MAPPING OUR RIVERS.
AN ACCURACY ASSESSMENT OF FEATURE EXTRACTION

Daniel Mostrales, Ma. Teresa Ignacio, Ruben Tudio, Florife Liwanag and Vincent Malales[1]
[1]Mindanao State University – Iligan Institute of Technology, A. Bonifacio Ave., Tibanga, Iligan City, Philippines
Email: daniel.mostrales@g.msuiit.edu.ph

KEY WORDS: Stream networks, LiDAR, digital elevation model, feature extraction, accuracy assessment

ABSTRACT
Mapping of rivers and streams and their watersheds are prerequisite in assessing water resources in an area. Mapping our resources is greatly facilitated with the acquisition of LiDAR data in our country. However, LiDAR data coverage is still limited and not available for a complete detailed feature extraction for most of our watersheds. This study aims to provide an accuracy assessment that would indicate relative accuracy that will be expected if one has limited access to DEMs with higher resolution. Five stream delineations were performed using a) 1-m resolution LiDAR-derived DEM, b) 5-m resampled LiDAR-derived DEM c) 10-m resampled LiDAR-derived DEM, d) 10-m resolution SAR DEM, and e) digitized streamlines using Google Earth images. Resulting streamlines are evaluated for comparative relative accuracy to a reference streamline. ArcHydro toolset of ArcGIS 10.2.2 was used to delineate watersheds and stream networks on a uniform threshold area of 1 sq. km. Sample lengths of streamlines were selected in two similar watersheds of Northern Mindanao, Philippines. While LiDAR-derived DEMs provides greater accuracy in streamline delineation, Google Earth images provides more realistic positions of streamlines than when using a 10-meter resolution SAR DEM. This paper also reports preliminary results of extracting hydrologic features of watersheds and stream networks in two provinces of Northern Mindanao.

1. INTRODUCTION

Hydrologic simulation models and water resources planning tools often use hydrographic datasets (stream network polylines, watershed boundaries, etc.) which can be derived from gridded (raster) DEMs using well established terrain analysis techniques. DEMs are, in turn, derived from a number of sources including 30-m resolution ASTER GDEM, 10-m SAR and in recent years in the country, a new source of data from LiDAR-derived DEMs at a higher 1-m resolution are available.

1.1 LiDAR Technology

LiDAR technology offers a relatively efficient way to produce DEMs for a variety of large-scale, high-accuracy mapping applications. LiDAR sensors are capable of receiving multiple laser pulse “returns” which, when combined with precision GPS location data can provide highly accurate and dense point sample measurements of terrain 3-D position. In this way, LiDAR can be used to define a detailed representation of the earth’s surface horizontally as well as vertically, making the LiDAR data source increasingly important for surface structure derivation and giving particular appeal to its use in hydrographic feature extraction. Consequently, streamlines extracted from a LiDAR-derived DEM have been shown to have a more complex morphology and correspond better with field-mapped networks than those derived from a conventionally produced DEM (Barber and Shortridge, 2005).

In the Philippines, LiDAR data has found applications not only in flood hazard mapping but also in the assessment of natural resources including agriculture, coastal, forestry, hydrologic datasets and the identification of potential hydropower generation sites. Available LiDAR-derived DEMs have a relative accuracy of 0.2 to 0.5 m.

1.2 Focus of Study

This study investigates the relative accuracy of using less detailed source of data for extracting stream networks. Comparison technique that is used is the longitudinal root-mean-square-error (LRMSE), to assess the horizontal variance between two polyline data sets representing reference and derived networks.

A reference streamline is assumed to be the best representation of a particular streamline. From this best available representation of streamlines, one can estimate the error contained within other streamline by comparing them to the reference streamline. In some cases, it may not be critical that the modeled or derived dataset perfectly matches the reference dataset, as long as it is a better match than is another dataset.

The reference streamline that is used in this study is a smoothed streamline extracted using 1-m LiDAR-derived DEM. Nearest distances are then calculated at random points created along the reference streamline to streamlines extracted using a) 1-m resolution LiDAR-derived DEM (denoted herein as L1), b) 5-m resampled LiDAR-derived DEM (L5), c) 10-m resampled LiDAR-derived DEM (L10), d) 10-m resolution SAR DEM (SAR), and e) digitized streamlines using Google Earth images (GE). 5-m and 10-m resolution LiDAR-derived DEMs were resampled from 1-m LiDAR-derived DEM.

Selected portions of streamlines in two different watersheds of Northern Mindanao and of three different average streamline slopes (Slope 1: 0.006, Slope 2: 0.02, and Slope 3: 0.06) were investigated for relative accuracy to a reference dataset.

2. BACKGROUND

Studies have been conducted on the effect of DEM resolution on hydrology-related. Vaze and Teng (2007b) present results from an investigation in which they re-sampled a 1 m LiDAR-derived DEM in steps (2, 5, 10, and 25 m) and compared the different spatial indices derived from these different resolution DEMs against the ones derived from the 1 m LiDAR-derived DEM. They reported that re-sampling to coarser grid cell sizes, which is equivalent to averaging across increasingly larger domains, results in an increasing loss of detail in the topography.

Techniques for generating DEM data from LiDAR have been greatly improved. With respect to the use of the LiDAR-derived DEMs for hydrologic modeling, Murphy et al. (2007) compared stream network modeling results using LiDAR and photogrammetric derived digital elevation which reveals that a flow network modeled from the LiDAR-derived DEM was most accurate.

RMSE is the square root of the average of the set of squared differences between dataset coordinate values and coordinate values from an independent source of higher accuracy for identical points. Positional errors, also known as displacements or distortions, are understood as the differences between the measured and the assumed true coordinates. RMSE is a useful index of errors in continuous variables. For n points with errors ε_i (i = 1, 2, …, n), observed as the differences in coordinates between the data sets to be tested and the more accurate reference data, the RMSE is

$$RMSE = \left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\right)^{1/2} \quad (1)$$

where the errors ε_i is the distance between a test or modeled data point (X_i, Y_i) and a corresponding reference data point (Xo_i, Yo_i). In other words, for Cartesian coordinates,

$$\varepsilon_i = d_i = [(Xo_i - X_i)^2 + (Yo_i - Y_i)^2]^{1/2} \quad (2)$$

We have chosen to define LRMSE as the horizontal RMSE computed between a number of paired sets of points located along both derived and reference stream network polylines. Thus,

$$LRMSE = \left(\frac{1}{n}\sum_{i=1}^{n}[(Xo_i - X_i)^2 - (Yo_i - Y_i)^2]^{1/2}\right)^{1/2} \quad (3)$$

Several commands in ArcGIS 2.2.2 could calculate distance and additional proximity information between the input features and the closest feature in another layer or feature class. The distance between any two features is calculated as the shortest to each other. Calculating LRMSE between two polyline data sets that are selected within the GIS environment considers one as the derived stream network and the other would be the reference stream network. For each polyline in the reference data, the reference polyline is divided into m number of segments between n evenly spaced points, where $m = n - 1$. Then, for each reference point, the nearest point on the derived polyline is identified and the (d_i) from that point on the derived polyline to the current point on the reference polyline is calculated. LRMSE is then calculated as

$$LRMSE = \left(\frac{1}{n}\sum_{i=1}^{n}d_i^2\right)^{1/2} \quad (4)$$

Google Earth now hosts high-resolution imagery that is rapidly expanding, cost-free and largely unexploited resource for scientific inquiry. Google Earth imagery has a horizontal accuracy that is sufficient for assessing moderate-resolution remote sensing products across most of the world's peri-urban areas (Potere, 2008).

3. STUDY AREA

One of the major components of an ongoing nationwide program using LiDAR data is the establishment of a national hydrologic dataset that includes gathering information on watershed boundaries, stream networks, inland wetlands and irrigation systems. The acquisition of LiDAR data was primarily started for flood hazard mapping and is mostly along the coastal area and does not cover the whole area of most of our watershed areas.

The image below shows some of the major river basins that are covered in Northern Mindanao. The shaded area indicates the extent of LiDAR coverage in this part of the country.

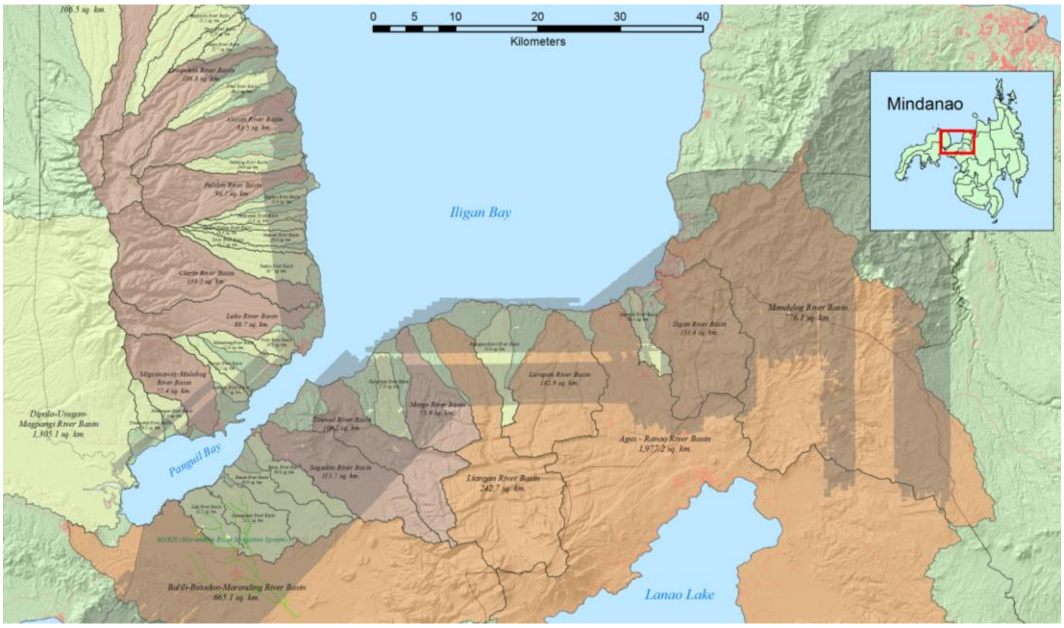

Figure 1. Some watersheds and extent of LiDAR coverage of the area.

For most of the watersheds, fusion of coarser resolution of elevation data (10-m resolution SAR DEM) was needed to completely process a watershed including its boundary and river networks. It has been observed that the use of LiDAR data greatly enhances the accuracy of delineation both on the watershed boundary and its stream networks.

However, in other areas where LiDAR data is available, there is a need to ascertain the level of accuracy when other sources of data, particularly Google Earth imagery is used to delineate stream networks. This study is an initial assessment on the accuracy of using other sources, including the use of resampled LiDAR-derived DEMs.

4. METHODOLOGY

4.1 Case Study Sites

Segments of streamlines were taken from two watersheds of about the same watershed area are shown below on an elevation map.

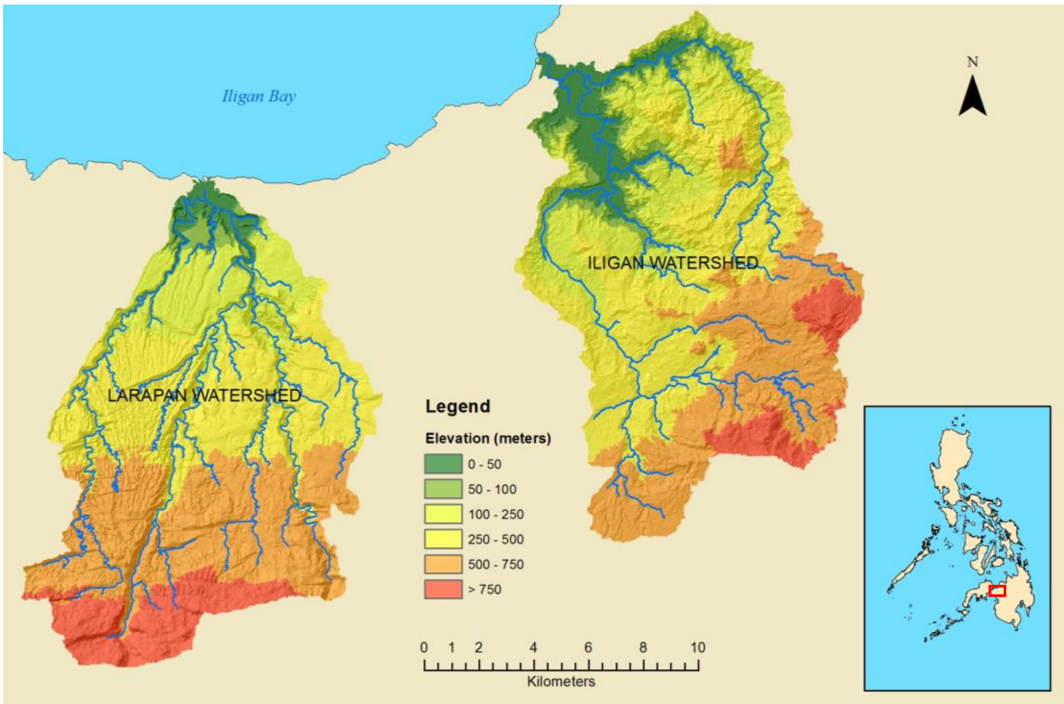

Figure 2. Location and elevation map of the case study sites.

Iligan Watershed

Iligan watershed is the second watershed from the right as shown in Figure 1. It is about 154 sq. km in area and ranges up to a maximum elevation of about 1,000 m. It drains towards a relatively urban area of Iligan City.

Larapan Watershed

Larapan watershed is the fourth major watershed from the right as shown in Figure 1. It is about 143 sq. km. in area and ranges up to a maximum elevation of about 1,040 m.

4.2 Assessment Method

Three segments of streamlines were used from the two watersheds having different slope conditions. Segments were chosen as extracted from available LiDAR data and recognizable from Google Earth images. The following table shows notations used and some characteristics of streamline segments.

Table 1. Notations used for reference.

Location	*Notations Used*	*DEM Used*
Iligan Watershed	IL_REF	
	IL_L1	1-m LiDAR-derived DEM
	IL_L5	5-m resampled LiDAR DEM
	IL_L10	10-m resampled LiDAR DEM
	IL_GE	(digitized from Google Earth)
	IL_SAR	10-m SAR DEM
Larapan Watershed	LP_REF	
	LP_L1	1-m LiDAR-derived DEM
	LP_L5	5-m resampled LiDAR DEM
	LP_L10	10-m resampled LiDAR DEM
	LP_GE	(digitized from Google Earth)
	LP_SAR	10-m SAR DEM

Table 2. Some properties of streamline segments used for accuracy assessment.

Location		Length, m	Slope (rise/run)
Iligan Watershed	Segment 1	3,150	0.0055
	Segment 2	8,291	0.0164
	Segment 3	1,072	0.1120
Larapan Watershed	Segment 1	4,340	0.0060
	Segment 2	3,911	0.0231
	Segment 3	1,030	0.0322

For each segment, there is an equivalent length of streamlines extracted from different sources of DEMs including the digitized streamline from Google Earth. Reference segments that are considered in this study are derived from smoothed streamlines from 1-m resolution LiDAR-derived DEM.

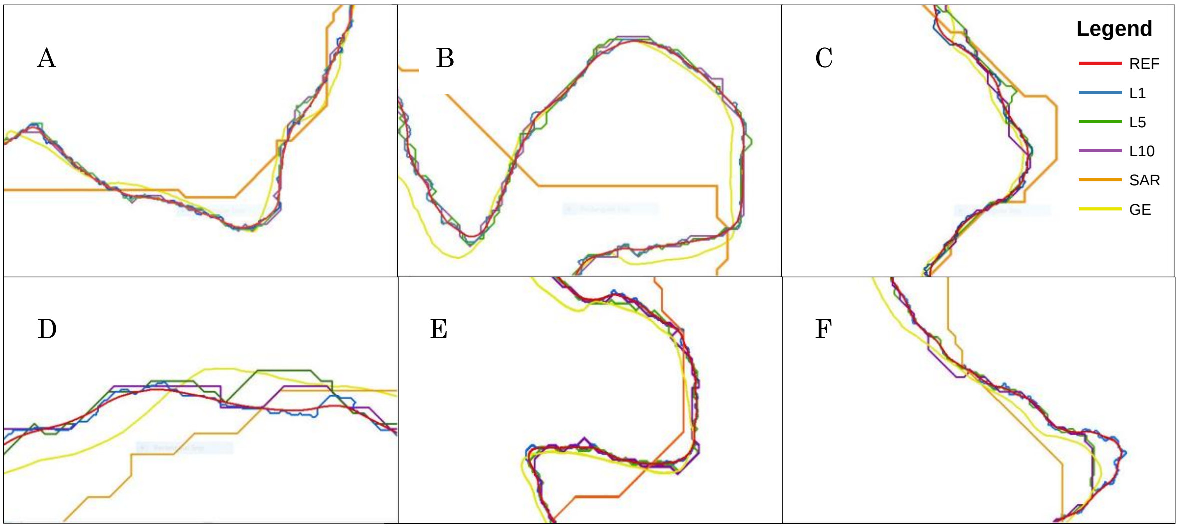

Figure 3. Samples of reaches along selected segments of streamlines. In Iligan watershed (A – Segment 1, B – Segment 2, C – Segment 3) and Larapan watershed (D – Segment 1, E – Segment 2, F – Segment 3).

GENERATE NEAR TABLE command of ArcGIS 10.2.2 and an equivalent conversion command to Excel file were used to generate a table of distances calculated from random points located on the reference segment. Corresponding LRMSEs were calculated for each streamline to the reference streamline.

5. RESULTS AND DISCUSSIONS

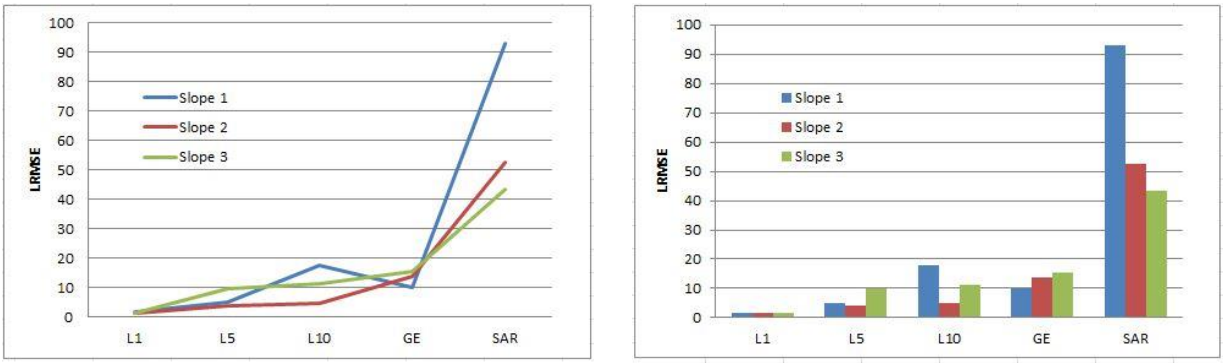

Figure 4. Results of LRMSE calculations from selected streamlines of Iligan watershed.

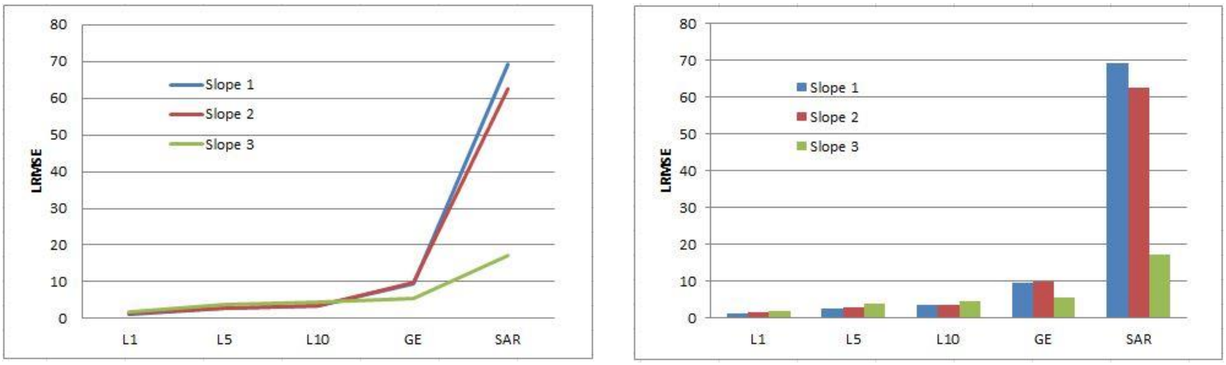

Figure 5. Results of LRMSE calculations from selected streamlines of Larapan watershed.

It is expected in this study that minimum LMRSE values with streamlines derived from 1-m LiDAR-derived DEM since the reference streamline was directly derived from this. The accuracy of reference streamlines could have been improved with other ortho-rectified aerial images.

The use of resampled Lidar DEMs shows consistent retention of streamline features even at 10-m resolution. The use of resampled DEMs facilitates faster processing and lesser computer memory requirement.

Streamlines derived from 10-m resolution SAR DEM shows the largest values of LRMSEs. Visual inspection indicates that it misses the position of the actual streamlines with large differences especially in flat terrain areas. Results of the study indicates that lesser LRMSE are possible with streamlines located at areas of greater slope.

Streamlines digitized from Google Earth images have a much less LRMSE (when compared to that of SAR DEM). In some cases, this has equivalent LRMSE values to streamlines extracted from 10-m LiDAR-derived DEM. There are limitations, however, in areas where streams are continuously covered with trees and foliage.

6. CONCLUSIONS

From the results of this study, the use of LiDAR-derived DEMs including those that are resampled to a coarser resolution facilitates extraction of streamlines that are closer to the reference streamline. This study could be improved with the use of more enhanced reference streamlines.

The use of Google Earth imagery provides an alternative process in extracting streamlines where no LiDAR data is available. Limitations include difficulty of recognizing hidden stretches of streamlines in areas where it is covered with trees and heavy foliage.

A more detailed and accurate extraction of streamlines in our country could be greatly improved if LiDAR data is made available to cover all of our countryside.

ACKNOWLEDGEMENTS

This paper was made through the involvement of the presenters in a nationwide Phil-LiDAR 1 and 2 program that is making use of LiDAR data for flood hazard mapping and detailed resources assessment and mapping activities which is supported and funded by the Department of Science and Technology.

REFERENCES

Anderson, D.L., 2012. Detailed Hydrographic Feature Extraction from High-LiDAR Data. Retrieved from https://inldigitallibrary.inl.gov/sti/5553757.pdf

Barber, C.P., and A. Shortridge, 2005. LiDAR Elevation Data for Surface Hydrologic Modeling. Resolution and Representation Issues. Cartography and Geographic Information Science, pp. 401-410.

ESRI, 2016. ArcGIS Online Help. Environmental Systems Research Institute, Inc., Available on http://desktop.arcgis.com/en/arcmap/

Foote, K.E. and D.J. Huebner, 1995. Error, Accuracy, and Precision. The Geographer's Craft Project, Department of Geography, University of Texas at Austin. Retrieved from http://www.colorado.edu/geography/gcraft/notes/error/error.html

Murphy, P. N. C., J. Ogilvie, F. Meng, and P. Arp (2007), Stream network modeling using Lidar and photogrammetric digital elevation models: A comparison and field verification, Hydrological Processes, 22(12), 1747–1754.

Potere, David, 2008. Horizontal Positional Accuracy of Google Earth's High-Resolution Imagery Archive. Retrieved from www.mdpi.com/journal/sensors/sensors-08-09973.pdf

Smith, J., 2010. Optimizing Remote Sensing for Disaster Monitoring, Retrieved from http://www.remotesening.net/index.htm

Vaze, J., and J. Teng (2007), Impact of DEM Resolution on Topographic Indices and Hydrological Modelling Results. Retrieved from http://www.mssanz.org.au/MODSIM07/papers/12_s27/ImpactofDEM_s27_Vaze_.pdf