Automated transformation of slope and surface curvatures to avoid long tails in frequency distributions

Ovidiu Csillik
Department of Geography
West University of Timişoara
Timişoara, Romania
cskovi@yahoo.com

Ian S. Evans
Department of Geography
Durham University
Durham City, England
i.s.evans@durham.ac.uk

Lucian Drăguţ
Department of Geography
West University of Timişoara
Timişoara, Romania
lucian.dragut@fulbrightmail.org

Abstract — **Automated procedures are developed to change scales so that long tails in frequency distributions of morphometric variables are avoided. They minimize the skewness of slope gradient frequency distributions, and modify the kurtosis of profile and plan curvature distributions towards that of the Gaussian (normal) model. Box-Cox (for slope) and arctangent (for curvature) transformations are tested on nine digital elevation models (DEMs) of varying origin and resolution, and different landscapes, and shown to be effective. Our results show considerable improvements over those for previously recommended slope transformations (sine, square root of sine, and logarithm of tangent). By avoiding long tails and outliers, they permit parametric statistics such as correlation, regression and principal component analysis to be applied, with greater confidence that requirements for linearity, additivity and even scatter of residuals (constancy of error variance) are likely to be met. It is suggested that such transformations should be routinely applied in all parametric analyses of long-tailed variables. Our Box-Cox and curvature automated transformations are based on a Python script, implemented as an easy-to-use script tool in ArcGIS.**

I. INTRODUCTION

For most types of statistical analysis, it is important to check the shape of the frequency distribution of each variable. Many statistical approaches assume that the variables are normally distributed, and a violation of this assumption can lead to errors in analysis. 'Long tails' of values at either extreme or both are the main problem. This is tackled by changing (transforming) the measurement scale. Unfortunately, many environmental science publications overlook the need to apply transformation in this way.

Most slope frequency distributions have the mean and mode usually closer to 0° than to the upper limit of 90°. The lower tail is limited and the upper tail is commonly more extended, giving widespread positive skew. This is common also because, even in mountain or hill regions consisting mainly of slopes, deposition in fans, floodplains and lakes produces extra areas of low gradient, 'fattening' frequencies below the mean. Where these features are absent, however, distributions may be symmetrical or, where high relief pushes gradient toward a limiting value for slope stability, negatively skewed – with a tail extending toward lower values.

Given this natural diversity between regions, there have inevitably been different transformations proposed to rectify slope skewness [1] [2] [3]. Evans [4] favoured no single transformation, but later he inclined to use of the square root of sine [5].

For real-world DEMs, the distribution of curvatures measured in degrees per unit length is always strongly peaked at the mode of zero, and both tails are very long. The presence of extremely positive and negative values makes calculations extremely sensitive to outliers. There is less work on transformation to rectify kurtosis than on skewness: transformation of curvatures to normality is difficult, but can be achieved using a two-sided function such as the arctangent [6].

We propose general solutions to the transformation of surface derivatives, specifically slope gradient and curvatures, so that estimates of statistics such as correlation, regression, analysis of variance and principal component analysis will not be distorted by extreme values. Automated procedures are described for reducing skewness and kurtosis to the parameters of a normal, Gaussian distribution.

II. DATA AND METHODS

A. *Data used*

Tests were conducted on 9 Digital Elevation Models (DEMs) differing in spatial resolution, extent, altitude range, type and landscape. Spatial resolutions ranged from 1m to 90m. The

In: Geomorphometry for Geosciences, Jasiewicz J., Zwoliński Zb., Mitasova H., Hengl T. (eds), 2015. Adam Mickiewicz University in Poznań - Institute of Geoecology and Geoinformation, International Society for Geomorphometry, Poznań

TABLE I. DEM CHARACTERISTICS.

Test Area	Name	Spatial resolution	Method	Scene size (pixels)	Altitude range (m)	Mean altitude (m)	Region	Courtesy of
A	**Slovinec**	1 m	Photogrammetric	306 x 300	141 - 244	187	Slovakia	University of Bratislava
B	**Fishcamp**	2.5 m	LiDAR	638 x 318	1443 - 1833	1658	USA	USGS National Map seamless server
C	**Boschoord**	5 m	LiDAR	1108 x 1079	2 - 20	6	Netherlands	Universiteit van Amsterdam
D	**Țarcu Mountains**	10 m	Topo	905 x 871	1045 - 2195	1706	Romania	
E	**Ebergotzen**	25 m	Topo	398 x 398	159 - 429	272	Germany	State Authority for Mining, Energy and Geology, Germany
F	**Baranja Hill**	25 m	Topo	145 x 147	85 - 244	158	Croatia	Croatian State Geodetic Department
G	**Zlatibor**	30 m	Topo	148 x 98	851 - 1174	991	Serbia	Geodetic Governmental Authority of Serbia
H	**Apuseni Mountains**	90 m	SRTM	412 x 411	404 - 1824	1054	Romania	USGS – Shuttle Radar Topography Mission
I	**Banat Plain and Hills**	90 m	SRTM	1185 x 604	65 - 588	106	Romania	USGS – Shuttle Radar Topography Mission

spatial extent of test areas ranged between 306 × 300 m and 106.7 × 54.4 km. DEM types include photogrammetric, LiDAR, SRTM and those derived from topographic maps (Table 1). Landscape and geomorphologic characteristics vary from very low relief, mixed plain and hilly landscape, hilly areas to mountainous areas.

B. *Automated normalization of LSVs*

Box-Cox transformation [7] is one of the most widely used methods to transform data to approximate the bell-shaped normal (Gaussian) frequency distribution model. It identifies an exponent (lambda, λ) to which all the values should be raised in order to acquire the above-mentioned shape. This is in line with Tukey's 'ladder of transformations'. Note that for λ = 0 slope values are not raised to the power of 0 (because this would be 1 for every value) but a logarithmic transformation is applied (Table 2).

TABLE II. BOX–COX TRANSFORMATIONS: LAMBDA (λ) VALUES USED TO TRANSFORM INITIAL VALUES OF SLOPE (x, IN DEGREES) INTO POSSIBLY NORMALIZED SLOPE (y).

λ	−2	−1	−0.5	0	0.5	1	2
y	$1/x^2$	$1/x$	$1/\sqrt{x}$	$\log x$	$\sqrt{x}$	x	x^2

Applying the Box-Cox transformation for slope angle, the initial skewness (with λ = 1) was compared with that for λ equal to 2 or 0.5. Further, we kept the lowest skewness and compared this iteratively with the next λ value in the Box-Cox transformation scale. Thus we selected the λ to be used in the normalization of slope. To avoid an indeterminate logarithm (λ =0) or division by 0 (λ <0) we added a constant value, (1− *min*), to each value of slope prior to applying the transformations, where *min* is the minimum value of slope. This moves the minimum value of the distribution to 1°, changing only the mean, while keeping standard deviation, skewness and kurtosis.

In order to deepen the analysis, we computed three other slope rasters using formulas available in the literature (eq. 1 – [3]; eq. 2 – [6]; eq. 3 – [4]) and compared the results with those of the Box-Cox transformation:

$$\text{TransformedSlope} = \sin(\text{slope}) \quad (1)$$

$$\text{TransformedSlope} = \text{sqrt}(\sin(\text{slope})) \quad (2)$$

$$\text{TransformedSlope} = \ln(\tan(\text{slope})) \quad (3)$$

For both profile and plan curvature, the formula proposed by Evans [7] was applied:

$$\text{TransformedCurvature} = \arctan(k \times \text{curvature}) \quad (4)$$

where k is a parameter to give normalized curvature with kurtosis close to 0. Arctangent transformation preserves the sign and positive, zero and negative curvatures remains such. Long tails are pulled in symmetrically, depending on the value of k: the higher the k is, the more pulling in. The selection of k values is still a trial-and-error approach and differs from one dataset to another, depending on their characteristics. The automated workflow to select the appropriate value of k starts from k = 0.1. If kurtosis of transformed curvature is less than the initial kurtosis, the iteration continues until a value of k producing kurtosis close to 0 is found.

III. RESULTS

In terms of skewness, the Box-Cox transformation gave the least skewed result for all test areas except one where sine transformation gives skewness of 0.006 instead of 0.224, but here no transformation is needed any way. The log-tangent (Eq. 3) seriously over-transformed seven areas, producing negative skewness. The sine (Eq. 1) is a weak transformation and usually makes little difference. The square-root of sine (Eq. 2) over-transformed the four distributions with skewness < 0.4, but it did improve distributions for two test areas. All slope gradient distributions had |skewness| < 0.28 after Box-Cox transformation, so this can be recommended for general use even if it is not quite optimal in two test areas (Fig. 1).

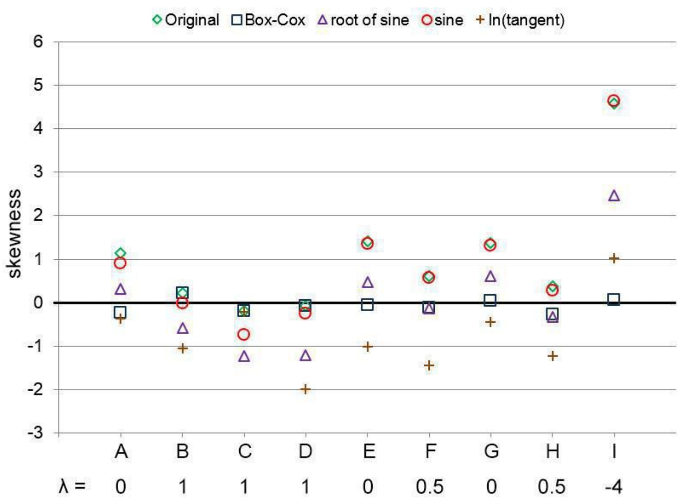

Figure 1. Comparison of skewness (vertical axis) of original slope and transformations according to different formulas for test areas A to I. Lambda (λ) values detected for Box-Cox are displayed below each test area.

For profile curvature, the initial kurtosis varies between 2.46 and 46.24: the higher the kurtosis, the sharper the mode around 0 and the longer the tails. The arctangent transformation brings in both tails, giving finite values mainly between −1.5 and +1.5. All the transformed distributions now have kurtosis values between −0.14 and 0.01, negligibly different from normality, suggesting that the automatically identified k values are meaningful (Fig. 2). This does not necessarily make them optimal, as other aspects of histogram shape should be considered. Use of $2k$ and $k/2$, however, showed much worse results both in kurtosis (Fig. 2) and in other aspects of shape: judged by visual inspection of histograms, the automatically selected k values seem optimal. The skewness is acceptable.

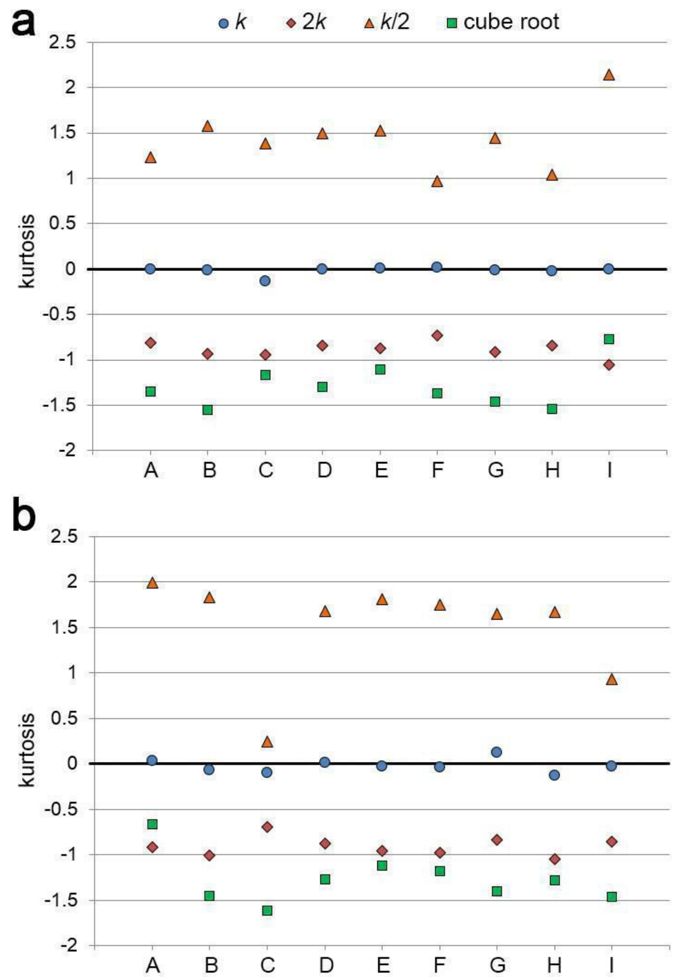

Figure 2. Kurtosis for (a) profile curvature and (b) plan curvature: values after transformation using k and (for comparison) after transformation using $2k$, $k/2$ and cube root.

Depending on the size, spatial resolution and relief characteristics of each test area, the values of k varied between 0.6 and 46.4. Cox [8] suggested that a cube root transform would reduce kurtosis while avoiding the need to select a k value. For these DEMs, however, it produced bimodal histograms with negative kurtosis (Fig. 2).

Plan curvature histograms have longer tails and initial kurtosis ranges from 0.43 to 82.99. As for profile curvature, the arctangent transformation reduced the range of values to between −1.5 and +1.5, the mode is around 0 and only one test area presents important secondary modes. All the transformed histograms of frequency distribution are more nearly Gaussian, with kurtosis between −0.13 and 0.13 with k values between 0.1

and 3.2 (Fig. 2). The higher the kurtosis, the higher the *k* value needed for normalization.

IV. Discussion and Conclusions

We developed a Python tool to perform the transformation of slope gradient and curvatures to near-Gaussian distribution shapes. This study showed that Box-Cox transformation is effective in identifying the appropriate transform of slope gradient in a given area, so that slope skewness can be automatically rectified. The algorithm for arctangent transformation of curvatures is based on the formula proposed by Evans [6], and replaces a trial-and-error determination of a data-dependent parameter *k*, by an iterative tuning towards kurtosis close to 0. Thus, we provide a 'push-the-button' solution to prepare these surface derivatives for statistical (parametric) analysis. Use of a tool such as this is important in any terrain-based environmental analysis where slope gradient and curvatures are statistically related to other variables using parametric techniques (e.g. correlation or regression).

For further information about transformation (normalization) of slope gradient and surface curvatures, please refer to Csillik et al. [9]. The ArcGIS toolbox is available on the page: http://research.enjoymaps.ro/downloads.

Acknowledgment

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS - UEFISCDI, project number PN-II-ID-PCE-2011-3-0499. We are very grateful to Nicholas J. Cox for numerous insightful comments on a draft.

References

[1] Strahler, A.N., 1950. "Equilibrium theory of erosional slopes approached by frequency distribution analysis; Part II". American Journal of Science, 248(11), 800-814.

[2] Chorley, R.J., 1966. "The application of statistical methods to geomorphology", In: G.H. Dury (Ed.), Essays in Geomorphology. Heinemann, London, pp. 275-387.

[3] Speight, J.G., 1971. "Log-normality of slope distributions". Zeitschrift für Geomorphologie, 15, 290-311

[4] Evans, I.S., 1975. "The Effects of Resolution on Gradients Calculated from an Altitude Matrix", Department of Geography, University of Durham.

[5] Evans, I.S., 1977. "Frequency Distributions of Gradient", Department of Geography, University of Durham.

[6] Evans, I.S., 1984. "Properties of the Wessex Land Surface: An Investigation of Dimensionality and the Choice of Key Variables", Department of Geography, University of Durham.

[7] Box, G.E.P., and D.R. Cox, 1964. "An analysis of transformations". Journal of the Royal Statistical Society. Series B (Methodological), 26(2), 211-252.

[8] Cox, N.J., 2011. "Stata tip 96: cube roots" Stata J. 11 (1), 149–154.

[9] Csillik, O., Evans, I.S., Drăguţ, L., 2015. "Transformation (normalization) of slope gradient and surface curvatures, automated for statistical analyses from DEMs". Geomorphology, 232, 65-77.