Landscape Ecol
DOI 10.1007/s10980-015-0248-7

RESEARCH ARTICLE

Surface metrics: scaling relationships and downscaling behavior

Amy E. Frazier

Received: 21 November 2014 / Accepted: 20 July 2015

Abstract
Context Considerable research has examined scale effects for patch-based metrics with the ultimate goal of predicting values at finer resolutions (i.e., downscaling), but results have been inconsistent. Surface metrics have been suggested as an alternative to patch-based metrics, although far less is known about their scaling relationships and downscaling potential. If successful, downscaling would enable integration of disparate datasets and comparison of landscapes using different resolution datasets.
Objectives (1) Determine how surface metrics scale as resolution changes and how consistent those scaling relationships are across landscapes. (2) Test whether these scaling relationships can be accurately downscaled to predict metric values for finer resolutions.
Methods Various scaling functions were fit to 16 surface metrics computed for multiple resolutions for a set of landscapes. Best-fitting functions were then extrapolated to test downscaling behavior (i.e., predict metric value for a finer resolution) for an independent set of validation landscapes. Relative error was assessed between the predicted and true values to determine downscaling robustness.
Results Seven surface metrics (*Sa*, *Sq*, *S10z*, *Sdq*, *Sds*, *Sdr*, *Srwi*) fit consistently well ($R^2 > 0.99$) with a 3rd order polynomial or power law. Of those, the scaling functions for *Sa*, *Sq*, and *S10z* were able to predict metric values at a finer resolution within 5 %. Three metrics, (*Ssk*, *Sku*, *Sfd*) were also notable in terms of fit and downscaling.
Conclusions Many metrics exhibit consistent scaling relations across resolution, and several are able to accurately predict values at finer resolutions. However, prediction accuracy is likely related to the amount of information lost during aggregation.

Keywords Scaling · Grain size · Resolution · Gradient landscapes · Tree canopy cover · Impervious surface area

Electronic supplementary material The online version of this article (doi:10.1007/s10980-015-0248-7) contains supplementary material, which is available to authorized users.

A. E. Frazier (✉)
Department of Geography, Center for Applications of Remote Sensing, Oklahoma State University, 337 Murray Hall, Stillwater, OK 74078, USA
e-mail: amy.e.frazier@okstate.edu

Introduction

The issue of scaling is central to ecology, particularly landscape ecology (Levin 1992; Urban 2005). Specifically, it is well known in landscape ecology that patch-based landscape patterns are spatially correlated and scale dependent (Wu 2004). These scale dependencies have been studied thoroughly, most notably with respect to the effects of changing grain size (i.e.,

Published online: 28 July 2015

pixel resolution) on spatial pattern indices (Turner et al. 1989; Milne 1991; Benson and MacKenzie 1995; Wickham and Riitters 1995; Jelinski and Wu 1996; O'Neill et al. 1996; Cain et al. 1997; Neel et al. 2004; Uuemaa et al. 2005; Na and Li 2013). A major outcome of these explorations has been the identification of consistent scaling relationships (i.e., linear or power law) for a handful of patch-based landscape metrics. These scaling relationships can describe the behavior of the metric as map resolution changes (Turner et al. 1989; Wu et al. 2000, 2002; Shen et al. 2004; Wu 2004; Saura 2004; Castilla et al. 2009; Alhamad et al. 2011; Li et al. 2011; Argañaraz and Entraigas 2014) and have the potential to help us understand how landscapes change across different scales. However, apart from the knowledge that scaling relationships exist, they have seen limited operational use or application.

A major interest in investigating scaling relationships is to ultimately develop ways to estimate metrics at a finer (i.e., higher) resolution than the original data were collected (Saura and Castro 2007). The ability to estimate landscape patterns at fine resolutions from coarser resolution data would enable detailed investigations without costly aerial or satellite data acquisitions (Riitters 2005). It would also allow the integration of disparate datasets (Atkinson 2012) and comparisons of landscapes at different points in time using different resolution data, without sacrificing information (Argañaraz and Entraigas 2014). Theoretically, estimation at finer resolutions can be completed via the scaling relationships described above, in which a statistical relationship (i.e., scaling function) is established from coarse resolution maps, and metric values are predicted for a finer resolution using that function. This process is known as 'downscaling'. Several attempts have been made to downscale patch-based landscape metrics (Riitters 2005; Garcia-Gigorro and Saura 2005; Saura and Castro 2007; Gardner et al. 2008; Argañaraz and Entraigas, 2014; Frazier 2014), but results have not been consistent. Proposed theories as to why there are inconsistencies include the 'scale domain' concept by Wiens (1989), which hypothesizes that there may be certain regions of the scale spectrum where pattern changes are predictable. Other possible explanations for the inconsistencies include issues matching the scale of observation with the scale of analysis (Karl and Maurer 2010) and also the varying levels of heterogeneity between different landscapes (Frazier 2014), but the true causes remain unknown.

Meanwhile, researchers are realizing that traditional patch-mosaic models of the landscape do not adequately represent continuous spatial heterogeneity (McIntyre and Barrett 1992; McIntyre and Hobbs 1999; Manning et al. 2004; McGarigal and Cushman 2005; Fischer and Lindenmayer 2006; McGarigal et al. 2009; Frazier and Wang 2013). Thus, there has been recent impetus to move beyond patch-mosaic models and incorporate additional landscape heterogeneity through the use of continuous landscape surfaces. Continuous surfaces represent phenomena that vary progressively across space, and they are characterized by a continuum rather than discrete values. An example of a continuous surface is a remotely sensed image that has been classified using the normalized difference vegetation index (NDVI). Each pixel is assigned a value ranging from -1.0 to $+1.0$ depending on the amount of live, green vegetation. Other types of continuous surfaces include elevation rasters and soft, or fuzzy, classifications of land cover where each pixel contains the proportion of a particular land cover type.

Continuous surfaces cannot be analyzed in the same manner as patch-mosaics because they do not contain discrete land cover boundaries (Frazier and Wang 2011). As a result, the large body of knowledge that has been dedicated to understanding and developing scaling relationships for patch-based metrics is not applicable to continuous surfaces. Instead, new metrics are being developed and tested for these landscapes including the recent adoption of surface metrology techniques (McGarigal et al. 2009) in landscape ecology. Despite several barriers to accessing and using surface metrics (e.g., software availability and cost), these metrics have already been successfully applied for modeling continuous spatial heterogeneity (Moniem and Holland 2013), and their use is expected to grow considerably once software becomes more widely available. However, little is known about the effects of changing spatial scale on continuous landscapes or the scaling relationships for the surface metrics being used to analyze them.

Since surface metrics are specifically designed to quantify the heterogeneity of continuous landscapes, and heterogeneity is known to be scale dependent (Wu 2004), it is expected that many surface metrics will be spatially correlated and scale dependent, much like their patch-based counterparts. Additionally, surface metrics are expected to exhibit more accurate downscaling capabilities than patch-based metrics. This is because surfaces can be statistically aggregated (e.g., based on mean) instead of relying on majority rules aggregation, which has been cited as a major reason why downscaling attempts for traditional patch-based metrics have been inconsistent (Frazier 2014). However, in-depth investigation is needed to determine (1) the exact nature of scaling relationships for surface metrics, and (2) the ability to extrapolate those scaling functions to accurately predict surface metric values for finer resolutions.

The objective of this research is to examine scaling relationships for a suite of surface metrics to determine whether or not these scaling relationships are consistent across landscapes and are robust enough to be extrapolated to predict metric values at finer, unmeasured spatial resolutions. The specific aims are to (1) examine how surface metrics scale (e.g., power law, polynomial, etc.) as resolution changes and determine how consistent those relationships are across different landscapes, and (2) test whether these scaling relationships are robust for accurately predicting metric values at finer resolutions through downscaling. Successful identification of consistent scaling functions that are able to accurately predict metric values for finer resolution surfaces would greatly increase the operational value and applicability of these metrics for ecological analyses.

Data and methods

The study area includes a variety of different landscapes in the United States. Two types of surface data were acquired from the Multi-Resolution Land Characteristics Consortium (MRLC) National Land Cover Database (NLCD) (Jin et al. 2013) in order to test realistic landscape data used in ecological applications. The MRLC produces continuous gradient surfaces of (1) percent impervious surface area (ISA), and (2) percent tree cover canopy (TCC) across the U.S. at 30 m resolution. Both datasets are derived from Landsat imagery with other ancillary data sources (Xian et al. 2011; Coulston et al. 2012, 2013).

Thirty different surfaces (15 ISA and 15 TCC) were selected to represent a variety of regions across the continental U.S. in order to capture a wide range of diverse urban and natural landscapes (Fig. 1). Since the focus of this study is on scaling behavior at different spatial grains, all landscapes were clipped to a spatial extent of approximately 20 × 20 km in an effort to reduce any ancillary scaling effects that might be introduced by variable extents (Turner et al. 1989; Wu 2004). Fifteen surfaces, including a mixture of TCC and ISA landscapes, were randomly selected to be used for calibrating the scaling functions (noted in Fig. 1). Hereafter, these surfaces are referred to as 'calibration' surfaces. The remaining fifteen landscapes were reserved to independently test the accuracy of the selected scaling for downscaling robustness. These surfaces are referred to as 'validation' surfaces.

To derive the set of coarse resolution surfaces needed to calibrate the scaling relationships, the original 30 m TCC and ISA surfaces for each landscape were aggregated to five additional coarse resolutions (60, 120, 180, 240, and 360 m) using statistical mean aggregation. In mean aggregation, the average of the four contributing 30 m pixels is assigned to the larger, aggregated 60 m pixel, and so on for each resolution. Mean aggregation has been found to retain a greater amount of information than other statistical aggregation techniques (Bian and Butler 1999). Next, 16 surface metrics (Table 1) were computed for each of the six resolutions of each landscape using Scanning Probe Image Processor ($SPIP^{TM}$) software. Metric computations used a plane correction for the overall surface mean, following the method used in McGarigal et al. (2009).

The six metric values for each landscape were then plotted as a scalogram, and four different scaling functions (power law, 1st order, 2nd order, and 3rd order polynomials) were fit to each scalogram to determine the best-fitting curve. Linear and power law scaling functions have been found to be characteristic of many patch-based scaling relationships (Wu 2004; Argañaraz and Entraigas 2014). The 2nd and 3rd order polynomial functions were included after preliminary assessment of scaling characteristics. A separate curve was fit to each calibration landscape using the Matlab Curve Fitting Toolbox (The Mathworks Inc., 2012). Goodness-of-fit for each scaling function was assessed using R^2. A separate R^2 value was computed for each metric, for each scaling function, for each of the 15 calibration landscapes ($16 \times 4 \times 15 = 960$ R^2 values). Values were then averaged across the 15

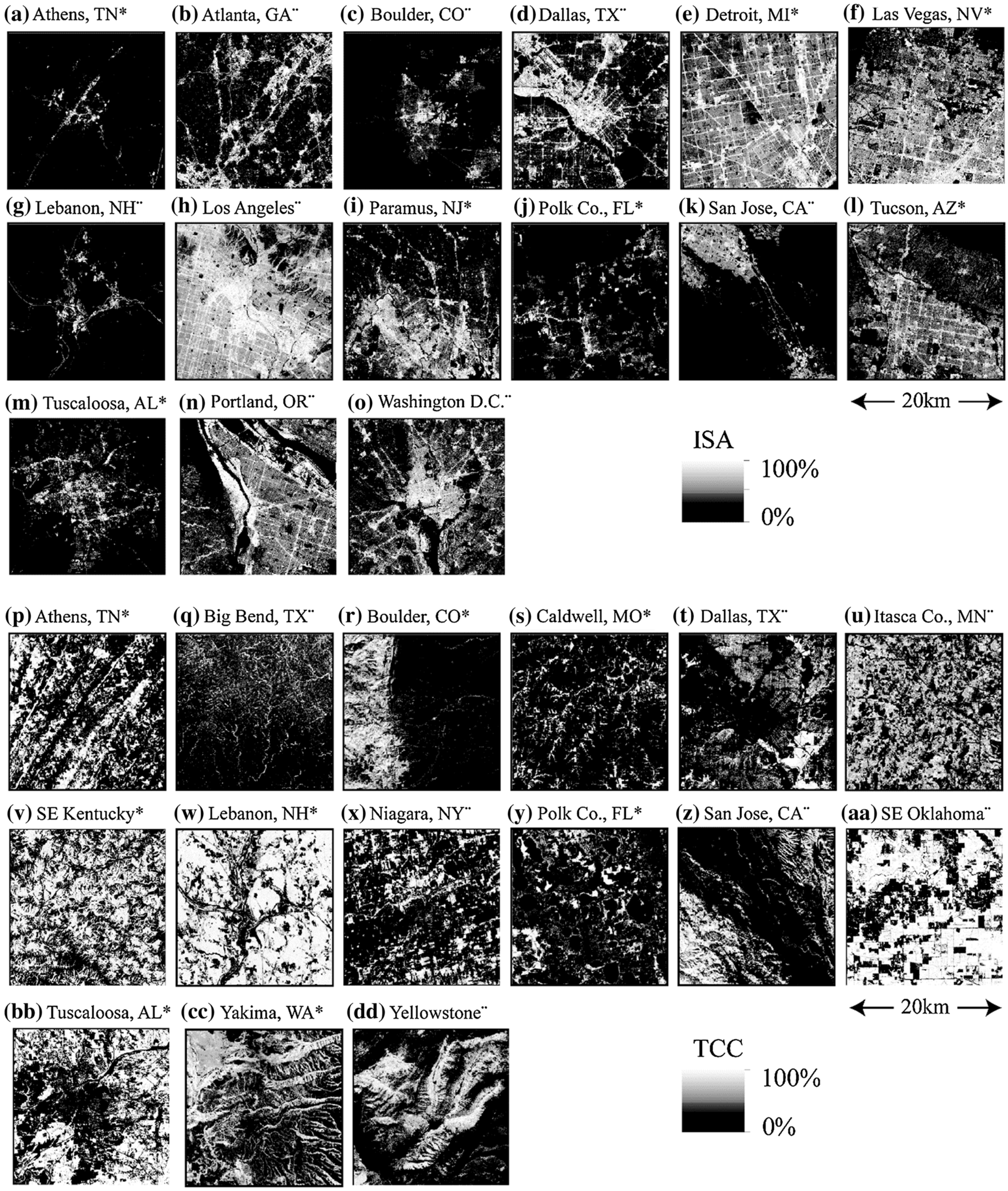

Fig. 1 Study areas (60 m resolution). *Top* group are impervious surface area (ISA) surfaces and *bottom* group are tree canopy cover (TCC) surfaces. Surfaces with (*asterisk*) indicate calibration landscapes, and surfaces with (*double dots*) indicate validation landscapes

calibration landscapes to produce a single μR^2 value for each metric at each scaling function ($16 \times 4 = 64$ μR^2 values reported). A μR^2 value of 0.99 was selected as the cutoff for considering the function a 'consistent good fit'. The use of a high R^2 threshold is based on prior research that found consistent scaling

Table 1 Surface metrics tested

Surface metric	Symbol[a]	Description
Amplitude metrics		
Roughness average	*Sa*	Statistical average of surface heights
Root mean square	*Sq*	Root mean square of surface heights
Surface skewness	*Ssk*	Skewness of surface heights
Surface kurtosis	*Sku*	Kurtosis of surface heights
Ten point height	*S10z*	The average height of the five highest local maximas plus average height of five lowest local minimas
Root mean square gradient	*Sdq*	The root mean square value of the surface slope within the sampling area
Surface bearing metrics		
Surface bearing index	*Sbi*	Uses the Abbott Curve to determine the area between 5 % and the maximum height.
Core fluid retention index	*Sci*	Void volume (area above the bearing area curve) in the core zone
Spatial metrics		
Summit density	*Sds*	The number of local maximas per area
Surface area ratio	*Sdr (%)*	Increment of the interfacial surface area relative to the area of the projected, flat, x, y plane
Texture direction index	*Stdi*	A measure of the dominance of the dominating texture direction
Radial wave index	*Srwi*	A measure of the dominance of the dominating radial wavelength
Fractal dimension	*Sfd*	The rate at which surface height increases with the scale of observation
Texture direction	*Std*	The angle of the dominating texture in the image
Texture aspect ratio	*Str20* *Str37*	Ratio of fastest to slowest decay to correlation 20 and 37 % of the autocorrelation function, respectively

[a] Symbols and descriptions based on Scanning Probe Image Processing (SPIP™) software. See Appendix—Supplemental Electronic Material for further details

relationships will typically have R^2 values above 0.99 (Frazier 2014).

To test the robustness of each scaling function for predicting an accurate metric value at a finer resolution (i.e., downscaling), the best-fitting scaling function determined from the calibration landscapes was applied to the validation landscapes. For this step, the 30 m resolution surface was set aside prior to fitting the scaling function, and the scaling function was fit to only the five coarsest surfaces (Fig. 2). The fitted scaling function was then extrapolated to 30 m (also using Matlab) to predict a metric value. The value of the metric at 30 m as predicted by the scaling function (i.e., the 'predicted' value) was then compared to the metric computed for the actual 30 m surface (i.e., the 'true' value) through a measure of relative error:

$$E_{rel}(\%) = \left|\left(M_p - M_t\right)/M_t\right| \times 100 \quad (1)$$

where E_{rel} is the relative error for a particular metric, M_p is the predicted metric based on the scaling function, and M_t is the true metric value calculated for the original 30 m surface. Low E_{rel} values indicate that the scaling function accurately predicts the metric value for a finer resolution surface. A separate E_{rel} value was computed for each metric for each of the 15 validation landscapes, and values were averaged across the 15 validation landscapes to produce a single μE_{rel} value for each metric. A μE_{rel} threshold of 5 % was selected as the cutoff for determining whether the scaling function accurately predicted the true value of the surface metric based on coarser resolutions. Metrics with scaling functions that satisfied this downscaling accuracy criteria were considered 'robust'. While the 5 % threshold is somewhat stringent, it is important that downscaling produce highly accurate results, particularly if the results are to be used as input into other types of analysis (e.g., super-resolution mapping). Alternatively, users may choose to downscale less-robust metrics if their applications do not require such stringent accuracy standards.

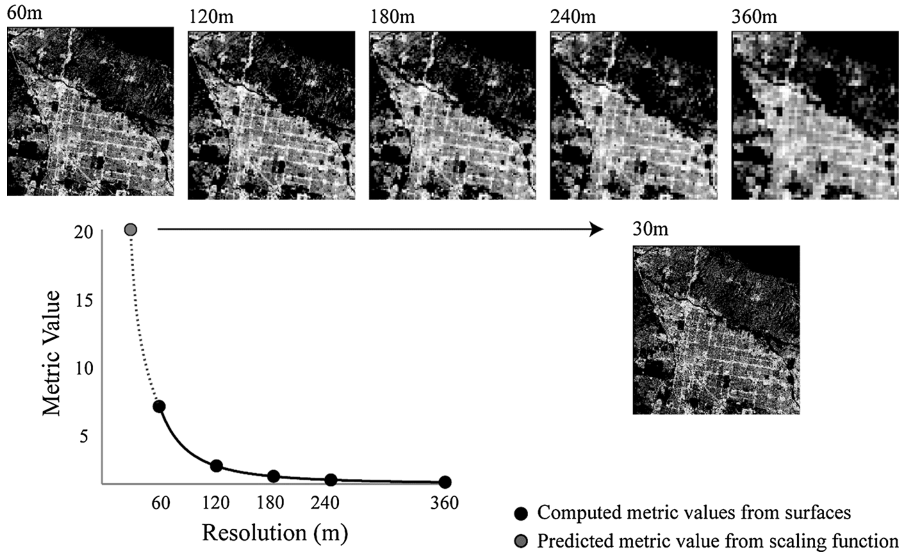

Fig. 2 Example of scaling function fit to surface metrics computed for Tucson, AZ (ISA). Scaling function is extrapolated to predict metric value at finer resolution (30 m)

A typology similar to that introduced by Wu (2004) for patch-based metrics was developed here for surface metrics. In Wu's (2004) typology, Type I metrics exhibit consistent scaling relationships, Type II metrics exhibit staircase responses, and Type III metrics behave erratically. Type I is further broken down into Type I_A and Type I_B for landscapes exhibiting robust similarity of scaling relationships between different patch types within the same landscape. Since patches do not exist in surface landscapes, this conception of robust is not appropriate. Instead, robustness for surface metrics refers to the ability of the scaling function to predict metric values at a finer resolution.

For surface metrics, Type I is assigned to metrics that exhibit consistent scaling functions ($\mu R^2 > 0.99$). Type I_A metrics are also robust in terms of the ability for a scaling function to correctly predict the metric value at the finer resolution to within 5 % μE_{rel}. Type I_B metrics exhibit a consistent scaling function but are unable to be accurately downscaled ($\mu E_{rel} > 5$ %). Type II metrics exhibit less consistent scaling functions ($\mu R^2 < 0.99$) that are less robust for downscaling ($\mu E_{rel} > 5$ %) but are still noteworthy. Type III metrics exhibit inconsistent scaling behaviors ($\mu R^2 \ll 0.99$).

Results

Average goodness-of-fit (μR^2) results for the four scaling functions fit to the calibration landscapes (Table 2) show that eight metrics were fit consistently well ($\mu R^2 > 0.99$) by at least one scaling function. Five of those metrics (*Sa*, *Sq*, *Ssk*, *S10z*, *Srwi*) were best fit with a 3rd order polynomial function, while the remaining three metrics (*Sdq*, *Sds*, *Sdr*) were best fit with a power law function. However, *Ssk* was only marginally consistent when considering its standard deviation. Furthermore, *Sku* can also be considered marginally consistent when considering its mean (0.9863) plus standard deviation (0.034). The 1st order polynomial (linear) scaling function did not fit any of the calibration surfaces well, with μR^2 values ranging from 0.4286 to 0.9595, and it will not be discussed further. The 2nd order polynomial fit several metrics fairly well (*Sa*, *Sq*, *Srwi*) with μR^2 values less than or just below 0.99, but it was always outperformed by the 3rd order polynomial.

The best-fitting scaling function for each metric was applied to the 15 validation surfaces to test downscaling. Results for the validation surfaces show eight metrics (*Sa*, *Sq*, *Ssk*, *Sku*, *S10z*, *Sci*, *Stdi*, *Sfd*) met the $\mu E_{rel} < 5$ threshold (Table 3). Again, *Ssk,* met the μE_{rel} threshold, but with the high variability exhibited by its standard deviation, it can be concluded that it is not a particularly strong candidate for downscaling. *Sku* was also marginal when considering its standard deviation along with the mean. The other six metrics had low μE_{rel} values even when considering their SD values. However, it should be noted that of the six metrics that were found to be robust for downscaling, only three were fit consistently with a scaling function

Table 2 Average goodness of fit (μR^2) values with standard deviations (SD) for the four scaling functions computed for the 15 calibration datasets

Metrics	Power Law μR^2 (SD)	1st Order Polynomial μR^2 (SD)	2nd Order Polynomial μR^2 (SD)	3rd Order Polynomial μR^2 (SD)
Sa	0.9754 (0.023)	0.9217 (0.049)	0.9898 (0.012)	**0.9983** (0.003)
Sq	0.9760 (0.020)	0.9286 (0.042)	0.9907 (0.042)	**0.9984** (0.003)
Ssk	0.8448 (0.183)	0.8657 (0.177)	0.9587 (0.062)	**0.9906** (0.018)
Sku	0.7795 (0.263)	0.8317 (0.224)	0.9742 (0.040)	**0.9863** (0.034)
S10z	0.7524 (0.213)	0.8999 (0.115)	0.9811 (0.013)	**0.9947** (0.004)
Sdq	**0.9977** (0.004)	0.6875 (0.097)	0.9067 (0.035)	0.9735 (0.001)
Sbi	0.7765 (0.212)	0.5814 (0.251)	0.8532 (0.142)	**0.9375** (0.132)
Sci	0.7867 (0.281)	0.7393 (0.355)	0.9019 (0.153)	**0.9619** (0.060)
Sds	**0.9972** (0.002)	0.5439 (0.049)	0.8383 (0.039)	0.9541 (0.017)
Sdr	**0.9986** (0.002)	0.5093 (0.061)	0.8051 (0.056)	0.9361 (0.028)
Stdi	0.6518 (0.310)	0.6147 (0.356)	0.8346 (0.143)	**0.8973** (0.129)
Srwi	0.9984 (0)	0.9892 (0.006)	0.9995 (0)	**0.9996** (0.057)
Sfd	0.7812 (0.334)	0.7007 (0.327)	0.9262 (0.113)	**0.9629** (0.064)
Std[a]	0.6215 (0.269)	0.7255 (0.272)	0.8401 (0.200)	**0.8861** (0.199)
Str20[a]	0.5060 (0.302)	0.5034 (0.290)	0.7176 (0.262)	**0.8698** (0.158)
Str37[a]	0.3697 (0.394)	0.3037 (0.278)	0.5767 (0.321)	**0.6928** (0.270)

Best-fit function for each metric is bolded

[a] Not all landscapes produced values for these metrics

Table 3 Average relative error (μE_{rel}) values with standard deviations (SD) for the 15 validation datasets using the best fit scaling function determined from the calibration datasets (Table 2)

Metrics	Best fit	μE_{rel} (SD)
Sa	3rd Order	**1.7** (2.35)
Sq	3rd Order	**1.37** (1.11)
Ssk	3rd Order	**2.93** (4.5)
Sku	3rd Order	**1.87** (2.73)
S10z	3rd Order	**1.71** (1.79)
Sdq	Power	5.13 (2.74)
Sbi	3rd Order	18.5 (15.3)
Sci	3rd Order	**1.40** (1.09)
Sds	Power	27.4 (13.9)
Sdr	Power	13.6 (6.22)
Stdi	3rd Order	**1.72** (1.42)
Srwi	3rd Order	17.8 (1.74)
Sfd	3rd Order	**0.46** (0.352)
Std(o)[a]	3rd Order	20.6 (36.3)
Str20[a]	3rd Order	8.72 (9.55)
Str37[a]	3rd Order	7.67 (4.85)

$\mu E_{rel} < 5$ are bolded

[a] Not all landscapes produced values for these metrics

as determined by the calibration landscapes (*Sa*, *Sq*, *S10z*). The typology for the metrics is included in Table 4.

Scalograms, or plots of grain size versus metric value, were created for Type I and Type II metrics (Fig. 3). Only data from the calibration surfaces are shown for clarity, however both the calibration and validation surfaces exhibited similar behavior. In particular, the scalograms for *S10z*, *Srwi,* and *Sku* (Fig. 3c, g, i) illustrate why there is a need for a higher order polynomial scaling function. *S10z* (Fig. 3c) exhibits an undulating patterns for many of the surfaces that would not be adequately captured by a power law or linear curve. Similarly, the plots for *Sku* (Fig. 3d) show varying directional trends. The scalograms for the Type I_B metrics *Sdq*, *Sds*, and *Sdr* clearly have consistent power law scaling relationships, but the fitted functions were not robust for prediction.

Discussion

The aims of this study were to (1) determine how surface metrics scale as resolution changes and how consistent those scaling relationships are across different landscapes, and (2) test whether fitted scaling relationships can be extrapolated to accurately predict metric values for finer resolution surfaces (i.e., downscaled). The major finding from the first objective was that many surface metrics do follow consistent scaling relationships as resolution changes, with

Table 4 Metric types determined from consistency ($\mu R^2 > 0.99$) and robustness ($\mu E_{rel} < 5$ %) threshold for downscaling

Type	Description	Metrics
Type I_A	Consistent, and robust	*Sa*, *Sq*, *S10z*
Type I_B	Consistent, not robust	*Sdq*, *Sds*, *Sdr*, *Srwi*
Type II	Less consistent, less robust	*Ssk*, *Sku*
Type III	Not consistent	*Sci*[a], *Stdi*[a], *Sfd*[a], *Sbi*, *Std*, *Str20*, *Str37*

[a] Metrics were not consistent but were robust for downscaling

ten metrics classified as Type I or Type II. All of the amplitude metrics (Table 1: *Sa*, *Sq*, *S10z*, *Ssk*, *Sku*, *Sdq*), which measure the vertical characteristics of the surface deviations, exhibited consistent scaling relations. Most of these metrics are non-spatial (except *Sdq*) and measure landscape composition through variations in surface height and height distribution statistics. According to McGarigal et al. (2009), *Sa*, *Sq*, and *S10z* are analogous to diversity measures in the patch-mosaic paradigm (e.g., Shannon's Diversity Index), and *Ssk* and *Sku*, while having no strong analogs, show some correlations with Simpson's Evenness Index and Largest Patch Index. It is not surprising then that these amplitude metrics scale consistently across resolutions because they measure simple variations in surface height that will naturally scale with aggregation (see Appendix-Electronic Supplementary Material), and several analogous patch-based composition measures have also been shown to exhibit consistent scaling relations (Wu 2004).

Several spatial metrics (*Sds*, *Sdr*, *Srwi*), which take into account the spatial distribution of the vertical profile, also were fit consistently. *Sdr* is analogous to patch-based edge density metrics (McGarigal et al. 2009), which scale consistently (Wu 2004), so this finding is expected. Lastly, the surface bearing metrics (*Sbi*, *Sci*), which are based on the Abbott curve of the cumulative height distribution, did not exhibit consistent scaling relations. In general, the findings from the first objective are important because they (i) confirm that many landscape composition metrics scale consistently across resolutions for gradient surfaces, and (ii) indicate that many of the same arguments developed for landscape ecology based on scaling knowledge of patch-based metrics may also apply when gradient conceptions of the landscape are adopted.

For the second objective, this study found that, in general, scaling functions fit to amplitude metrics can be downscaled accurately to predict the metric value for a finer resolution. Since most amplitude metrics do not consider spatial heterogeneity (McGarigal et al. 2009), they are not affected by the difficulties associated with capturing heterogeneity across scales (see Frazier 2015). This likely explains why these metrics showed good downscaling results (*Sdq*, the one amplitude metric that does consider spatial heterogeneity, was marginally less robust). The difficulties associated with modeling heterogeneity across scales may also explain why the three spatial metrics (*Sds*, *Sdr*, *Srwi*) did not have good downscaling success even though they were fit consistently with a scaling function.

Interestingly, three metrics (*Sci*, *Stdi, Sfd*) showed robust downscaling despite inconsistent scaling function fits (Table 4). Both *Sci* and *Stdi* are constrained in terms of their range of possible values (see Appendix—Electronic Supplementary Material), which may have contributed to better downscaling accuracy despite poor fits, although no limits were specified for prediction values. The metric *Sfd* is also interesting because even the poorest fitting function (i.e., linear with an μR^2 value of 0.7347) had exceptional downscaling accuracies (e.g., $\mu E_{rel} = 0.72$). Since *Sfd* represents the fractal dimension of the surface, it is possible that the inherent scaling nature of the metric played a role in the high prediction success. Additionally, the metric transforms the data to log–log curves, which may have contributed to high downscaling accuracies, particularly for the linear function.

Implications for scaling

The 3rd order polynomial was found to be the best fit for many metrics, but with higher order functions,

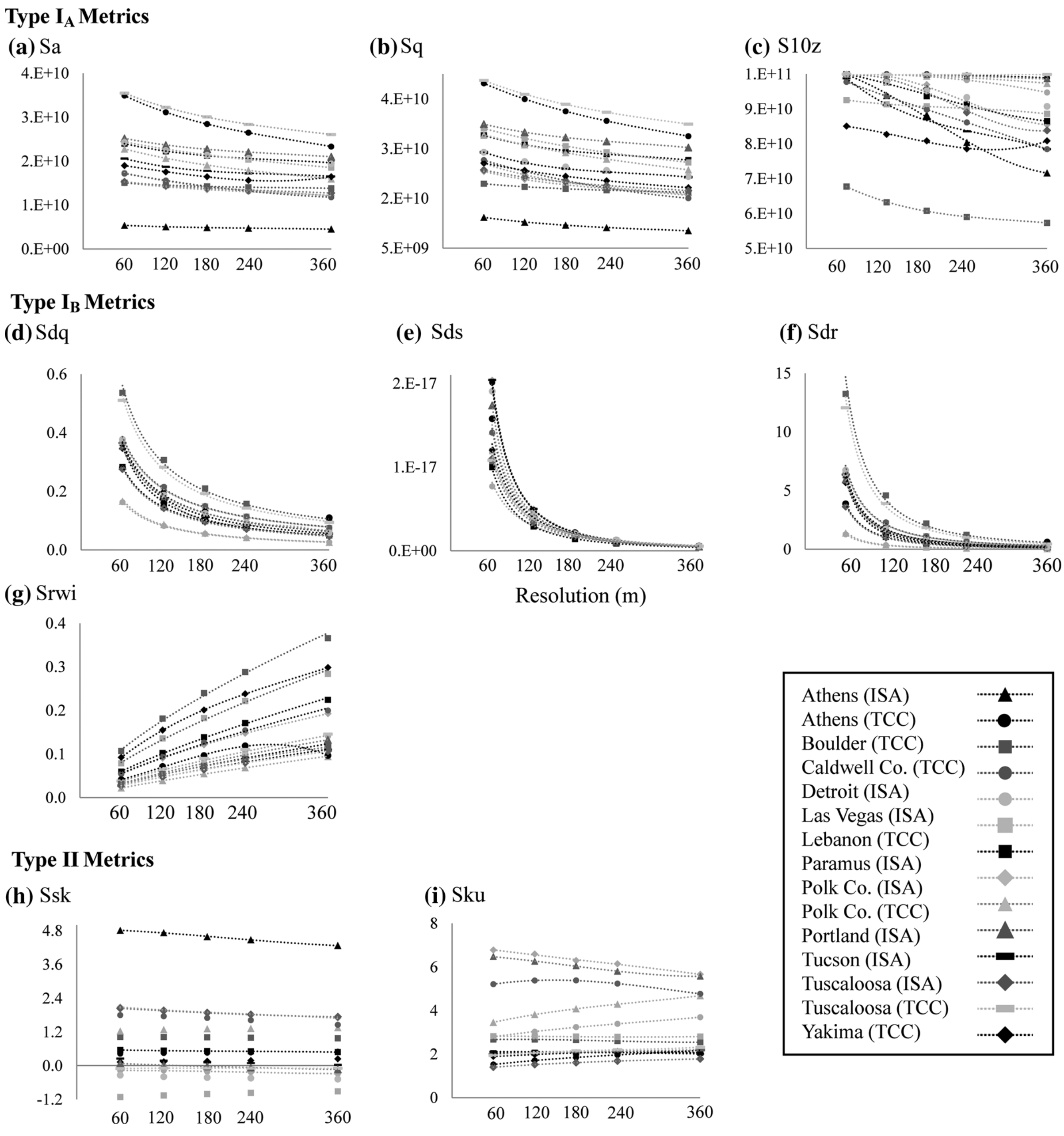

Fig. 3 Scalograms for Type I and Type II metrics using the 15 calibration surfaces

there is the risk of over-fitting curves, especially when the model contains too many parameters relative to the number of data points. For the 3rd order polynomial, six observations were modeled with four terms, which may have led to over-fitting in some cases. One way to test for over-fitting is to determine if the higher order models are less accurate for downscaling. Results using this technique suggest that two of the 16 metrics were over-fit. *Srwi* can be fit with either a 2nd order polynomial or a power law within the accuracy threshold (Table 2), and downscaling for the 2nd degree polynomial outperformed the 3rd degree polynomial indicating it was indeed over-fit, although neither results were within the 5 % error threshold.

The metric *Sq* can also be fit with a 2nd order model since the μR^2 for that function also satisfied the 0.99 threshold. Downscaling, while slightly less accurate than the 3rd order polynomial, did satisfy the 5 % threshold with a value of 2.52 % (SD = 1.7) indicating *Sq* can be fit with a 2nd order polynomial and still achieve acceptable modeling and downscaling results. For the remaining metrics, satisfactory R^2 values were occasionally obtained for individual landscapes using the 2nd order polynomial, but it was very rare for the 2nd order function to outperform the 3rd order function for downscaling. This implies that the 3rd order function is appropriate for scaling many surface metrics.

When fitting models to data points, there is also a tradeoff between the complexity of the model and it's goodness-of-fit, and the best-fitting model may not always be the best choice. Measures such Akaike's Information Criterion (AIC) (Akaike 1974) provide an estimate of this tradeoff and can be used to select the optimal model, but the nature of scalogram creation presents an interesting challenge. For example, measures such as 'leave one out' cross-validation (akin to AIC for linear regressions) or bootstrapping fit the model to various sets of points by iteratively leaving out one or more samples. These measures are problematic for scalogram curve fitting because the number of resolutions and scaling factors (i.e., resolution increment for aggregation) are selected at the discretion of the analyst and can be altered at any time. A base surface can produce an unlimited number of aggregations using non-integer scaling factors, so relative fit measures may easily be manipulated by overinflating the number of scalogram points.

A more useful measure for scalograms would be to determine an optimal set of resolutions and scaling factors for each scaling function. For instance, Šímová and Gdulová (2012) found that the shape of the scaling function for patch-based metrics is dependent on the range of pixel resolutions as well as the increment between each resolution. Their findings suggest that the exact data points chosen for the scalogram will influence the best-fit model, and thus downscaling accuracy. Since only a single set of resolutions with fixed increments was tested in this research, it is not clear whether the shape of the scaling function would change if the range of resolutions was altered or the increments between resolutions were changed. The use of only a few data points is supported by Wu et al. (2002), who found that extrapolation and interpolation of Type I patch-based metrics can be done accurately based on a few data points. However, given the intricacies of model selection for scalograms, determining the optimal resolution ranges and scaling factors for curve fitting would be more productive than comparing measures of relative quality for user-defined numbers of scalogram points.

Perhaps of larger interest is understanding why many surface metrics behave in a complex manner across resolutions. From the scalograms (Fig. 3), it can be observed that there is often one resolution (usually 360 m) where the metric value is "out of sync" with the rest of the values. This is most apparent for Athens (TCC) for *Srwi* (Fig. 3h) where the 360 m value actually decreases despite the increasing trend as resolution increases. The presence of these anomalies illustrate why the 3rd order polynomial outperforms other models, and understanding why surface metrics exhibit these characteristics is vital if we wish to use their scaling relations for future applications.

The concept of minimum mapping unit (MMU) offers one explanation for this behavior. In the patch-mosaic paradigm, MMU is the minimum size a land cover patch must be in order to be detected (Castilla et al. 2009). MMU will vary based on the different types of patches and land covers within a scene, but it has been suggested that map resolution must be 2–5 times smaller than a patch in order for it to be detected (O'Neill et al. 1996). Similarly, Saura (2002) found that sparse and fragmented land cover classes can be misrepresented when mapping resolution is increased and the classes occupying a large portion of the map begin to dominate. Surface metrics are likely experiencing similar biases when certain intensity values dominate the landscape. For example, the Big Bend, TX (TCC) landscape was the only validation landscape that did not meet the 5 % E_{rel} downscaling threshold for the metric *Sa*. Its E_{rel} value was 9.5, whereas all other landscapes were below 3.0. The Big Bend site is dominated by very low intensity values (Fig. 1q), and at 30 m resolution, the highest intensity value is only 83 whereas most other 30 m surfaces comprise pixels ranging from 0 to 100. At the 360 m aggregation, the highest pixel value is only 32 whereas the other surfaces have values in the 80 and 90 s indicating they lost a smaller range of values as aggregation proceeded. This observation suggests that the findings made by Saura (2002) are also manifesting for surface metrics whereby sparse and fragmented intensity values become misrepresented when more intense areas of land cover dominate aggregation. Moving forward, it may be possible to identify thresholds of data reduction during aggregation beyond which a scaling function is unable to predict landscape patterns for finer resolutions. This concept is similar to Wiens' (1989) scale domains concept discussed earlier, and investigating this threshold of information dilution (Karl and Maurer 2010) should be the focus of future explorations.

Lastly, the findings of this research have implications for patch-based metrics. Selection of 'best fit' models for patch-based metrics has traditionally been based on visual examination. However, recent studies have reported different 'best' scaling relationships when multiple types of scaling functions are analyzed and selection is instead based on R^2 values (Bar Massada et al. 2008; Argañaraz and Entraigas 2014). This study supports the use of R^2 values instead of visual examination for selecting scaling relationships as scalograms often appeared linear or power law but were indeed better fit with power law or 3rd order polynomials, respectively. This study also found that a 3rd order polynomial provided the best fit for many surface metrics, which has potential implications for patch-based metrics since studies of patch-based metric scaling have yet to investigate higher order polynomials. The results found here suggest it may be possible to improve downscaling results for patch-based metrics by using higher order scaling relationships in place of linear or power law functions, but further study is needed to investigate this potential.

Limitations

There are several limitations of this study that deserve mention. First, only a limited number of total surfaces (n = 30) were used to develop and test the scaling relationships, and more robust statistical generalizations would require a larger number of landscapes. Additionally, some landscapes contributed both TCC and ISA surfaces, so the number of independent landscapes was nominally less (see Fig. 1). A more robust analysis with a larger sample size is needed if the typology of surface metrics is to be formalized. Also, not every surface metric available through the SPIP software was tested in this study, so some metrics are not represented in the typology.

Second, the ground 'truth' surfaces are not completely independent from the coarse resolution surfaces since the 30 m resolution surfaces were originally used to derive the 60 m surface. However, the autonomy of the 'ground truth' surfaces is somewhat supported by the fact that many metrics still showed extremely high errors when predicting the 30 m values when utilizing scaling functions with high R^2 values. Additionally, while mean aggregation is the least biased statistical aggregation for creating coarser surface, the method selected by the user (e.g., mean, center pixel value, etc.) will impact results (Bian and Butler 1999), and it is possible that using different aggregation methods will lead to different results.

Lastly, only a single scaling factor was tested in this research. Scaling factor refers to the ratio between the finest known resolution (60 m) and the target resolution for downscaling (30 m). Therefore, the only scaling factor tested in this study was 2. In general, downscaling accuracy decreases as scaling factor increases (Frazier forthcoming) since the distance between the known and unknown values increases. Most downscaling studies do not test multiple resolutions, but this is an area for future consideration.

Conclusions

This paper explored scaling functions for 16 surface metrics and investigated their ability to predict metric values for finer resolutions (i.e., downscaling). By plotting the metric values for a series of coarse resolutions and then selecting the best-fitting functions to test downscaling, this study found that it may be possible to accurately predict metric values at finer resolutions for several surface metrics. Key findings are as follows:

- Three surface metrics (*Sa*, *Sq*, *S10z*) were fit consistently well ($\mu R^2 > 0.99$) with 3rd order polynomials that were robust for predicting a surface metric value for a finer resolution to within 5 % of the true value ($\mu E_{rel} < 5$ %). These metrics can be considered consistent and robust and have been designated as Type I_A.
- Four metrics (*Sdq, Sds*, *Sdr*, *Srwi*) were fit consistently well with a 3rd order polynomial or a power law, but their scaling functions were not able to accurately predict the surface metric value

for a finer resolution. These metrics can be considered consistent but not robust and have been designated as Type I_B.

- Two metrics (*Ssk*, *Sku*) were less consistently well fit and achieved less robust downscaling success, but they remain notable and have been designated as Type II.
- Seven metrics (*Sbi*, *Sci*, *Std*, *Sfd*, *Stdi*, *Str20*, *Str37*) did not exhibit consistent scaling relationships and have been designated as Type III despite robust downscaling success in some cases.

Additionally, results indicate that surface metrics and their scaling functions may be able to better predict values at finer resolutions (i.e., downscaling) than patch-based metrics, and the ability for a scaling function to predict the surface metric value at a finer resolution through downscaling may be related to information loss during aggregation.

Acknowledgments This research was supported by the National Science Foundation Geography and Spatial Sciences Program under award number 1303086. The author would also like to thank the anonymous reviewers for their comments, which helped improve the manuscript.

References

Akaike H (1974) A new look at the statistical model identification. IEEE T Automat Control 19(6):716–723

Alhamad MN, Alrababah MA, Feagin RA, Gharaibeh A (2011) Mediterranean drylands: the effect of grain size and domain of scale on landscape metrics. Ecol Indic 11:611–621

Argañaraz JP, Entraigas I (2014) Scaling functions evaluation for estimation of landscape metrics at higher resolutions. Ecol Inform 22:1–12

Atkinson PM (2012) Downscaling in remote sensing. Int J Appl Earth Obs Geoinf 22:106–114

Bar Massada A, Gabay O, Perevolotsky A, Carmel Y (2008) Quantifying the effect of grazing and shrub-clearing on small scale spatial pattern of vegetation. Landscape Ecol 23:327–339

Benson BJ, MacKenzie MD (1995) Effects of sensor spatial resolution on landscape structure parameters. Landscape Ecol 10(2):113–120

Bian L, Butler R (1999) Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data. Photogramm Eng Remote Sens 65(1):73–84

Cain DH, Riitters K, Orvis K (1997) A multi-scale analysis of landscape statistics. Landscape Ecol 12(4):199–212

Castilla G, Larkin K, Linke J, Hay GJ (2009) The impact of thematic resolution on the patch-mosaic model of natural landscapes. Landscape Ecol 24:15–23

Coulston JW, Moisen GG, Wilson BT, Finco MV, Cohen WB, Brewer CK (2012) Modeling percent tree canopy cover: a pilot study. Photogramm Eng Remote Sens 78(7):715–727

Coulston JW, Jacobs DM, King CR, Elmore IC (2013) The influence of multi-season imagery on models of canopy cover: a case study. Photogramm Eng Remote Sens 79(5):469–477

Fischer J, Lindenmayer DB (2006) Beyond fragmentation: the continuum model for fauna research and conservation in human-modified landscapes. Oikos 112:473–480

Frazier AE (2014) A new data aggregation technique for improving landscape metric downscaling. Landscape Ecol 29:1261–1276

Frazier AE (2015) Landscape heterogeneity and scale considerations for super-resolution mapping. Int J Remote Sens 36(9):2395–2408

Frazier AE, Wang L (2011) Evaluation of soft classifications for characterizing spatial patterns of invasive species. Remote Sens Environ 115:1997–2007

Frazier AE, Wang L (2013) Modeling landscape structure response across a gradient of land cover intensity. Landscape Ecol 28(2):233–246

Garcia-Gigorro S, Saura S (2005) Forest fragmentation estimated from remotely sensed data: is comparison across scales possible? Forest Sci 51(1):51–63

Gardner RH, Lookingbill TR, Townsend PA, Ferrari J (2008) A new approach for rescaling land cover data. Landscape Ecol 23:513–526

Jelinski DE, Wu J (1996) The modifiable areal unit problem and implications for landscape ecology. Landscape Ecol 11(3):129–140

Jin S, Yang L, Danielson P, Homer C, Fry J, Xian G (2013) A comprehensive change detection method for updating the National Land Cover Database to circa 2011. Remote Sens Environ 132:159–175

Karl JW, Maurer BA (2010) Multivariate correlations between imagery and field measurements across scales: comparing pixel aggregation and image segmentation. Landscape Ecol 25:591–605

Levin SA (1992) The problem of pattern and scale in ecology. Ecology 73:1943–1967

Li X, Du Y, Ling F, Wu S, Feng Q (2011) Using a sub-pixel mapping model to improve the accuracy of landscape pattern indices. Ecol Indic 11:1160–1170

Manning AD, Lindenmayer DB, Nix HA (2004) Continua and umwelt: novel perspectives on viewing landscapes. Oikos 104:621–628

McGarigal K, Cushman SA (2005) The gradient concept of landscape structure. In: Wiens J, Moss M (eds) Issues and perspectives in landscape ecology. Cambridge University Press, Cambridge, pp 112–119

McGarigal K, Tagil S, Cushman SA (2009) Surface metrics: an alternative to patch metrics for the quantification of landscape structure. Landscape Ecol 24:433–450

McIntyre S, Barrett GW (1992) Habitat variegation, an alternative to fragmentation. Conserv Biol 6:146–147

McIntyre S, Hobbs RJ (1999) A framework for conceptualizing human effects on landscapes and its relevance to management and research models. Conserv Biol 13:1282–1292

Milne BT (1991) Heterogeneity as a multiscale characteristic of landscapes. In: Kolasa J, Pickett TA (eds) Ecological heterogeneity. Springer-Verlag, New York, pp 69–84

Moniem HEMA, Holland JD (2013) Habitat connectivity for pollinator beetles using surface metrics. Landscape Ecol 28:1251–1267

Na Z, Li H (2013) Sensitivity and effectiveness of landscape metric scalograms in determining characteristics scale of a hierarchically structured landscape. Landscape Ecol 28(2):343–363

Neel MC, McGarigal K, Cushman SA (2004) Behavior of class-level landscape metrics across gradients of class aggregation and area. Landscape Ecol 19:435–455

O'Neill RV, Hunsaker CT, Timmins SP, Jackson BL, Jones KB, Riitters KH, Wickham JD (1996) Scale problems in reporting landscape patterns at the regional scale. Landscape Ecol 11(3):169–180

Riitters KH (2005) Downscaling indicators of forest habitat structure from national assessments. Ecol Indic 5(4):273–279

Saura S (2002) Effects of minimum mapping unit on land cover data spatial configuration and composition. Int J Remote Sens 23(22):4853–4880

Saura S (2004) Effects of remote sensor spatial resolution and data aggregation on selected fragmentation indices. Landscape Ecol 19:197–209

Saura S, Castro S (2007) Scaling functions for landscape pattern metrics derived from remotely sensed data: are their subpixel estimates really accurate? Int J Photogramm 62:201–216

Shen W, Jenerette D, Wu J, Gardner RH (2004) Evaluating empirical scaling relations of pattern metrics with simulated landscapes. Ecography 27:459–469

Šímová P, Gdulová K (2012) Landscape indices behavior: a review of scale effects. Appl Geogr 34:385–394

Turner MG, O'Neill RV, Gardner RH, Milne BT (1989) Effects of changing spatial scale on the analysis of landscape pattern. Landscape Ecol 3:153–162

Urban DL (2005) Modeling ecological processes across scales. Ecology 86(8):1996–2006

Uuemaa E, Roosaare J, Mander U (2005) Scale dependence of landscape metrics and their indicatory value for nutrient and organic matter losses from catchments. Ecol Indic 5:350–369

Wickham JD, Riitters KH (1995) Sensitivity of landscape metrics to pixel size. Int J Remote Sens 17(18):3585–3594

Wiens JA (1989) Spatial scaling in ecology. Funct Ecol 3:385–397

Wu J (2004) Effects of changing scale on landscape pattern analysis: scaling relations. Landscape Ecol 19:125–138

Wu J, Jelinski DE, Luck M, Tueller PT (2000) Multiscale analysis of landscape heterogeneity: scale variance and pattern metrics. Geogr Inf Sci 6:6–19

Wu J, Shen W, Sun W, Tueller PT (2002) Empirical patterns of the effects of changing scale on landscape metrics. Landscape Ecol 17:761–792

Xian G, Homer C, Dewitz J, Fry J, Hossain N, Wickham J (2011) The change of impervious surface area between 2001 and 2006 in the conterminous United States. Photogramm Eng Remote Sens 77(8):758–762