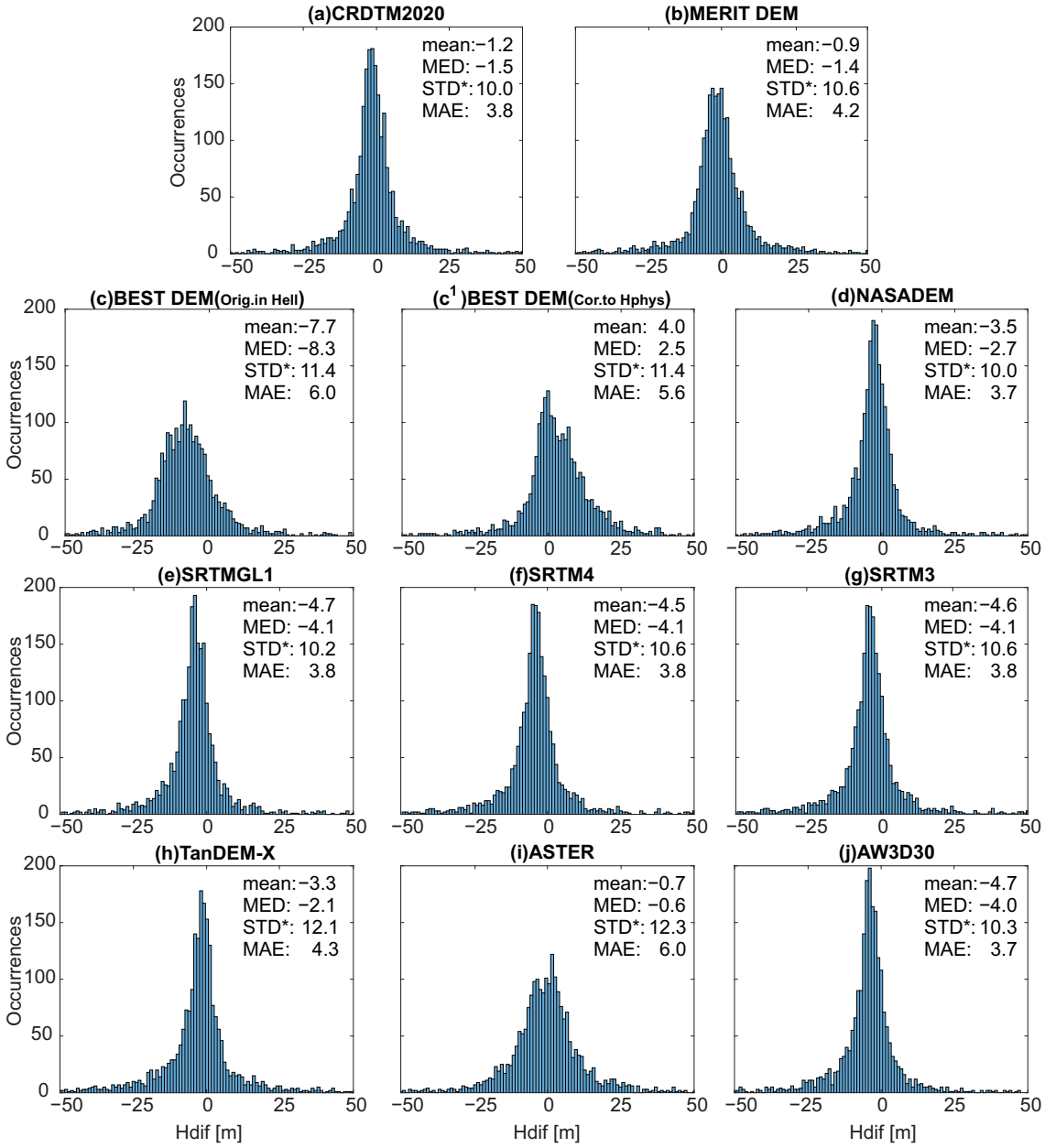

Figure S3: Histograms of height differences (H_{dif}) between elevations of levelling points and ten DEMs, all referred to the the same height reference surface, the EGM96 geoid. For visual interpretation, histograms (and STD*) were computed for height differences within a chosen threshold of ± 50 m and presented with the same vertical axis. Figure (c) presents results for BEST DEM evaluation with its original elevations, which we believe to be ellipsoidal heights for all cells corrected for VB. Figure (c^1) presents results for BEST DEM evaluation after all elevations were corrected to physical heights above EGM96 geoid. As can be observed, the bi-modal distribution, in this case, is separated for bare-earth points whose peak is around zero and the second peak on the right-hand side for points in vegetated areas. The statistics is provided in terms of the mean difference, Median Difference (MED), Standard Deviation (STD), and Median Absolute Difference (MAD).

S5 Vegetation bias computed for the extended area

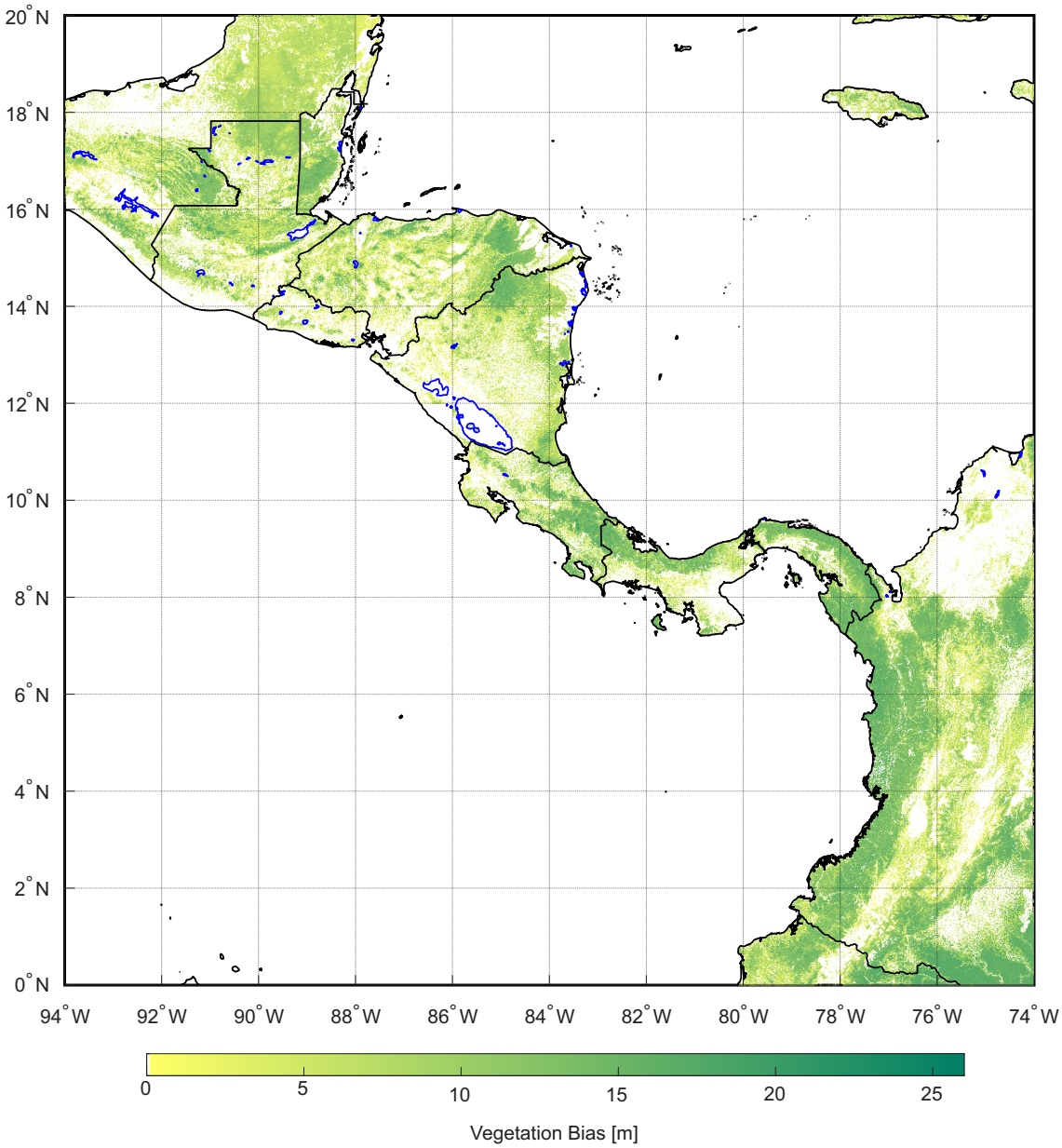

Figure S4: Vegetation Bias that was eliminated from NASADEM is computation of CRDTM2020plus.

S6 CRDTM2020plus visualisation

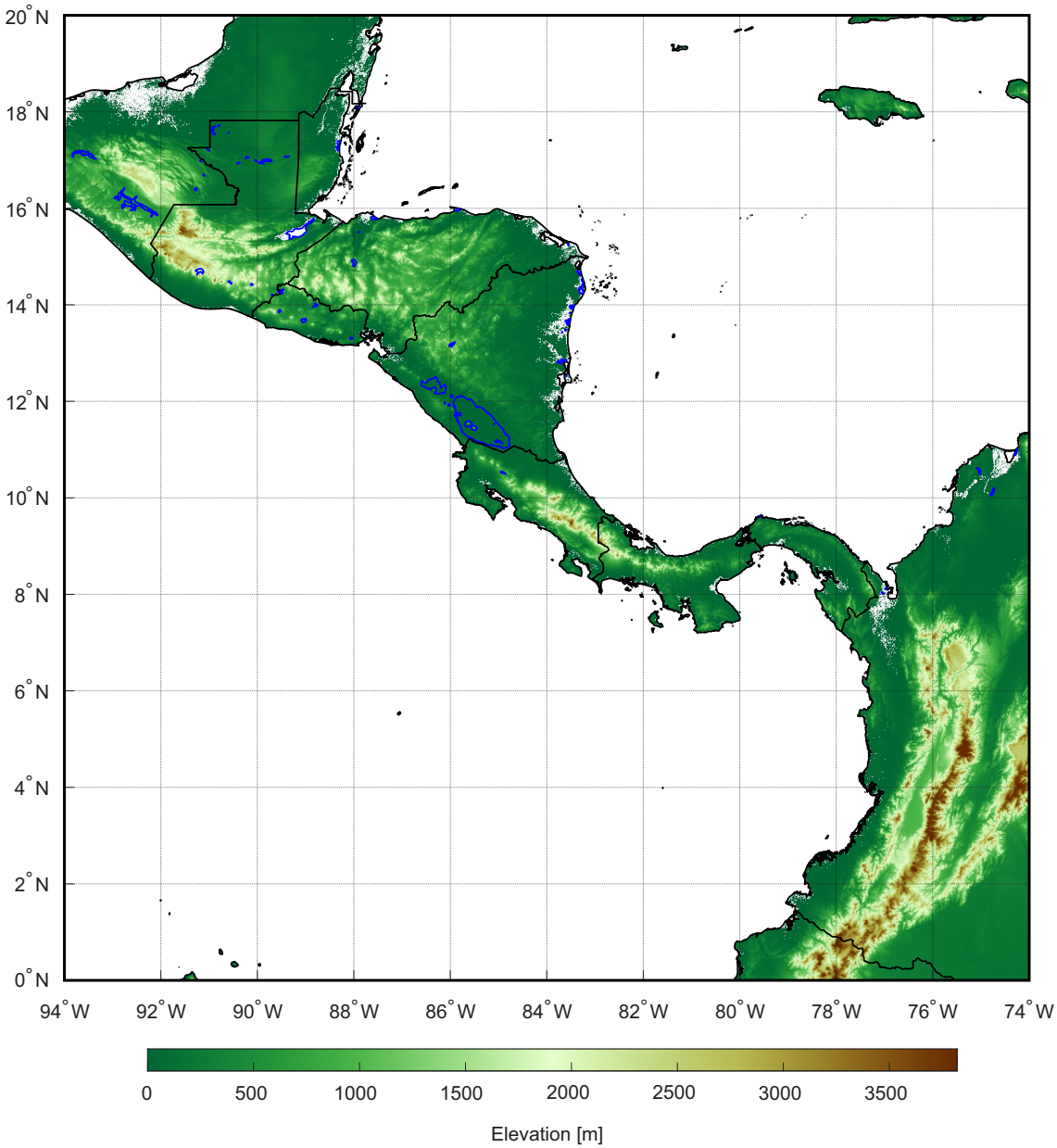

Figure S5: CRDTM2020plus elevations with respect to EGM2008 geoid model.