Annual Academic Sessions 2012
ISSN 2012-9912; ©2012 Open University of Sri Lanka
Open University Of Sri Lanka

FITTING GENERALIZED LAMBDA DISTRIBUTION TO THE DIFFERENCE OF SPATIAL ELEVATIONS DATA IN SRI LANKA

S. R. Gnanapragasam[1,20] and K. Perera[2]

[1]*Department of Mathematics and Computer Science, The Open University of Sri Lanka*

[2]*Department of Engineering Mathematics, University of Peradeniya*

INTRODUCTION

Two spatial elevation data sets Shuttle Radar Topography Mission (SRTM) and Contour are used to evaluate their difference in elevation. Both have four different datasets, which are corresponding to *Paddhiruppu*, *Kegalle*, *Badulla*, and *Katharagama* locations in Sri Lanka. SRTM data follow the shape of the actual ground level but not the actual elevation and it is pushed up or down. SRTM data has more elevation data, and it is available for all locations in Sri Lanka.The Contour dataset is obtained from 1: 50,000 maps of Survey Department created based on the actual ground survey data using ArcGIS software. SRTM data are available in 90m×90m squared grid and therefore other data are also prepared in the same way. Extensive faulting and erosion over time have produced a wide range of topographic features. Three clusters are distinguishable by elevation: the Central Highlands which elevation is over 200m above sea level, the Plains between 30 and 200 meters, and the Coastal Belt which elevation is less than 30m. Most of the island's surface consists of plains (Wikipedia, 2012).

The Generalized Lambda Distribution (GLD) is a generalization of three-parameter lambda distribution and it is a four- parameter family that has been used for fitting distributions to a wide variety of datasets. Ramberg and Schmeiser (1974) generalized Tukey's lambda distribution by introducing one more parameter. They used the under-lying distribution to include unimodal asymmetric distributions and to provide an algorithm for generating asymmetric random variables. Ramberg *et al.* (1979) indicated the usefulness of the distribution for representing data, especially when the underlying model is unknown.
The objective of this study is defining the Generalized Lambda Distribution for the difference of both SRTM and Contour data. This is to find out whether there exists any pattern or shape in common within a cluster. If the difference value is known in a particular location in a cluster then SRTM value can be added to get the estimated contour value.

METHODOLOGY

SRTM and Contour are secondary datasets, which are arranged in matrix form. The value in a cell represents the elevation of the 90m×90m squared area. The formal three standard tests; Shapiro-Wilk test, Kolmogorov-Smirnov test and Anderson-Darling test, are applied to test the normality of difference of elevations. . If they do not follow the normal distribution, fitting the GLD will be more appropriate, since GLD covers different curve shapes.

Percentile Method to estimate the parameters in Generalized Lambda Distribution

The percentile method is used for parameter estimation in GLD. For a given dataset, $X_1, X_2, ..., X_n$, let u be a number between 0 and $\frac{1}{4}$ and $\hat{\pi}_p$ denote the $(100p)^{th}$ percentile of the data. It defines the four sample statistics, $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$, $\hat{\rho}_4$ by

[20] Correspondences should be addressed to Mr. S. R. Gnanapragasam, Department of Mathematics and Computer Science, The Open University of Sri Lanka (email: reyalt@ou.ac.lk)

$$\hat{\rho}_1 = \hat{\pi}_{0.5} \; ; \quad \hat{\rho}_2 = \hat{\pi}_{1-u} - \hat{\pi}_u \; ; \quad \hat{\rho}_3 = \frac{\hat{\pi}_{0.5} - \hat{\pi}_u}{\hat{\pi}_{1-u} - \hat{\pi}_{0.5}} \; ; \quad \hat{\rho}_4 = \frac{\hat{\pi}_{0.75} - \hat{\pi}_{0.25}}{\hat{\rho}_2} \quad \ldots\ldots (1)$$

Also

$$\lambda_2 = \frac{(1-u)^{\lambda_3} - u^{\lambda_4} + (1-u)^{\lambda_4} - u^{\lambda_3}}{\hat{\rho}_2} \ldots\ldots (2) \qquad \lambda_1 = \hat{\rho}_1 - \frac{\left(\frac{1}{2}\right)^{\lambda_3} - \left(\frac{1}{2}\right)^{\lambda_4}}{\lambda_2} \ldots\ldots (3)$$

Here for ease of discussion $u = 0.1$ was assumed momentarily. Karian and Dudewicz (1999) proposed the following algorithm to estimate the parameters of *GLD.*

Algorithm for fitting a GLD, Percentile method

(i). Use (1) to compute $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4$.

(ii). Find the entry point in the table (Tables for GLD Fits: Method of Percentiles) closest to $(\hat{\rho}_3, \hat{\rho}_4)$; if $\hat{\rho}_3 > 1$, use $\left(1/\hat{\rho}_3 , \hat{\rho}_4 \right)$ instead of $(\hat{\rho}_3, \hat{\rho}_4)$.

(iii). Using the entry point from Step (ii), extract $\hat{\lambda}_3$ and $\hat{\lambda}_4$ if $\hat{\rho}_3 > 1$, interchange $\hat{\lambda}_3$ and $\hat{\lambda}_4$.

(iv). Use $\hat{\lambda}_3$ for λ_3 and $\hat{\lambda}_4$ for λ_4 in (2) to determine $\hat{\lambda}_2$ (Remembering that u should be set 0.1 in (2) for use of the tables).

(v). Use $\hat{\lambda}_2$ for λ_2, $\hat{\lambda}_3$ for λ_3 and $\hat{\lambda}_4$ for λ_4 in (3) to obtain $\hat{\lambda}_1$.

RESULTS AND DISCUSSION

Hereafter throughout this paper dataset 51 will refer *Paddhiruppu* dataset, dataset 53 will refer *Kegalle* dataset, dataset 69 will refer *Badulla* dataset and dataset 83 will refer *Katharagama* dataset. The Table 1 consists of the p-values of three normality tests, used to test the null hypothesis Ho: difference of both elevation datasets follow the normal distribution.

Dataset	Kolmogorov- Smirnov (D) (p- value)	Shapiro-Wilk (W-sq) (p- value)	Anderson- Darling (A-sq) (p- value)
51	<0.0100	<0.0050	<0.0050
53	<0.0100	<0.0050	<0.0050
69	<0.0100	<0.0050	<0.0050
83	<0.0100	<0.0050	<0.0050

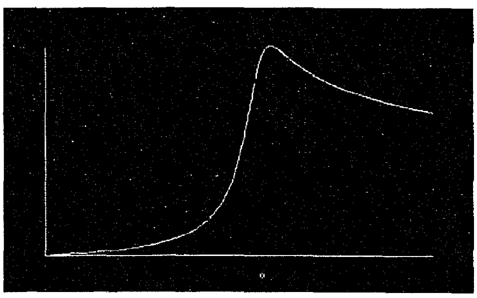
Table 1 p- values of the normality test

According to the Table 1, the p–values of all three tests are less than 0.05. Therefore Ho is rejected at 5% significance level. Since they do not follow the normal distribution, fitting the Generalized Lambda Distribution will be more appropriate.

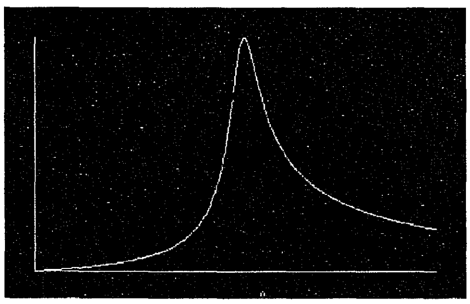
The Table 2 consists of quantile estimates, sample statistics and parameter estimates of *GLD* for the difference of two elevation data in all four datasets. According to these parameter estimates, the probability density functions are plotted using the Java Applet which was programmed by King (2000).

Dataset	Quantile Estimates		Sample Statistics		Parameter Estimates	
51	10%	-13.00	$\hat{\rho}_1$	-3.00	$\hat{\lambda}_1$	-3.02
	25%	-8.89	$\hat{\rho}_2$	20.00	$\hat{\lambda}_2$	0.45
	50%	-3.00	$\hat{\rho}_3$	1.00	$\hat{\lambda}_3$	1.31
	75%	2.23	$\hat{\rho}_4$	0.56	$\hat{\lambda}_4$	20.62
	90%	7.00				
53	10%	-15.00	$\hat{\rho}_1$	0.17	$\hat{\lambda}_1$	-5.12
	25%	-7.032	$\hat{\rho}_2$	35.00	$\hat{\lambda}_2$	0.03
	50%	0.170	$\hat{\rho}_3$	0.76	$\hat{\lambda}_3$	2.65
	75%	8.762	$\hat{\rho}_4$	0.45	$\hat{\lambda}_4$	11.50
	90%	20.00				
69	10%	-50.30	$\hat{\rho}_1$	-6.95	$\hat{\lambda}_1$	-0.33
	25%	-23.09	$\hat{\rho}_2$	59.44	$\hat{\lambda}_2$	0.01
	50%	-6.95	$\hat{\rho}_3$	2.70	$\hat{\lambda}_3$	18.85
	75%	2.05	$\hat{\rho}_4$	0.42	$\hat{\lambda}_4$	3.43
	90%	9.14				
83	10%	-23.12	$\hat{\rho}_1$	6.51	$\hat{\lambda}_1$	-4.86
	25%	-8.17	$\hat{\rho}_2$	66.65	$\hat{\lambda}_2$	0.02
	50%	6.51	$\hat{\rho}_3$	0.80	$\hat{\lambda}_3$	2.46
	75%	23.13	$\hat{\rho}_4$	0.47	$\hat{\lambda}_4$	11.61
	90%	43.53				

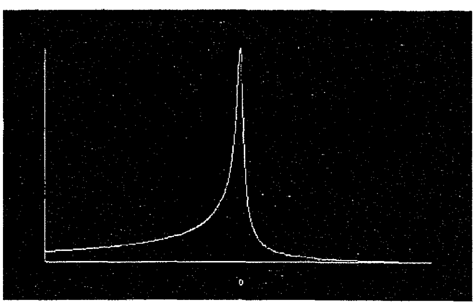
Table 2 Estimates in all four datasets



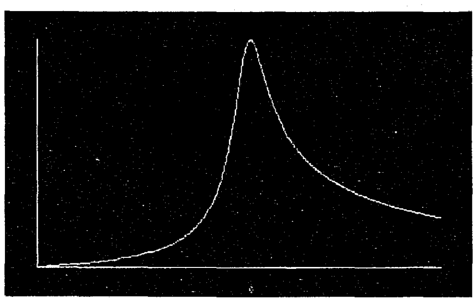
GLD (-3.02, 0.05, 1.31, 20.62)



GLD (-5.12, 0.03, 2.65, 11.50)



GLD (-0.33, 0.01, 18.85, 3.43)



GLD (-4.86, 0.02, 2.46, 11.61)

Figure 1 The graphs of probability density function for the difference in datasets SRTM and

Contour

Figure 1 is the probability density functions; *GLD* (-3.02, 0.05, 1.31, 20.62), *GLD* (-5.12, 0.03, 2.65, 11.50), *GLD* (-0.33, 0.01, 18.85, 3.43), *GLD* (-4.86, 0.02, 2.46, 11.61), of the Generalized Lambda Distribution of difference in elevations data for the dataset *Paddhiruppu*, dataset *Kegalle*, dataset *Badulla* and dataset *Katharagama* respectively.

Identifying the clusters for the datasets

The mean of SRTM data of the dataset 51 is 29.88m, which is less than 30m, it belongs to Coastal Belt cluster. Also the mean of *SRTM* data of the datasets 53 and 83 are 106.54 and 63.30 meters respectively, which are in between 30m and 200m. Therefore, those two datasets belong to Plain cluster. But the mean of SRTM data of the dataset 69 is 1279.10m, which is greater than 200m. Therefore, this dataset belongs to Central Highlands.

CONCLUSIONS/ RECOMMENDATIONS

From the probability density functions in the Figure 1, the graph *GLD* (-3.02, 0.05, 1.31, 20.62) is for dataset *Paddhiruppu* and belong to cluster Coastal Belt, the graph *GLD* (-0.33, 0.01, 18.85, 3.43) is for dataset *Badulla* and belong to cluster Central Highlands.

The graphs *GLD* (-5.12, 0.03, 2.65, 11.50) and *GLD* (-4.86, 0.02, 2.46, 11.61) are for datasets *Kegalle* and *Katharagama* respectively and belong to cluster Plains. Since those two graphs almost follow the same shape and the corresponding data belong to the same cluster. It is concluded that within a cluster the probability density function of the Generalized Lambda Distribution of difference in elevation data follow the same shape and pattern.

This study could further be used to extract the difference data from *GLD* of relevant cluster and could be added with the corresponding SRTM data to predict the contour data in the particular cluster.

REFERENCES

Karian, Z. A. and Dudewicz, E. J. (1999). Fitting the Generalized Lambda Distribution to data: A method based on Percentiles. *COMMUN. STATIST.=SIMULA.*, 28(3), 793-819.

King. R, (2000, November 01). *The Generalised Lambda Distribution Also known as the Asymmetric Lambda Distribution*. Retrieved from http://tolstoy.newcastle.edu.au/~rking/gld/

Ramberg, J. S. and Schmeiser, B. W. (1974). An approximate Method for Generating Asymmetric Random Variable, *Comm. ACM*, 17, 78-82.

Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R. and Mykytka, E. F. (1979). A probability Distribution and Its uses in Fitting Data, *Technometric*, 21, 201-214.

Wikipedia, (2012, February 13). Geography of Sri Lanka. Retrieved from http://en.wikipedia.org/wiki/Geography_of_Sri_Lanka

ACKNOWLEDGEMENTS

The authors express their gratitude to Dr. U. R. Ratnayake, Senior Lecturer, Department of Civil Engineering, University of Peradeniya for providing the datasets to carry out the research work.