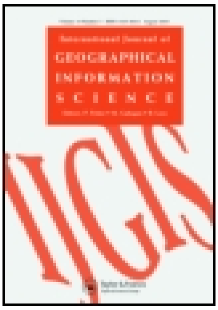


International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgis20

Neighborhood size and spatial scale in raster-based slope calculations

Jing Gao [a], James E. Burt [a] & A-Xing Zhu [a b]

[a] Department of Geography, University of Wisconsin-Madison, Madison, WI, USA

[b] State Key Laboratory of Resources and Environmental Information Systems, Chinese Academy of Sciences, Beijing, PR China

Available online: 27 Feb 2012

To cite this article: Jing Gao, James E. Burt & A-Xing Zhu (2012): Neighborhood size and spatial scale in raster-based slope calculations, International Journal of Geographical Information Science, DOI:10.1080/13658816.2012.657201

To link to this article: http://dx.doi.org/10.1080/13658816.2012.657201

International Journal of Geographical Information Science
i*First*, 2012, 1–20

Taylor & Francis
Taylor & Francis Group

Neighborhood size and spatial scale in raster-based slope calculations

Jing Gao[a]*, James E. Burt[a] and A-Xing Zhu[a,b]

[a]*Department of Geography, University of Wisconsin–Madison, Madison, WI, USA;* [b]*State Key Laboratory of Resources and Environmental Information Systems, Chinese Academy of Sciences, Beijing, PR China*

(*Received 2 August 2010; final version received 9 January 2012*)

Raster-based slope estimation is routine in GIS. Like many other terrain attributes, the slope at a location is determined from elevations of surrounding cells. This spatial extent – 'neighborhood size' – is often treated as the 'spatial scale' of the calculation. In fact, neighborhood size and spatial scale are two connected yet different concepts, but few studies have investigated the relationship between them. The distinction is important because neighborhood size is under user control whereas spatial scale is merely implicit in the computational method. This article attempts to clarify and provide a more precise meaning of the two terms by considering slope operators from the standpoint of the frequency (or wavenumber) domain. This article derives analytical expressions for the amplitude response functions of four popular slope estimators. These are used to characterize the individual methods and also to show that the neighborhood size and spatial scale of a slope calculation are not numerically the same. In fact, because there is no single spatial scale that can be unambiguously associated with a given neighborhood size, neighborhood size cannot be an adequate indicator of spatial scale. Furthermore, this article shows that different indices of 'scale' yield different impressions about the action of a slope estimator and its response to changing neighborhood size. Therefore, it is necessary to examine the amplitude response function when investigating the spatial scale. The article also provides guidance for GIS practitioners when selecting a slope estimation method.

Keywords: neighborhood size; spatial scale; window size; slope; terrain analysis

1. Introduction

Slope is a widely used terrain characteristic and serves as a key input for a diverse array of environmental analysis and modeling. At present, the dominant way to calculate slope is through raster-based terrain analysis, which generates terrain derivatives in orthogonal directions from a gridded digital elevation model (DEM). These derivatives are calculated from elevations within a 'neighborhood' of surrounding cells, and for that reason the neighborhood size affects the spatial scale of features appearing in the resulting raster. Neighborhood size is explicit and under user control, whereas the spatial scale is implicit and neither readily apparent by inspection of the neighborhood nor unambiguously measured. However, spatial scale is far more important than neighborhood for most purposes, which suggests landscape analysts should have a clear understanding of the scales associated with various slope estimation methods. This is especially important given the

*Corresponding author. Email: jgao3@wisc.edu

ISSN 1365-8816 print/ISSN 1362-3087 online

http://dx.doi.org/10.1080/13658816.2012.657201
http://www.tandfonline.com

increasing availability of high-resolution DEMs, whose scales[1] are often much smaller than those of relevant physical processes (Zhu 2008, Zhu *et al*. 2008). Slope estimators, which have the potential to amplify small-scale features, are of particular concern in this regard. The purpose of this article is to clarify the distinction between neighborhood and scale and to compare the action of several popular slope estimation methods in terms of their effect on landscape features having various spatial scales.

From a mathematical standpoint, slope is the gradient vector and, for analytical surfaces, is obtained by maximizing the directional derivative. However, throughout this article 'slope' means the magnitude of the gradient and thus the direction (slope aspect) in what follows is not considered. Most raster GIS processing involves the use of local methods whose extent is typically described in terms of a 'roving window size' (e.g., ESRI® 2010). The roving window is a rectangular area centered on the pixel of interest that moves through the input DEM pixel by pixel during the slope estimation. So, 'roving window size' and 'neighborhood size' convey similar information. However, the former is usually (and in this article) expressed as the number of pixels in each dimension, while the latter is expressed in metric units. Therefore neighborhood size numerically equals the product of input DEM resolution and roving window size; for example, a 3×3 window means a square area with edges of three pixels in length and would correspond to a neighborhood of 30 m when used with a DEM having 10 m resolution. Hereafter 'window size' is used to mean the width and height of the roving window.

The 'spatial scale' of a landform feature is the horizontal extent of the feature. The spatial scale of a rill might be 2 m, while the spatial scale of a hill could be 200 m. Typical landscapes contain multiple scales, but for a particular application only some scales are relevant, and the ideal slope computation would capture those and ignore the others. As is well known, a DEM or any other matrix can be represented using a discrete Fourier series, which decomposes the surface into a finite number of sinusoids of progressively shorter wavelength. In the spatial domain the reciprocal of wavelength is 'wavenumber' with units of inverse length (e.g., m^{-1}). Wavenumber is directly analogous to frequency in the time domain. Although 'frequency' should not technically be applied to spatial fields, this article will follow convention and use the term throughout the discussion. The reader should remember that frequency here means cycles per unit distance rather than cycles per unit time. For the examples above, the frequency of the rill is $1/2\ \mathrm{m}^{-1}$, and the frequency of the hill is one hundred times lower, $1/200\ \mathrm{m}^{-1}$.

If one accepts that Fourier series is an appropriate representation of the land surface, any particular slope calculation will treat information at various scales differently, suppressing some frequencies, amplifying others, and leaving the rest untouched. Therefore, a slope operator is like a digital filter of the input DEM and the 'spatial scale' of a slope calculation refers to the frequencies that pass through the filter and are reflected in the resulting slope matrix. Clearly, the definitions of 'neighborhood size' and 'spatial scale' are very different.

In addition, various slope operators may exhibit different filtering effects on the input information even when using the same neighborhood. To give an example, Figure 1 is a LIDAR-generated DEM of a small part of Mt. Erebus, Ross Island, Antarctica. The spatial resolution is 2 m, and the image contains 106×106 pixels. Figure 2 shows slopes as estimated by four methods discussed in detail later, where Horn's and Evans' methods have a fixed 3×3 window, and 3dMapper®'s and Wood's methods allow user-defined window size. Figure 2 is based on 13×13 and 37×37 windows, respectively. The first two slope maps contain many small-scale features with slopes above 30%, whereas the others give very different impressions. If from a process standpoint one was interested in

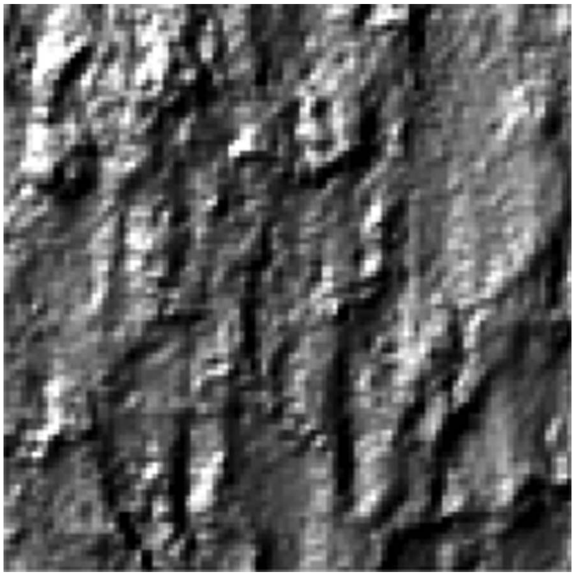


Figure 1. DEM of a part of Mt. Erebus, Ross Island, Antarctica.

scales larger than 25 m, the first two methods would clearly be inappropriate. That said, would 13 $\times$ 13 be the appropriate size? Should one pick a window somewhat larger or smaller than the desired scale? Further, as is seen in Figure 2, given the same window size, different methods provide different estimations. What method–neighborhood pair fits the desired scale the best? Answering these questions requires a more formal analysis of the relation between neighborhood size and scale.

In some ways this article follows Hodgson (1995) who empirically tested the 'best represented cell size' (scale) of three slope calculation methods using a 3 $\times$ 3 window. The major drawback of an empirical methodology, as discussed by Hodgson (1998), is that the method used to determine true values has important implications and the comparisons are therefore inevitably biased. Also, new slope estimators, which allow user-defined neighborhood size, have been developed in the past 10 years. Such methods are motivated by an understanding that relevant spatial scales are not necessarily commensurate with the DEM resolution, but no studies have characterized methods from the scale perspective. This article does so by considering slope estimators from the frequency standpoint. Besides the obvious conceptual appeal, the frequency viewpoint offers analytical tools that can provide new insights on these issues. The article derives analytical expressions for the filtering effects of four popular slope calculation methods and aims to address the following questions: (1) What is the relationship between neighborhood size and spatial scale in commonly used slope estimators? (2) How should one select an appropriate slope operator for a particular GIS application? (3) What constitutes an ideal slope estimator?

2. Slope estimation from frequency perspective

2.1. Frequency approach

As mentioned in the introduction, any function $F(t)$ can be disassembled into simple functions of different frequencies – a Fourier series, provided that (1) $F(t)$ is single-valued and

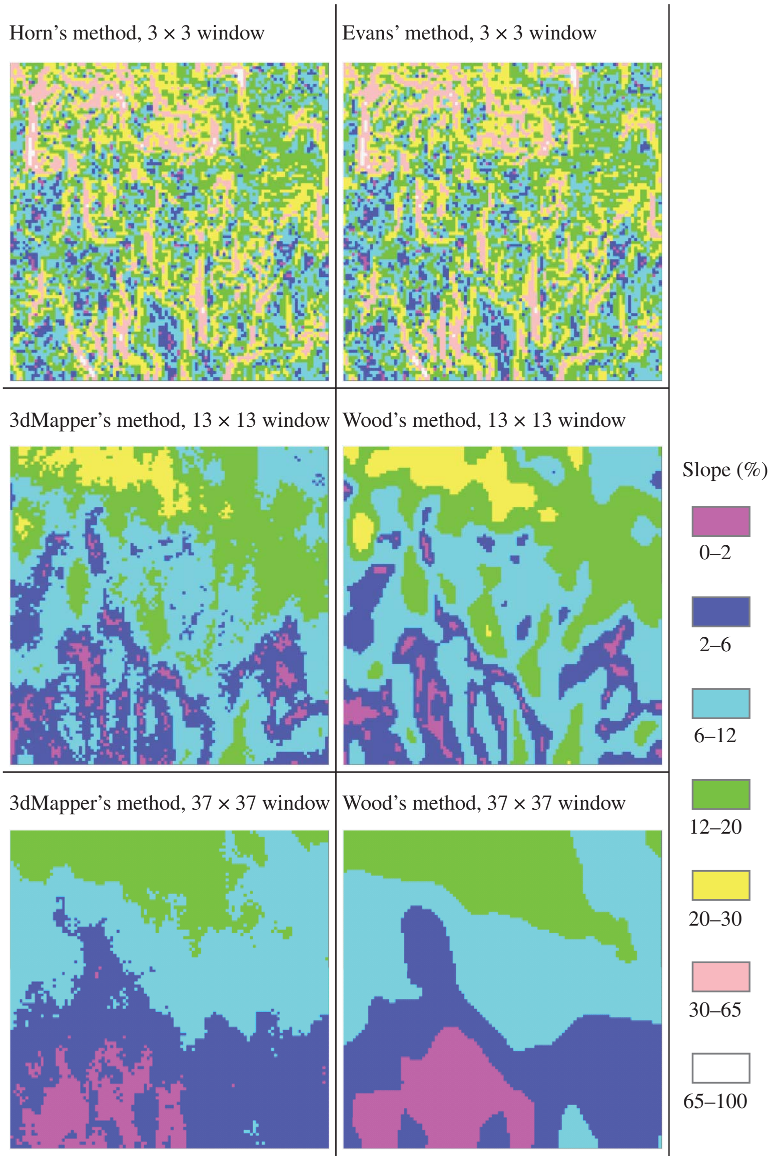


Figure 2. Slopes of Mt. Erebus, using various methods and roving window sizes.

finite, (2) $F(t)$ is defined for every point over some basic interval, and (3) $F(t)$ has a finite number of maxima, minima, and discontinuities in that interval (Rayner 1971, Pinkus and Zafrany 1997, Morgan 2005). For a DEM, the domain is obviously discrete, with F – the elevation – defined at the nodes of the (x, y) grid. Equally obviously, the DEM meets the three requirements. Also, no feature smaller than twice the resolution h of the DEM can be represented in the DEM, thus the highest frequency in the corresponding Fourier series is always $1/(2h) = 0.5f_{\mathrm{DEM}}$ (the so-called Nyquist frequency). Therefore, letting $z_{j,k}$ be the elevation at row j column k, the DEM has a complex representation as

$$z_{j,k} = \sum_{f_x=0}^{0.5f_{\mathrm{DEM}}} \sum_{f_y=0}^{0.5f_{\mathrm{DEM}}} c\left[f_x,f_y\right] \exp\left(2\pi i(jf_x + kf_y)\right) \tag{1}$$

where $i = \sqrt{-1}$, f_x and f_y are frequencies in the x- and y-directions, and $c\left[f_x,f_y\right]$ are complex numbers. Using the Euler relation $c\exp(-i\theta) = c(\cos\theta + i\sin\theta)$ one sees that the exponentials are sinusoids with amplitude and phase determined by the leading coefficients $c\left[f_x,f_y\right]$. The absolute value $|c\left[f_x,f_y\right]|$ is the amplitude of $z_{j,k}$ at $\left[f_x,f_y\right]$ and the argument $\arg(c\left[f_x,f_y\right])$ is the phase. That is, if $c\left[f_x,f_y\right] = a + bi$,

$$amplitude\left[f_x,f_y\right] = (a^2 + b^2)^{1/2}, \quad phase\left[f_x,f_y\right] = \arctan(b/a)$$

Methods for determining the coefficients $c\left[f_x,f_y\right]$ are of no particular concern here. The important idea is that any DEM can be thought of as containing features at many scales, and through the coefficients $c\left[f_x,f_y\right]$, Equation (1) provides a way to compare the amplitudes of those features.

2.2. *Digital filters and amplitude response*

Hamming (1989) and other texts define a digital filter as an arbitrary linear operation. In raster-based terrain analysis, slopes are estimated ultimately as weighted sums of elevations surrounding the pixel of interest (central pixel):

$$S_{j,k} = \sum_{r=-n}^{n} \sum_{s=-n}^{n} w_{r,s} z_{j+r,k+s}$$

That is, the slope estimator can be written as a matrix of constants **w**, with each element paired with and multiplied by an element of the DEM. This is an arbitrary linear operation, so raster-based slope estimators are digital filters.

The amplitude response function (R) is one of the tools used to characterize the action of digital filters. The amplitude response at a given frequency gives the proportion of that frequency's amplitude that remains after the filter is applied (see, for example, Burt *et al*. 2009):

$$R\left[f_x,f_y\right] = \frac{amplitude\ \left[f_x,f_y\right]\ of\ output}{amplitude\ \left[f_x,f_y\right]\ of\ input}$$

If the value is unity, the filter does not affect the features having that frequency. An amplitude response value of zero means that frequency is completely removed. A filter's pattern of different behaviors (removing, exaggerating, etc.) at different frequencies is its amplitude response function.

By finding the amplitude response function of a slope estimator, one can see what scales emerge in the slope field and what scales are effectively blocked. This allows comparison of output scales with neighborhood size; and the relationship between the two may be examined.

2.3. Amplitude response function for the mathematical definition of slope

Slope has a precise mathematical definition for continuous analytical surfaces and can be expressed as an angle (θ) or in a dimensionless ratio form. In particular, the slope is found from the partial derivatives:

$$S\_math = \tan\theta = \left[\left(\frac{\partial Z}{\partial X}\right)^2 + \left(\frac{\partial Z}{\partial Y}\right)^2\right]^{1/2}$$

where Z is the elevation and X and Y are the coordinate axes.

To deduce the amplitude response function of the mathematical slope filter, one should first look at a simplified input terrain, which consists of only one frequency in the X dimension, f_x, and only one frequency in the Y dimension, f_y, and has amplitude unity and phase zero:

$$z_{j,k} = \exp\left(2\pi i(jf_x + kf_y)\right)$$

Then, the mathematical slope can be calculated as follows:

$$\begin{aligned}
\left(\frac{\partial Z}{\partial X}\right)_{j,k} &= 2\pi i f_x \exp\left(2\pi i(jf_x + kf_y)\right)\\
\left(\frac{\partial Z}{\partial Y}\right)_{j,k} &= 2\pi i f_y \exp\left(2\pi i(jf_x + kf_y)\right)\\
S_math_{j,k} &= 2\pi i(f_x^2 + f_y^2)^{1/2} \cdot \exp\left(2\pi i(jf_x + kf_y)\right)
\end{aligned}$$

In this slope expression, the frequency term is the same as the frequency term in the elevation expression, but its coefficient is changed from the real number 1 to the imaginary number $2\pi i(f_x^2 + f_y^2)^{1/2}$. So, the amplitude of $\left[f_x, f_y\right]$ in the slope surface is $2\pi(f_x^2 + f_y^2)^{1/2}$ and the phase is $\pi/2$. The slope calculation caused a phase change of $\pi/2$, because it ultimately is a differential operation. A simple way to understand this is to realize that, given a sinusoid function $\sin f$, its derivative is $\cos f = \sin(f + \pi/2)$; therefore, the differential operation adds a $\pi/2$ phase shift to the original frequency. All of the slope estimators considered in this article produce the appropriate phase shift, and thus only changes in amplitude are relevant. According to the definition, the amplitude response at $\left[f_x, f_y\right]$ is

$$\begin{aligned}
R_math\left[f_x, f_y\right] &= \frac{amplitude\left[f_x, f_y\right] of\, S_math_{j,k}}{amplitude\left[f_x, f_y\right] of\, z_{j,k}} = \frac{2\pi(f_x^2 + f_y^2)^{1/2}}{1}\\
&= 2\pi(f_x^2 + f_y^2)^{1/2}
\end{aligned}$$

Let f_x and f_y be independent variables; then this expression becomes the amplitude response function of the mathematical slope calculation. As mentioned in Section 2.1, the meaningful range of f_x and f_y is $[0, 0.5f_{\mathrm{DEM}}]$. For the sake of simplicity, f_x and f_y are normalized by f_{DEM}, so that no matter what DEM resolution is used, the meaningful range of f_x and f_y is always [0, 0.5]. This makes f_{DEM} the unit of frequency domain. Because $f_{\mathrm{DEM}} = 1/h$, this also makes the DEM resolution h the unit of spatial domain. As a result window size is the only factor that influences neighborhood size. Using this convention $f_x \in [0, 0.5]$, $f_y \in [0, 0.5]$. It is important to note that although using the sampling frequency f_{DEM} to normalize the frequency domain is a long-held convention in digital filter analysis, the choice of normalizing factor is arbitrary. However, because normalization merely scales the horizontal axes of the amplitude response function, conclusions about amplitude response patterns are unaffected by the normalization value used.

In this particular case, we see that *R_math* is an increasing function of frequency and it would be so for any normalizing value. Clearly, the mathematical derivative amplifies high frequencies (small scales). If applied to the hill and rill examples, the mathematical slope filter increases the amplitude of the rill by about 4.44 times, whereas the hill will be attenuated by the factor 0.044. Thus small-scale features will be more dominant in the resulting slope field.

3. Four raster-based slope gradient calculation methods

Many (perhaps all) raster-based slope estimators are based on approximations to the partial derivatives of Z; thus they are approximations of one sort or another to *S_math*. Two strategies are commonly used: finite difference methods and polynomial methods. Among the four popular slope estimators that this article investigates, Horn's estimator (1981) employs the finite difference strategy and Evans' (1979, 1980), Wood's (1996), and the 3dMapper®'s (2004) estimators use the polynomial method.

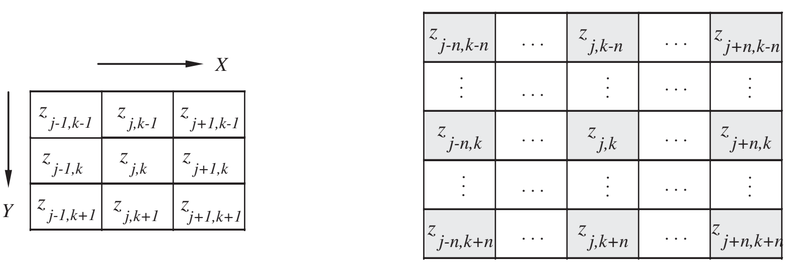
3.1. Horn's method

Using finite differences to approximate the derivatives, Horn (1981) developed a widely used estimator based on a 3 × 3 window (see Figure 3a). The partial derivatives are approximated as follows:

$$\left(\frac{\partial Z}{\partial X}\right)_{j,k} = \frac{(z_{j+1,k+1} + 2z_{j+1,k} + z_{j+1,k-1}) - (z_{j-1,k+1} + 2z_{j-1,k} + z_{j-1,k-1})}{8h}$$

$$\left(\frac{\partial Z}{\partial Y}\right)_{j,k} = \frac{(z_{j+1,k+1} + 2z_{j,k+1} + z_{j-1,k+1}) - (z_{j+1,k-1} + 2z_{j,k-1} + z_{j-1,k-1})}{8h}$$

where h is again the DEM cell size in units of z. This method is used by ArcGIS® and therefore might be the most widely applied in GIS applications.



(a) Horn's method (b) 3dMapper's method

Figure 3. Roving windows.

Derivation of the amplitude response function for Horn's method is similar to the derivation of *R_math*. Start again with the simplified input terrain. The partial derivatives are estimated as follows:

$$\left(\frac{\partial Z}{\partial X}\right)_{j,k} = \{[\exp(2\pi i(f_x + f_y)) + 2\exp(2\pi i f_x) + \exp(2\pi i(f_x - f_y))] - [\exp(2\pi i(-f_x + f_y)) + 2\exp(-2\pi i f_x) + \exp(2\pi i(-f_x - f_y))]\} \cdot \exp(2\pi i(jf_x + kf_y))/8$$

$$\left(\frac{\partial Z}{\partial Y}\right)_{j,k} = \{[\exp(2\pi i(f_x + f_y)) + 2\exp(2\pi i f_y) + \exp(2\pi i(-f_x + f_y))] - [\exp(2\pi i(f_x - f_y)) + 2\exp(-2\pi i f_y) + \exp(2\pi i(-f_x - f_y))]\} \cdot \exp(2\pi i(jf_x + kf_y))/8$$

Using the Euler identity $\sin x = [\exp(ix) - \exp(-ix)]/(2i)$ and the identity $\sin(x + y) = \sin x \cos y + \cos x \sin y$, the partial derivatives are simplified:

$$\left(\frac{\partial Z}{\partial X}\right)_{j,k} = \frac{i}{2}\sin 2\pi f_x(1 + \cos 2\pi f_y)\exp(2\pi i(jf_x + kf_y))$$

$$\left(\frac{\partial Z}{\partial Y}\right)_{j,k} = \frac{i}{2}\sin 2\pi f_y(1 + \cos 2\pi f_x)\exp(2\pi i(jf_x + kf_y))$$

The amplitude of $[f_x, f_y]$ in the slope surface can be written accordingly; and after dividing by the amplitude in the elevation surface, the amplitude response function of Horn's slope estimator can be written as

$$R\_horn = \frac{1}{2}\left\{[\sin 2\pi f_x(1 + \cos 2\pi f_y)]^2 + [\sin 2\pi f_y(1 + \cos 2\pi f_x)]^2\right\}^{1/2}$$

where here and throughout the remainder of the article $f_x \in [0, 0.5], f_y \in [0, 0.5]$.

3.2. Evans' method

Evans (1979, 1980) proposed using least squares to compute the coefficients of a quadratic polynomial which best fits the elevations in a 3 × 3 window, and then using the partial derivatives of that polynomial to calculate terrain derivatives for the central pixel of the window. Evans' approximation of a local surface is in the form $z = ax^2 + by^2 + cxy + dx + ey + f$. The partial derivatives for x and y are then as follows:

$$\frac{\partial Z}{\partial X} = 2ax + cy + d$$

$$\frac{\partial Z}{\partial Y} = 2by + cx + e$$

Since only the slope at the central point of the quadratic surface is of interest, by adopting a local coordinate system with the origin located at the point of interest, the partial derivatives can be simplified as follows (given $x = y = 0$):

$$\frac{\partial Z}{\partial X} = d, \quad \frac{\partial Z}{\partial Y} = e$$

To solve for the six coefficients of the polynomial, Evans (1979) noted that the least squares fit for a 3 × 3 window is greatly simplified by the arrangement of data on the square grid and gave the expressions for the six coefficients. For slope estimation:

$$d = \frac{(z_{j+1,k+1} + z_{j+1,k} + z_{j+1,k-1}) - (z_{j-1,k+1} + z_{j-1,k} + z_{j-1,k-1})}{6h}$$

$$e = \frac{(z_{j+1,k+1} + z_{j,k+1} + z_{j-1,k+1}) - (z_{j+1,k-1} + z_{j,k-1} + z_{j-1,k-1})}{6h}$$

Evans' estimator inspired a family of polynomial methods (Hengl and Reuter 2009). One well-known variation was proposed by Zevenbergen and Thorne (1987), who argued that if the modeled surface does not coincide with the nine original elevations, it does not represent the land surface accurately. They then suggested, as an improvement to Evans' polynomial, a partial quadratic equation which requires nine coefficients. With the nine data points in a 3 × 3 window, a unique solution of the coefficients ensures fidelity at each measured pixel; however, this fidelity makes the method sensitive to elevation errors. Florinsky (1998) found that Evans' method among three other methods, including Zevenbergen and Thorne's, is least affected by elevation errors and is the most accurate, assuming *S_math* is the true slope.

The amplitude response function of Evans' slope estimator can be found similarly as before, resulting in

$$R\_evans = \frac{1}{3}\left\{[\sin 2\pi f_x(1 + 2\cos 2\pi f_y)]^2 + [\sin 2\pi f_y(1 + 2\cos 2\pi f_x)]^2\right\}^{1/2}$$

3.3. Wood's method

Using a fixed 3 × 3 window, the neighborhood size of all the techniques reviewed above is constrained by the resolution of input DEM. Wood (1996) argued that since DEM resolution is often arbitrarily defined and not necessarily related to the required scale of analysis, a fixed 3 × 3 window may not always produce appropriate results. Given the utility of quadratic parameterization, Wood (1996) then proposed a generalization of Evans' method where the window is $m \times m$ and under user control. His method uses all the pixels within the $m \times m$ window to fit the six coefficients of Evans' quadratic polynomial and then calculates terrain derivatives analytically from the polynomial. The coefficients are solved through six simultaneous equations, which after simplification are

$$\begin{bmatrix} \sum x_j^4 & \sum x_j^2y_j^2 & 0 & 0 & 0 & \sum x_j^2 \\ \sum x_j^2y_j^2 & \sum x_j^4 & 0 & 0 & 0 & \sum x_j^2 \\ 0 & 0 & \sum x_j^2y_j^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sum x_j^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sum x_j^2 & 0 \\ \sum x_j^2 & \sum x_j^2 & 0 & 0 & 0 & m^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = \begin{bmatrix} \sum z_jx_j^2 \\ \sum z_jy_j^2 \\ \sum z_jx_jy_j \\ \sum z_jx_j \\ \sum z_jy_j \\ \sum z_j \end{bmatrix}$$

where a to f are the coefficients, x_j and y_j are the horizontal coordinates of the center of each pixel in the window, and z_j is the elevation of each pixel in the window.

To calculate slope gradient, only coefficients d and e are needed. Without solving the entire matrix, they can be simply obtained as follows:

$$d = \frac{\sum z_jx_j}{\sum x_j^2}, \; e = \frac{\sum z_jy_j}{\sum x_j^2}$$

Wood's method has proven effective in multiscale terrain analysis (Schmidt *et al*. 2003, Schmidt and Andrew 2005). It is implemented in the Geographic Resources Analysis Support System (GRASS GIS®), module r.param.scale (GRASS Development Team 2009). As Wood's method is a generalization of Evans' method, the two give exactly the same results when using a 3×3 window ($n = 1$).

To deduce the amplitude response function for Wood's slope estimator, the procedure used above gives

$$R\_wood = \frac{1}{(2n+1)\sum_{j=1}^{n} j^2} \left\{ \left[\left(1 + 2\sum_{j=1}^{n} \cos 2\pi j f_y \right) \sum_{j=1}^{n} (j \sin 2\pi j f_x) \right]^2 + \left[\left(1 + 2\sum_{j=1}^{n} \cos 2\pi j f_x \right) \sum_{j=1}^{n} (j \sin 2\pi j f_y) \right]^2 \right\}^{1/2}$$

where n determines the window size ($m \times m = (2n+1) \times (2n+1)$)

3.4. 3dMapper®'s method

Wood's method incorporates the consideration of multiscale analysis, but it is computationally consuming. Based on its definition, the amount of computation is proportional to n^2. In GRASS GIS®, the maximum window size is limited to 69. However, when working with high-resolution DEMs, the desired scale may demand a bigger window, and the computation cost will be extravagant.

3dMapper® (Burt and Zhu 2004), a terrain visualization and mapping program, uses a method that is both computationally efficient and allows variable neighborhood sizes. Utilizing Evans' polynomial for terrain derivative estimation, 3dMapper® lets the user set a neighborhood size, calculates a window accordingly, and uses pixels at the window corners and edge mid-points to fit the six coefficients of Evans' polynomial (see Figure 3b). Derivatives are then calculated for the central pixel of the window.

The two coefficients needed for slope estimation are

$$d = \frac{(z_{j+n,k+n} + z_{j+n,k} + z_{j+n,k-n}) - (z_{j-n,k+n} + z_{j-n,k} + z_{j-n,k-n})}{6nh}$$

$$e = \frac{(z_{j+n,k+n} + z_{j,k+n} + z_{j-n,k+n}) - (z_{j+n,k-n} + z_{j,k-n} + z_{j-n,k-n})}{6nh}$$

For 3dMapper®'s method, the amount of computation is constant and independent of n. The computational requirements are so low that the estimation can be done on the fly in an interactive environment. As with Wood's method, the 3dMapper® method is identical to Evans' method for a 3×3 window ($n = 1$).

Using the same procedure, one can deduce the amplitude response function for 3dMapper®'s slope estimator:

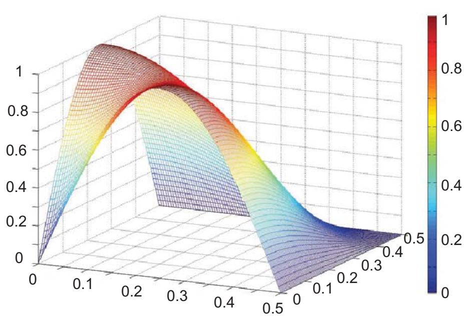
$$R\_3dMapper = \frac{1}{3n}\left\{[\sin 2\pi nf_x(1 + 2\cos 2\pi nf_y)]^2 + [\sin 2\pi nf_y(1 + 2\cos 2\pi nf_x)]^2\right\}^{1/2}$$

4. Results and discussion

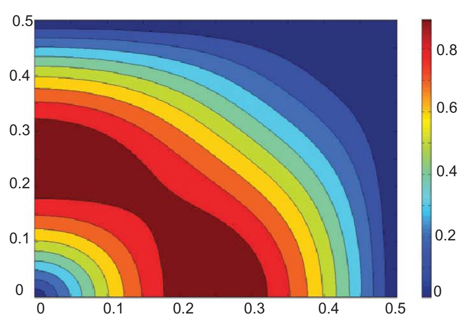
4.1. Neighborhood size and spatial scale

Figures 4–6 show the amplitude response functions determined in the previous section for the four slope estimators. In each figure, the two horizontal axes indicate f_x and f_y, respectively; the vertical axis indicates the amplitude response; and R is plotted as a surface. As seen in the figures, for all four slope estimators the amplitude response varies strongly with frequency, and there are spectral regions of high response that appear as one or more bands or ridges on the surface. The frequencies falling into these bands pass through the filter more than other frequencies, so they emerge more strongly in the resulting slope raster. These frequencies therefore correspond to the spatial scale of the calculation.

The figures also show that all four methods depart substantially from the mathematical definition of slope, with high frequencies attenuated rather than amplified, but there is no consistency in the amount of attenuation. It should be mentioned that 3dMapper® was designed for visualization, and that by default the DEM is smoothed by varying amounts before slope estimates are computed. Because the DEM seen by the program will not have

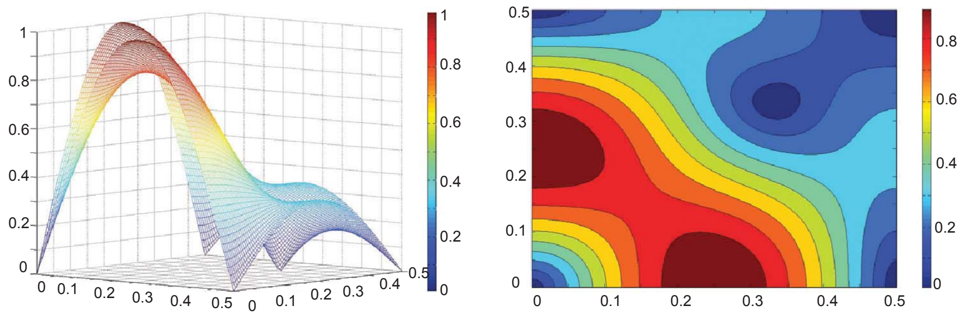


(a) perspective



(b) contour (interval = 0.1)

Figure 4. Amplitude response surface of Horn's method.



(a) perspective

(b) contour (interval = 0.1)

Figure 5. Amplitude response surface of Evans' method.

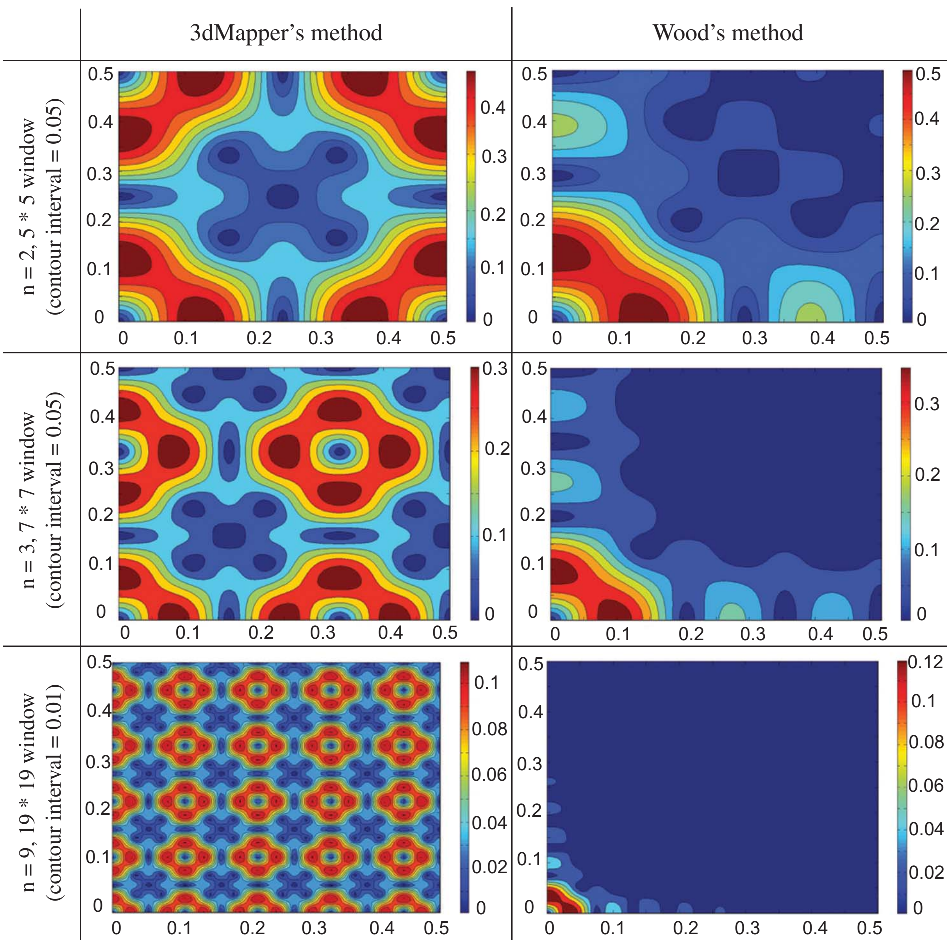


Figure 6. Amplitude response surfaces of 3dMapper®'s method and Wood's method. (When $n = 1$, i.e., 3×3 window, both methods have the same amplitude response surface as Evans' method.).

variance at small scales, the high-frequency peaks in *R_3dMapper* are somewhat misleading when the 3dMapper® *program itself* is used. The plots are, however, a correct depiction of the slope estimation *method alone*.

To investigate the relationship between neighborhood size and spatial scale, the frequency of neighborhood size and the passing bands of the slope estimators are compared. The frequency of a neighborhood size is the reciprocal of the neighborhood size, which is the same in X and Y owing to the square shape of the window:

$$f_{NH-x} = f_{NH-y} = \frac{1}{Neighborhood\ Size} = \frac{1}{Window\ Size \times h} = \frac{1}{Window\ Size} f_{\mathrm{DEM}}$$

Because the frequency domain has been normalized by f_{DEM}, the normalized frequency of a neighborhood is simply $1/Window\ Size$.

Figure 7 contrasts the frequency of neighborhood size and the $f_x = f_y$ transects of different slope estimators' amplitude response surfaces. The $f_x = f_y$ transect is a good indicator of the amplitude response function, because (1) it shows roughly the same trend as cross-sections in other directions; (2) in so doing one can compare the amplitude response functions of different methods in one figure; and (3) the frequency of neighborhood size is on this transect.

As shown in Figure 7, for none of the analyzed slope estimators does the neighborhood size fall into a major passing band(s). Therefore, the neighborhood size and the spatial scale are not numerically the same. It is also important to note that neighborhood size corresponds to only one frequency/scale, while all the frequencies/scales that fall into the

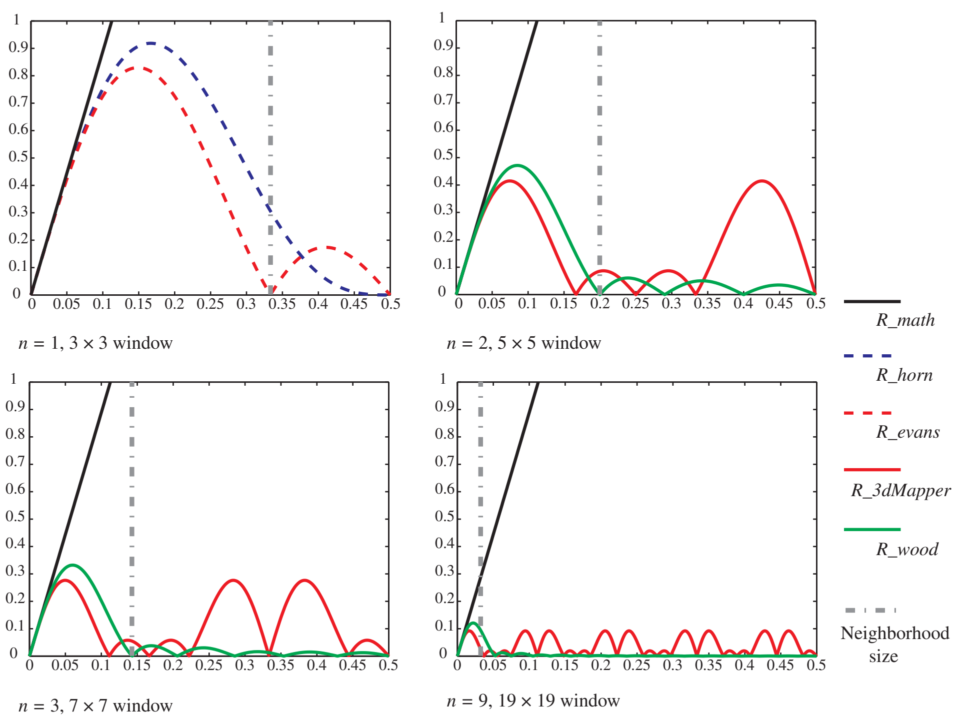


Figure 7. Amplitude response functions at different window sizes, when $f_x = f_y$.

passing band(s) are part of the 'scale' of the slope estimator. Thus there is no single scale that can be associated with an estimator.

4.2. Neighborhood size and scale indices

In order to understand the relationship between neighborhood size and spatial scale, it is necessary to compare the passing bands of different slope estimators (see Figure 8).

The comparison remains difficult, because the definition of 'passing' is application-specific and can be fuzzy. For some situations, people might want to define passing as amplitude response greater than 0.5, while for some other situations, 0.8 might be an appropriate threshold. However, the peaks (maxima) of a curve are objectively defined, so here the peaks on the $f_x = f_y$ amplitude response transect (f_{peak}) are used as an admittedly incomplete indicator for the passing bands of a slope estimator.

Another index for passing scales is suggested by the fact that all of the investigated estimators depart progressively more from the mathematical derivative as f increases (Figure 8). One could argue that if a method's amplitude response is less than half of the mathematical derivative's response function, the method is not passing those frequencies. Therefore one can use the transition frequency f_{trans}, where $R(f)/R\_math(f) < 0.5$ for all $f > f_{\text{trans}}$, as another indicator for the passing band of a slope estimator.

Table 1 lists the frequencies and scales corresponding to these indicators of passing bands on the amplitude response transects in Figure 8.

As shown in Figure 7 and Table 1, when using the same window size, the peak frequency (f_{peak}), transition frequency (f_{trans}), and widths of passing bands all vary among different slope estimation methods. When considering f_{peak}, the methods differ in the number of passing bands. There is obviously no constant multiplier that can convert the window size to the peak passing scale(s). Even if one considers only f_{trans}, notable differences

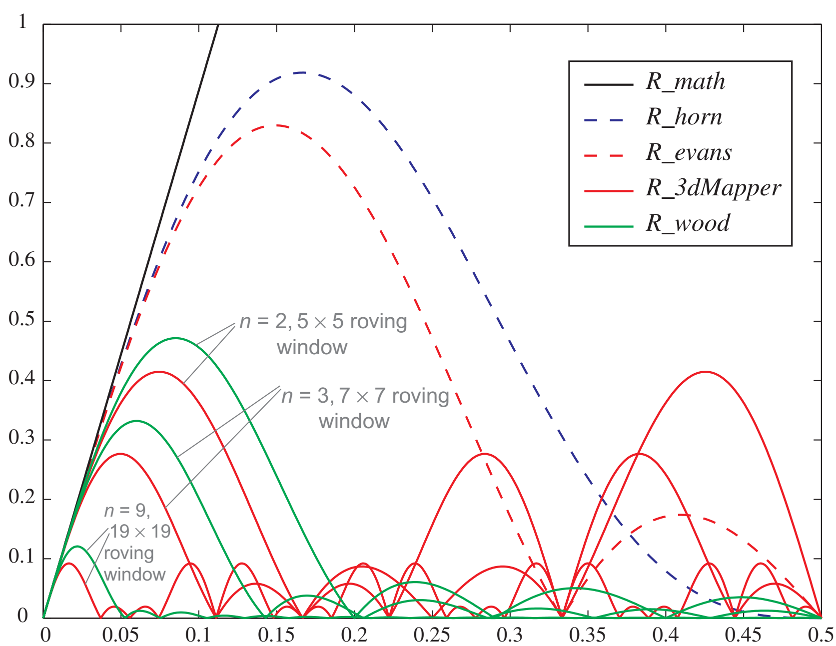


Figure 8. Amplitude response functions of all four methods, when $f_x = f_y$.

Table 1. Window size versus indicators of passing bands on the $f_x = f_y$ transect.

Method	Window size	Neighborhood size	f_{peak}	Scale of f_{peak}	f_{trans}	Scale of f_{trans}
Horn	3×3	$3\,h$	$0.168\,f_{\text{DEM}}$	$5.95\,h$	$0.199\,f_{\text{DEM}}$	$5.03\,h$
Evans	3×3	$3\,h$	$0.149\,f_{\text{DEM}}$	$6.71\,h$	$0.179\,f_{\text{DEM}}$	$5.59\,h$
3dMapper	5×5 $n = 2$	$5\,h$	$0.075\,f_{\text{DEM}}$ $0.425\,f_{\text{DEM}}$	$13.33\,h$ $2.35\,h$	$0.089\,f_{\text{DEM}}$	$11.24\,h$
	7×7 $n = 3$	$7\,h$	$0.051\,f_{\text{DEM}}$ $0.282\,f_{\text{DEM}}$ $0.384\,f_{\text{DEM}}$	$19.61\,h$ $3.55\,h$ $2.60\,h$	$0.060\,f_{\text{DEM}}$	$16.67\,h$
	19×19 $n = 9$	$19\,h$	$0.018\,f_{\text{DEM}}$ $0.093\,f_{\text{DEM}}$ $0.129\,f_{\text{DEM}}$ $0.204\,f_{\text{DEM}}$ $0.240\,f_{\text{DEM}}$ $0.315\,f_{\text{DEM}}$ $0.351\,f_{\text{DEM}}$ $0.426\,f_{\text{DEM}}$ $0.462\,f_{\text{DEM}}$	$55.56\,h$ $10.75\,h$ $7.75\,h$ $4.90\,h$ $4.17\,h$ $3.17\,h$ $2.85\,h$ $2.35\,h$ $2.16\,h$	$0.020\,f_{\text{DEM}}$	$50.00\,h$
Wood	5×5 $n = 2$	$5\,h$	$0.085\,f_{\text{DEM}}$	$11.76\,h$	$0.102\,f_{\text{DEM}}$	$9.80\,h$
	7×7 $n = 3$	$7\,h$	$0.060\,f_{\text{DEM}}$	$16.67\,h$	$0.072\,f_{\text{DEM}}$	$13.89\,h$
	19×19 $n = 9$	$19\,h$	$0.022\,f_{\text{DEM}}$	$45.45\,h$	$0.026\,f_{\text{DEM}}$	$38.46\,h$

between methods are seen. It is interesting to note that for all methods the spatial scale of the lowest f_{peak} is somewhat larger than or very close to twice the neighborhood size and f_{trans} is not too different from the lowest f_{peak}. Of course, if one had used a transition threshold different than 0.5, this would not be so. All of the above shows the inability of window size to express the complicated nature of spatial scale.

In addition, the same neighborhood size can result from different combinations of DEM resolution and window size. For example, using a 3×3 window on a 9 m resolution DEM and using a 9×9 window on a 3 m resolution DEM both generate a neighborhood size of 27 m. However, it has been shown that a 3×3 window and a 9×9 window can have very different amplitude response functions, even for the same estimation method. Thus neither window size nor the equivalent neighborhood size can quantitatively indicate the scale of slope estimation.

Furthermore, different slope filters react differently to change in window size. First, look at f_{peak}. For Wood's method, increasing the window size moves the major (peak) passing band to lower frequencies as shown in Figures 7 and 8. Table 2 confirms this and reveals that the scale of the peak frequency is close to $2h(2n + 1)$, that is, the scale is nearly twice the neighborhood size for this range of n. Indeed, plotting the peak scale values (Figure 9) suggests a linear relationship and fitting a least squares line gives

$$\textit{peak passing scale}_{(\text{Wood}, f_x = f_y)} = h(4.832n + 2.152) = 2h(2.416n + 1.076)$$

As expected, the greater the window size the greater is the peak passing scale of the estimation.

Table 2. Window size versus indicators of passing bands on the $f_x = f_y$ transect for Wood's method.

Window size	f_{peak}	Scale of f_{peak}	f_{trans}	Scale of f_{trans}
$3 \times 3, n = 1$	$0.149\,f_{\mathrm{DEM}}$	$6.71\ h$	$0.179\,f_{\mathrm{DEM}}$	$5.59\ h$
$5 \times 5, n = 2$	$0.085\,f_{\mathrm{DEM}}$	$11.76\ h$	$0.102\,f_{\mathrm{DEM}}$	$9.80\ h$
$7 \times 7, n = 3$	$0.060\,f_{\mathrm{DEM}}$	$16.67\ h$	$0.072\,f_{\mathrm{DEM}}$	$13.89\ h$
$9 \times 9, n = 4$	$0.046\,f_{\mathrm{DEM}}$	$21.74\ h$	$0.056\,f_{\mathrm{DEM}}$	$17.86\ h$
$11 \times 11, n = 5$	$0.038\,f_{\mathrm{DEM}}$	$26.32\ h$	$0.045\,f_{\mathrm{DEM}}$	$22.22\ h$
$13 \times 13, n = 6$	$0.032\,f_{\mathrm{DEM}}$	$31.25\ h$	$0.038\,f_{\mathrm{DEM}}$	$26.32\ h$
$15 \times 15, n = 7$	$0.028\,f_{\mathrm{DEM}}$	$35.71\ h$	$0.033\,f_{\mathrm{DEM}}$	$30.30\ h$
$17 \times 17, n = 8$	$0.024\,f_{\mathrm{DEM}}$	$41.67\ h$	$0.029\,f_{\mathrm{DEM}}$	$34.48\ h$
$19 \times 19, n = 9$	$0.022\,f_{\mathrm{DEM}}$	$45.45\ h$	$0.026\,f_{\mathrm{DEM}}$	$38.46\ h$
$21 \times 21, n = 10$	$0.020\,f_{\mathrm{DEM}}$	$50.00\ h$	$0.024\,f_{\mathrm{DEM}}$	$41.67\ h$

For 3dMapper®'s method, as the window size increases, the amplitude response function shows more peak passing bands; all of them are of the same height and are evenly distributed throughout the domain of R (see Figures 6 and 7). Therefore, larger window sizes are not associated with greater peak scales and there is no single dominant peak.

However, using f_{trans} to indicate the passing band gives a different picture. In particular, for both Wood's and 3dMapper®'s methods, the scale of f_{trans} increases monotonically with n (see Figure 9 and Tables 2 and 3), and the increasing rates are not greatly different for the two estimators.

These results demonstrate that, with either indicator of passing bands, different slope filters react differently to changing window size. So, both the window size and method must be specified in order to know the spatial scales passed by the filter. At the same time, using different indicators of passing bands gives different impressions about the scale associated with a particular neighborhood–method combination. Therefore, no single number will wholly describe the scale, and one must examine the amplitude response function to completely understand the matter.

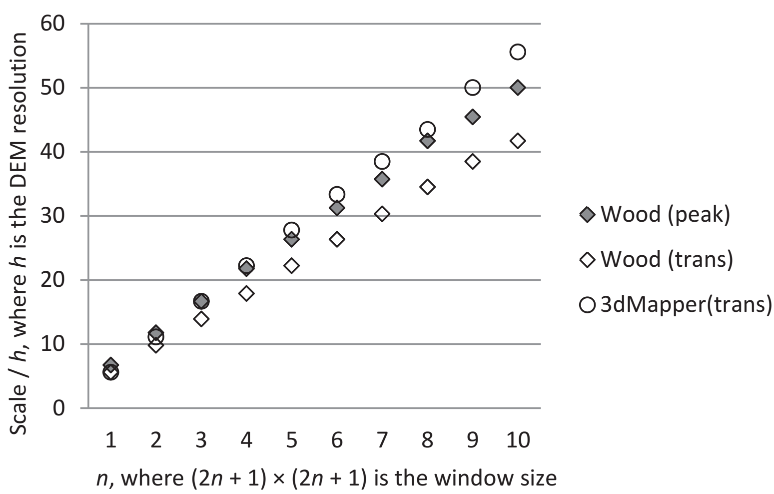


Figure 9. Scales of passing band indicators on the $f_x = f_y$ transect with different window sizes.

Table 3. Window size versus f_{trans} on the $f_x = f_y$ transect for 3dMapper®'s method.

Window size	f_{trans}	Scale of f_{trans}
$3 \times 3, n = 1$	$0.179 f_{\text{DEM}}$	$5.59\ h$
$5 \times 5, n = 2$	$0.090 f_{\text{DEM}}$	$11.11\ h$
$7 \times 7, n = 3$	$0.060 f_{\text{DEM}}$	$16.67\ h$
$9 \times 9, n = 4$	$0.045 f_{\text{DEM}}$	$22.22\ h$
$11 \times 11, n = 5$	$0.036 f_{\text{DEM}}$	$27.78\ h$
$13 \times 13, n = 6$	$0.030 f_{\text{DEM}}$	$33.33\ h$
$15 \times 15, n = 7$	$0.026 f_{\text{DEM}}$	$38.46\ h$
$17 \times 17, n = 8$	$0.023 f_{\text{DEM}}$	$43.48\ h$
$19 \times 19, n = 9$	$0.020 f_{\text{DEM}}$	$50.00\ h$
$21 \times 21, n = 10$	$0.018 f_{\text{DEM}}$	$55.56\ h$

For a particular application, analysts should pick the indicator(s) that best fit their needs. The following sections of this article use the actual value of the amplitude response in defining passing bands. Alternatively, an analyst might prefer to use the relative definition.

4.3. Selecting an appropriate estimator

Each of the four amplitude response functions contains three independent variables: f_x, f_y, and the window size n. In practice the GIS analyst usually has a target scale or scales in mind. By calculating the reciprocal of the desired scale(s), one derives the corresponding desired frequency(s). With these and the DEM resolution, the normalized f_x and f_y are known. Inspection of Figure 8 and/or 4–6 will show the suitability of a given estimator for analysis at the chosen scale(s).

For example, if the desired normalized frequency is $f_x = f_y = 0.05$, the best window size for 3dMapper®'s method is 7×7, but this estimator may amplify undesirable high-frequency information if not smoothed beforehand. By contrast, Wood's method with a 7×7 window passes the desired frequency reasonably well and notably reduces high frequencies; however, the computational cost is much higher.

One should also consider the scales of variability within the DEM. The amplitude of a frequency in the slope surface is the product of the amplitude of that frequency in the input DEM and the amplitude response of the slope filter at that frequency. If either is near zero, the product will obviously be trivial. Therefore, if the input DEM contains little high-frequency information, one could exploit 3dMapper®'s computational advantage without worrying about having multiple passing bands. On the other hand, if the input DEM contains considerable small-scale variability, Wood's method is justified despite its higher computational cost.

To give another example, none of the estimators is satisfactory for a desired peak passing frequency of 0.1. Using larger windows will not lead to better results, because the major passing band of Wood's method will shift to lower frequencies, and 3dMapper®'s method will introduce more peak passing bands at high frequencies. Remembering that for a 3×3 window Wood's and 3dMapper®'s methods are equivalent to Evans' method, and seeing that there is no possible window size between 3×3 and 5×5, the desired scale $f_x = f_y = 0.1$ cannot be the peak passing scale with these existing slope estimators. In other words, for some scales no variant of these methods can produce satisfactory results.

In light of the limitations seen with existing estimators, it seems natural to consider the question of what constitutes an ideal method.

4.4. Ideal slope filter

In Section 2.3, the amplitude response function of the mathematical slope filter was derived. Along the $f_x = f_y = f$ transect, the response is

$$R\_math_{(f_x=f_y)} = 2\sqrt{2}\pi f$$

As shown, the amplitude response increases linearly with frequency, and frequencies greater than $(2\sqrt{2}\pi)^{-1} \approx 0.11$ are amplified. For $f = 0.25$, the amplification is greater than a factor of 2. Thus the mathematical derivative can be considered as a high-pass filter. Although *R_math* provides useful hints for how a differential filter should behave, one might not want to completely match it with an actual estimator, because high frequencies are often considered noise in measurements. So, an ideal slope estimator might look something like the filter seen in Figure 10. The estimator would closely approximate the derivative for frequencies less than a certain value f_{cutoff} and remove the frequencies greater than f_{cutoff} (Hamming 1989).

None of the four examined estimators has quite the shape of the ideal slope filter. Horn's method has the same general pattern as the ideal slope filter, but the transition band from the peak passing frequency to the cutoff frequency is far too wide. Evans' method has a narrower primary transition band, but as a trade-off it has a secondary passing band at higher frequencies. 3dMapper®'s method has multiple peak passing bands and they all reach the same response height. Wood's method is the most similar to the ideal among the four: as the window size increases, the peak passing band moves to lower frequencies and the transition band becomes narrower.

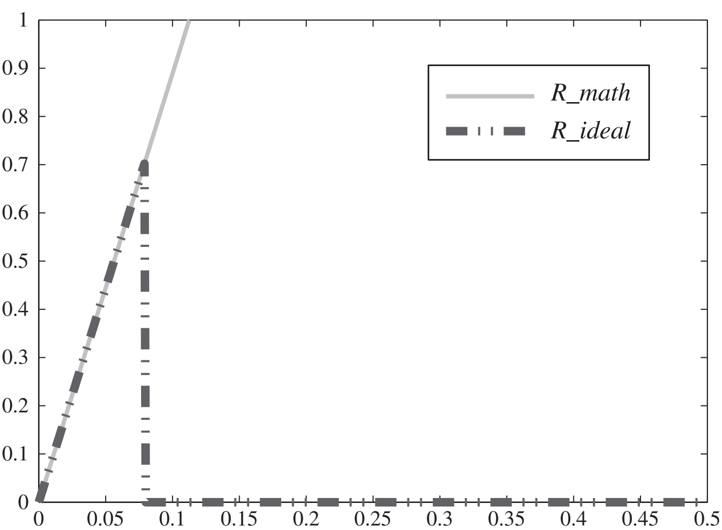


Figure 10. $f_x = f_y$ transect of the amplitude response function of an 'ideal' slope estimator.

In addition, none of the four methods has complete flexibility in changing the cutoff frequency. Some applications might call for the estimator to be near the mathematical definition's response function for only a range of frequencies in the middle of the spectrum, with frequencies below and above that range removed. None of the four methods discussed come close to exhibiting such behavior.

Ideally one would be able to define a desired response function (i.e., the relevant spatial scales) and also specify the associated computational burden one is willing to pay. In the case of window methods, this amounts to specifying the neighborhood size *n*. This is the province of digital filter design, for which there exists a rich literature (see, for example, Parks and Burrus 1987, Schlichthärle 2000, De Freitas 2005). Applying that theory to GIS slope estimation seems a worthy endeavor.

5. Conclusions

This research demonstrated that analyzing the amplitude response function of a slope estimator is an effective way to investigate the relationship between the window size (which is associated with the neighborhood size) and the implicit scale of analysis. Examination of four popular existing slope estimators showed that neighborhood size and spatial scale of slope estimation are not numerically the same and that there is no uniform scaling from one to another. In fact, the definition of passing band (scale) is application-specific, and different indicators may give different impressions. Thus there is no single number that can completely describe a method–neighborhood combination. One must examine the amplitude response function to fully understand the estimator. That said, in comparison with the mathematical definition of 'slope', all of the investigated methods suppress small scales, but there is no consistency with regard to relative attenuation. It was also shown that knowing the response signature of different method–neighborhood combinations can be helpful in selecting an estimator to accomplish a particular goal. However, given the limitations of existing estimation methods, tools should be developed that will allow an analyst to devise a slope operator that comes as close as possible to a target response function at an affordable computational cost.

Note

1. Note that in this article 'scale' is defined as in Earth sciences and physical sciences, for example, as the characteristic length scale in turbulence theory, or as 'downscaling' in climate studies. This differs from cartographic use, where 'scale' is the ratio of map to Earth distance.

References

Burt, J.E., Barber, G.M., and Rigby, D.L., 2009. *Elementary statistics for geographers*. 3rd ed. New York: Guilford Press.

Burt, J.E. and Zhu, A.X., 2004. *3dMapper®* [online]. Madison, WI: Terrain Analytics LLC. Available from: http://www.terrainanalytics.com [Accessed 25 June 2010].

De Freitas, J.M., 2005. *Digital filter design solutions*. Boston, MA: Artech House.

ESRI®, 2010. *Slope* [online]. ArcGIS® desktop 9.3 help. Available from: http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Slope [Accessed 25 June 2010].

Evans, I.S., 1979. *An integrated system of terrain analysis and slope mapping*. U.S. Army European Research Office, London. England: University of Durham, Final report on contract DA-ERO-591-73-G0040.

Evans, I.S., 1980. An integrated system for terrain analysis for slope mapping. *Zeitschrift fur Geomorphologie*, 36, 274–295.

Florinsky, I.V., 1998. Accuracy of local topographic variables derived from digital elevation models. *International Journal of Geographical Information Science*, 12 (1), 47–61.

GRASS Development Team, 2009. *r.param.scale - Extracts terrain parameters from a DEM* [online]. GRASS GIS® 6.4.0svn reference manual. Available from: http://grass.osgeo.org/grass64/manuals/html64_user/r.param.scale.html [Accessed 25 June 2010].

Hamming, R.W., 1989. *Digital filters*. Englewood Cliffs, NJ: Prentice-Hall.

Hengl, T. and Reuter, H.I., eds., 2009. *Geomorphometry: concepts, software, applications*. Amsterdam: Elsevier.

Hodgson, M.E., 1995. What cell size does the computed slope/aspect angle represent? *Photogrammetric Engineering & Remote Sensing*, 61 (5), 513–517.

Hodgson, M.E., 1998. Comparison of angles from surface slope/aspect algorithms. *Cartography and Geographic Information Systems*, 25 (3), 173–185.

Horn, B.K.P., 1981. Hill shading and the reflectance map. *Proceedings of the IEEE*, 69 (1), 14–47.

Morgan, F., 2005. *Real analysis*. Providence, RI: American Mathematical Society.

Parks, T.W. and Burrus, C.S., 1987. *Digital filter design*. New York: Wiley.

Pinkus, A. and Zafrany, S., 1997. *Fourier series and integral transforms*. Cambridge: Cambridge University Press.

Rayner, J.N., 1971. *An introduction to spectral analysis*. London: Pion.

Schlichthärle, D., 2000. *Digital filters: Basics and design*. Berlin: Springer-Verlag.

Schmidt, J. and Andrew, R., 2005. Multi-scale landform characterization. *Area*, 37 (3), 341–350.

Schmidt, J., Evans, I.S., and Brinkmann, J., 2003. Comparison of polynomial models for land surface curvature calculation. *International Journal of Geographical Information Science*, 17 (8), 797–814.

Wood, J.D., 1996. *The geomorphological characterisation of digital elevation models*. Thesis (PhD). University of Leicester. Available from: http://www.soi.city.ac.uk/~jwo/phd [Accessed 25 June 2010].

Zevenbergen, L.W. and Thorne, C.R., 1987. Quantitative analysis of land surface topography. *Earth Surface Processes & Landforms*, 12 (1), 47–56.

Zhu, A.X., 2008. Spatial scale and neighborhood size in spatial data processing for modeling nature environment. *In*: N.J. Mount, G.L. Harvey, P. Aplin, and G. Priestnall, eds. *Innovations in GIS 13: representing, modeling, and visualizing the natural environment*. Boca Raton, FL: CRC Press, 147–165.

Zhu, A.X., *et al*., 2008. The impact of neighborhood size on terrain derivatives and digital soil mapping. *In*: Q. Zhou, B. Lees, and G. Tang, eds. *Advances in digital terrain analysis*. Berlin: Springer-Verlag, 333–348.