ICA
ACI
International Cartographic Association
Association Cartographique Internationale

Enhancing the vertical accuracy of Copernicus digital elevation model using tree-based machine learning models

Chukwuma Okolie[a,b,c]*, Jon Mills[c], Adedayo Adeleke[d], Julian Smit[e]

[a]Division of Geomatics, University of Cape Town, South Africa, [b]Department of Surveying & Geoinformatics, University of Lagos, Nigeria, [c]School of Engineering, Newcastle University, United Kingdom, [d]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa; [e]AfriMap Geo-Information Services, Cape Town, South Africa
c.j.okolie2@newcastle.ac.uk

* Corresponding author

Keywords: Copernicus, digital elevation model, Random forest, AdaBoost, XGBoost, LightGBM, CatBoost.

Abstract:

The mapping, visualisation and representation of the Earth's topography is an important problem confronting cartographers, and open-access global Digital Elevation Models (DEMs) are increasingly being used for these purposes. Moreover, the limited availability of high-resolution DEMs means that open-access DEMs are relevant for updating topographic maps, particularly in data-sparse regions. Nonetheless, such open-access DEMs are known to suffer from vertical accuracy defects caused by a myriad of problems. Hence, techniques for enhancing or improving their vertical accuracies are an active area of research (e.g., Bagheri et al., 2018). The aim of this study is to assess the capability of tree-based supervised machine learning algorithms for enhancing the vertical accuracy of the 30m Copernicus DEM.

Five models are compared with a 30m Copernicus-derived DEM in Cape Town, South Africa. The models used were Random forest (RF), AdaBoost (Adaptive boosting), XGBoost (Extreme gradient boosting), LightGBM (Light gradient boosting machine), and CatBoost (Categorical boosting). Initially, the Copernicus DEM was compared to a reference 2m aerial LiDAR dataset, and height errors (Δh) were derived. Subsequently, the models were trained at five different sites using Copernicus-derived terrain parameters (elevation, slope, aspect, surface roughness, topographic position index, terrain ruggedness index, terrain surface texture, and vector ruggedness measure) and land cover parameters (urban footprints, percentage tree cover, and percentage bare ground cover) to predict the height errors. Prior to the extraction of training (80%) and test (20%) samples, all datasets were pre-processed through datum harmonisation, co-registration and resampling to ensure coincidence in the pixel-to-pixel analysis. The extraction of parameters was performed within a GIS environment. Subsequently, the training and test samples were converted to csv format and imported into Google Colab where the models were implemented using python scripting and other open-source packages with their baseline default parameters (i.e., without hyperparameter tuning). At the test sites, analysis of the predicted versus actual height errors showed a reasonable predictive capability of Random Forest, XGBoost, LightGBM and CatBoost, except for AdaBoost which performed below expectation at several sites. Figure 1 shows the coefficients of determination (R^2) between the predicted and actual height errors at the cultivated sites.

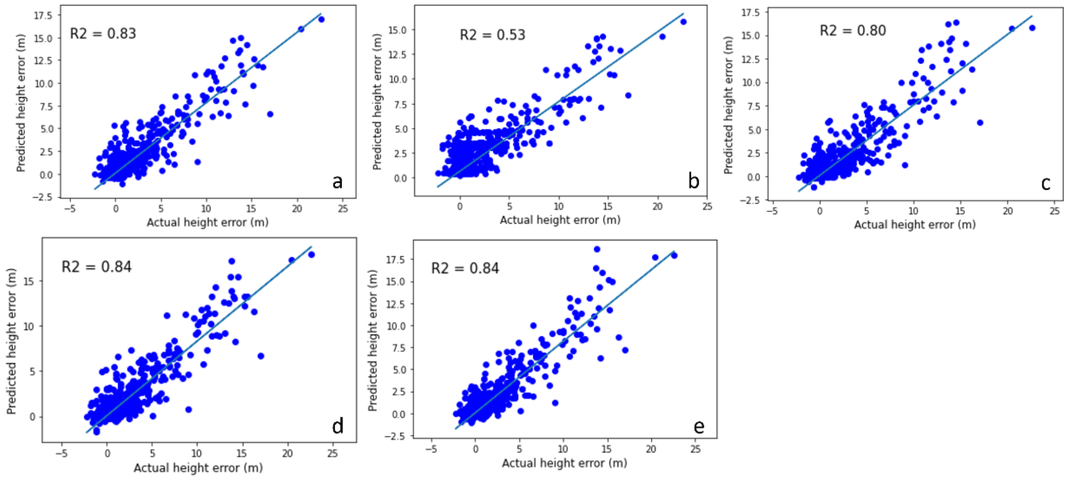

Figure 1. R-square of predicted versus actual height errors at the cultivated sites, (a) Random Forest (b) AdaBoost (c) XGBoost (d) LightGBM (e) CatBoost.

Abstracts of the International Cartographic Association, 6, 193, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa.
https://doi.org/10.5194/ica-abs-6-193-2023

Thereafter, the trained models were applied for predicting height errors and deriving corrected DEMs (i.e., $DEM_{Corrected} = DEM_{Original} - \Delta h$) at five independent sites with similar terrain characteristics. The accuracy measures are presented in Table 1.

Land type	Mean absolute error (m)						Root mean square error (m)					
	Original DEM	Corrected DEM					Original DEM	Corrected DEM				
		RF	Ada Boost	XG Boost	Light GBM	Cat Boost		RF	Ada Boost	XG Boost	Light GBM	Cat Boost
Urban/ Industrial	1.09	**0.60**	0.66	0.63	0.70	0.70	1.48	**0.86**	0.96	0.88	0.95	0.97
Cultivated Field	0.42	**0.35**	0.59	0.36	0.36	0.38	1.23	0.79	1.06	**0.78**	0.81	0.86
Mountain	4.85	5.01	12.28	**4.82**	5.08	5.10	10.24	9.20	17.43	**9.05**	9.61	9.34
Peninsula	0.35	0.34	0.42	0.34	0.36	**0.31**	0.52	0.53	0.63	0.54	0.62	**0.49**
Grassland/ Shrubland	0.50	0.29	0.26	0.28	**0.24**	0.29	0.56	0.46	0.37	0.41	**0.36**	0.42

Table 1. Comparative analysis of Copernicus DEM after correction

Figure 2 shows visualisations of the corrected DEMs at an urban/industrial site. There is a remarkable reduction in the vertical errors achieved by the tree-based models. For example, the MAE of the original DEM in the grassland/shrubland area was improved by 53% with LightGBM. Similarly, the RMSE and MAE in the urban/industrial area improved by 42% and 45% respectively using Random Forest. The MAE improvement by XGBoost was 42% (urban/industrial) and 15% (cultivated fields). The models' performance varies based on the environment, e.g., the highest reduction in RMSE was by XGBoost in the Table Mountain, LightGBM in grassland/shrubland and CatBoost in the Cape Peninsula areas.

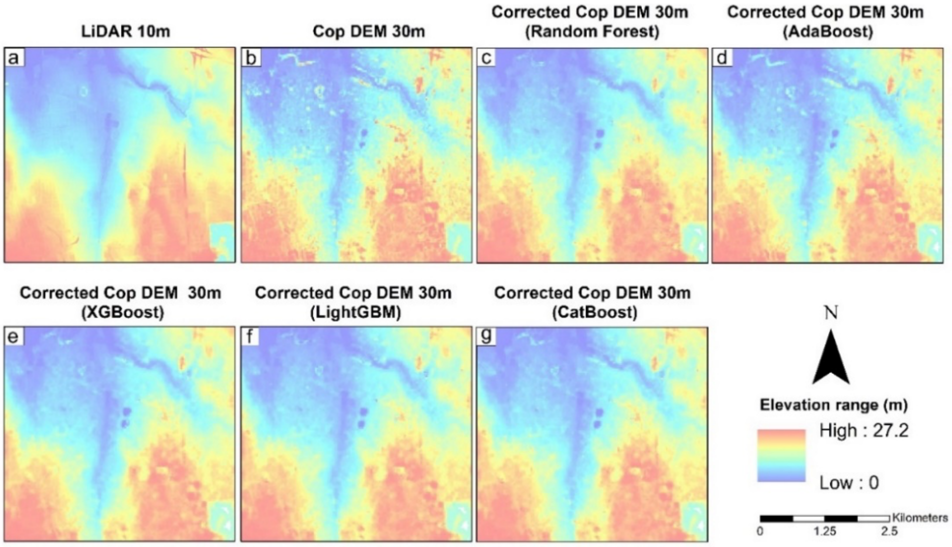
LiDAR 10m
Cop DEM 30m
Corrected Cop DEM 30m (Random Forest)
Corrected Cop DEM 30m (AdaBoost)
a
b
c
d
Corrected Cop DEM 30m (XGBoost)
Corrected Cop DEM 30m (LightGBM)
Corrected Cop DEM 30m (CatBoost)
e
f
g
N
Elevation range (m)
High : 27.2
Low : 0
0 1.25 2.5 Kilometers

Figure 2. Visual comparison of the corrected Copernicus DEM in the Cape Town urban/industrial area

Visually, the corrected DEMs from Random Forest, XGBoost, LightGBM and CatBoost show several topographic improvements such as smoothing of rough edges, enhanced stream channel conditioning and diminution of coarse/grainy pixels. This shows the capability of tree-based ML models for improving the vertical accuracy of coarse resolution DEMs, with further gains in topographic mapping. This low-cost approach can be adopted by national mapping organisations with budgetary constraints to enhance wide-area DEMs for producing more accurate topographic maps and cartographic products. Further work to improve the accuracy gains is ongoing.

Acknowledgements

The Commonwealth Scholarship Commission UK, the University of Cape Town Postgraduate Funding Office, and the Information and Knowledge Management Department, City of Cape Town for supporting this research. Special thanks to Ikechukwu Maduako, Hossein Bagheri, Tom Komar, Shidong Wang, and Maria Peppa for their helpful comments.

References

Bagheri, Hossein, Schmitt, M., & Zhu, X. X. (2018). Fusion of TanDEM-X and Cartosat-1 elevation data supported by neural network-predicted weight maps. *ISPRS Journal of Photogrammetry and Remote Sensing*, *144*, 285–297.