GEOGRAPHICALLY WEIGHTED REGRESSION FOR SURFACE ELEVATION ESTIMATION USING GLOBAL DEMs AND SPOT HEIGHT DATA

Le Van Bien[1], Venkatesh Raghavan[1], Vinayaraj Poliyapram[1], Tran Thi An[2]

[1]Graduate School for Creative Cities, Osaka City University, Japan
[2]University of Science and Education, The University of Danang, Vietnam.
Email: ttan@ued.udn.vn

ABSTRACT

Global DEMs with the advantages of free and open access have been used widely in many applications including terrain related and hydrological studies. However, the utilization of global DEMs in some cases leads to inaccurate results due to the limitation in airborne elevation data collection methods. This study has explored the advantages of Geographically Weighted Regression (GWR) for estimation of an improved elevation model using three global DEMs (ALOS AW3D30, ASTER GDEM, SRTM) and the field survey spot height reference points. The global DEMs have been used as multiple independent variables (explanatory variables) and reference points are as dependent variable in the GWR model. As a result, the GWR based estimated DEM shows better accuracy compared to individual global DEM in almost all statistical parameters. The Mean of Absolute Error (MAE), Root Mean Square Error (RMSE) and Standard Deviation (SD) of the estimated DEM were much improved compared to the global DEMs. The RMSE was reduced from 7.0m in ALOS-30, 8.3m in the GDEM and 6.4m in SRTM to 4.1m in the GWR DEM. The MAE was also reduced from 5.2m in ALOS-30, 5.4m in GDEM and 4.6m in SRTM to 1.9m in the estimated DEM. The GWR DEM and reference data also shows the significant correlation, with R^2 = 0.9994. Results from this study reveal that GWR for surface elevation estimation can provide significant enhancement on quality of DEMs derived from various sources.

1. INTRODUCTION

Digital Elevation Model (DEM) which is a 3D model representation of a terrain surface is nowadays considered as one of the essential inputs for various research fields. The quality of a DEM directly affects the accuracy of the application research. In almost studies, the free access global DEMs such as GDEM, SRTM and ALOS-30 DEM are used frequently. However, global DEMs with its inherent limitations sometimes affect the quality of application results. This study proposes a method for improving quality of global DEMs and estimating a combined elevation surface using Geographically Weighted Regression.

Geographically Weighted Regression (GWR) is a local form of linear regression used to explore and model spatially varying relationships in regression models of geo-referenced data (Fotheringham *et al*., 2002). The GWR approach for spatial modeling was becoming an important tool which provides technique to deal with spatial non-stationarity in multivariate regression and estimates regression coefficients locally using spatially dependent weights (Brunsdon *et al*., 1996). GWR has been applied more widely in urban geographical, economic and environmental studies.

GWR is aimed to estimate a combined parameter based on dependent and explanatory variables in a given neighborhood (bandwidth). Results of GWR model show both the estimation and the spatial distribution of error compared to the observational data. Based on the spatial heterogeneity, GWR may generate a separate weights and bias for the multiple linear regression equation. Weights and bias of the multiple linear regression equation will be updated for each point according the correlation of points in a defined bandwidth. This study has taken the advantage of GWR for generating a new topographic surface using reference elevation points and three global DEMs. Hitherto, there still not much study on elevation estimation using GWR. This study has first time developed a GWR model for improving DEM estimation using a limited number of reference points.

2. METHODOLOGY

2.1 Study area

The study area is located in Danang City (15º55′N to 16º14′N and 107º18′E to 108º20′E), an Eastern Sea's coastal city in middle of Central Vietnam (Figure 1). The study area of 950 square km covers inland area of Danang city. The characteristic of topography is varied with elevation ranging from 0m to 1664m above mean sea level and spreading from mountain in the west to flat region in the east of city. The varying topography is one of reason causes the different representation in elevation on each global DEM data type. Thus, ALOS AW3D30 and ASTER GDEM generated by traditional optical stereo matching technique, as well as SRTM generated by Interferometric Synthetic Aperture Radar (InSAR) technique still contain inherent anomalies that need to be detected and minimized.

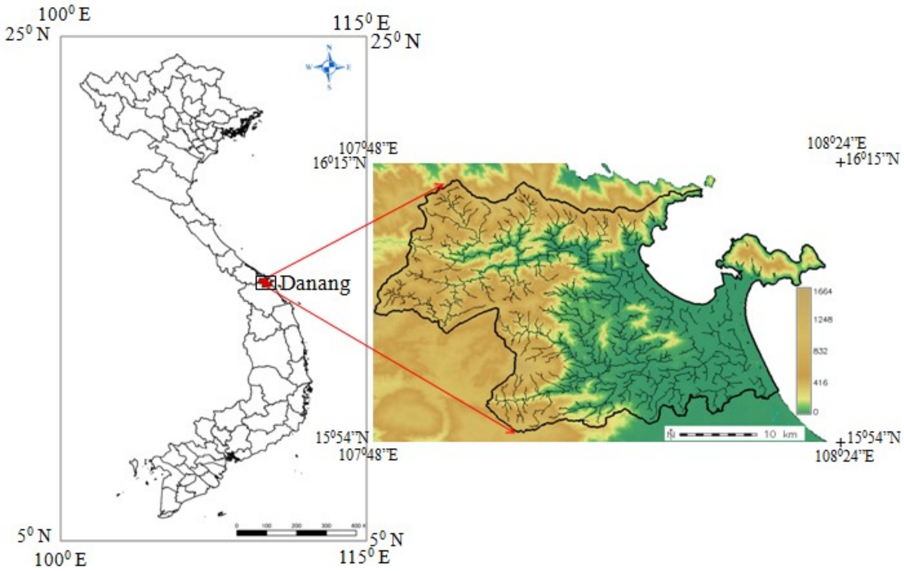

Figure 1. Location of study area and topographic overview.

2.2 DEM Datasets

The data used in this study includes ALOS-30 Version 1.1, Aster GDEM Version 2 and SRTM-30 Version 3. ALOS-30 (ALOS AW3D30) has been released by JAXA (Japan) in 2015 and can be downloaded free of charge from http://www.eorc.jaxa.jp/ ALOS/en/aw3d30. GDEM version 2 which was derived from ASTER optical satellite images has been released on October 2011 (ASTER GDEM Validation Team, 2011). SRTM for this study is SRTM V3 (or SRTM Plus) that has resolution of 30m (1 arc second) and provides globally in September, 2014. This version is an enhancement resolution from SRTM 90m (3 arc seconds) which was released publicly in 2003. Both GDEM V2 and SRTM V3 can be downloaded freely from http://earthexplorer.usgs.gov. Three global DEMs used in this study are in the same geographic coordinate system, with the World Geodetic System 1984 (WGS84) horizontal datum and the Earth Gravitational Model 1996 (EGM96) vertical datum.

The reference elevation data which was provided by Department of Natural Resources and Environment of Danang city, Vietnam with more than 130,000 elevation points is the spot height elevation data surveyed in 2009. This data was converted to raster format in order to be comparable to the global DEMs. The data are projected in a Vietnamese projection named VN2000 and Vietnamese vertical datum named Hon Dau, Hai Phong. The accuracy of reference data is 0.1m. Both global DEMs and reference data have been converted to WGS84 zone 49N projection before assessment and estimation.

2.3. Pre-processing

Due to the varying topography of study area as well as the characteristics of generation technique of each global DEM, anomalies and voids are still noticed in ASTER GDEM and SRTM. GDEM has some anomalies in the western mountainous parts of Danang city, with significantly higher elevation values (ranging from over 2000m to 8016m). These artifacts may be caused due to cloud coverage that is very common in optical satellite data. Artifacts of GDEM cover about 0.05% of Danang area effecting 1031 pixel. In SRTM, there are some voids with no data located in the western mountain area and small areas in the coast. 7872 pixels are voids appearing as white areas in SRTM covering 0.35% Danang and comprising 7 square kilometers. These voids maybe caused by the limitation of InSAR technique used in SRTM for the high relief areas. Elimination of anomalies in GDEM and filling voids in SRTM are necessary before further processing. Considering ALOS-30m DEM, there is no anomalies as well as voids in this DEM. Therefore, ALOS-30m value was used to replace the areas with voids in SRTM as well as anomalies in GDEM. As the result, most of the voids and anomalies in SRTM and GDEM were removed. Additionally, the root mean square error (RMSE) of two DEMs was reduced significantly compared to reference data.

Both global DEMs and reference data have been converted to the same projection that is UTM-WGS84 Zone 49N. Regarding the vertical datum, there is difference between three global DEMs and reference data because of the difference in data collection method. Global DEMs use the EGM96 vertical datum, while reference data uses Vietnamese vertical datum named Hon Dau, Hai Phong that is related to MSL in Hon Dau Island, Hai Phong city, Vietnam. Nguyen and Le (2002) showed that the global EGM96 model is almost similar to the Vietnamese vertical datum, 97% of data shows the height difference around 1.5m, only 3% of data shows higher than 1.5m. Therefore, the vertical datum of global DEMs was converted from EGM96 to Vietnamese vertical datum by subtracting 1.5m elevation from the global DEMs (Nguyen and Le, 2002). After correcting vertical datum, the RMSE and mean absolute error (MAE) of ALOS-30, GDEM and SRTM compared to reference data show better than before. In general, among the three DEMs, SRTM showed the highest accuracy with MAE of 4.5m and RMSE of 6.4m. The ALOS-30m is next to SRTM in terms of accuracy, with MAE and RMSE values of 5.2 and 7.0 m, respectively. Finally, GDEM presented the lowest accuracy with MAE and RMSE of 5.4 and 8.3m, respectively.

2.4. DEM Estimation Using Geographically Weighted Regression (GWR)

GWR is a local form of linear regression which is a method for exploring the relationship between a dependent variable (Y) and independent variables (X). Subsequently, it will predict the dependent variable based on independent variables. In contrast to global regression, instead of using same equation coefficient for entire study area and one independent variable, GWR can use multiple independent variables and can build a separate equation for local area in the dataset within given bandwidth (neighborhood size). That means GWR can consider geographical differences in the estimation. The equation of GWR model is based on Fotheringham *et al.*, 2002.

This study has explored the advantages of GWR for estimation of a new DEM using three global DEMs (ALOS AW3D30, ASTER GDEM, SRTM) and referent elevation points. The global DEMs have been used as multiple independent variables (explanatory variables) and reference points are dependent variable in the GWR model. Firstly, reference points with more than 130,000 elevation points were separated randomly into two groups. One group was used for estimation DEM and the remaining group was used for validation of estimated DEM. This was done using the *r.random* function in GRASS GIS. Many cases with different number of reference points were applied for running GWR in DEM estimation. The different estimated DEMs were evaluated in relation with slope, landform, and landuse. As the results, estimated DEM used random 50.000 input points showed the best result. Therefore, we selected random 50,000 reference points for DEM estimation and the remaining points for validation. The workflow and simulation of GWR model used in this study are shown in Figure 2 and Figure 3.

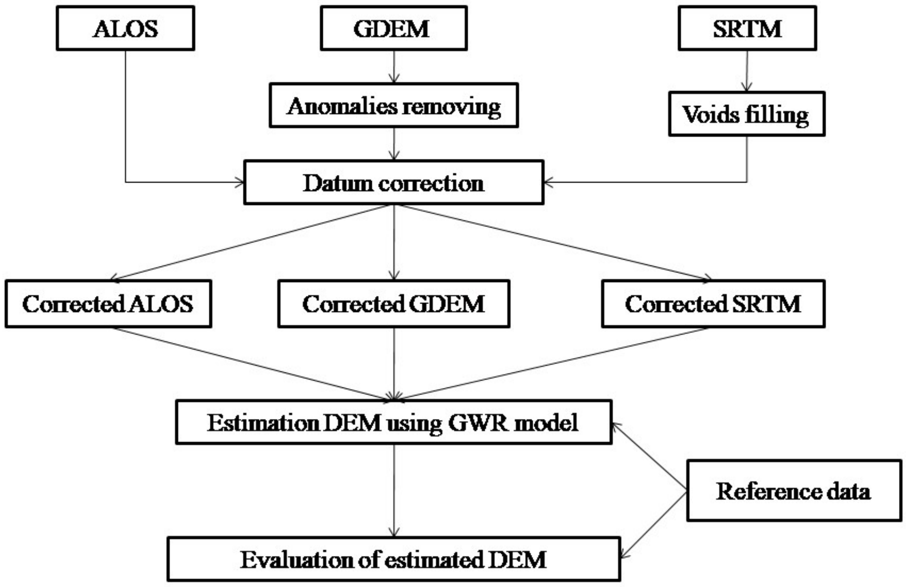

Figure 2. The study workflow.

In GWR model, the terms of bandwidth and kernel function have effect on the estimation result. Various conditions of bandwidth and kernel were applied and bi-square function (equation 1) with fixed bandwidth size 50 showed the best estimated DEM.

$$w_p = (1 - (d/bw)^2)^2 \quad (1)$$

Where, *bw* is the bandwidth, *d* is the distance from a pixel to the current pixel, and w_p is the weight assigned to a pixel. GWR is available as module *r.gwr* in GRASS GIS. As a result, the estimated DEM which derived from three global DEMs and the number of 50,000 reference elevation points was generated (Figure 3).

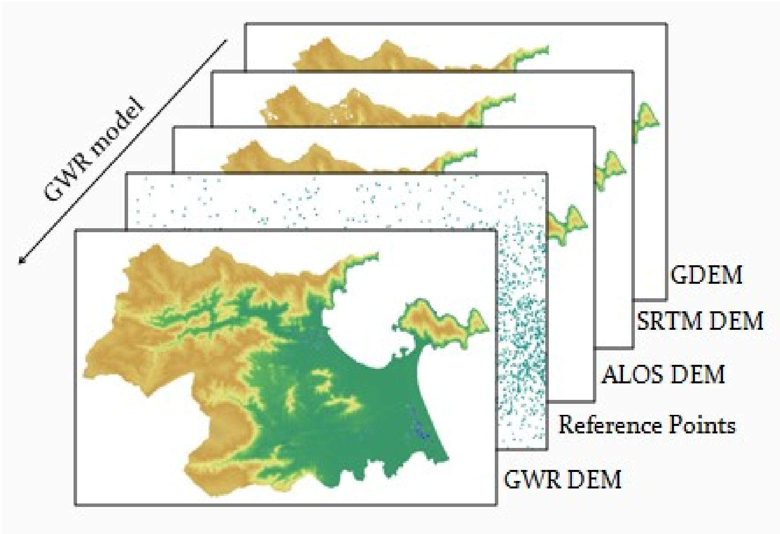

Figure 3. GWR model used in DEM estimation.

3. RESULTS AND DISCUSSIONS

The remaining group of reference elevation points including 80000 points was used for evaluation of the GWR derived DEM. The accuracy of the estimated DEM can be determined by statistical analysis such as MAE, RMSE, Standard Deviation (SD) and correlation coefficient (R^2). The MAE, RMSE and SD of the estimated DEM were much improved compared to the global DEMs. The RMSE was reduced from 7.0 m in ALOS-30 to 4.1m in the GWR DEM. For GDEM and SRTM, the RMSE was reduced from 8.3m in the GDEM and 6.1 m in SRTM to 4.1m in the GWR DEM. The MAE was also reduced from 5.2 m in ALOS-30, 5.4 m in GDEM and 4.6 m in SRTM to 1.9 m in the estimated DEM (Table 1).

The correlation coefficient (R^2) between the GWR DEM and reference data also shows the significant correlation, with $R^2 = 0.9994$ (Figure 4). Comparing to global DEMs has some outliers with R^2 for ALOS, GDEM and SRTM of 0.9985, 0.9973 and 0.9987, respectively. It can be, therefore, concluded that the estimated DEM shows better correlation with the reference data. Statistical comparison of vertical accuracy of ALOS-30, GDEM, SRTM and the estimated DEM is shown in Table 1. The MAE, RMSE and SD of the estimated DEM show much improvement compared to ALOS, GDEM and SRTM before estimation.

Table 1. Summary statistics for the error of global DEMs and GWR DEM

DEM	MAE (m)	RMSE (m)	SD (m)	R^2
ALOS-30	5.2	7.0	5.9	0.9985
GDEM2	5.4	8.3	8.2	0.9973
SRTM	4.6	6.4	5.5	0.9987
GWR DEM	1.9	4.1	4.1	0.9994

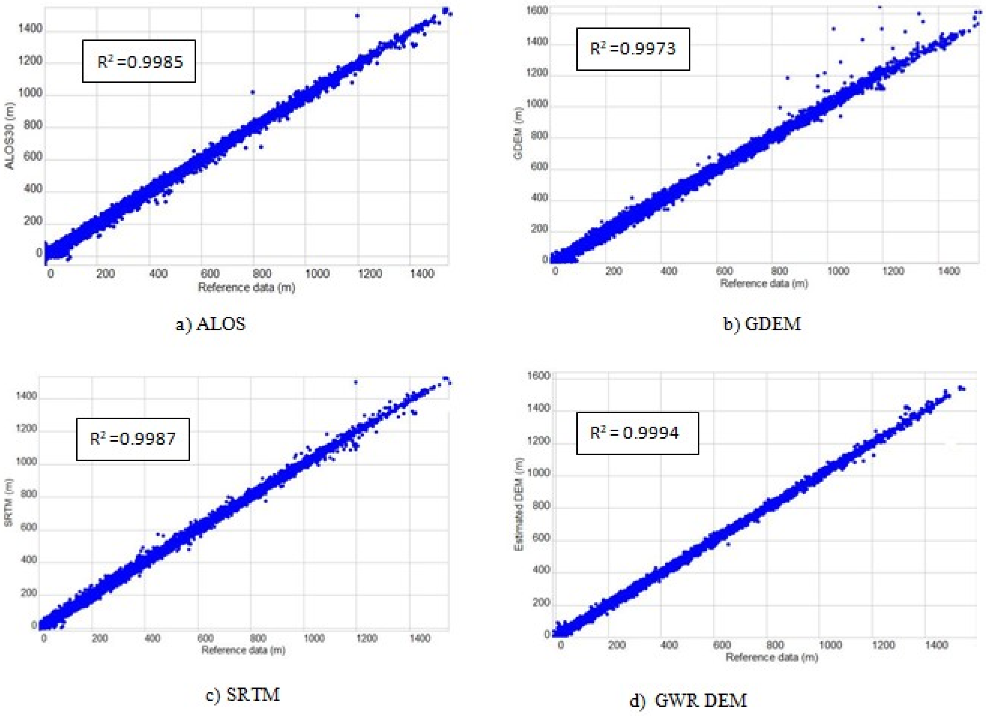

Figure 4. Correlations between the estimated DEM, global DEMs and reference data.

4. CONCLUSIONS

In this study, GWR model was applied successfully for surface elevation estimation for Danang area. Results indicate that the estimated DEM has improved accuracy compared to individual global DEMs. Therefore, estimation of an elevation model based on field survey and satellite-derived elevation data using GWR is effective for area with limited number of survey point data. Fixed bandwidth type was used for running GWR in this case and gave good result. However, adaptive bandwidth type should be applied and compared to the result of fixed bandwidth to find out the better method.

5. REFERENCES

ASTER GDEM Validation Team (2011) ASTER Global Digital Elevation Model Version 2 – Summary of Validation Results August 31, 2011

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E. (1996) Geographically Weighted Regression: a method for exploring spatial non-stationarity. *Geographical Analysis, Volume 28,* pp. 281-298

Fotheringham, AS., Brunsdon, C. and Charlton, M., (2002) *Geographically Weighted Regression: the analysis of spatially varying relationships,* Wiley, Chichester, pp. 13-15.

Nguyen, L. N. and Le, H. V. (2002) Methods for positioning ellipsoid over the Vietnam area based on EGM96 model, *Ho Chi Minh University of Technology*, Ho Chi Minh, Vietnam, pp. 1–9, 2002.