Environmental Modelling & Software 89 (2017) 19–30

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Towards a framework for terrain attribute selection in environmental studies

Vincent Lecours [a, *], Rodolphe Devillers [a], Alvin E. Simms [a], Vanessa L. Lucieer [b], Craig J. Brown [a, c]

[a] Department of Geography, Memorial University of Newfoundland, 232 Elizabeth Avenue, St. John's, Newfoundland and Labrador, A1B 3X9, Canada
[b] Institute for Marine and Antarctic Studies, University of Tasmania, 20 Castray Esplanade, Battery Point, Tasmania, 7004, Australia
[c] Applied Ocean Technology Research Department, Nova Scotia Community College, 80 Mawiomi Place, Dartmouth, Nova Scotia, B2Y 0A5, Canada

ARTICLE INFO

Article history:
Received 30 May 2016
Received in revised form
28 November 2016
Accepted 29 November 2016
Available online 23 December 2016

Keywords:
Bathymetry
Geomorphometry
Surrogacy
Terrain analysis
Terrain attributes
Topography

ABSTRACT

Terrain attributes (*e.g.* slope, rugosity) derived from digital terrain models are commonly used in environmental studies. The increasing availability of GIS tools that generate those attributes can lead users to select a sub-optimal combination of terrain attributes for their applications. Our objectives were to identify sets of terrain attributes that best capture terrain properties and to assess how they vary with surface complexity. 230 tools from 11 software packages were used to derive terrain attributes from nine surfaces of different topographic complexity levels. Covariation and independence of terrain attributes were explored using three multivariate statistical methods. Distinct groups of correlated terrain attributes were identified, and their importance in describing a surface varied with surface complexity. Terrain attributes were highly covarying and sometimes ambiguously defined within software documentation. We found that a combination of six to seven particular terrain attributes always captures more than 70% of the topographic structure of surfaces.

1. Introduction

Combining georeferenced species data with environmental datasets has become common practice in environmental studies both in the terrestrial (Elith and Leathwick, 2009) and marine realms (Brown et al., 2011). Exploring species-environment relationships is important for habitat mapping, biogeographical classification, conservation, and management (Harris and Baker, 2012). Research in these fields has been fueled by progresses in remote sensing and Geographic Information Systems (GIS), along with the increase in data availability and computing power (Wiersma et al., 2011; Vierod et al., 2014). In parallel, these elements have motivated the development of geomorphometry (Bishop et al., 2012; Zhou and Zhu, 2013), the field that helps quantitatively describe digital terrain models (DTM) using terrain attributes such as slope, orientation or rugosity (Pike, 1995). Terrain attributes have been found to be linked with the distribution of many terrestrial and marine species in different types of environments (*e.g.* forests, agroecosystems, deep-sea, continental shelf) and are now routinely integrated in environmental studies (Bouchet et al., 2015; Lecours et al., 2016a). Other environmental disciplines that make use of terrain attributes include hydrology, soil mapping, vegetation mapping, geomorphology, meteorology and agriculture (Florinsky and Kuryakova, 1996; Florinsky et al., 2002; Lacroix et al., 2002; Hengl and Reuter, 2009; Schwanghart and Heckmann, 2012; Bispo et al., 2016).

Led by the increasing availability of different types of intuitive GIS tools that "automatically" derive terrain attributes from DTMs (Bishop et al., 2012; *e.g.* Klingseisen et al., 2008; Han et al., 2012; Rigol-Sanchez et al., 2015) - either digital elevation (DEM) or bathymetric (DBM) models - ecologists and other GIS users often select a small subset of terrain attributes to perform their analyses. Non-expert GIS users do not always understand the underpinnings of the numerous options available (Bishop and Shroder, 2004; Bouchet et al., 2015), and a lack of guidance can lead them to select an arbitrary and sub-optimal set of terrain attributes. Such selections are often based on the availability and simplicity of the

* Corresponding author. Fisheries and Aquatic Sciences, School of Forest Resources & Conservation, University of Florida, 1745 McCarty Drive Gainesville, FL, 32611-0410, USA.

E-mail addresses: vlecours@mun.ca (V. Lecours), rdeville@mun.ca (R. Devillers), asimms@mun.ca (A.E. Simms), vanessa.lucieer@utas.edu.au (V.L. Lucieer), craig.brown@nscc.ca (C.J. Brown).

http://dx.doi.org/10.1016/j.envsoft.2016.11.027
1364-8152/

GIS tools rather than on statistical grounds or ecological, biological, or geomorphological relevance. An inappropriate selection of terrain attributes can however produce results that do not accurately represent the observed phenomenon, fail to capture the key properties of the terrain relevant to the question or problem, and influence subsequent analysis (*e.g.* species-environment relationship measurements).

A same terrain attribute derived using different algorithms can also produce significantly different outcomes. For instance, Dolan and Lucieer (2014) demonstrated that five different slope algorithms derived from DBMs resulted in different slope surfaces, confirming previous work performed on DEMs (Jones, 1998a) and artificial surfaces (Jones, 1998b). Since the algorithms used by GIS tools are not always made explicit within the software, users are often left with little choice on which one to use and sometimes are not free to decide the details of particular parameters such as the neighbourhood size. These elements, combined with the lack of explicit statements in the ecological literature of algorithms and parameters used for deriving terrain attributes (Dolan and Lucieer, 2014), may lead to misleading and incorrect comparisons of results from different studies. To add to the confusion, geomorphometry is a field recognized for its ambiguous terminology (Bishop et al., 2012), where terrain attributes measuring a same terrain characteristic can be named differently depending on the source or software.

Finally, a poor selection of terrain attributes may cause covariation between variables. Being all derivatives of the same DTM, terrain attributes are likely to covary and induce redundancy in the analysis (Pittman et al., 2009), violating the basic assumptions of many statistical analysis methods used. For instance, Rooper and Zimmermann (2007) calculated a correlation of 0.90 between their measures of slope and rugosity. Assessing covariation between variables is however rarely performed (Graham, 2003), despite being recognized to obscure the influence of individual drivers on a response variable, and to impact statistical models, species distribution models and regression analyses (Hijmans, 2012; Dormann et al., 2013).

Selecting a suitable set of independent variables, including terrain attributes, is essential to ensure robust analyses and increase reliability of results in environmental studies (King and Jackson, 1999). A theoretical and operational framework to geomorphometric analysis is still to be defined (Pike, 1995), and “the use of quantitative geomorphological knowledge must be revisited in an analytical framework” (Bishop et al., 2012, p.6). This paper bridges geomorphometry and environmental studies by proposing an operational framework that addresses the common issue of terrain attribute selection in environmental applications like ecology. It aims to identify combinations of available terrain attributes that minimize covariation between attributes and optimize the information given on the characteristics of a terrain. The specific objectives are to 1) explore existing GIS software to compute available local terrain attributes, 2) identify groups of local terrain attributes that represent unique morphological terrain characteristics, 3) and explore the relationship between the importance of these groups and terrain complexity.

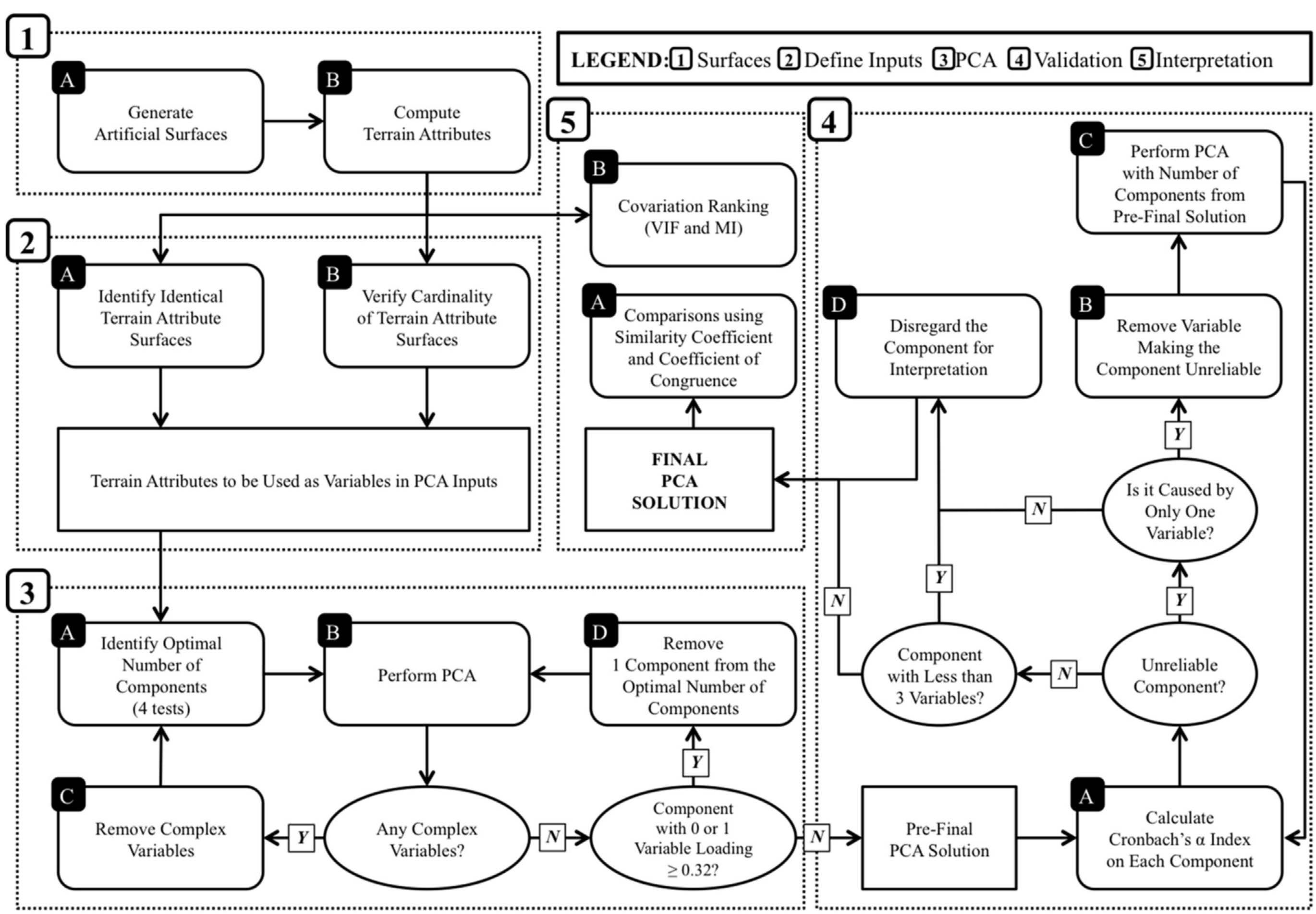


Fig. 1. Conceptual model of the analysis performed on each artificial surface.

2. Materials and methods

A summary of the methods is presented in Fig. 1. First, DTMs were generated from which terrain attributes were derived. Then, three iterative statistical methods were used to explore independence and both linear and non-linear relationships amongst terrain attributes: Principal Component Analysis (PCA), Variable Inflation Factor (VIF), and Mutual Information (MI). Because of the intricacy of the methods used in this study, more details on how and why they were used are provided in Appendix A to allow potential replication and generalization.

2.1. Surfaces and terrain attributes

2.1.1. Artificial surfaces

Artificial surfaces have proven to be valuable tools in ecology (With and King, 1997) and geomorphometry (Jones, 1998b). Since terrain attributes are sensitive to DTM errors and uncertainty (Raaflaub and Collins, 2006; Kinsey-Henderson and Wilkinson, 2013), artificial surfaces provide a controlled environment in which to test hypotheses (Halley et al., 2004). Nine artificial surfaces of 106 m × 106 m, with a 1 m spatial resolution, and presenting different complexity levels were created using spectral synthesis in LandSerf 2.3 (step 1A in Fig. 1; Fig. 2). The spectral synthesis technique, developed by Peitgen and Saupe (1988), is one of the methods that provide the most realistic artificial landscapes (Chipperfield et al., 2011). It generates surfaces so that their correlative characteristics, which are studied in this paper, are present at all spatial scales (Keitt, 2000). LandSerf requires a user-specified fractal dimension to create surfaces. Fractal dimension expresses and encodes the degree of complexity of objects (Dimri, 2005) and is a "useful simple summary of roughness distribution" that can be used to compare DTM characteristics (Lloyd, 2014, p.152). Fractal dimension was chosen to differentiate surfaces because of its scale-invariance properties (Pfeifer, 1984). These properties and the preservation of the correlative characteristics of the surfaces make results from this study generalizable to other surfaces with the same fractal dimension at different scales (see Appendix A). The fractal dimension of the computed surfaces ranges from 2.06 to 2.79 (Fig. 2), covering most complexity levels found in real terrain (2.20–2.60, Hofierka et al. (2009)) and other natural elements (2.28–2.61 in coral reefs, Zawada and Brock (2009)).

2.1.2. Terrain attributes and GIS tools

For each of the nine surfaces, 230 terrain attribute surfaces were derived (step 1B in Fig. 1; Appendix B) using 11 different commercial and open-source software (Table 1). A 3 × 3 analysis window was used as several software packages use this as default and do not allow modifying it. As commonly performed in ecology, measures of aspect (orientation) were transformed into northerness and easterness to remove circularity in the data (Olaya, 2009). To eliminate edge contamination, the outer 3 m were removed from all resulting attribute surfaces, reducing the extent to 100 m × 100 m surfaces.

2.1.3. Preparation of the data

To identify terrain attributes computed using the same algorithm, each attribute was tested against the others to detect pairs giving identical results (step 2A, Fig. 1). For each set of duplicates, only one of the two attributes was kept for further analysis to account for each algorithm only once. To keep the analysis objective in terms of software used, the attributes removed were added back before interpreting the results, assuming that their behaviour through the analyses would have been the same as their duplicate.

Cardinality (*i.e.* the number of different values for a variable) was assessed for each terrain attribute (step 2B, Fig. 1). Terrain attributes with low cardinality were identified and not used as input in PCA analyses (Section 2.2): such variables are known to complicate PCA solution as they account for only a negligible amount of the total variance (Tabachnick and Fidell, 2014). They were however carried over to the VIF and MI analyses (Section 2.3 below).

2.2. Principal component analysis

The first statistical analysis performed on the datasets of terrain attributes for each surface was an iterative PCA (step 3, Fig. 1). PCA was chosen to address the second objective as it helps removing collinearity and redundancy, and regrouping variables into uncorrelated groups of correlated variables (*i.e.* components) while minimizing information loss (Tabachnick and Fidell, 2014). PCA is often used in ecology to explore patterns in large multidimensional datasets as it can handle variable dependency where other multivariate techniques cannot (King and Jackson, 1999; McGarigal et al., 2000). PCA were performed using IBM SPSS Statistics software v.22, correlation matrices with standardized data, and a Varimax orthogonal rotation with Kaiser Normalization (Kaiser, 1958) (step 3B in Fig. 1; Appendix A). The orthogonal rotation of the solution allows for a "theoretically more meaningful" component structure that is easier to interpret (McGarigal et al., 2000, p.59), and the Varimax method is the most widely used and often performs better than others (Bhattacharyya, 1981; Henson and Roberts, 2006).

2.2.1. Identifying the optimal number of components

An over-extraction or under-extraction of components in ecological studies can lead to inferential issues (Franklin et al., 1995; Wood et al., 1996) (Appendix A). Several methods exist to estimate the appropriate number of PCA components to use (Zwick and Velicer, 1986). Since these methods are known to often give different results (Zwick and Velicer, 1986; O'Connor, 2000), Thompson and Daniel (1996, p.200) state that it is "appropriate and often desirable" to simultaneously use multiple methods. Using O'Connor (2000) SPSS programs, four methods were combined to determine the statistically optimal number of components to retain (step 3A, Fig. 1): Minimum Average Partial Correlation (MAP) (Velicer, 1976), Parallel Analysis (PA) (Horn, 1965), and modifications of MAP and PA respectively proposed by Velicer et al. (2000) and O'Connor (2015). The mode of the four results was chosen as the appropriate number of components to retain. Since this number depends on the number of PCA input variables, this step was performed before each PCA computation.

2.2.2. Complexity of variables

Our third objective required comparing components from different solutions, which is only possible if solutions reach a simple structure. A simple structure has an invariance property (Kaiser, 1958) that allows generalization (Rummel, 1970). When a simple structure is reached, PCA solutions from different computations will always find the same components regardless of the insertion of other variables in the dataset (Rummel, 1970). A simple structure is deemed to be reached when most components have "marker" variables, *i.e.* those that load strongly on only one component. Variables that load strongly on more than one component are called "complex" variables. They are redundant, do not contribute to the model, and have to be removed for the solution to reach a simple structure (Tabachnick and Fidell, 2014). After removing complex variables (step 3C, Fig. 1), steps 3A and 3B were performed again, and more terrain attributes were removed if

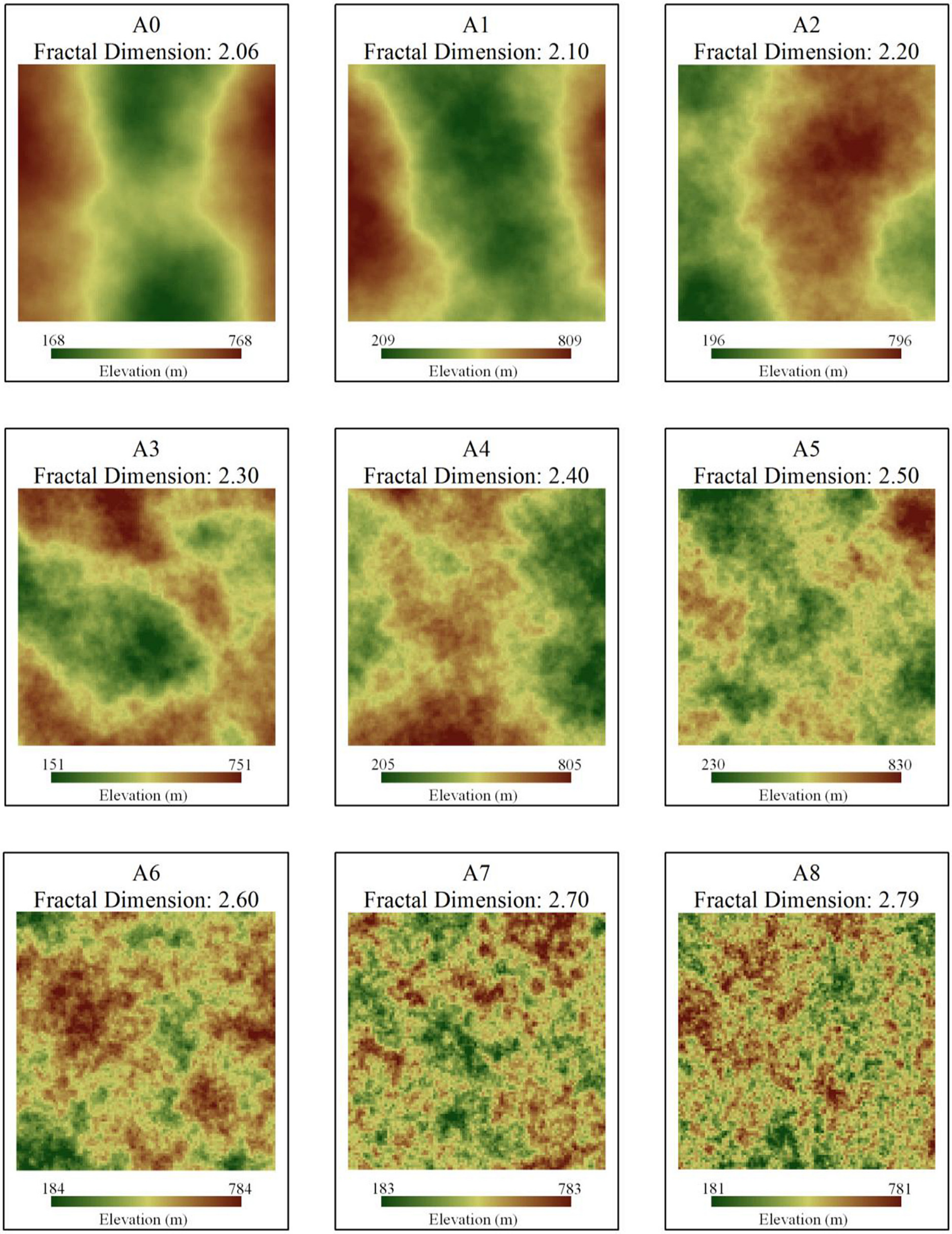


Fig. 2. Artificial surfaces computed and analysed in this study, from least complex (A0) to most complex (A8).

new complex variables appeared. Iterations stopped once the PCA ceased to isolate complex variables.

2.2.3. Towards a final solution

After removing complex variables, the previously obtained optimal number of components may need to be adjusted. If the last component had no or only one marker variable loading on it, the PCA was re-run with one less component (step 3D, Fig. 1). Once there were no more complex variable and no more components with less than two variables, the solutions had reached a

Table 1
List of software used, number of terrain attributes that were computed using each of them, and percentage of these terrain attributes that reached the final PCA solution for each surface.

Software and versions	Number of attributes computed	Percentage of terrain attributes in the final solution								
		A0	A1	A2	A3	A4	A5	A6	A7	A8
ArcGIS 10.2.2 with Python 2.7.8	22	86	91	82	77	73	77	77	82	82
ArcGIS 10.2.2 with DEM Surface Tools (v.2.1.399)	17	82	76	76	76	71	65	65	71	65
ArcGIS 10.2.2 with Benthic Terrain Modeler 3.0 rc3	12	83	92	83	83	83	83	83	83	67
Diva-GIS 7.5.0	7	86	86	100	100	100	100	100	100	100
Idrisi Selva 17.0	7	86	86	71	71	86	86	86	86	86
LandSerf 2.3	12	75	58	58	58	58	50	50	42	42
Quantum GIS 2.4.0 Chugiak	13	100	100	100	100	100	100	100	100	85
SAGA GIS 2.0.8	96	76	74	80	80	78	58	71	68	66
TNTmips Free 2014 (MicroImages)	25	100	100	80	84	96	56	92	84	80
uDig 1.4.0b	9	89	100	78	67	78	78	78	56	56
Whitebox GAT 3.2.1 Iguazu	10	90	90	100	90	90	70	80	80	70
Total Number of Attributes:	230	83	83	81	80	81	67	77	74	70

simple structure and were ready for validation.

2.2.4. Validation

Components' internal consistency was assessed in SPSS using Cronbach's α coefficient of reliability (Cronbach, 1951) (step 4A, Fig. 1; Appendix A). Unreliable components (α index below 0.6) were made reliable when possible (*i.e.* if unreliability was caused by a single variable, Appendix A) by removing the variable, and PCA was re-run (steps 4B and 4C, Fig. 1). When unreliable components could not be made reliable, they were not considered in the interpretation of results.

Components with less than three variables are considered weak and unstable (Costello and Osborne, 2005) and do not meet the criteria of replicability (Gorsuch, 1983). Variables in these components have a higher probability to be grouped by chance (Gorsuch, 1983). Components demonstrating this characteristic were thus not carried over to the interpretation (step 4D).

2.2.5. Comparisons

In step 5A (Fig. 1), a quantitative comparison of the general configuration of the solutions (*i.e.* one per surface) was performed using two measures that together represent a “geometrically important and meaningful” way of comparing components (Rummel, 1970, p.462): a similarity coefficient (SC) (Harman, 1967) and a coefficient of congruence (CC) (Tucker, 1951) (see Appendix A). These coefficients measure different underlying characteristics of the PCA solution. SC quantifies similarity of components in terms of magnitude and direction of the components, while CC looks at pattern and magnitude of the components through the shared variance of the components (Cattell, 1978).

2.2.6. Interpretation

Rummel (1970) identified five elements to look at when interpreting and comparing PCA solutions: number of components, variance, complexity, communality, and configuration. The number of components is an indication of the convergence level of the dataset to a certain dimensionality. The variance indicates the relative importance of each component and helps investigate if a component's importance is specific to a particular surface or is generalizable (Gorsuch, 1983). Communality is a measure of the variance of a variable that is accounted for by the sum of the components. It can help identifying which variables are unique: if the communality of a particular variable is very low, this variable does not match very well the solution and needs to be further explored (see Section 2.3). The configuration corresponds to the pattern and magnitude of loadings of variables on a solution. Finally, the complexity represents the behaviour of a variable, *i.e.* if it shifts from a component to another in a different solution. The PCA results are presented below and interpreted according to these five elements.

2.3. Covariation assessment: variable inflation factor and mutual information

Two drawbacks of PCA are that it does not account for non-linear relationships and that if a variable does not covary with any others (potentially being unique), it will not load on any component. To address these issues, two independent stepwise measures of covariation were computed in the statistical software R v. 3.1.1: the Variable Inflation Factor (VIF) and Mutual Information (MI) (step 5B in Fig. 1; Appendix A). The VIF is one of the most commonly used methods to detect covariation due to its simplicity (Belsey et al., 2004) and is recommended over other methods by Dormann et al. (2013) to assess covariation in datasets. MI is a measure of co-expression widely used in information theory that has the capacity of detecting non-linear relationships and is often used to generalize the correlation structure of a dataset through a Mutual Information Matrix (MIM) (Steuer et al., 2002; Song et al., 2012). For each measure, the variables were ranked from least covarying to most covarying and an average of the two rankings was made (Appendix A). This average ranking was used to confirm PCA results and explore the uniqueness of the variables that did not load on any component, had a very low amount of variance accounted for by the sum of the components (*i.e.* low communality), or were not considered in the PCA because of their low cardinality.

3. Results

3.1. Terrain attributes and software

Of the 230 computed terrain attributes, 48 were found to be identical to another one and were removed from further analysis (Appendix B). Eight other terrain attributes were not considered for the PCA analyses because of their low cardinality (Appendix C). SAGA GIS offered the most terrain attributes ($n = 96$), while Diva-GIS and Idrisi Selva offered the least ($n = 7$) (Table 1). However, more terrain attributes computed from a software package did not necessarily result in more of them being useful in describing surface variability. For instance, the 13 attributes computed using Quantum GIS were all found in the final PCA solutions of the eight least complex surfaces (A0 to A7). LandSerf was the software that had the lowest percentage of terrain attributes reaching the final solutions, with an average of 55% (Appendix C).

3.2. Principal component analysis

3.2.1. Number of components, variance and communality

A general pattern did not emerge from the number of components extracted with regards to surface complexity (Appendix C). A slight decreasing trend of total variance accounted for could be observed with increasing surface complexity (Appendix C). The solution corresponding to the lowest complexity surface accounted for 91.67% of the topographic structure, while the solution of the highest complexity surface accounted for 86.56%. A minimum was reached with surface A6 (85.35%). The average percentage of variance was 88.91%. When only considering reliable components, the percentage of variance ranged from 80.88% (A7) to 87.37% (A0), with an average of 84.34%.

Information on the communalities of each variable for each PCA solution can be found in Appendix C. 72% of the terrain attributes showed a negative trend where communality decreased with increasing surface complexity, which is also shown by the decreasing proportion of high communality terrain attributes as complexity increases (Appendix C). Over 90% of the variables of the least complex surfaces (A0 to A3) have high communalities. Surfaces A4 and A8 have the highest percentage of variables with low communalities, with 2.7% and 2.4% respectively.

3.2.2. Configuration

Several terrain attributes were found to be consistently grouped on the same components across solutions. These groups are presented in Fig. 3A and their category, or interpretation, was based on the variables with the highest loadings in each of them (McGarigal et al., 2000). The locations of these groups on the different solutions are represented in the configuration summary presented in Fig. 3B.

A visual interpretation of Fig. 3B allows identifying the groups that loaded on the same component for more than one surface, thus indicating similarity between the solutions. This similarity was confirmed quantitatively by the SC and CC values, measured on reliable components of consecutive solutions: the pairs of solutions A0-A1, A2-A3, A4-A5, and A5-A6 are all considered identical, with a one to one match of their components. Three other pairs of solutions were considered very similar, with only one inversion in components: A1 and A2 are the same except for the inversion of the second and fifth components, A6 and A7 show the same configuration but the inversion of the statistical parameters and slope attributes (fifth and sixth components), and A7 and A8 would be considered identical if it were not for the inversion of the fourth and fifth components.

Only one pair of consecutive solutions, A3-A4, is considered significantly different. Only three out of ten pairs of components were found to be matching: the first, fifth and sixth components. The overall mismatch is caused by the decreasing relative importance of the two groups of rugosity indices (groups 2 and 7), and the increased relative importance of the second group of slope algorithms (group 5B).

3.2.3. Complexity of variables

Fig. 3B shows how the complexity of the major groups of terrain attributes presented in Fig. 3A vary with surface complexity. Only Group 1 (interpreted as curvatures and relative position) does not vary in complexity across solutions. In general, the aspect (Groups 3A-3B) and local statistical attributes (Group 4) shift to a higher component as surface complexity increases. On the opposite, slopes (Groups 5A-5B) shift to a lower component as surface complexity increases. Group 2 does not demonstrate any particular pattern.

Table 2 presents a summary of the solutions according to variable's behaviour (*e.g.* complex or marker variables), and Appendix C helps assessing changes in behaviour with changing surface complexity. Some of the variables that are not included in the major groups of terrain attributes previously identified show particular patterns. For instance, a small number of northerness measures (ID104, ID105, ID107, ID108, ID109) are marker variables in solutions from low complexity surfaces, and become complex variables as surface complexity increases. A similar pattern is observed for measures of total curvature. On the other hand, some terrain attributes, such as a few measures of profile curvature (ID135, ID141, ID142, ID144) are complex variables for low complexity surfaces and become marker variables as surface complexity increases.

3.3. Variable inflation factor and mutual information

The average rank of each variable according to VIF and MI is presented in Appendix C. 114 variables became more covarying (*i.e.* ranked lower) with increasing surface complexity, while 116 variables became less covarying with increasing surface complexity. Variables from Group 1 were generally losing some positions in the ranking, while variables from Group 2 and 7 were gaining ranks.

Of the eight variables whose cardinality values were too low to enter the PCA, three of them consistently ranked in the top 50 least covarying variables: center versus neighbours variability (ID2), and one measure of easterness (ID42) and northerness (ID101). Of the variables that had low communality for any surface, mean of residuals (ID70), representativeness (ID158), standard deviation of slope (ID195), and one measure of profile curvature (ID146), plan curvature (ID120), tangential curvature (ID204), easterness (ID42) and northerness (ID104) were consistently falling in the top 50 least covarying variables.

4. Discussion

4.1. Surface complexity and importance of terrain attributes

Results indicate that the relative importance of terrain attributes in capturing relevant information on terrain characteristics varies with surface complexity. For instance, the relative importance of slope diminishes as surfaces become more complex, while local statistical attributes importance increases. Also, the importance of differentiating the algorithms used to compute a terrain attribute seems to vary with surface complexity. For instance, most slope algorithms are grouped into one component for the least complex surfaces, indicating that they are all highly correlated, but get separated across several components as surface complexity increases, indicating that some of them are not correlated anymore. An opposite trend can be observed for vector ruggedness measures (VRM) algorithms. For low complexity surfaces, VRMs computed with Sappington et al., (2007) method are relatively important (they load on the second component), while the VRMs computed with SAGA GIS' algorithms load on the tenth component. VRMs however all load together on the tenth component of the most complex surfaces, with no difference between the algorithms used.

Also, both the quantitative (SC and CC) and qualitative (Fig. 3B) comparative assessments of the solutions' configuration allowed finding an important break when surface complexity reaches a fractal dimension of 2.40 (surface A4). This is likely to correspond to a threshold where a significant number of terrain attributes' behaviour changes in relation to terrain characteristics. This threshold is located somewhere between a fractal dimension of 2.30 and 2.40, representing approximately the middle of the range of complexities tested.

a)

MAJOR GROUPS	Group 1	Group 2	Group 3A	Group 3B	Group 4	Group 5A	Group 5B
Category	Curvatures and Relative Position	Rugosity Indices	Orientation (Aspect)		Local Statistical Attributes	Slope	
Terrain Attributes	Topographic Position Index, General Curvature, Deviation from Mean Value, Plan Curvature, Convergence Index, Minimum Curvature, Maximum Curvature, Mean Curvature, Morphometric Position Index	Terrain Ruggedness Index, Surface Ratio, Standard Deviation, Melton Ruggedness Index, Roughness, Range	Easterness	Northerness	Mean, Median, Minimum, Maximum	Slopes (Horn's Method)	Slopes (4-Cell Method, Zevenbergen & Thorne Method)

MINOR GROUPS	Group 5C	Group 5D	Group 5E	Group 6A	Group 6B	Group 7	Group 8
Category	Slope			Vector Ruggedness Measures		Rugosity Indices	
Terrain Attributes	Slopes (Maximum Slope Algorithm and Maximum Triangle Slope Algorithm)	Slopes (Quadratic Surfaces, Least-Square Fit Algorithm)	Other Slope Algorithms	Vector Ruggedness Measures (Sappington's Method)	Vector Ruggedness Measures (SAGA GIS Algorithms)	Planimetric-to-Surface Ratio, Ruggedness Index	Representativeness, Residual at Centre

MINOR GROUPS	Group 9A	Group 9B	Group 9C	Group 9D	Group 9E	Group 9F	Group 9G
Category	Curvatures						
Terrain Attributes	Profile Curvatures (TNT mips Algorithms)	Plan Curvatures (TNT mips Algorithms)	Profile & Plan Curvatures (SAGA GIS Algorithms)	Curvatures (DEM Surface Tools' Algorithms)	Curvatures (WhiteBox GIS Algorithms)	Curvatures (Haralick's Algorithm)	Curvatures (uDig Algorithms)

b)

COMPONENTS	SURFACES' PCA SOLUTIONS									COMPONENTS
	A0	A1	A2	A3	A4	A5	A6	A7	A8	
1	Group 1									1
	Cross-sectional Curvature									
		Profile Curvature								
	Tangential Curvature									
2	Groups 5A, 5B, 5D, 5E & 6A, Fractal Dimension, Surface Roughness Index		Groups 2 & 7, Variance, Coefficient of Variation		Group 3A					2
	Group 5C									
3	Group 3A				Group 3B					3
4	Group 3B				Groups 2 & 7, Coefficient of Variation			Group 4		4
					Variance					
5	Groups 2 & 7, Variance		Groups 5A & 5E					Group 4	Group 2, Coefficient of Variation	5
			Group 5B				Group 5D			
6	Group 4							Group 5A		6
7	Group 9A				Group 5B, Fractal Dimension					7
8	Group 9B				Group 9A					8
9	*Groups 9E & 9G*			Group 5B	Group 9B	*Groups 9E & 9G*	Group 9B			9
10	Group 6B				*Groups 9E & 9G*		Groups 6A & 6B			10
11	*Total Curvature, Slope Variability*			Groups 9E & 9G	Group 6B		Group 5C	Group 5C	Group 9F	11
12			*Group 9C*	*Total Curvature*	*Group 9C*		*Groups 9E & 9G*	Group 5D	Groups 9E & 9G	12
13			*Group 8*	*Group 9C*	Group 5C			Group 9F	Group 5D	13
14				*Group 8*	Group 5D			Group 6A	Group 9D	14
15					*Group 8*			Group 9D		15
16								Group 9E		16

Fig. 3. A: Main groups of covarying terrain attributes that consistently loaded together on the different components.B: Summary of the PCA solutions' configuration. The different groups are detailed in Fig. 3A. Characters in italic represent the unreliable components as assessed by Cronbach's α, and underlined characters indicate components with less than three variables.

4.2. Terrain attributes, algorithms and software

The diversity and amount of variables loading on the same components demonstrate high level of covariation amongst terrain attributes. For low complexity surfaces for instance, measures of slope covary with local fractal dimension, measures of roughness, and VRM, attributes often used together in ecological studies (Harris and Baker, 2012). Group 1 also shows high variable diversity (Fig. 3A). About one third of the attributes were found to correlate with several components (*i.e.* the complex variables), thus being very collinear and holding low potential for ecological applications. For instance, measures of longitudinal curvature were complex

Table 2
Percentage of the 230 variables that formed each solution or were removed during the iterative PCA. Complex variables that loaded equally on more than one component were removed, while complex variables that loaded more strongly on one component were kept (Appendix A).

	In the final solution		Removed during iterations		
	Marker variables	Complex variables		Reliability	Cardinality
A0	66.1	17.4	13.0	0.0	3.5
A1	63.0	19.6	13.9	0.0	3.5
A2	62.2	19.1	14.8	0.4	3.5
A3	63.5	17.0	16.1	0.0	3.5
A4	65.2	15.7	15.2	0.4	3.5
A5	63.9	3.0	29.1	0.4	3.5
A6	70.4	7.0	18.3	0.9	3.5
A7	63.9	10.9	20.9	0.9	3.5
A8	62.2	8.3	24.8	1.3	3.5

variables for all surfaces, and their high level of covariation was confirmed by their VIF-MI ranking. Of the 230 computed attributes, 79% were found to be unique (with no identical attribute surfaces), confirming and extending the conclusions from Dolan and Lucieer (2014) and Jones (1998a; 1998b) that same terrain attributes derived using different algorithms or software can give different results.

The extent to which different slope algorithms are divided into sub-groups as surface complexity increases and how they rank on the different components is particularly interesting. Skidmore (1989) argued that Zevenbergen and Thorne (1987) method was better than Horn (1981) even if results for the latter were not very different from those of the former. Our results demonstrate that while this is the case for low complexity surfaces, Horn's method captures more variance in more complex surfaces. Skidmore (1989) conclusions may then be limited by, and dependent on, the surfaces that were used. On the other hand, Hodgson (1995) argued that Zevenbergen and Thorne (1987) method worked best on rough surfaces, while Horn (1981) worked best on smooth surfaces, without further distinction on what are rough and smooth surfaces. Our results suggest that Horn (1981) method (Group 5A) captures more variance than others for rougher surfaces (fractal dimension ≥ 2.30) and that there are no differences for low complexity, or smoother, surfaces (fractal dimension ≤ 2.20). Jones (1998b) found that the performance of slope algorithms depended on the relative importance of random noise versus mean smooth elevation difference: in general the 4-cell method and Horn's method performed better than others, but when a surface was very noisy, Sharpnack and Akin (1969) method, from which Zevenbergen and Thorne (1987) method is derived, performs better.

Some ambiguity in the names and types of terrain attributes was found during the analysis, confirming previous observations (Bishop et al., 2012). For instance, topographic position index (TPI) measures from SAGA GIS (ID214, ID217) were found identical to "difference from mean values" (ID23, ID26) attributes. Some attributes identified simply as "curvature" (ID14, ID16) were found to be duplicates of measures of profile (ID143, ID145) and plan curvature (ID123, ID125); profile and plan curvatures are however supposed to describe different terrain characteristics. Solely using "curvature" to describe a measure could mislead a user into believing that it is a measure of general or total curvature. Also, different types of curvatures were often found to be inter-correlated when computed with the same software, but uncorrelated to those generated from other software. For instance, Groups 9C, 9D, 9E, 9F and 9G all combine correlated measures of curvature that should theoretically be uncorrelated, but that are generated from the same software. Finally, several terrain attributes from SAGA GIS computed from three different methods (no distance weighting, inverse of the distance, and squared inverse of the distance) often gave identical results. Since these measures are directly dependent on the window size used (distance-based algorithms), it is possible that these attributes would give different outcomes if a bigger window size would be used. Further work will be necessary to assess the behaviour of terrain attributes with changing surface complexity using different window sizes.

4.3. Suitable subset of terrain attributes

The five major groups of terrain attributes presented in Fig. 3A were always the ones accounting for the most variance, regardless of terrain complexity. These groups alone accounted for an average of 75% of the overall topographic structure (Appendix C). Based on psychometric methods, a suitable subset of variables in a PCA consists of selecting one variable from each component (Gorsuch, 1983). Since these five groups always loaded on different but the highest components for all levels of topographic complexity, a suitable subset of terrain attributes would include one variable from each of the groups 1, 2, 3A, 3B, 4, and 5A.

While any attributes from each of these groups could be a good choice, we recommend a specific combination that, in addition to reducing covariation and redundancy, also reduces ambiguity: relative difference to mean value (Group 1), local standard deviation (Group 2), easterness and northerness (Groups 3A-3B), local mean (Group 4), and a measure of slope preferably computed with Horn's method (Group 5A, Table 2 in Appendix B). Relative difference to mean value is a measure of relative position, local standard deviation is a measure of rugosity, and easterness and northerness are measures of orientation derived from aspect. Local mean may however not be required if users include elevation or depth in the analysis, as the local mean will be highly correlated with the input DTM. However, local mean could potentially be more reliable than the initial DTM if a surface is noisy, as using the mean may filter out some of the noise. This recommendation of six attributes increases replicability and generality as it selects terrain attributes that are easily computed from any software over terrain attributes that are only available in some of them. For instance, we recommend using local standard deviation in Group 2 over terrain ruggedness index or roughness as there is no ambiguity on how to compute standard deviation and all software have focal statistics tools that can compute it. To help potential users of this method, we provide with this paper a toolbox developed for ArcGIS, named TASSE (Terrain Attribute Selection for Spatial Ecology), that automatically generates the six proposed terrain attributes (Lecours, 2015).

While using the proposed terrain attributes helps optimize the information extracted from the terrain, it does not necessarily mean that all of the proposed attributes will be useful for a given ecological application (Lecours et al., 2016b). For instance, using them in species distribution modelling exercises will not necessarily result in these six terrain attributes being drivers or surrogates of species distribution. Rather, it means that most of the terrain properties, or topographic structure, and the variation in these properties will be accounted for when performing the analysis and modelling. An analyst that includes such selection can be assured that most of the topographic structure is considered in the analysis, and can then focus on the integration of the selection with other environmental data (*e.g.* remotely sensed data, climate data, oceanographic data, vegetation data). Also, some of these terrain attributes could be correlated in a particular application, despite extracting different information from the surface. For instance, if high slopes are found in regions of a study area with high rugosity, the slope and standard deviation terrain attributes would be spatially correlated, despite extracting different information from the surface.

It is important to note that this combination of attributes and the alternatives (if selecting one different attribute from each group) are mainly based on the PCA analysis. However, the exploration of communality combined with the covariation assessment using VIF and MI allowed identifying unique variables that do not correlate highly with any other but could potentially give information on unique terrain characteristics. Some of these variables may improve the proposed selection of terrain attributes, and we recommend further exploration of these unique attributes in future work (see Lecours et al., 2016b).

5. Conclusion

If the use of terrain attributes in environmental studies is now common, selecting an appropriate set of terrain attributes is a challenge rarely approached carefully in these studies. This can have potential consequences that can invalidate the analyses and not make the best use of terrain data. This study conducted an extensive analysis to suggest a suitable selection of terrain attributes that optimizes the information derived from DTMs. Three methods were used to iteratively assess the relationships between 230 terrain attribute surfaces computed from nine artificial surfaces of different complexity using 11 different software packages. Results confirm that (1) different algorithms computing a same terrain attribute (*e.g.* slope, curvature) can produce different outcomes, that (2) terrain attributes are highly covarying, that (3) there is some ambiguity in the denomination of terrain attributes, and that (4) their selection for any application needs to be carefully performed. Our conclusions highlight the importance to explicitly report the software, algorithms, and parameters used to generate terrain attributes in ecological studies to allow careful interpretation of the results and comparison between studies. We also encourage software and tools developers to be explicit in their documentation or metadata about the methods or algorithms used by their products.

Based on our analysis, we recommend the use of six local terrain attributes that optimize the local topographic structure accounted for when performing environmental studies: (1) relative difference to mean value (a measure of relative position), (2) local standard deviation (a measure of rugosity), (3) easterness and (4) northerness (measures of orientation), (5) local mean, and (6) slope (preferably computed with Horn's method). The proposed selection reduces redundancy, covariation and ambiguity, and improves generality and replicability, and can be applied across a wide range of terrain complexity. While the six proposed attributes can easily be computed using any GIS package, an ArcGIS toolbox was provided with this paper to easily generate these attributes (Lecours, 2015). Our work has also identified unique terrain attributes that were not considered by the PCA analysis but that showed potential to represent different characteristics of a terrain and that needs to be further explored.

Our recommendations provide an operational framework to any users willing to incorporate topography or bathymetry and their derivatives (*i.e.* terrain attributes) in environmental models and analyses. While the six proposed attributes may not necessarily be useful for all applications and should be tested separately, for instance as effective surrogates, their combination ensures that the amount of topographic structure accounted for is optimized in the analysis. The proposed operational framework can help users make more robust analyses and bridge geomorphometry with disciplines like ecology, biogeography, habitat mapping and distribution modelling. An application of our recommended terrain attributes and their comparison to other subsets is presented in a real ecological application, namely a marine benthic habitat mapping exercise, in Lecours et al. 2016b.

Software availability

A toolbox for ArcGIS (TASSE: Terrain Attribute Selection for Spatial Ecology) and the Python scripts used to build it are provided with this paper, and can be downloaded for free at www.marinegis.com.

Acknowledgements

Many thanks are due to Dr. Arnaud Vandecasteele for his help with the generation of data, and to Emma LeClerc and Dr. Evan Edinger for their critical comments on the manuscript. This work was supported by the Natural Sciences and Engineering Council of Canada (NSERC).

Appendices

Appendix A: Detailed material and methods.

Appendix B: Artificial surfaces and list of terrain attribute derived, with software, algorithms, references, and duplicates.

Appendix C: Extended results of the iterative principal component analysis (PCA), variable inflation factor (VIF), and mutual information (MI), for all surfaces and computed terrain attributes.

Complete appendices, including Appendix B and Appendix C, can be downloaded here: https://www.dropbox.com/s/h7o5c574dj282sr/Lecours_etal_EMS_appendices.pdf.

Appendix A. . Detailed materials and methods

Surfaces and terrain attributess

Artificial surfaces. Despite not being perfectly fractal (Halley et al., 2004), real terrains often demonstrate fractal-like properties (Milne, 1992; With and King, 1997) and several authors indicate that such fractal-based surfaces are appropriate to develop “null hypotheses” (Evans and McClean, 1995; Tate, 1998; Halley et al., 2004). Because of the scale-dependency of terrain attributes and topography (Tate and Wood, 2001), a scale-invariant measure was essential to characterize the complexity of the representation of the surface (*i.e.* the DTM), rather than the complexity of the terrain itself. For instance, if a rough terrain is represented using a broad-resolution DTM, it may appear smooth: since terrain attributes are dependent on the DTM and not the real terrain, a suitable subset of terrain attributes to characterize this particular DTM would be one that is appropriate for smooth surfaces. This information would be captured by the fractal dimension of the DTM. If the DTM had a higher spatial resolution, it would represent better the roughness of the terrain, the fractal dimension would be higher, and the appropriate subset of terrain attributes would be chosen accordingly. A surface's fractal dimension can theoretically range from 2.0 (very smooth) to 2.9 (very complex) (Peterson, 1984).

Principal component analysis

Identifying the optimal number of components. Minimum Average Partial Correlation (MAP) (Velicer, 1976) and Parallel Analysis (PA) (Horn, 1965) are commonly used and recommended by statisticians as they tend to give better results than other methods (Zwick and Velicer, 1986; Henson and Roberts, 2006). MAP is however recognized to sometimes extract too few components (O'Connor, 2000) while PA is recognized to sometimes extract too many components (Buja and Eyuboglu, 1992). The revised MAP

(Velicer et al., 2000) raises the partial correlations used in the calculation to the fourth power rather than squared. The modified PA (O'Connor, 2015) uses the raw dataset (*i.e.* the actual values of terrain attributes) rather than theoretical values to determine the number of components. This method is considered very accurate and relevant for datasets that are not normally distributed, which is often the case in environmental datasets (Austin, 1987; O'Connor, 2015).

Complexity of variables. The complex variables that loaded significantly more on one component than the others were kept in the analysis as they may help exploring the behaviour of variables during interpretation.

Validation. A component was found unreliable when its Cronbach's α index (Cronbach, 1951) was lower than 0.6 (Nunnally, 1978). Although, during the computation of α, SPSS also computes the potential values of α if each individual variable is removed from the component. This allowed checking if only one specific variable was making the component unreliable. When this was the case, this specific variable was removed from the component (step 4B, Fig. 1 of main article) and the PCA was re-run (step 4C, Fig. 1 of main article).

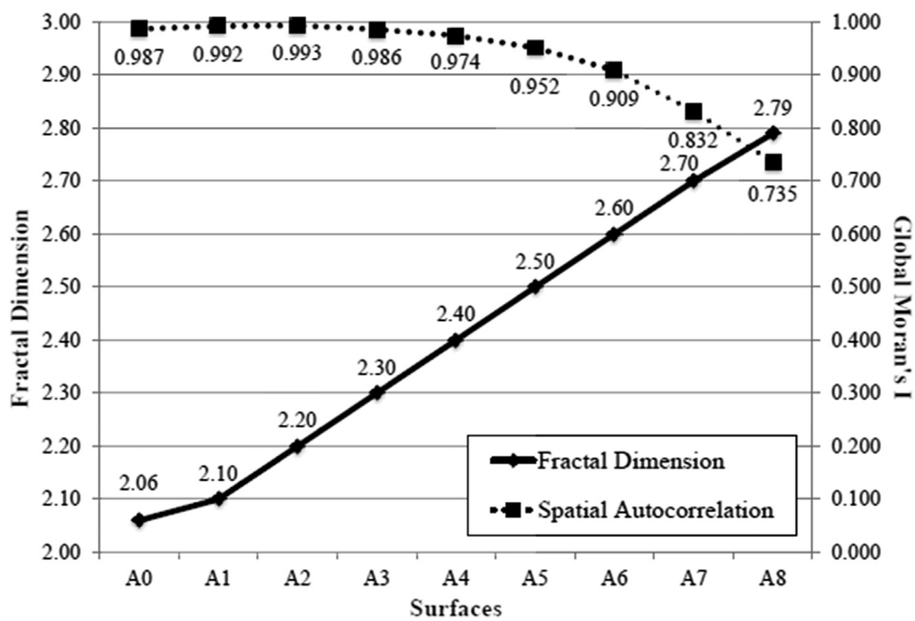
Covariation assessment: Variable inflation factor and mutual information

These measurements lack meaningful thresholds to separate the variables that demonstrate covariation to those that do not (Belsey et al., 2004), but allow ranking the attributes from least covarying to most covarying. Such ranking is often performed in machine-learning as a pre-processing step (Kohavi and John, 1997), and has proven to be computationally efficient and statistically robust (Guyon and Elisseeff, 2003).

Stepwise calculations of VIF and MI were necessary because of the changing levels of co-association with variables being removed from the datasets. The two stepwise algorithms were computed in the statistical software R v. 3.1.1 and work the same way: they (1) rank terrain attributes based on the calculation of the VIF or MI measure of each of them, (2) remove the most redundant or least informative terrain attribute, (3) save it in a separate list, and (4) repeat the process until all the attributes are ranked in the list.

Adequacy of Methods: Replicability, Reliability and Generalization

The Global Moran's I values of the artificial surfaces (Figure A4)



show high spatial autocorrelation of the elevation values, which is expected in environmental data (Legendre, 1993). We thus believe that the artificial surfaces are good surrogates of natural terrain. In addition, the fractal dimension tested cover most of what can be find in natural environment (Hofierka et al., 2009; Zawada and Brock, 2009).

The comparability of the study design applied on each surface (Fig. 1 of main article) allowed us to compare PCA solutions and draw conclusions from the comparisons (Rummel, 1970). The 10,000 pixels included in the statistical analyses are considered enough to obtain generalizable and replicable results and produce more accurate solutions (Barrett and Kline, 1981; Costello and Osborne, 2005); Comrey and Lee (1992) advise that more than 1000 samples is excellent, but that 10 observations per variables is also good. 10,000 samples is thus significantly more than the 1740 samples that would have been necessary (230 variables − 8 low cardinality variables − 48 duplicates = 174 variables in the PCA).

According to Gorsuch (1983), communality is important for replicability as it assesses the appropriateness of the PCA model and consequently serves to validate the method. In the current study, the high average communalities of each solution are an indication that the iterative PCA were appropriate, stable, and replicable (Cliff and Pennell, 1967).

The relatively constant variance of the first component across the solutions is an indication of the invariance property of Group 1 of terrain attributes (Kaiser, 1958). An invariant component is highly reliable and replicable as is indicates that the importance of the variables loading on it do not exclusively belong to these solutions, but are inherent to any similar datasets (Gorsuch, 1983).

The results of VIF and MI were compared to make sure that one of the two methods was not consistently ranking variables higher or lower than the other method. For each surface, about half of the variables were ranked higher by one of the two methods, thus indicating that none of them influenced the average more than the other. Since they both measure covariation from different factors, we believe that their average is a good indication of the overall covariation behaviour of the variables.

Finally, the high loadings on each component (Appendix C), high correlations between variables loading on the components (Appendix C), and the meaningfulness of each component (Fig. 3A of the main article) are indications of the appropriateness of the method (Wood et al., 1996). These are characteristics of a simple structure solution that allows for generalization of results (Kaiser, 1958): the same clusters of variables are consistently found (Fig. 3A and B).

References

Austin, M.P., 1987. Models for the analysis of species' response to environmental gradients. Vegetatio 69, 35–45.
Barrett, P.T., Kline, P., 1981. The observation of variable ratio in factor analysis. Personality Study Group Behav. 1, 23.
Belsey, D.A., Kuh, E., Welsch, R.E., 2004. Regression Diagnostics: Identifying Influential Data and Sources of Multicollinearity. Wiley, Hoboken.
Bhattacharyya, H., 1981. Theory and methods of factor analysis and principal components. In: Capen, D.E. (Ed.), The use of multivariate statistics in studies on wildlife habitat, pp. 72–79. U.S. Forest Service General Technical Report RM-87.
Bishop, M.P., Shroder Jr., J.F., 2004. GIScience and mountain geomorphology: overview, feedbacks, and research directions. In: Bishop, M.P., Schroder Jr., J.F. (Eds.), Geographic Information Science and Mountain Geomorphology. Springer-Praxis, Chichester, pp. 1–31.
Bishop, M.P., James, L.A., Shroder Jr., J.F., Walsh, S.J., 2012. Geospatial technologies and digital geomorphological mapping: concepts, issues and research. Geomorphology 137, 5–26.
Bispo, P.C., Santos, J.R., Valeriano, M.M., Graça, P.M.L., Balzter, H., França, H., Bispo, P.C., 2016. Predictive models of primary tropical forest structure from geomorphometric variables based on SRTM in the Tapajos Region, Brazilian Amazon. PLoS One 11, e0152009.
Bouchet, P.J., Meeuwig, J.J., Kent, C.P.S., Letessier, T.B., Jenner, C.K., 2015. Topographic determinants of mobile vertebrate predator hotspots: current knowledge and future directions. Biol. Rev. 90, 699–728.
Brown, C.J., Smith, S.J., Lawton, P., Anderson, J.T., 2011. Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuar. Coast. Shelf Sci. 92, 502–520.
Buja, A., Eyuboglu, N., 1992. Remarks on parallel analysis. Multivar. Behav. Res. 27, 509–540.
Cattell, R.B., 1978. The Scientific Use of Factor Analysis in Behavioral and Life Sciences. Plenum Press, New York.
Chipperfield, J.D., Dytham, C., Hovestadt, T., 2011. An updated algorithm for the generation of neutral landscapes by spectral synthesis. PLos ONE 6, e17040.
Cliff, N., Pennell, R., 1967. The influence of communality, factor strength, and loading size on the sampling characteristics of factor loadings. Psychometrika 32, 309–326.
Comrey, A.L., Lee, H.B., 1992. A First Course in Factor Analysis. Erlbaum, Hillsdale.
Costello, A.B., Osborne, J., 2005. Best practices in explanatory factor analysis: four recommendations for getting the most from your analysis. Pract. Assess. Res. Eval. 10, 1–9.
Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334.
Dimri, V.P., 2005. Fractals in geophysics and seismology – an introduction. In: Dimri, V.P. (Ed.), Fractal Behaviour of the Earth System. Springer-Verlag, Berlin, pp. 1–22.
Dolan, M.F.J., Lucieer, V.L., 2014. Variation and uncertainty in bathymetric slope calculations using geographic information systems. Mar. Geod. 37, 187–219.
Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J.R., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46.
Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697.
Evans, I.S., McClean, C.J., 1995. The land surface is not unifractal: variograms, cirque scale and allometry. Z. für Geomorphol. 101, 127–147.
Florinsky, I.V., Kuryakova, G.A., 1996. Influence of topography on some vegetation cover properties. Catena 27, 123–141.
Florinsky, I.V., Eilers, R.G., Manning, G.R., Fuller, L.G., 2002. Prediction of soil properties by digital terrain modelling. Environ. Model. Softw. 17, 295–311.
Franklin, S.B., Gibson, D.J., Robertson, P.A., Pohlmann, J.T., Fralish, J.S., 1995. Parallel analysis: a method for determining significant principal components. J. Veg. Sci. 6, 99–106.
Gorsuch, R.L., 1983. Factor Analysis, second ed. Erlbaum, Hillsdale.
Graham, M.H., 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84, 2809–2815.
Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.
Halley, J.M., Hartley, S., Kallimanis, A.S., Kunin, W.E., Lennon, J.J., Sgardelis, S.P., 2004. Uses and abuses of fractal methodology in ecology. Ecol. Lett. 7, 254–271.
Han, W., Di, L., Zhao, P., Shao, Y., 2012. DEM Explorer: an online interoperable DEM data sharing and analysis system. Environ. Model. Softw. 38, 101–107.
Harman, H.H., 1967. Modern Factor Analysis, revised edition. University of Chicago Press, Chicago.
Harris, P.T., Baker, E.K., 2012. Seafloor Geomorphology as Benthic Habitat: GeoHab Atlas of Seafloor Geomorphic Features and Benthic Habitats. Elsevier, Amsterdam.
Hengl, T., Reuter, H.I., 2009. Geomorphometry: Concepts, Software, Applications. Developments in Soil Science, 33. Elsevier, Amsterdam.
Henson, R.K., Roberts, J.K., 2006. Use of explanatory factor analysis in published research: common errors and some comment on improved practice. Organ. Res. Methods 7, 191–205.
Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial bias and calibration with a null model. Ecology 36, 27–46.
Hodgson, M.E., 1995. What cell size does the computed slope/aspect angle represent? Photogrammetric Eng. Remote Sens. 61, 513–517.
Hofierka, J., Mitášová, H., Neteler, M., 2009. Geomorphometry in GRASS GIS. In: Hengl, T., Reuter, H.I (Eds.), Geomorphometry: Concepts, Software, Applications. Elsevier Science, Amsterdam, pp. 387–410.
Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30, 179–185.
Horn, B.K.P., 1981. Hill shading and the reflectance map. Proc. IEEE 69, 14–47.
Jones, K.H., 1998a. A comparison of algorithms used to compute hill slope as a property of the DEM. Comput. Geosciences 24, 315–323.
Jones, K.H., 1998b. A comparison of two approaches to ranking algorithms used to compute hill slopes. GeoInformatica 2, 235–256.
Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika 23, 187–200.
Keitt, T.H., 2000. Spectral representation of neutral landscapes. Landsc. Ecol. 15, 479–493.
King, J.R., Jackson, D.A., 1999. Variable selection in large environmental data sets using principal components analysis. Environmetrics 10, 67–77.
Kinsey-Henderson, A.E., Wilkinson, S.N., 2013. Evaluating shuttle radar and interpolated DEMs for slope gradient and soil erosion estimation in low relief terrain. Environ. Model. Softw. 40, 128–139.
Klingseisen, B., Metternicht, G., Paulus, G., 2008. Geomorphometric landscape analysis using a semi-automated GIS-approach. Environ. Model. Softw. 23, 109–121.
Kohavi, R., John, G., 1997. Wrappers for feature selection. Artif. Intell. 97, 273–324.
Lacroix, M.P., Martz, L.W., Kite, G.W., Grabrecht, J., 2002. Using digital terrain analysis modeling techniques for the parametrization of a hydrologic model. Environ. Model. Softw. 17, 127–136.
Lecours, V., 2015. Terrain Attribute Selection for Spatial Ecology (TASSE), V. 1.0. www.marinegis.com.
Lecours, V., Dolan, M.F.J., Micallef, A., Lucieer, V.L., 2016a. A review of marine geomorphometry, the quantitative study of the seafloor. Hydrology Earth Syst. Sci. 20, 3207–3244.
Lecours, V., Brown, C.J., Devillers, R., Lucieer, V.L., Edinger, E.N., 2016b. Comparing selections of environmental variables for ecological studies: a focus on terrain attributes. PLoS One. http://dx.doi.org/10.1371/journal.pone.0167128.
Legendre, P., 1993. Spatial autocorrelation: trouble of new paradigm? Ecology 74, 1659–1673.
Lloyd, C.D., 2014. Exploring Spatial Scale in Geography. Wiley, United States.
McGarigal, K., Cushman, S.A., Stafford, S.G., 2000. Multivariate Statistics for Wildlife and Ecology Research. Springer-Verlag, New York.
Milne, B.T., 1992. Spatial aggregation and neutral models in fractal landscapes. Am. Nat. 139, 32–57.
Nunnally, J.C., 1978. Psychometric Theory, second ed. McGraw-Hill, New York.
Olaya, V., 2009. Basic land-surface parameters. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry: Concepts, Software, Applications. Elsevier Science, Amsterdam, pp. 141–169.
O'Connor, B.P., 2000. Using parallel analysis and Velicer's MAP test. Behav. Res. Methods, Instrum. Comput. 32, 396–402.
O'Connor, B.P., 2015. SPSS, SAS, MATLAB, and R Programs for Determining the Number of Components and Factors Using Parallel Analysis and Velicer's MAP Test. https://people.ok.ubc.ca/brioconn/nfactors/nfactors.html [accessed 21 January 2015].
Peitgen, H.O., Saupe, D., 1988. The Science of Fractal Images. Springer-Verlag, New York.
Peterson, I., 1984. Ants in the labyrinth and other fractal excursions. Sci. News 125, 42–43.
Pfeifer, P., 1984. Fractal dimension as working tool for surface-roughness problems. Appl. Surf. Sci. 18, 146–164.
Pike, R.J., 1995. Geomorphometry: progress, practice, and prospect. Z. für Geomorphol. 101, 221–238.

Pittman, S.J., Costa, B.M., Battista, T.A., 2009. Using lidar bathymetric and boosted regression trees to predict the diversity and abundance of fish and corals. J. Coast. Res. 53, 27–38.
Raaflaub, L.D., Collins, M.J., 2006. The effect of error in gridded digital elevation models on the estimation of topographic parameters. Environ. Model. Softw. 21, 710–732.
Rigol-Sanchez, J.P., Stuart, N., Pulido-Bosch, A., 2015. ArcGeomorphometry: a toolbox for geomorphometric characterisation of DEMs in the ArcGIS environment. Comput. Geosciences 85, 155–163.
Rooper, C.N., Zimmermann, M., 2007. A bottom-up methodology for integrating underwater video and acoustic mapping for seafloor substrate classification. Cont. Shelf Res. 27, 947–957.
Rummel, R.J., 1970. Applied Factor Analysis. Northwestern University Press, Evanston.
Sappington, J.M., Longshore, K.M., Thompson, D.B., 2007. Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave desert. J. Wildl. Manag. 71, 1419–1426.
Schwanghart, W., Heckmann, T., 2012. Fuzzy delineation of drainage basins through probabilistic interpretation of diverging flow algorithms. Environ. Model. Softw. 33, 106–113.
Sharpnack, D.A., Akin, G., 1969. An algorithm for computing slope and aspect from elevations. Photogramm. Eng. 35, 247–248.
Skidmore, A.K., 1989. A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. Int. J. Geogr. Inf. Syst. 3–4, 323–334.
Song, L., Langfelder, P., Horvath, S., 2012. Comparison of co-expression measures: mutual information, correlation, and model based indices. BMC Bioinforma. 13, 328.
Steuer, R., Kurths, J., Daub, C.O., Weise, J., Selbig, J., 2002. The mutual information: detecting and evaluating dependencies between variables. BMC Bioinforma. 18, S231–S240.
Tabachnick, B.G., Fidell, L.S., 2014. Using Multivariate Statistics, sixth ed. Pearson Education Limited.
Tate, N.J., 1998. Estimating the fractal dimension of synthetic topographic surfaces. Comput. Geosciences 24, 325–334.
Tate, N.J., Wood, J., 2001. Fractals and Scale Dependencies in Topography. In: Tate, N., Atkinson, P. (Eds.), Scale in Geographical Information Systems. Wiley, Chichester, pp. 35–51.
Thompson, B., Daniel, L.G., 1996. Factor analytic evidence for the construct validity of scores: a historical overview and some guidelines. Educ. Psychol. Meas. 56, 197–208.
Tucker, L.R., 1951. A method for synthesis of factor analysis studies. Personnel Research Section Report 984. Department of the Army, Washington D.C.
Velicer, W.F., 1976. Determining the number of components from the matrix of partial correlations. Psychometrika 41, 321–327.
Velicer, W.F., Eaton, C.A., Fava, J.L., 2000. Construct explication through factor or component analysis: a review and evaluation of alternative procedures for determining the number of factors or components. In: Goffin, R.D., Helmes, E. (Eds.), Problems and Solution in Human Assessment. Kluwer, Boston, pp. 41–71.
Vierod, A.D.T., Guinotte, J.M., Davies, A.J., 2014. Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. Deep-Sea Res. II 99, 6–18.
Wiersma, Y.F., Huettmann, F., Drew, C.A., 2011. Introduction. landscape modeling of species and their habitats: history, uncertainty, and complexity. In: Drew, C.A., Wiersma, Y.F., Huettmann, F. (Eds.), Predictive Species and Habitat Modeling in Landscape Ecology. Springer, New York, pp. 1–6.
With, K.A., King, A.W., 1997. The use and misuse of neutral landscape models in ecology. Oikos 79, 219–229.
Wood, J.M., Tataryn, D.J., Gorsuch, R.L., 1996. Effects of under- and overextraction on principal axis factor analysis with Varimax rotation. Psychol. Methods 1, 354–365.
Zawada, D.G., Brock, J.C., 2009. A multiscale analysis of coral reef topographic complexity using lidar-derived bathymetry. J. Coast. Res. 53, 6–15.
Zevenbergen, L.W., Thorne, C.R., 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Landforms 12, 47–56.
Zhou, Q., Zhu, A.X., 2013. The recent advancement in digital terrain analysis and modeling. Int. J. Geogr. Inf. Sci. 27, 1269–1271.
Zwick, W.R., Velicer, W.F., 1986. Comparison of five rules for determining the number of components to retain. Psychol. Bull. 99, 432–442.