DIGITAL ELEVATION MODEL CORRECTION IN URBAN AREAS USING EXTREME GRADIENT BOOSTING, LAND COVER AND TERRAIN PARAMETERS

Chukwuma Okolie[1,2,3*], Jon Mills[3], Adedayo Adeleke[4], Julian Smit[5]

[1]Division of Geomatics, University of Cape Town, South Africa; oklchu002@myuct.ac.za
[2]Department of Surveying & Geoinformatics, University of Lagos, Nigeria; cjohnokolie@gmail.com
[3]School of Engineering, Newcastle University, United Kingdom; jon.mills@newcastle.ac.uk
[4]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa; adedayo.adeleke@up.ac.za
[5]Cape Peninsula University of Technology, South Africa; drjlsmit@gmail.com

KEY WORDS: Data fusion, Digital elevation model, Copernicus, ALOS World 3D, Extreme gradient boosting, Bayesian optimisation, Terrain parameters, Urban footprints.

ABSTRACT:

The accuracy of digital elevation models (DEMs) in urban areas is influenced by numerous factors including land cover and terrain irregularities. Moreover, building artefacts in global DEMs cause artificial blocking of surface flow pathways. This compromises their quality and adequacy for hydrological and environmental modelling in urban landscapes where precise and accurate terrain information is needed. In this study, the extreme gradient boosting (XGBoost) ensemble algorithm is adopted for enhancing the accuracy of two medium-resolution 30-metre DEMs over Cape Town, South Africa: Copernicus GLO-30 and ALOS World 3D (AW3D). XGBoost is a scalable, portable and versatile gradient boosting library that can solve many environmental modelling problems. The training datasets are comprised of eleven predictor variables including elevation, urban footprints, slope, aspect, surface roughness, topographic position index, terrain ruggedness index, terrain surface texture, vector roughness measure, forest cover and bare ground cover. The target variable (elevation error) was calculated with respect to highly accurate airborne LiDAR. After training and testing, the model was applied for correcting the DEMs at two implementation sites. The corrections achieved significant accuracy gains which are competitive with other proposed methods. There was a 46 – 53% reduction in the root mean square error (RMSE) of Copernicus DEM, and a 72 - 73% reduction in the RMSE of AW3D DEM. These results showcase the potential of gradient-boosted decision trees for enhancing the quality of global DEMs, especially in urban areas.

1. INTRODUCTION

For quantitative assessments of environmental processes and hazards in urban areas, one of the critical requirements are reliable digital elevation models (DEMs). Important urban applications of DEMs include change detection, urban monitoring (Sirmacek et al. 2010), site selection and suitability analysis, and flood simulation and modelling. However, DEMs are known to suffer accuracy defects in urban areas due to sensor distortions, source data attributes, and errors inherent in the DEM production methods. Moreover, building artefacts in global DEMs cause artificial blocking of surface flow pathways in hydrological modelling (Liu et al., 2021). These errors compromise their quality and adequacy for hydrological and environmental applications (e.g., flood and watershed modelling) where precise and accurate terrain information is needed. High/very high-resolution DEMs e.g., from airborne light detection and ranging (LiDAR) are often prohibitively expensive at the city scale. Several space-borne LiDAR missions have been launched in recent years (e.g., ICESat, ICESat-2, GEDI), but do not provide wall-to-wall elevation coverage. Consequently, medium-resolution synthetic aperture radar (SAR) or photogrammetric global DEMs are a viable option, especially in data-sparse regions.

Several methods including machine learning have been proposed to improve DEM accuracy in urban areas (e.g., Liu et al. 2021; Olajubu et al. 2021; Xu et al. 2021; Hawker et al. 2022). For example, Kim et al. (2020) integrated Sentinel-2 multispectral imagery with an artificial neural network (ANN) for improving the accuracy of 30 m shuttle radar topography mission (SRTM) DEM in dense urban cities. Similarly, Liu et al. (2021) adopted the random forest model for the correction of building artefacts in the MERIT DEM using publicly available datasets such as global population density, satellite night-time lights and OpenStreetMap buildings. Hawker et al. (2022) applied random forest for the removal of forests and building offsets from the 30 m Copernicus DEM to produce a globally corrected DEM product referred to as FABDEM.

Tree-based ensemble machine learning algorithms have become prominent in the machine learning community, and they have proven to be very reliable. Decision trees enable researchers to gain a straightforward understanding of the connections between objects at varying levels of detail (Miao et al., 2012). Moreover, decision trees are reputedly tolerant to multicollinearity (Climent et al 2019; Han et al., 2019; Pham and Ho, 2021). Despite the appeal of tree-based ensemble algorithms, the remote sensing community is yet to harness their full potential for DEM correction and/or enhancement. Several important terrain conditioning factors and parameters were not considered in some previous studies, and there are still some unknowns regarding the interdependence between terrain parameters and their specific contributions to machine learning predictions, when applied to DEM correction.

Among the tree-based ensembles in use, gradient tree boosting which uses decision trees as weak learners has performed excellently in numerous machine learning tasks and applications (e.g., Samat et al., 2020; Bentéjac et al., 2021; Łoś et al., 2021). In this study, the extreme gradient boosting (XGBoost) algorithm is adopted. XGBoost is a scalable end-to-end tree boosting system that is commonly used by data scientists,

* Corresponding author

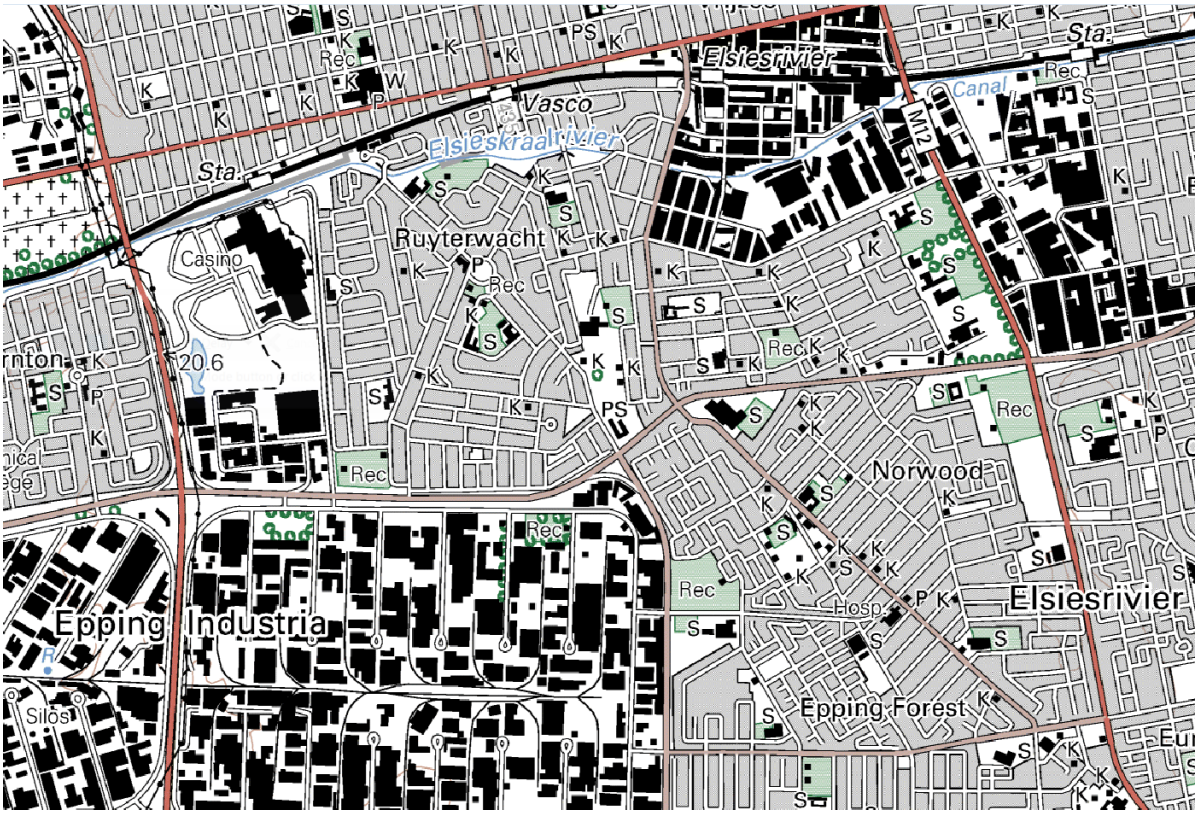

Figure 1. A typical urban/industrial setting in Cape Town (Source: 1:50,000 South Africa topographic map series, published by the Chief Directorate: National Geo-spatial Information, © 2015)

provides state-of-the-art results on many problems, and has excelled in numerous data mining and machine learning challenges (Chen and Guestrin, 2016). One of the reasons that XGBoost has been very successful is because it is scalable in different scenarios. This study aims to integrate XGBoost with terrain and land cover parameters for urban correction of Copernicus and ALOS World 3D DEMs in a section of Cape Town, South Africa.

2. METHODS

2.1 Study Area

The City of Cape Town in South Africa is a large urban area with an intense movement of people, goods and services, high population density, extensive developments, industrial areas and multiple business districts (Western Cape Government, 2022).

Figure 2. Satellite image close-up view showing buildings in the urban area of Cape Town

Cape Town is a complex and diverse city, and the second-largest city in South Africa, with a population of over four million people (Smit, 2020). The sites for this study are selected from urban/industrial districts in Cape Town (Figures 1 – 2).

2.2 Datasets

2.2.1 Digital elevation models: Two 30 m global digital elevation models (DEMs) are considered in this study, the Copernicus GLO-30 DEM and ALOS World 3D DEM (AW3D30). Copernicus GLO-30 is derived from the WorldDEM data. The WorldDEM data product is based on the radar satellite data which was acquired during the TanDEM-X Mission (ESA, 2020). AW3D30 was released by the Japan Aerospace Exploration Agency (JAXA). It was generated from the earlier ALOS DEM which was produced at a spatial resolution of 5 m with an accuracy of 5 m (standard deviation) (JAXA 2017).

Copernicus GLO-30 and AW3D v3.2 were adopted in this study. The City of Cape Town (CCT) airborne LiDAR-derived DEM is used as the reference dataset. It was acquired from the Information and Knowledge Management Department of the City of Cape Town. The height accuracy of the point cloud used for generating the DEM is 0.15 m. The LiDAR DEM is spatially referenced to the Hartebeesthoek94 horizontal co-ordinate system and vertically referenced to the South Africa (SA) Land Levelling Datum (SAGEOID2010).

2.2.2 Global urban footprint: The Global Urban Footprint (GUF) is a human settlement layer that was created from the global synthetic aperture radar (SAR) dataset that was acquired during the TanDEM-X (TDM) mission. The methodology for deriving the GUF is presented in Esch et al. (2010, 2013). The

high-resolution 0.4 arc-second (~12 m) GUF 2012 product was adopted for this study.

2.2.3 Terrain parameters and land cover data: To characterise the influence of the terrain on the elevation error, the following additional input variables which are known influencers of DEM error were selected: elevation, slope, aspect, surface roughness, topographic position index (TPI), terrain ruggedness index (TRI), terrain surface texture (TST), vector ruggedness measure (VRM), percentage forest cover and percentage bare ground cover. The elevation errors or differences (ΔH) between the DEMs and reference LIDAR were calculated as follows:

$$\Delta H = H_{Global\ DEM} - H_{RefDEM} \quad (1)$$

Where,

H_{RefDEM} = elevations from LiDAR DEM.

$H_{Global\ DEM}$ = individual elevations from the global DEMs (i.e., Copernicus and AW3D)

2.3 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a "scalable end-to-end tree boosting system" (Chen and Guestrin, 2016). It "sequentially builds short and simple decision trees as each tree tries to improve the performance of the previous one" (Safaei et al., 2022).

Gradient boosting combines weak learners into strong learners in an iterative fashion (Friedman, 2001; Hastie et al., 2009; Deng et al., 2019; Can et al., 2021). The core algorithm of XGBoost is its optimisation of the objective function, $O(\theta)$, which consists of the training loss and regularisation (Deng et al., 2019; Can et al., 2021):

$$O(\theta) = L(\theta) + \Omega(\theta) \quad (2)$$

Where:
L – training loss function
Ω – regularisation term

The training loss enables the evaluation of model performance on training samples, while the regularisation term helps in addressing the model complexity. The regularisation term for a decision tree can be defined as (Deng et al., 2019):

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (3)$$

Where,
T - the number of leaves in a decision tree
w - the vector of scores on leaves
γ - the complexity of each leaf
λ - a parameter to scale the penalty

The objective function for calculating the structure score of XGBoost is derived as (Deng et al., 2019):

$$O = \sum_{j=1}^{T} \left[g_j w_j + \frac{1}{2}(h_j + \lambda) w_j^2 \right] + \gamma T \quad (4)$$

where w_j are independent with each other. The form $g_j w_j + \frac{1}{2}(h_j + \lambda) w_j^2$ is quadratic and the best w_j for a given structure is $q(x)$.

2.4 Data Preparation

To harmonise the horizontal datums, the global DEMs were transformed from the geographic to the Universal Transverse Mercator (UTM) projection in WGS84. Similarly, the LiDAR DEM was transformed from Hartebeesthoek94 to UTM WGS84. The vertical datum of the global DEMs and airborne LiDAR were harmonised to EGM2008.

Global DEM	No of points (Training, validation and testing)	No of points (Implementation site A)	No of points (Implementation site B)
Copernicus	573377	23041	22988
AW3D	572374	23041	22988

Table 1. Distribution of model training, validation, test and implementation sites data

To derive the elevation errors (ΔH), the LiDAR elevations were subtracted from the corresponding elevations of the global DEMs at specific points. The elevation values, along with the values of the elevation error and terrain/land cover parameters were extracted from the rasters to csv files, and sorted. This resulted in the final set of points used for model training, validation and testing split into 80% for training and validation, and 20% for testing. After training and testing, the models were independently evaluated at external sites referred to as model implementation sites A and B respectively. The model implementation provides an opportunity for evaluation of the prediction capability and accuracy of the trained models. Table 1 shows the data distribution.

2.5 Model Implementation

The model was trained using the elevation, urban footprints, slope, aspect, surface roughness, topographic position index, terrain ruggedness index, terrain surface texture, vector ruggedness measure, percentage forest canopy and percentage bare ground cover as input parameters, and the elevation error as the target variable (or predictand).

The training was carried out using the default hyperparameters and tuned hyperparameters. For the hyperparameter tuning, Bayesian optimisation was adopted. The theoretical background of Bayesian optimisation is already well documented in the extant literature. Essentially, it "provides a principled technique based on Bayes Theorem to direct a search of a global optimization problem that is efficient and effective" (Brownlee, 2019). The explanations of the XGBoost hyperparameters are provided in the XGBoost library documentation (XGBoost, 2022). Summarily, the adopted hyperparameters and the search space are shown in Table 2. After training and testing, the models were saved, loaded and implemented for predicting the height errors at implementation sites (A and B) with similar terrain characteristics. The predicted elevation errors were applied for deriving the corrected DEMs (i.e., $DEM_{Corrected} = DEM_{Original} - \Delta H_{Predicted}$).

2.6 Model Performance and Accuracy Assessment

Learning curves are used as a diagnostic tool for evaluating the performance of XGBoost on the training and validation datasets. In the following analysis, the learning curves are reviewed to assess the representativeness of the training and validation datasets, and any possible problems with underfitting or overfitting. Two learning curves are analysed: (i) training curve: to give an idea of how well the model is learning, and (ii) validation curve which gives an idea of how well the model is generalising. Furthermore, the training and test error were compared using the root mean square error (RMSE) metric. Going further, the accuracy of the corrected DEMs at the implementation sites was also assessed using the RMSE. The RMSE is very common, and is considered an excellent general-purpose error metric for assessing numerical predictions (Christie and Neill, 2022). The model explainability is addressed using feature importance plots.

3. RESULTS

3.1 Model Diagnostics

In the learning curves for both comparisons (default vs. Bayesian; Figure 3), the plot of the training and validation error decreases with successive epochs. In the models with default hyperparameters, the gap between the training and validation curves is minimal within the 100 epochs shown. The Bayesian tuning of XGBoost increased the number of estimators (n_estimators) significantly, thus leading to a higher number of training epochs, that ranged from 500 – 800. Even with the Bayesian optimised model validation accuracy plateau-ing early, it is still better than the accuracy of the default model. Early stopping was implemented for both models to stop training at any point where there was no further improvement in the RMSE after 10 rounds. Expectedly, the training errors at both sites are lower than the validation error because the models are fit on the training data. The test errors (RMSEs) using

Name of hyperparameter	Name/alias	Search space
No of trees	n_estimators	(0, 2000)
Max depth of tree	max_depth	(1, 10)
Learning rate	learning_rate	(0.001, 1)
Regularisation parameters	reg_alpha	(0.001, 10)
	reg_lambda	(0.001, 10)
Others	subsample	(0.001, 1.0)
	colsample_bytree	(0.001, 1.0)
	min_child_weight	(0.001, 10)
	gamma	(0.001, 10)

Table 2. The uniquely defined hyperparameter search space for Bayesian optimization of XGBoost

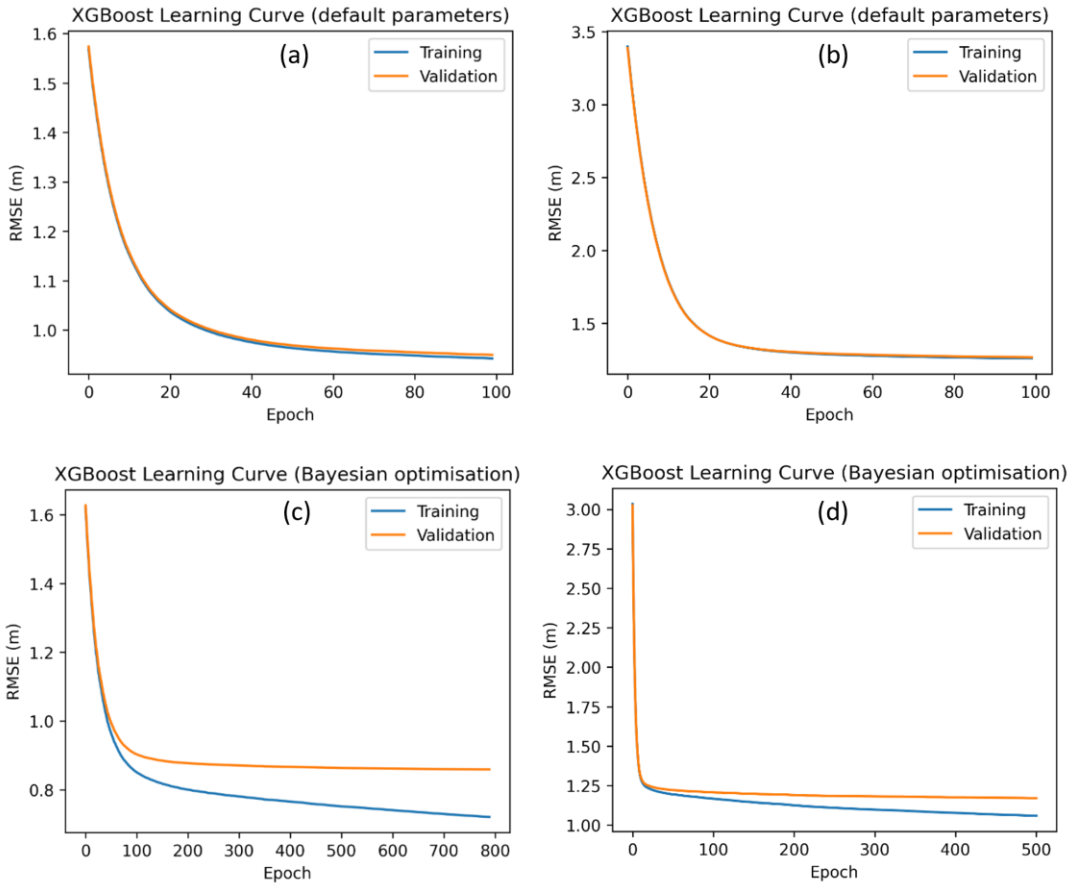

Figure 3. Learning curves showing error loss at successive training epochs, (a) Copernicus, default (b) AW3D, default (c) Copernicus, Bayesian-optimised (d) AW3D, Bayesian-optimised

default hyperparameters are 0.95 m (Copernicus) and 1.27 m (AW3D). With Bayesian-optimised hyperparameters, the RMSEs reduced to 0.86 m (Copernicus) and 1.17 m (AW3D).

3.2 Accuracy of Corrected DEMs

Table 3 shows the accuracy metrics of the original DEMs versus corrected DEMs, with default and Bayesian-optimised hyperparameters. The corrections achieved more realistic terrain representations in both Copernicus and AW3D. The height error map (Figure 4) shows a significant reduction in the elevation error, after correction. Notable improvements in accuracy are observed in Copernicus and AW3D DEMs using both the default and optimised models. The optimised model outperformed the default model in some cases. At site A, the correction to Copernicus DEM achieved a reduction in RMSE from 1.859 m to 0.877 m and 0.888 m in the default and optimised models respectively.

At site B, the correction reduced the RMSE of Copernicus DEM from 1.470 m to 0.788 m and 0.771 m in the default and optimised models respectively. Similarly, there were improvements in the accuracy of AW3D (RMSE reduction from 4.783 m to 1.324 m and 1.316 m respectively at site A; and reduction from 3.628 m to 0.987 m and 1.005 m at site B, respectively). This represents an improvement factor of 46 – 53% in Copernicus, and 72 – 73% in AW3D. Other positive changes in the corrected DEMs include:

- Diminution of surface artefacts and discontinuities.
- Minimisation of building artefacts.

Implementation Site	DEM	Original DEM RMSE (m)	Corrected DEM RMSE (m)	
			Default hyperparameters	Bayesian Optimisation
A	Copernicus	1.859	**0.877**	0.888
	AW3D	4.783	1.324	**1.316**
B	Copernicus	1.470	0.788	**0.771**
	AW3D	3.628	**0.987**	1.005

Table 3. Accuracy of original DEMs versus corrected DEMs, default hyperparameters *vs* Bayesian optimisation

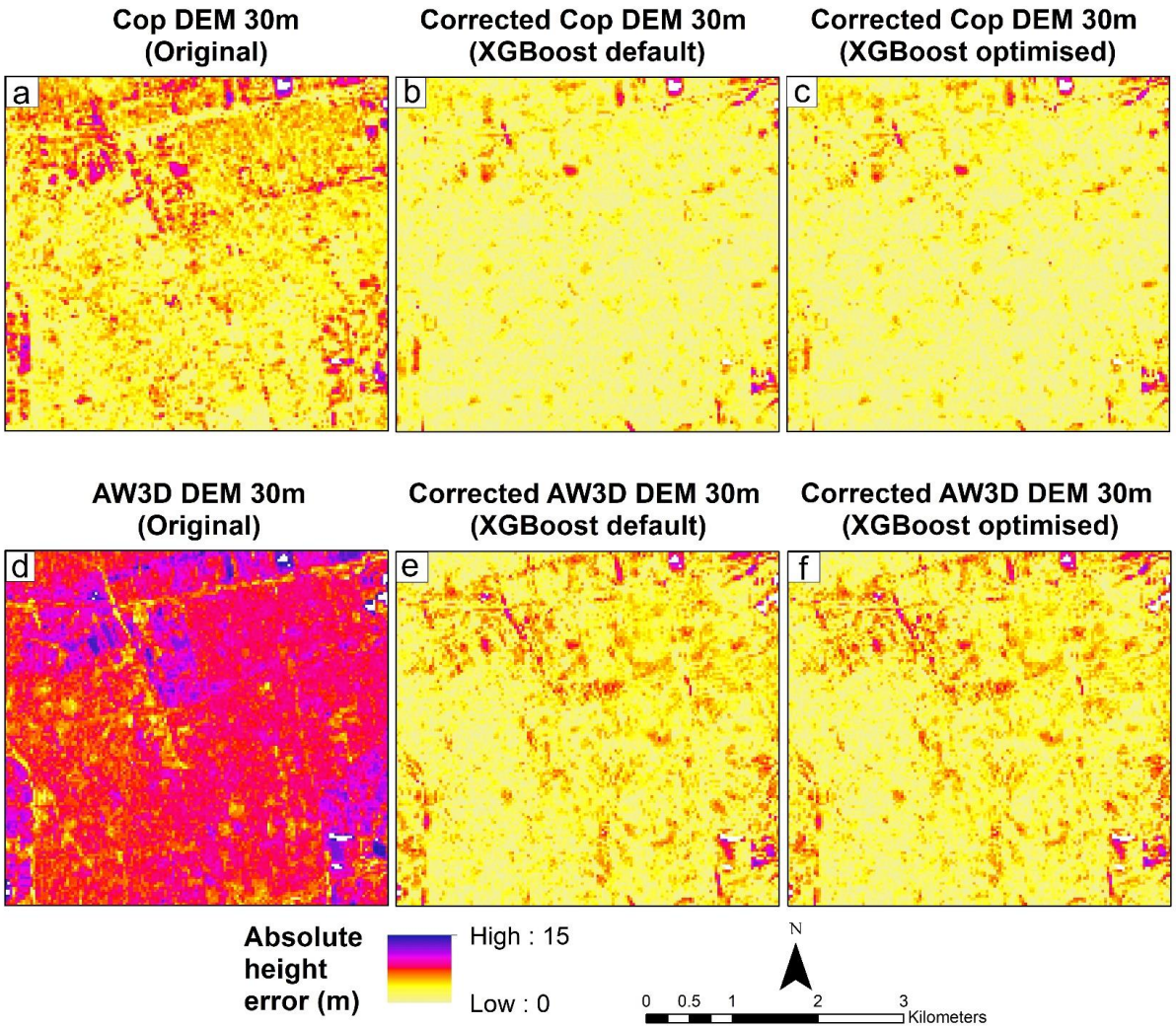

Figure 4. Height error maps of original *vs* corrected DEMs at site A, Copernicus (a – c) and AW3D (d – f)

- Significant improvement in the representation of terrain details.
- Minimisation of elevation error dispersion.

for diverse applications in the national, regional and global domain.

This research proposes a machine learning approach for the

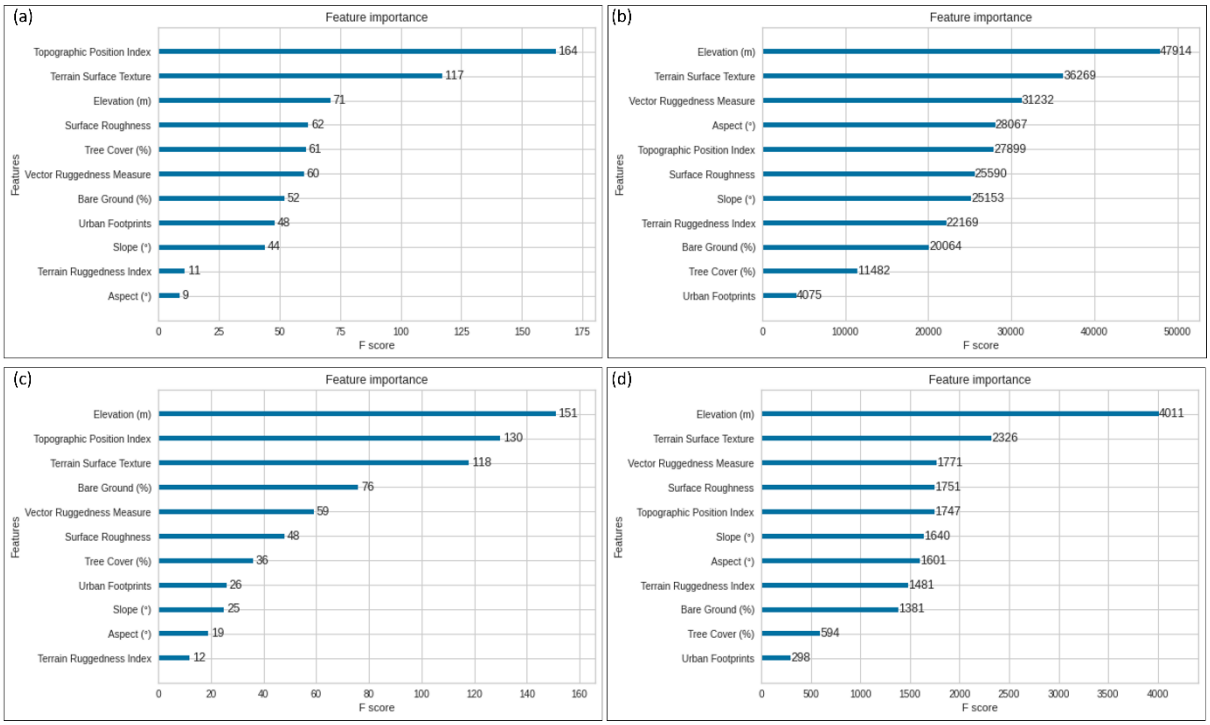

Figure 5. Feature importance plots of the input variables for height error prediction, (a) Copernicus, default (b) Copernicus, Bayesian-optimised (c) AW3D – default (d) AW3D, Bayesian-optimised

3.3 Analysis of Feature Importance

The topographic position index (TPI) and elevation are some of the most important features influencing the prediction of elevation error by the default and Bayesian-optimised models, respectively (Figure 5). With the default model, the top three most influential parameters are the TPI, terrain surface texture (TST) and elevation, whereas with the Bayesian-optimised model the top three most influential features are the TST, VRM and elevation.

Generally, the F-scores of the features in the Bayesian-optimised version are orders of magnitude higher than the default mode, suggesting the capability of Bayesian optimisation to exploit more complex interaction between the variables for better predictions. However, it should be noted that the influence or relevance of features in the outcome of the machine learning predictions is also terrain dependent.

4. DISCUSSION AND CONCLUSION

For decades, the issue of positional errors in satellite remote sensing datasets has been a major concern of scientists and researchers, and remains a recurring theme in the remote sensing community all around the world. The presence of vertical errors in global DEMs has been a challenge and a source of concern to end-users of satellite remote sensing products, geomatics practitioners and the remote sensing community at large. It compromises the utility of global DEMs

enhancement of global DEMs in urban/industrial settings. The backbone of the framework is an XGBoost-based feature-level fusion of terrain and land cover parameters that incorporates Bayesian optimisation for DEM error prediction and correction.

Generally, the corrected DEMs show several topographic improvements such as the diminution of terrain offsets. The elevation error dispersion has also been reduced in the corrected DEMs. There was a 46 – 53% reduction in the root mean square error (RMSE) of Copernicus DEM, and a 72 - 73% reduction in the RMSE of AW3D DEM. This reduction in the elevation bias is very significant and competitive, when compared with the achievements of previous studies focused on urban areas (e.g., Olajubu et al., 2021; Hawker et al., 2022).

The results also show that hyperparameter tuning with Bayesian optimisation can yield appreciable gains in accuracy. Thus, tuning the hyperparameters of tree-based models is recommended as a measure to improve the accuracy of predictions. The topographic position index and elevation were some of the most influential features in the default and optimised models.

The methodology presented in this study is simple, low-cost and easy-to-follow. Moreover, the ensemble framework can learn non-linear and multi-variate spatial patterns in urban environments. The corrections are implemented on a point-by-point basis, in contrast to other techniques that only address the global bias. The introduced methodology based on the integration of XGBoost, land cover and terrain parameters

shows great potentials for improved hydrological modelling in urban catchments.

Supervised machine learning regression is a powerful and effective approach for modelling complex and non-linear terrain parameters even in challenging, inaccessible and difficult landscapes. Tree-based ensemble machine learning and recent implementations of gradient boosting are very powerful for reducing the uncertainty in digital elevation datasets. Given the sheer amount of environmental, hydrological and geological applications that rely on free global DEMs, DEM correction will remain a strategic research mandate in the remote sensing community.

The proposed enhancement scheme can be adopted by remote sensing research consortia for producing the next urban-corrected global DEM. Since machine learning algorithms are likely to be biased to the landscape characteristics fed into them, future research can explore the performance of the proposed approach in different landscapes.

While high resolution DEMs are available in many advanced countries, low and middle income (LMICs) continue to struggle with funding constraints to pursue such high-resolution mapping. Thus, the proposed enhancement scheme is a low-cost and viable alternative for national mapping agencies in LMICs to improve the vertical accuracy of readily available global DEMs for use in their national geospatial infrastructure.

ACKNOWLEDGEMENTS

Special thanks to the Commonwealth Scholarship Commission UK, and the University of Cape Town Postgraduate Funding Office for funding support for this research. LIDAR data for the City of Cape Town was provided by the Information and Knowledge Management Department, City of Cape Town. Also, we appreciate Ikechukwu Maduako, Hossein Bagheri, Tom Komar, Maria Peppa, and Chima Iheaturu for their insightful feedback.

REFERENCES

Bentéjac, C., Csörgő, A. Martínez-Muñoz, G.A., 2021. Comparative analysis of gradient boosting algorithms. *Artif Intell Rev* 54, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Can, R., Kocaman, S., Gokceoglu, C.A., 2021. Comprehensive Assessment of XGBoost Algorithm for Landslide Susceptibility Mapping in the Upper Basin of Ataturk Dam, Turkey. *Appl. Sci.* 2021, 11, 4993. https://doi.org/10.3390/app11114993

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

Christie, D., Neill, S.P., 2022. Measuring and Observing the Ocean Renewable Energy Resource. In *Comprehensive Renewable Energy* (Second Edition) Volume 8, 2022, Pages 149-175. https://doi.org/10.1016/B978-0-12-819727-1.00083-2

Climent, F., Momparler, A., Carmona, P., 2019. Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. *Journal of Business Research*, 101, 885-896. https://doi.org/10.1016/j.jbusres.2018.11.015

Deng, S., Wang, C., Li, J., Yu, H., Tian, H., Zhang, Y., Cui, Y., Ma, F., Yang, T., 2019. Identification of Insider Trading Using Extreme Gradient Boosting and Multi-Objective Optimization. *Information* 2019, 10, 367. https://doi.org/10.3390/info10120367

Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenböck, H., Müller, A., Dech, S., 2013. Urban footprint processor - Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. In: *IEEE Geoscience and Remote Sensing Letters*. Vol 10. No. 6, pp. 1617-1621.

Esch, T., Thiel, M., Schenk, A., Roth, A., Muller, A., Dech, S., 2010. Delineation of Urban Footprints From TerraSAR-X Data by Analyzing Speckle Characteristics and Intensity Information. *IEEE Transactions on Geoscience and Remote Sensing* 48(2), 905-916, Feb. 2010, doi: 10.1109/TGRS.2009.2037144.

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* 2001, 29, 1189–1232.

Han, J., Fang, M., Ye, S., Chen, C., Wan, Q., Qian, X., 2019. Using Decision Tree to Predict Response Rates of Consumer Satisfaction, Attitude, and Loyalty Surveys. *Sustainability*, 11(8), 2306. http://dx.doi.org/10.3390/su11082306

Hastie, T., Tibshirani, R., Friedman, J.H., 2009. Boosting and Additive Trees. In The *Elements of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 2009, 337–384. ISBN 978-0-387-84858-7.

Hawker., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., Neal, J., 2022. A 30 m global map of elevation with forests and buildings removed. *Environ. Res. Lett*. 17 024016. DOI 10.1088/1748-9326/ac4d4f

Kim, D.E., Liu, J., Liong, S.-Y., Gourbesville, P., Strunz, G., 2021. Satellite DEM Improvement Using Multispectral Imagery and an Artificial Neural Network. *Water*, 13, 1551. https://doi.org/10.3390/w13111551

Liu, Y., Bates, P.D., Neal, J.C., Yamazaki, D., 2021. Bare-Earth DEM generation in urban areas for flood inundation simulation using global digital elevation models. *Water Resources Research*, 57, https://doi.org/10.1029/2020WR028516

Łoś, H., Sousa Mendes, G., Cordeiro, D., Grosso, N., Costa, H., Benevides, P., Caetano, M., 2021. Evaluation of Xgboost and Lgbm Performance in Tree Species Classification with Sentinel-2 Data, 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021, 5803-5806, doi: 10.1109/IGARSS47720.2021.9553031.

Miao, X., Heaton, J.S., Zheng, S., Charlet, D.A., Liu H., 2012. Applying tree-based ensemble algorithms to the classification of ecological zones using multi-temporal multi-source remote-sensing data, *International Journal of Remote Sensing*, 33,6, 1823-849, DOI: 10.1080/01431161.2011.602651

Olajubu, V., Trigg, M.A., Berretta, C., Sleigh, A., Chini, M., Matgen, P., Mojere, S., Mulligan, J., 2021. Urban correction of global DEMs using building density for Nairobi, Kenya. *Earth Sci Inform* 14, 1383–1398. https://doi.org/10.1007/s12145-021-00647-w

Pham, X.T.T., Ho, T.H., 2021. Using boosting algorithms to predict bank failure: An untold story. *International Review of Economics & Finance*, 76, 40-54. https://doi.org/10.1016/j.iref.2021.05.005

Safaei, N., Safaei, B., Seyedekrami, S., Talafidaryani, M., Masoud, A., et al., 2022. E-CatBoost: An efficient machine learning framework for predicting ICU mortality using the eICU Collaborative Research Database. *PLOS ONE* 17(5), e0262895. https://doi.org/10.1371/journal.pone.0262895

Samat, A., Li, E., Wang, W., Liu, S., Lin, C., Abuduwaili, J., 2020. Meta-XGBoost for Hyperspectral Image Classification Using Extended MSER-Guided Morphological Profiles. *Remote Sensing*. 2020; 12(12),1973. https://doi.org/10.3390/rs12121973

Sirmacek, B., d'Angelo, P., Krauss, T., Reinartz, P, 2010. Enhancing Urban Digital Elevation Models using Automated Computer Vision techniques. In: Wagner W., Székely, B. (eds.): *ISPRS TC VII Symposium* – 100 Years ISPRS, Vienna, Austria, July 5–7, 2010, IAPRS, Vol. XXXVIII, Part 7B. https://www.isprs.org/proceedings/xxxviii/part7/b/pdf/541_XXXVIII-part7B.pdf

Smit, W., 2020. Cape Town – Oxford Bibliographies. https://www.oxfordbibliographies.com/display/document/obo-9780190922481/obo-9780190922481-0036.xml (Date accessed: 9th March, 2023)

Western Cape Government, 2022. City of Cape Town: Overview. https://www.westerncape.gov.za/your_gov/33#:~:text=The%20City%20of%20Cape%20Town%20includes%20the%20Cape%20Metropolitan%20Council,Oostenberg%2C%20South%20Peninsula%20and%20Tygerberg. (Date accessed: 9th March, 2023)

Xu, K., Fang, J., Fang, Y., Sun, Q., Wu, C., Liu, M., 2021. The Importance of Digital Elevation Model Selection in Flood Simulation and a Proposed Method to Reduce DEM Errors: A Case Study in Shanghai. *Int J Disaster Risk Sci* 12, 890–902. https://doi.org/10.1007/s13753-021-00377-z