Filtering of Digital Terrain Models by Two-Dimensional Singular Spectrum Analysis

N.E. Golyandina 1, K.D. Usevich 1, I.V. Florinsky 2

1 Department of Statistical Modelling
St.-Petersburg State University
University pr. 28, Petrodvoretz, St.-Petersburg, 198504, Russia
e-mail: nina@gistatgroup.com

2 Institute of Mathematical Problems of Biology
Russian Academy of Sciences
Pushchino, Moscow Region, 142290, Russia
e-mail: iflor@mail.ru

ABSTRACT
Singular Spectrum Analysis (SSA) has been approved as a model-free technique to analyse time series. SSA can solve different problems such as decomposition into a sum of trend, periodicities, and noise, smoothing, and others. In this paper, we validate abilities of 2D-SSA (the extension of SSA to analyse two-dimensional scalar fields) to treat digital terrain models (DTMs). The study is exemplified by a 30-arc-second digital elevation model of a part of South America derived from GTOPO30. Results demonstrate that 2D-SSA is an efficient method to denoise and generalise DTMs. It can be also used to decompose a topographic surface into components of continental, regional, and local scales.

Keywords: Digital Elevation Model, Singular Spectrum Analysis, Image Processing, Filtering, Periodicity, Noise.


INTRODUCTION
A surface can be viewed as a sum of surfaces. This triviality forms the basis to solve a variety of problems using digital terrain analysis. The best-known tasks follow: (1) separating topographic components of different scales; (2) denoising digital terrain models (DTMs); and (3) generalising DTMs, that is, removing non-noise high-frequency components from DTMs. These tasks are usually attacked by multiple regressions (Chorley and Huggett, 1965; Tobler, 1969), weighted moving averages (Tobler, 1969), Fourier analysis (Rudy, 1989), Kalman filtering (Gallant, 2006), and isobase mapping (Grohmann et al., 2007). In this paper, we report the first results of evaluation of two-dimensional Singular Spectrum Analysis (2D-SSA) as a tool to denoise and generalise DTMs and to separate continental, regional, and local components of topography.

SSA was originated as a model-free technique to analyse one-dimensional time series (Elsner and Tsonis, 1996; Danilov and Zhigljavsky, 1997; Golyandina et al., 2001). The SSA can be used to decompose a time series into a sum of trend, oscillations, and noise, to detect periodicities, to smooth and denoise signals, to forecast time series, and to impute missing data (GistaT Group, 1997-2007).

There are multidimensional extensions of SSA. Multichannel SSA (MSSA) is intended to analyse simultaneously a set of time series with common features (Elsner and Tsonis, 1996; Danilov and Zhigljavsky, 1997; Golyandina and Stepanov, 2005). MSSA can be applied to
2D scalar fields if one dimension is considered as time. The 2D-SSA was specially designed to process 2D scalar fields (Danilov and Zhigljavsky, 1997). Unlike MSSA, 2D-SSA is invariant regarding field rotation. 2D-SSA shares common traits with 2D-ESPRIT (Rouquette and Najim, 2001) applied to a specific problem of estimation of frequencies and damping factors.

2D-SSA ALGORITHM

Let us consider a 2D discrete field \( f : \{1, \ldots, N_f\} \times \{1, \ldots, N_c\} \mapsto \mathbb{R} \) given by a matrix

\[
F = \begin{pmatrix}
  f(1, 1) & f(1, 2) & \ldots & f(1, N_c) \\
  f(2, 1) & f(2, 2) & \ldots & f(2, N_c) \\
  \vdots & \vdots & \ddots & \vdots \\
  f(N_f, 1) & f(N_f, 2) & \ldots & f(N_f, N_c)
\end{pmatrix}
\]  

Algorithm parameters are window sizes \((L_r, L_c)\), where \(1 \leq L_r \leq N_f, 1 \leq L_c \leq N_c, 1 < L_rL_c < N_fN_c\). Set \(K_r = N_f - L_r + 1\) and \(K_c = N_c - L_c + 1\). The algorithm includes two stages, decomposition and reconstruction, each of them consists of two steps (see details on matrix calculus elsewhere – Magnus and Neudecker, 1999, Chapter 1).

Decomposition

Embedding

The first step consists in the construction of the trajectory matrix of the field \( F \) by moving \( L_r \times L_c \)-windows. In the 1D case, one transforms a 1D object into a 2D matrix (Golyandina et al., 2001). Here, we embed a 2D object into a four-dimensional space. To flatten the 4D object, we transform moving windows

\[
F_{i,j} = \begin{pmatrix}
  f(i, j) & f(i, j + 1) & \ldots & f(i, j + L_c - 1) \\
  f(i + 1, j) & f(i + 1, j + 1) & \ldots & f(i + 1, j + L_c - 1) \\
  \vdots & \vdots & \ddots & \vdots \\
  f(i + L_r - 1, j) & f(i + L_r - 1, j + 1) & \ldots & f(i + L_r - 1, j + L_c - 1)
\end{pmatrix}
\]  

where \(1 \leq i \leq K_r, 1 \leq j \leq K_c\), to columns of the flattened trajectory matrix \( W \) \((F_{i,j} \) transfers to the \((i + (j - 1)K_r)\)th column). For example, if \(L_r = L_c = 2\), then the window

\[
\begin{pmatrix}
  f(1, 1) & f(1, 2) \\
  f(2, 1) & f(2, 2)
\end{pmatrix}
\]

is transformed into the first column \((f(1,1), f(2,1), f(1,2), f(2,2))^T\).

It is appropriate to use vectorising and matricizing operations: for a \(M \times N\)-matrix \(B\), \(\text{vec}B \in \mathbb{R}^{MN}\) is the vector constructed from stacked columns of \(B\). If we fix matrix sizes \(M\) and \(N\), then \((M,N)\)-matricizing will be opposite to vectorising: \(\text{matr vec}B = B\). Thus, the trajectory matrix \(W\) of the field \(F\) consists of \(K_rK_c\) columns \(\text{vec}F_{i,j}\), \(1 \leq i \leq K_r, 1 \leq j \leq K_c\). Furthermore, the matrix \(W\) of the size \(L_rL_c \times K_rK_c\) can be presented in a more structured form:
\[ \mathbf{W} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \ldots & \mathbf{H}_{K_c} \\ \mathbf{H}_2 & \mathbf{H}_3 & \ldots & \mathbf{H}_{K_c+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{L_c} & \mathbf{H}_{L_c+1} & \ldots & \mathbf{H}_{N_c} \end{pmatrix}, \]  

(4)

where

\[ \mathbf{H}_i = \begin{pmatrix} f(1,i) & f(2,i) & \ldots & f(K_r,i) \\ f(2,i) & f(3,i) & \ldots & f(K_r+1,i) \\ \vdots & \vdots & \ddots & \vdots \\ f(L_r,i) & f(L_r+1,i) & \ldots & f(N_r,i) \end{pmatrix}. \]  

(5)

The matrix \( \mathbf{W} \) has the block-Hankel structure with the same blocks along secondary diagonals. Each block \( \mathbf{H}_i \) is Hankel itself: it is the trajectory matrix of the 1D series \( f(\cdot,i) \) (the \( i \)th column of the initial field \( \mathbf{F} \)). The matrix \( \mathbf{W} \) will be further called the block-Hankel trajectory matrix of the field \( \mathbf{F} \). Note that there is the one-to-one correspondence between \( N_r \times N_c \) fields and block-Hankel matrices with \( L_r \times K_c \) Hankel blocks.

**Singular value decomposition**

This step is a singular value decomposition of the block-Hankel trajectory matrix:

\[ \mathbf{W} = \sum_{i=1}^{d} \mathbf{W}_i = \sum_{i=1}^{d} \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T, \]  

(6)

where \( \lambda_1, \ldots, \lambda_d \) are non-zero eigenvalues of the matrix \( \mathbf{WW}^T \) arranged in decreasing order of magnitudes (\( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d > 0 \)), \( \{\mathbf{U}_1, \ldots, \mathbf{U}_d\}, \mathbf{U}_i \in \mathbb{R}^{L_r \times L_c} \) is the corresponding orthonormal system of the eigenvectors, and \( \{\mathbf{V}_1, \ldots, \mathbf{V}_d\}, \mathbf{V}_i \in \mathbb{R}^{K_r \times K_c} \) is the orthonormal system of the corresponding factor vectors

\[ \mathbf{V}_i = \mathbf{W}_i / \sqrt{\lambda_i}. \]  

(7)

By analogy with principal component analysis, the vectors \( \sqrt{\lambda_i} \mathbf{V}_i \) are called principal component vectors. They are conveniently considered as matrices: the \( (L_r, L_c) \)-matricizing of an eigenvector is called an eigenfield, the \( (K_r, K_c) \)-matricizing of a factor vector is called a factor field, and the \( (K_r, K_c) \)-matricizing of a vector of principal components is called a principal component field. A set of square root of \( i \)th eigenvalue, \( i \)th eigenfield, and \( i \)th factor field is called \( i \)th eigentriple (ET).

**Reconstruction**

**Grouping**

This step consists in grouping of addends in the decomposition (Eq. 6), that is, the corresponding eigentriples. Let us divide the set \( \{1, \ldots, d\} \) into \( m \) disjoint subsets \( I_1, \ldots, I_m \). Summing \( \mathbf{W}_i, i \in I_k \), we come to the expansion

\[ \mathbf{W} = \sum_{k=1}^{m} \mathbf{W}_{I_k}. \]  

(8)

**Averaging**

Grouped matrices \( \mathbf{W}_{I_k} \) do not necessarily have block-Hankel form. Therefore, one needs an
additional step to transfer the decomposition (Eq. 8) of the block-Hankel trajectory matrix $W$ into a decomposition of the initial field $F$. This can be done by the orthogonal projection (in Frobenious norm) of the matrices $W_{k}$ on the set of block-Hankel matrix with Hankel blocks, like Eq. 4. After projection, we obtain

$$W = \sum_{k=1}^{m} \tilde{W}_{k},$$

where $\tilde{W}_{k}$, $k = 1, \ldots, m$, have a form of Eq. 4. Using the one-to-one correspondence between block-Hankel trajectory matrices and 2D fields, we come to the final decomposition of the initial field:

$$F = \sum_{k=1}^{m} \tilde{F}_{k}.$$  

Let us discuss the projection algorithm in details. Consider one of components of the decomposition (Eq. 8):

$$W_{i} = \begin{pmatrix} W_{1,1} & W_{1,2} & \cdots & W_{1,K_{c}} \\ W_{2,1} & W_{2,2} & \cdots & W_{2,K_{c}} \\ \vdots & \vdots & \ddots & \vdots \\ W_{c,1} & W_{c,2} & \cdots & W_{c,K_{c}} \end{pmatrix},$$

where $W_{i,j}$ are $L_{r} \times K_{c}$ blocks. The projection is equivalent to two sequential averaging procedures: ‘within block’ hankelization and ‘between blocks’ one.

1. ‘Within block’ hankelization. Averaging $(k,l)$-entries of $W_{i,j}$ with $k+l = s$, where $1 \leq s \leq N_{r}$, we obtain the Hankel matrices. Thus, we have a matrix with Hankel blocks:

$$W_{i} = \begin{pmatrix} W_{1,1}' & W_{1,2}' & \cdots & W_{1,K_{c}}' \\ W_{2,1}' & W_{2,2}' & \cdots & W_{2,K_{c}}' \\ \vdots & \vdots & \ddots & \vdots \\ W_{c,1}' & W_{c,2}' & \cdots & W_{c,K_{c}}' \end{pmatrix}.$$  

2. ‘Between blocks’ hankelization. The matrix $W_{i}'$ can be considered as a matrix of blocks. Therefore, hankelization by blocks lies in averaging of the blocks $W_{i,j}'$, $i + j = s$, where $1 \leq s \leq N_{c}$. The final form is block-Hankel with Hankel blocks:

$$\tilde{W}_{i} = \begin{pmatrix} \tilde{H}_{c} & \tilde{H}_{c+1} & \cdots & \tilde{H}_{K_{c}} \\ \tilde{H}_{c} & \tilde{H}_{c+1} & \cdots & \tilde{H}_{K_{c}+1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{H}_{c} & \tilde{H}_{c+1} & \cdots & \tilde{H}_{N_{c}} \end{pmatrix}.$$  

Notice that operations 1 and 2 are permutable.

**Comments**

The result of the method consists in a decomposition of the initial field into a sum of components. It is expected that if the field $F$ is a sum of a smooth surface, oscillations, and noise, so there exists such a grouping that the resultant decomposition (Eq. 10) is close to the
initial field decomposition. This gives an opportunity for smoothing, denoising, removing periodic noise, etc.

Rules to select 2D-SSA parameters are mostly similar to the 1D case (Golyandina et al., 2001). In particular, small window sizes produce adaptive smoothing. The most detailed decomposition and the better separation of the field components can be obtained if window sizes are close to \((N_x/2, N_y/2)\). However, the use of big window sizes can lead to the mixing problem caused by too detailed decomposition of components. Calculation cost can set additional restrictions on parameters, as one should find eigenvectors and eigenvalues of a large matrix \(L_x L_y \times L_x L_y\).

At the grouping step, selection of ET in Eq. 6 for the grouping stems from the fact that eigenfields and factor fields are similar to the component of \(F\), which generates them. In particular, smooth surface generates slowly varying factor fields.

**MATERIALS**

To validate 2D-SSA, we select a portion of the Northern Andes measuring 4º by 4º, located between 2º S and 2º N, and 78º30' W and 74º30' W (Fig. 1a). The area covers regions of Ecuador, Colombia, and Peru including parts of the coastal plain, the Andean range, and the Upper Amazon basin.

A digital elevation model (DEM) of the study area was extracted from GTOPO30, the 30 arc-seconds gridded global DEM (U.S. Geological Survey, 1996). The DEM measures 480 columns by 481 rows (viz., 230,880 points). It has a grid size of 30" (about 925 m). Elevation ranges from 5 m to 6085 m (Fig. 1b).

We selected this area and GTOPO30 on two counts. First, it is well known that this DEM incorporates a high-frequency noise caused by interpolation errors and inaccurate merging of topographic charts. Spatial distribution of noise in these DEMs is uneven and depends on the accuracy of cartographic sources. In particular, mountainous regions are marked by relatively low level of the noise, while the potent noise is typical for forested regions of Africa and South America. This is because reasonably detailed and accurate topographic data is unavailable for such regions, so the interpolation of sparse contours has been used to compile these portions of GTOPO30. Although DEM noise is no obstacle to produce realistic maps of elevation, it leads to derivation of noisy and unreadable maps of secondary topographic variables (e.g., curvatures). This is because their derivation is based on the calculation of the first and second partial derivatives of elevation that increases dramatically the noise (Florinsky, 2002). The study area, consisting of high mountains and forested foothills marked by different level of noise, is ideally suited for validating 2D-SSA as a tool to denoise DEMs. Second, the use of this DEM allows us to test possibilities of 2D-SSA to decompose a topographic surface into components of different scales under complex geomorphic conditions.

**DATA PROCESSING**

To reduce the huge range of elevations (6080 m), the initial DEM was transformed by taking the natural logarithm. Logarithmic DEMs were used in the further processing and mapping. We did not interpolate or smooth DTMs.

Using the window size of 30 by 30, the initial DEM was decomposed into 900 eigentriples
We evaluated various combinations of eigentriples to reconstruct DEMs. Finally, the best variants were selected. To denoise the DEM, a DEM was reconstructed from the ET 1-100. To exemplify DTM generalisation, two DEMs were reconstructed from the ET 1-50 and 1-25. To separate continental, regional, and local components of topography, six DEMs were reconstructed from the ET 1, 2, 3, 2-3, 4-25, and 51-100. To visualise a noise component of GTOPO30, a DEM was reconstructed from the ET 101-900. DEM processing was done by software 2D-SSA version 1.2 (© K. Usevich and N. Golyandina, 2005-2007).

Horizontal curvature ($k_h$), one of the most important and widely used topographic attributes, was derived from the initial and reconstructed DEMs by the method intended to process DEMs given by a spheroidal trapezoidal grid (Florinsky, 1998). $k_h$ values were also transformed by taking the natural logarithm:

$$k_h = \begin{cases} \ln(1+10^8 k_h), & k_h > 0 \\ 0, & k_h = 0 \\ -\ln(1-10^8 k_h), & k_h < 0 \end{cases}.$$  

(14)

To clarify an effect of denoising and generalisation, we mapped $k_h$ stratifying its values into two levels: $k_h > 0$ and $k_h < 0$ (areas of flow divergence and convergence, correspondingly). All DTMs produced had a grid size of 30". The plate carrée projection was used to map DTMs. Calculations and mapping was done with LandLord 4.0 (Florinsky et al., 1995). Initial and reconstructed DTMs were compared using Statgraphics Plus 3.0 (© Statistical Graphics Corp., 1994-1997).

RESULTS AND DISCUSSION

A visual comparison of elevation maps derived from the initial DEM (Fig. 3a) and two DEMs reconstructed from the ET 1-100 (Fig. 3b) and 1-50 (Fig. 3c) allows one to see nothing but marginal changes in image patterns. A cursory examination may lead to an underestimation of results of the DEM denoising and generalisation. $k_h$ maps give better insight into the results. Indeed, typical manifestation of interpolation errors – ‘tracks’ of contours (Florinsky, 2002) – can be found on the $k_h$ map derived from the initial DEM (Fig. 3d). These tracks are typical for the Andean foothills covered by dense rain forests. It is hardly probable that this map may be used for any application. However, there are no error tracks on $k_h$ maps derived from DEMs reconstructed from the ET 1-100 (Fig. 3e) and 1-50 (Fig. 3f). One can see so-called flow structures formed by convergence and divergence areas (black and white image patterns, correspondingly). These maps may be used for geomorphic and geological interpretation. A comparison between $k_h$ maps derived from different DEMs (Fig. 3d-f) shows a pronounced effect of the map generalisation. The less number of the eigentriples used to reconstruct a DEM, the more smooth and simplified image patterns obtained. Reducing the number of ET used for DEM reconstruction leads to the marked reduction of the range of $k_h$ values (Fig. 3d-f) but influences the range of elevation values only slightly (Fig. 3a-c). This is also demonstrated by histograms and scatter plots (we do not present them for reasons of space).

A DEM reconstructed from the ET 1 is marked by the highest level of generalisation. This DEM represents a generalised morphostructure of the continental scale, the Andean Range with foothills (Fig. 4a). Derivation of $k_h$ from this DEM allows us to reveal a system of near-NW-SE-striking lineaments (Fig. 4b), which may indicate strike-slip faults (Florinsky, 1996). Although there are no structures of this sort in the recent database of Quaternary faults (Eguez et al., 2003), this does not attest that the lineaments are of erosional origin. First, geology of the Upper Amazon basin is poorly known. Second, they may indicate pre-
Quaternary structures. Indeed, Chebanenko (1964) has described a system of deep-seated faults with the NW-SE strike situated to the southeast of the study area (Fig. 1a). The lineaments detected may be associated with northwestern extensions of the faults.

A DEM reconstructed from the ET 2-3 (Fig. 4c) represents landforms probably connected with regional tectonic structures. For a DEM reconstructed from the ET 2, patterns agree closely with the direction of the Andean range. A DEM reconstructed from the ET 3 represents features probably related to the $k_b$ lineaments (Fig. 4b). It is conceivable that near-NW-SE-striking structures control spatial distribution of river valleys of the Upper Amazon basin. A DEM reconstructed from the ET 4-25 (Fig. 4d) represents landforms associated with geomorphic processes of a regional scale. For example, one can find generalised valleys of tributaries of Amazon River.

A DEM reconstructed from the ET 51-100 (Fig. 5a) represents high-frequency components, which cannot be considered as noise. However, it is impossible to find familiar patterns of the drainage network on this map. These components might represent topographic manifestation of local geomorphic processes. A DEM reconstructed from the ET 101-900 (Fig. 5b) presents the noise inherent in GTOPO30 (Fig. 3a). This is precisely the noise, which rendered the $k_b$ map derived from the initial DEM (Fig. 3d) unsuitable for applications. Among other patterns, one can see a rectangular feature along the northeastern border of the map (Fig. 5b). This is a trace of the inaccurate merging of adjacent topographic charts during the compilation of GTOPO30. This map shows the difference between the initial DEM (Fig. 3a) and the DEM reconstructed from the ET 1-100 (Fig. 3b).

The DEM reconstructions from the ET 1, 2, 3 and 2-3 may be considered as application of low-pass filters to the initial DEM, while the DEM reconstruction from the ET 101-900 and 51-100 – as application of high-pass filters.

**CONCLUSIONS**

The results demonstrate that 2D-SSA is a powerful method to denoise DTMs. The 2D-SSA-based denoising of DEMs leads to extremely fine changes in elevations. These changes cannot be captured except by derivation of secondary terrain attributes. In fact, 2D-SSA can remove the noise without damage to the signal unlike usual smoothing by moving averages. This suggests that 2D-SSA is a method of exceptional importance for preliminary treatment of noisy DEMs including global ones. This opens the way to utilise noisy DEMs for derivation of important topographic variables, such as land surface curvatures.

2D-SSA can be used to decompose a topographic surface into components of continental, regional, and local scales. ET selection to reconstruct topographic components of different scales may be marked by non-uniqueness and ambiguity. The similar problem arises if a multiple regression or spectral filtering is used to extract regional or local components from DEMs. However, the problem is largely associated with the qualitative character of scale notions in geomorphology rather than with mathematical features of a decomposition technique. Moreover, as in the 1D case (Golyandina et al., 2001), 2D-SSA provides techniques to identify ET appropriate for a particular grouping (this was not used in the study). Finally, 2D-SSA is a model-free approach, that is, it does not use *a priori* assumption or statistical hypothesis about DEM structure as contrasted to multiple regressions.
REFERENCES


Florinsky, I.V., 1996, Quantitative Topographic Method of Fault Morphology Recognition, Geomorphology, 16, 103-119.


Fig. 1. The study area: (a) geographical location and some deep-seated faults (Chebanenko, 1964); and (b) elevation map.

Fig. 2. Principal component fields 1-9 of the DEM decomposition. Percentage reflects shares of the corresponding eigentriples in the singular value decomposition (Eq. 6). Logarithmic scale is used.
Fig. 3. DTM denoising. Elevation maps derived from the initial DEM (a), reconstructed from the ET 1-100 (b), and 1-50 (c); $k_h$ maps derived from the initial DEM (d), the DEM reconstructed from the ET 1-100 (e), and 1-50 (f).
Fig. 4. Low-frequency components. Elevation maps reconstructed from the ET 1 (a), 2-3 (c), and 4-25 (d); (b) $k_h$ map derived from the DEM reconstructed from the ET 1.

Fig. 5. High-frequency components. Elevation maps reconstructed from the ET 51-100 (a), and 101-900 (b).