Geomorphology 232 (2015) 65–77

Contents lists available at ScienceDirect

Geomorphology

journal homepage: www.elsevier.com/locate/geomorph

Transformation (normalization) of slope gradient and surface curvatures, automated for statistical analyses from DEMs

O. Csillik [a,*], I.S. Evans [b,c], L. Drăguţ [a]

[a] Department of Geography, West University of Timişoara, V. Pârvan Blv. 4, 300223 Timişoara, Romania
[b] Department of Geography, Durham University, South Road, Durham City DH1 3LE, England, United Kingdom
[c] Department of Geography, University of Suceava, Universităţii Street 13, 720229 Suceava, Romania

ARTICLE INFO

Article history:
Received 11 August 2014
Received in revised form 29 December 2014
Accepted 30 December 2014
Available online 8 January 2015

Keywords:
Normal distribution
Gaussian
Box–Cox
Histogram
Algorithm
Land surface variables

ABSTRACT

Automated procedures are developed to alleviate long tails in frequency distributions of morphometric variables. They minimize the skewness of slope gradient frequency distributions, and modify the kurtosis of profile and plan curvature distributions toward 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. Resulting histograms are illustrated and show considerable improvements over those for previously recommended slope transformations (sine, square root of sine, and logarithm of tangent). Unlike previous approaches, the proposed method evaluates the frequency distribution of slope gradient values in a given area and applies the most appropriate transform if required. Sensitivity of the arctangent transformation is tested, showing that Gaussian–kurtosis transformations are acceptable also in terms of histogram shape. Cube root transformations of curvatures produced bimodal histograms. The transforms are applicable to morphometric variables and many others with skewed or long-tailed distributions. By avoiding long tails and outliers, they permit parametric statistics such as correlation, regression and principal component analyses 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.

1. Introduction

For most types of statistical analysis, it is important to check the shape of the frequency distribution of each variable. The use of powerful parametric statistics based on ‘least-squares’ requires ‘well-behaved’ data, with relationships that show linearity, additivity and even scatter of residuals. This is interpreted initially as a need for ‘normal’ (i.e. Gaussian) univariate frequency distributions of each variable involved (Chorley, 1966; Stuart and Ord, 1994). Many statistical approaches assume that the variables are normally distributed, and a violation of this assumption can lead to errors in analysis. While this is often necessary for test results to be optimal, the important consideration for applied statistics is whether the deviation from normality is sufficient to distort the statistics calculated: thus, the establishment of linear relations and constancy of error variance is all-important, but is likely to be achieved if univariate distributions can be transformed to ‘normality’. In the extreme, such deviation may involve outliers — single or multiple values which are far removed from the others, or which appear to be inconsistent with the remainder of the data set. Their presence is suspected if values of skewness or kurtosis are high: it may be clear in exploratory graphics, and various precise statistical tests are available (Barnett and Lewis, 1994), once the possibility of data error has been excluded. Even moderately non-normal skewness or kurtosis, however, may distort results for correlation, regression, analysis of variance and principal component 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.

Here we consider the processing of slope gradients and surface curvatures, the first and second derivatives of the altitude field, to avoid misleading statistical results. These derivatives can be measured directly on the ground (Young, 1972), but are now usually calculated from the altitude point values of a square-gridded DEM (Hengl and Reuter, 2009). The increasing availability of high-quality, high-resolution DEMs has made it very easy to generate large data sets of slope and curvature, sampling the whole land surface in a representative and uniform way (Roering et al., 2013; Tarolli, 2014). The results are applied in land classification and many other fields, including hydrology, ecology, soil science and planning, so many scientists who are not geomorphometrists or statisticians analyze data on slope and curvature.

---

* Corresponding author. Tel.: +40 727 103310.
E-mail address: cskovi@yahoo.com (O. Csillik).

http://dx.doi.org/10.1016/j.geomorph.2014.12.038
0169-555X/

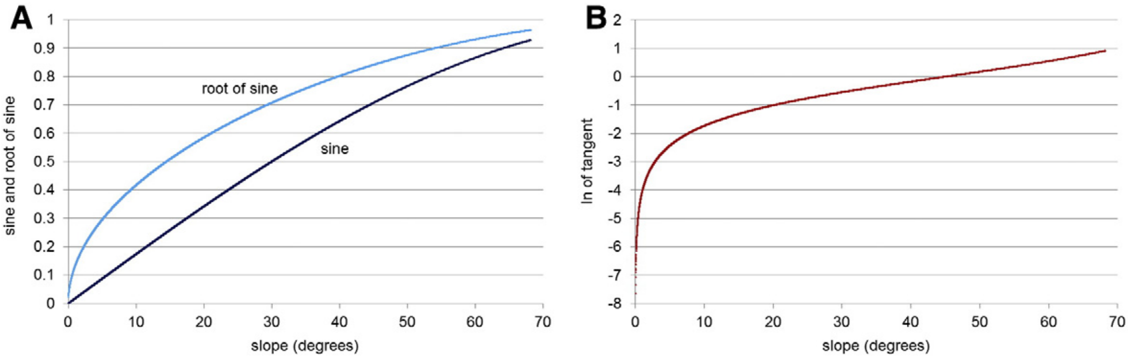

Fig. 1. Transformations proposed to rectify slope skewness: sine, root of sine (a) and ln(tangent) (b). These examples represent slope transformations for test area Slovinec.

First we briefly review previous work on expression of slopes, then of curvatures.

1.1. Previous work on slope gradients or angles

Most geomorphologists express hillslope gradients in degrees (Young, 1964, 1972), giving values with an upper as well as a lower bound. By contrast the percentage or tangent scale of slope gradient extends from zero to infinity (Tricart, 1965), a problem where very steep slopes occur. The tangent scale is used by those working on rivers, following the engineering convention, because steep slopes are rare. Strahler (1950) and Tricart (1965) also pointed out that sine of slope angle is mechanically more important, so it gives a natural gravity-related scale appropriate to process studies.

Most slope frequency distributions are affected by the limiting value of 0°, imposed by definition or by the operational procedure whereby slope is measured from a DEM. As the mean and mode are 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.

In mountain areas where mean and modal slopes exceed 30°, the instability of high slopes of > 35° provides an effective upper limit, and frequency distributions (expressed in degrees) tend to weaken negative skew. For example in the Karakoram and northwest Himalaya, where Burbank et al. (1996) calculated slope for nine regions from a 90 m mesh DEM, all six mountain regions show slight negative (left-) skew. The Deosai Plateau shows positive skew, and Ganugah and 'Edge of Deosai' are intermediate, near-normal. In the Northern Japanese Alps (Honshu), Katsube and Oguchi (1999) used a 50 m mesh DEM and calculated slopes within 500 m bands of altitude. Below 500 m, skew was slightly positive, but for all higher bands it was negative, increasingly so with altitude (the highest band was 2500–3000 m). Oguchi et al. (2011) found very similar relationships for the Central and Southern Japanese Alps and for the Taiwan Central Range: distributions are somewhat negatively skewed above 500 m, and modal slopes are around 35° above 1300 m. All these areas have narrow incised valleys: in areas of high or moderate relief where valleys are wider, with floodplains or terraces, slope frequency distributions are more positively skewed — as they are in lowlands with occasional hills (Evans, 1975, 1977).

Given this natural diversity between regions, there have inevitably been different transformations proposed to rectify slope skewness (Fig. 1). Serious studies of slope frequency distributions began with Strahler (1950), measuring mean and maximum slope on steep hillsides in the Verdugo and San Rafael Hills of southern California. His 'Law of Constancy of Slopes' states that "Within an area of essentially uniform lithology, soils, vegetation and stage of development, maximum slope angles tend to be normally distributed with low dispersion about a mean value…". He favored expression of slope on a sine scale and mapped this, the "magnitude of the downslope component of gravitational acceleration", but he also mapped slope tangents (Strahler, 1950). Chorley (1966) illustrated the positive skew of Strahler's slope tangents and the negative skew of sines, for the Perth Amboy Badlands: here and in some other examples, slope angles are fairly close to normal distribution (Chorley, 1966). Speight (1971) found widespread positive skew and favored the logarithm of tangent as usually providing near-normal distributions, although for his data the sine, tangent and degree scales were little different. The log(tangent) slightly over-transformed (to negative skew), but the square root of sine under-transformed. Evans (1975) favored no single transformation, but later he inclined to use the square root of sine (Evans, 1977). This was also used by Evans (1984), for 54 map sheets in southwest England. Minár et al. (2013)

Table 1
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 × 300	141–244	187	Slovakia	University of Bratislava
B	Fishcamp	2.5 m	LiDAR	638 × 318	1443–1833	1658	USA	USGS National Map seamless server
C	Boschoord	5 m	LiDAR	1108 × 1079	2–20	6	Netherlands	Universiteit van Amsterdam
D	Tarcu Mountains	10 m	Topo	905 × 871	1045–2195	1706	Romania	
E	Ebergotzen	25 m	Topo	398 × 398	159–429	272	Germany	State Authority for Mining, Energy and Geology, Germany
F	Baranja Hill	25 m	Topo	145 × 147	85–244	158	Croatia	Croatian State Geodetic Department
G	Zlatibor	30 m	Topo	148 × 98	851–1174	991	Serbia	Geodetic Governmental Authority of Serbia
H	Apuseni Mountains	90 m	SRTM	412 × 411	404–1824	1054	Romania	USGS — Shuttle Radar Topography Mission
I	Banat Plain and Hills	90 m	SRTM	1185 × 604	65–588	106	Romania	USGS — Shuttle Radar Topography Mission