Application of Machine Learning Methods to improve vertical accuracy of CARTOSAT DEM

Venkatesh Kasi 1, Pavan Kumar Yeditha 2, Maheswaran Rathinasamy 3, Chandramouli Sangamreddi 4

1 School of Infrastructure, Indian Institute of Technology Bhubaneswar, India. 2 IHE Delft Institute for Water Education, Water Science and Engineering, Delft Netherlands. 3 Department of Civil Engineering, Indian Institute of Technology Hyderabad, Telangana, India. 4 Department of Civil Engineering, MVGR College of Engg, Vizianagaram, 535005, India




PRESENTED AT:

INTRODUCTION

· Digital Elevation Model (DEM) is a 3-dimensional representation of the earth's surface. Its uses are diverse and extensive which includes flood modeling (Setti et al. 2018), land use studies (Sridhar et al. 2019), geomorphological studies (Maheswaran et al. 2016), watershed delineation (Kasi et al., 2020), glaciology (Cook et al. 2012), evaluation of natural hazards and geological applications.

· In the past, machine learning tools have gained a lot of attention is several research areas viz. hydrological forecasting, spatial interpolation, owing to their ability to capture nonlinear relationships (Yeditha et al., 2020).

· In this study, we propose to develop a framework based on Genetic Programming (GP) and Artificial Neural Network (ANN) to generate high quality, DEM by an amalgamation of the SRTM-30 m, CARTOSAT-30 m DEMs

STUDY AREA AND DATASETS

Study area:

a sub-basin (Fig. 1) of the Champavathi River Basin located on the east coast of India

· Catchment area is 868 km2.

· The location of the watershed is lies between the latitude of 18° 11′ North and longitude of 83° 40′ East.

· The physiography of the watershed is dominated by plain areas with occasional small mountain ranges.

· The altitude ranges from 26 m to 564 m.

· The major land cover and land use include agricultural tracts, urban areas, hill slope, grasslands, and water bodies.

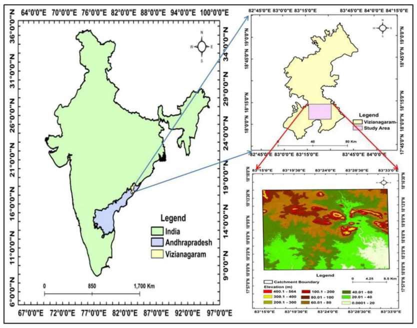

Fig. 1: Geographical location of the study area. The left panel shows the index map of the study area and the right bottom panel shows the terrain of the study area using SRTM DEM-30 m.

Datasets:

SRTM – 30M

CARTOSAT – 10M

CARTOSAT – 30M

Ground Control Points (DGPS) (Fig2)

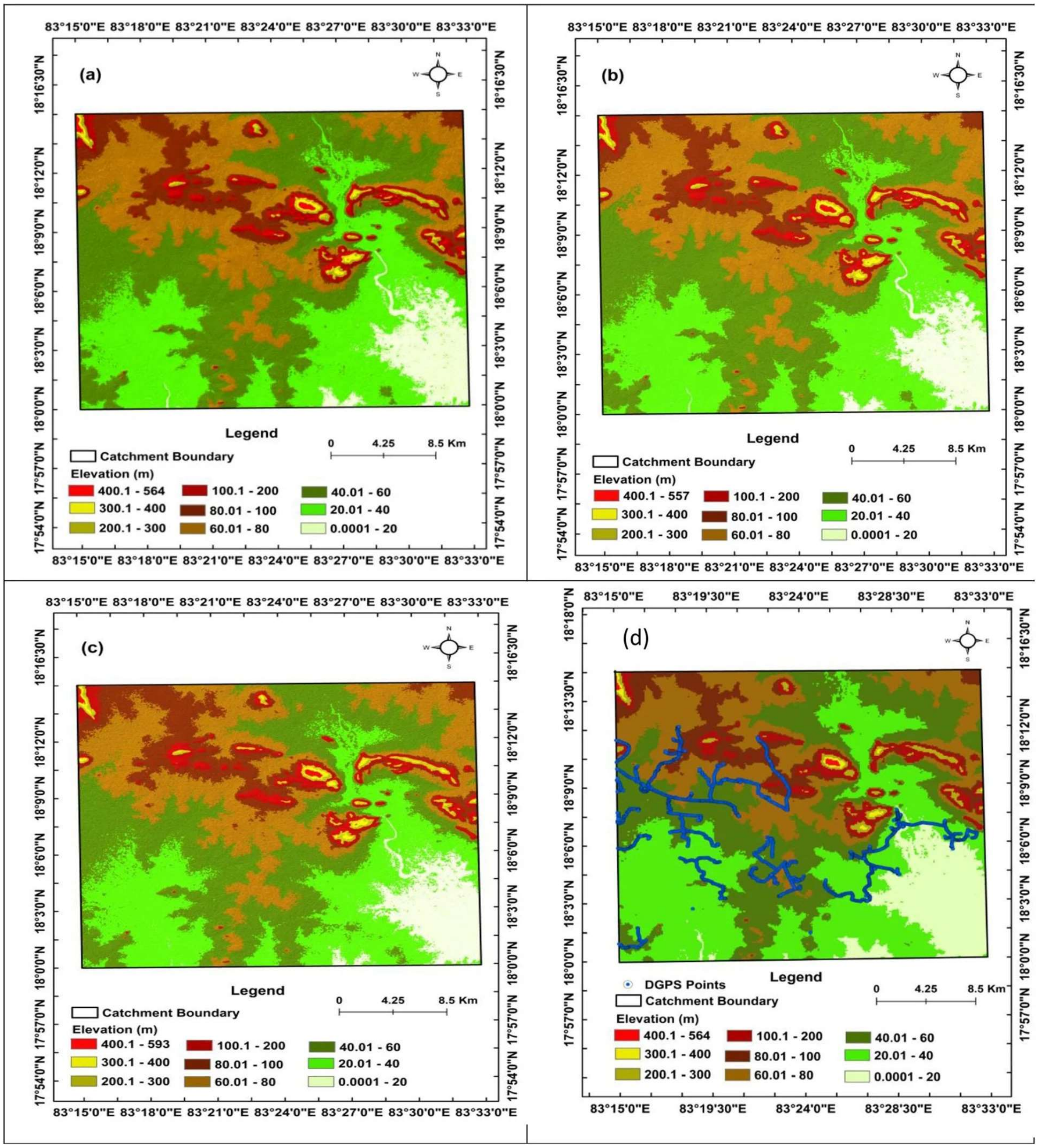

Fig. 2 a–c represents the Geographical location of the DGPS points in the study area.d represents the Topography of the study area obtained from CARTOSAT-10 m, CARTOSAT-30 m and SRTM-30 m, respectively and

METHODOLOGY

The proposed approach is categorized into three major steps as illustrated in Fig. 3.

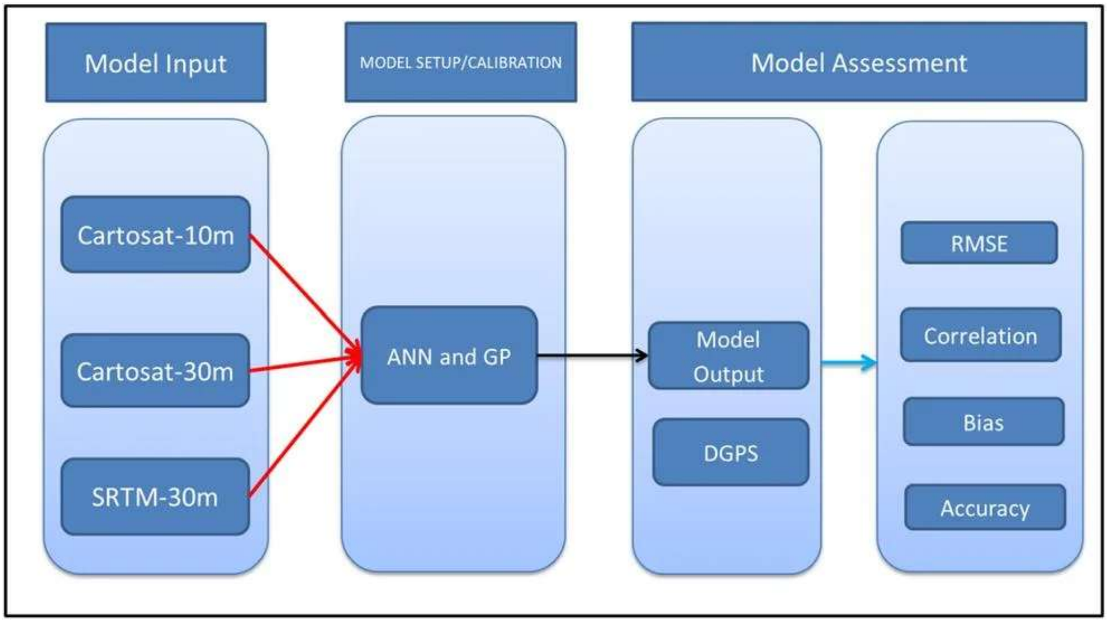

Fig. 3: Schematic of the proposed methodology

(i) Model Input Processing:

· Generally, DEM contains voids/gaps (null values) (no elevation value in the cells) the gaps occur in regions with the water surface, areas covered with low reflecting materials, areas with rough slopes.

· These voids are required to be removed or corrected.

· Further, before the analysis, we have projected all the data set to the same datum (WGS84) and projection system (UTM-44 N).

(ii) Model Setup

In this study for comparison, we have adopted two different models ANN and GP.

· For both the models, the input data sets are elevation from CARTOSAT-10 m, CARTOSAT-30 m and SRTM-30 m DEM and the model output was elevation at GCP points.

· Around 60% of the total GCP points were used for model training, validation was done with 20%and the testing is carried out with the remaining percentage.

· The accuracy of the DEMs was evaluated using the following measures: (a) RMSE (RootMean Square Error and (b) Bias.

$$\textbf{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \Sigma (z_{DGPS}^{i} - z_{DEM}^{i})^{2}}{N-1}} \qquad \text{(a)}$$

$$\textbf{Bias (\%)} = \frac{\sum_{i=1}^{N} (z_{DGPS}^{i} - z_{DEM}^{i})}{\sum_{i=1}^{N} (z_{DGPS}^{i})} \text{x100} \qquad \text{(b)}$$

Where: - Z_{DGPS}^{i} = Elevation obtained using DGPS,
Z_{DEM}^{i} = Elevation obtained using DEM,
N is the total no of sampling Points.

(iii) Post Processing and Application

After the models were trained and tested the models were applied to develop the improved DEM using the original DEM datasets

RESULTS AND DISCUSSION

Calibration of validation of the machine learning models

RMSE and correlation between the observed elevation (from GCP) and the modeled elevation values show that both models perform satisfactorily. (Figure 4)

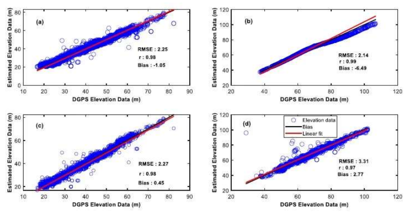

Fig. 4: Scatter plot between elevation values from GCPs and model estimated elevation valu3 from top panel**-** neural networks during a) calibration and b) validation, bottom panel **–** genetic programming during c) calibration d) validation

RESULTS AND DISCUSSION

Application of model for DEM generation

For the application of the calibrated models in DEM generation, we have extracted elevations from each of the original DEMs at 100048 points, randomly generated using the fishnet tool in ArcGIS.

From the simulated elevation values at these points, DEM was generated using different interpolation methods and corresponding results are shown in Figure 5

To verify the accuracy of the generated DEM, we have used the test data GCP points Figure 6 shows the elevations scatter plot for the 1000 testing Points

From the results, the ANN- IDW method performs better in terms of RMSE. The RMSE of the resulting DEM is significantly improved from 5.03 m of the original CARTOSAT to 3.25 m (i.e., 35% improvement), and performing far better than SRTM 30 m. Similar results were obtained from the GP-based DEMs.

The RMSE for the GP model results were found to be varying from 3.2–3.29 m and the bias between 1.48– 2.21 m.

It can very clearly have observed that in comparison with the original DRM, there is a considerable improvement in the vertical accuracy in all the models. This indicates the efficacy of the proposed approach.

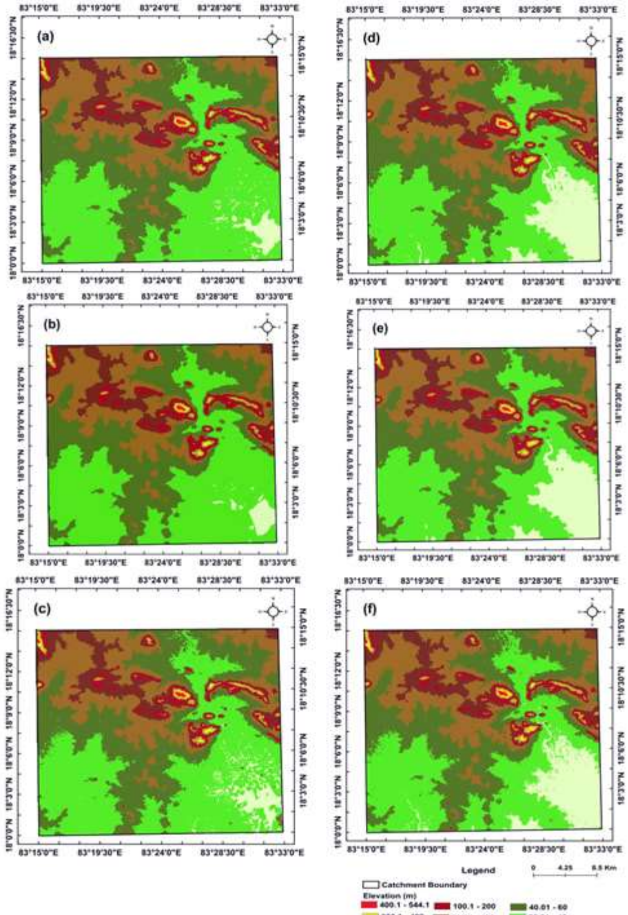

Fig. 5 Left column: improved DEM obtained from neural network models using

a) IDW
b) KRIGING
c) SPLINE interpolation methods

and right column shows DEM obtained from GP models using

d) IDW
e) KRIGING
f) SPLINE interpolation methods

CONCLUSION

· Overall, the proposed approach is shown to be able to improve the CARTOSAT DEM significantly in the selected study area in terms of vertical accuracy and bias.

· The proposed method is applicable in regions where there the spatial data is sparse due to cost, facility, and data sharing strategy.

· The proposed method makes use of reference elevation data from the area which is data-rich to help in developing DEM of better accuracy in data-sparse regions having similar terrain characteristics.

· Two different machine learning approaches –combined with three interpolation techniques results in six different models.

· The performance of the proposed approach was an encouraging and significant improvement in RMSE of the order of 35% when compared to the original CARTOSAT DEM.

· It is important to note that this methodology requires DGPS for areas where improvement is required.

· However, these can be replaced with GLAS/ ICES at data sets for training the model or even high-resolution data can be used if available.

AUTHOR INFORMATION

Venkatesh Kasi 1, Pavan Kumar Yeditha 2, Maheswaran Rathinasamy 3, Chandramouli Sangamreddi 4

1 School of Infrastructure, Indian Institute of Technology Bhubaneswar, India. 2IHE Delft Institute for Water Education, Water Science and Engineering, Delft Netherlands. 3Department of Civil Engineering, Indian Institute of Technology Hyderabad, Telangana, India. 4Department of Civil Engineering, MVGR College of Engg, Vizianagaram, 535005, India

ABSTRACT

In recent decades, the application of Digital Elevation Models (DEMs) has been widely used in various aspects such as land management and flood planning since it reflects the actual topographic characteristic on the Earth's surface. However, obtaining a high-quality DEM is often quite challenging because it is time-consuming, costly, and often confidential. This study presents an innovative approach to derive an improved vertical accuracy of CARTOSAT 10m DEM by blending it with publicly available SRTM (Shuttle Radar Topography Mission) DEM using machine learning methods such as Genetic Programming (GP) and Artificial Neural Networks (ANN). SRTM-1 DEM and CARTOSAT DEM in India are applied to GP and ANN to generate improved vertical accuracy high-quality DEM. The results revealed that the proposed approach improves the vertical accuracy by considering the reference as Ground control Points (GCPs) elevation from Differential Global Positioning System (DGPS) survey data. A significant improvement of 47 and 35% generated DEMs in RMSE compared to the SRTM-1 and CARTOSAT, respectively.

REFERENCES

1. Setti S, Rathinasamy M, Chandramouli S (2018) Assessment of water balance for a forest dominated coastal river basin in India using a semi distributed hydrological model. Model Earth Syst Environ 4: 127–140. https://doi.org/10.1007/s40808-017-0402-0 (https://doi.org/10.1007/s40808-017-0402-0)

2. Sridhar S, Maheswaran R, Radha D and Sridhar V (2020) Attribution of hydrologic changes in a tropical river basin to rainfall variability and land use change: A case study from India, Journal of Hydrol.Engg. 25(8).

3. Maheswaran R, Khosa R, Gosain AK, Lahari S, Sinha SK, Chahar BR, Dhanya CT (2016) Regional scale groundwater modelling study for Ganga River basin. J Hydrol 541:727–741. https://doi.org/10.1016/ (https://doi.org/10.1016/) j.jhydrol.2016.07.029

4. Kasi, V., Pinninti, R., Landa, S. R., Rathinasamy, M., Sangamreddi, C., Kuppili, R. R., & Dandu Radha, P. R. (2020). Comparison of different digital elevation models for drainage morphometric parameters: a case study from South India. *Arabian Journal of Geosciences*, *13*(19). https://doi.org/10.1007/s12517-020-06049-4

5. Yeditha, P. K., Kasi, V., Rathinasamy, M., & Agarwal, A. (2020). Forecasting of extreme flood events using different satellite precipitation products and wavelet-based machine learning methods. *Chaos (Woodbury, N.Y.)*, *30*(6), 063115. https://doi.org/10.1063/5.0008195.

6. Cook AJ, Murray T, Luckman A, Vaughan DG, Barrand NE (2012) A new 100-mdigital elevationmodel of the antarctic peninsula derived fromASTER globalDEM:methods and accuracy assessment. Earth Syst Sci Data 4:129–142. https://doi.org/10.5194/essd-4-129-2012