Remote Sensing of Environment 184 (2016) 615–627

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Spatial structure and scaling of agricultural networks

Daniel Sousa *, Christopher Small

Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form 19 June 2016
Accepted 31 July 2016
Available online xxxx

Keywords:
Scaling
Rank-size
Agriculture
Power law
Network

ABSTRACT

Considering agricultural landscapes as networks can provide information about spatial connectivity relevant for a wide range of applications including pollination, pest management, and ecology. At global scales, spatial networks of agricultural land use inferred from land cover products are well-described by power law rank-size distributions. However, regional analyses of agricultural land use typically focus on subsets of the total global network. In this paper, we seek to address the following questions: Does the globally observed scale-free property of agricultural networks hold over smaller spatial domains? Can similar properties be observed at kilometer to meter scales? Does the observed scale-free structure persist as agricultural networks evolve over the growing season? We analyze 9 intensively cultivated Landsat scenes on 5 continents with a wide range of vegetation distributions. We find that networks of vegetation fraction within the domain of each of these Landsat scenes exhibit substantial variability – but generally still possess similar scaling properties to the global distribution of agriculture. We also find similar results when comparing Landsat and Sentinel-2 imagery for 3 agricultural regions in Europe, as well as in an IKONOS image of an agricultural region of China. To illustrate an application of spatial network analysis, we show an example of network disruption. We compare two networks with similar rank-size distributions that behave differently when nodes are progressively removed. We suggest that treating agricultural land cover as spatial networks can provide a straightforward way of characterizing the connectivity and evolution of complex spatial distributions of agriculture across a wide range of landscapes and at spatial scales relevant for practical agricultural applications.

1. Introduction

The spatial distribution of agriculture in a landscape can provide information which is complementary to the properties of individually treated fields or political units. Pollination, insect diversity, and other ecosystem services are reliant on the spatial connectivity of an agricultural landscape (Diekötter et al., 2008; Ricketts et al., 2006). Outbreaks of pests and pathogens can sometimes be contained by breaking spatial adjacency between cropped areas (Gilligan, 2008). The ecology of native species populations can be altered by habitat fragmentation of natural landscapes by agriculture (Dixo et al., 2009; Luoto et al., 2003). However, the diversity of agricultural landscapes around the globe demands a tool which is flexible enough to accommodate a wide range of spatial distributions and connectivity patterns. Network theory provides the basis for a conceptually simple model which can represent a variety of processes with complex spatial structure.

Globally, maps of cropland extent have been observed to display an unexpected consistency in their size distributions (Small and Sousa, 2016). Despite considerable disagreement when compared directly in the same locations, 4 different global agriculture products possess the property that the sizes of contiguous patches of agricultural land diminish at the same rate that their frequency increases (Fig. 1). This property implies (nearly) uniform distributions of total agricultural area across a wide range of spatial scales. This implies that the sum of the area of the largest segments is equal to the sum of the area of the smallest segments, which is equal to the sum of the segments of any arbitrary size interval in between. The consistency of this observation across the 4 products is especially surprising given the substantial differences in the input data, assumptions, and algorithms used in each of the 4 products. The consistency of the observation at global scales begs the question of whether this pattern can also be observed at finer spatial scales.

The property of diminishing magnitude with increasing number is common in nature and is often referred to as a power law relationship. Power law relationships are also a defining characteristic of many networks – often referred to as "scale-free" because of the implied self-similarity and lack of a characteristic scale. Because networks are capable of representing processes with complex spatial structure (e.g. as reviewed by Barthélemy, 2011), and because many networks (such as those in Barabási and Albert, 1999) display similar power law relationships to those observed for agriculture on the global scale, we suggest that networks may be a useful tool to characterize agricultural landscapes and provide insight into processes reliant on agricultural connectivity.

* Corresponding author.
E-mail address: d.sousa@columbia.edu (D. Sousa).

http://dx.doi.org/10.1016/j.rse.2016.07.038
0034-4257/

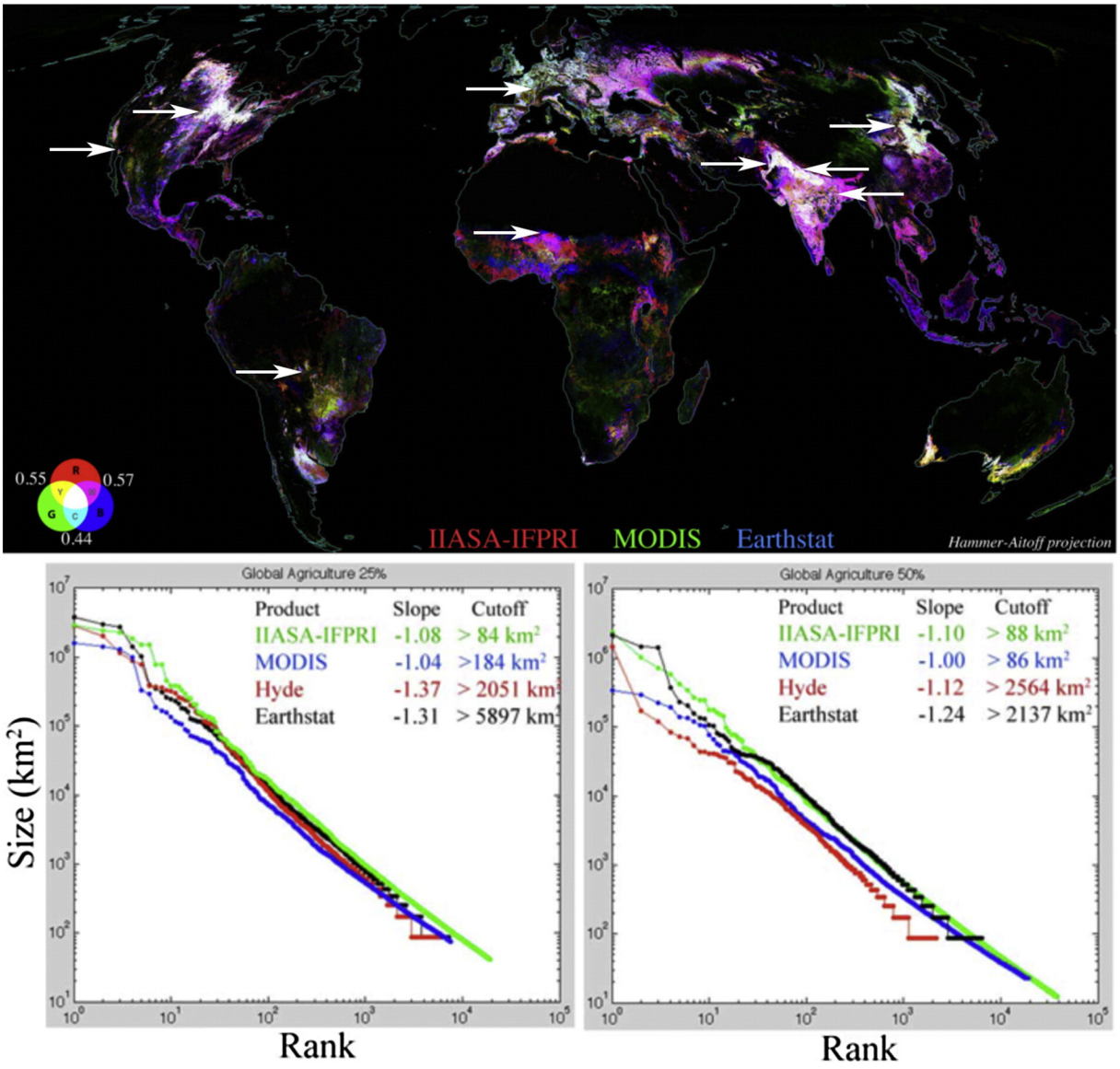

Fig. 1. Global comparison of agricultural land cover products (top) and corresponding rank size distributions of agricultural land area (bottom). Areal fractions of land under cultivation for three global products shown as red, green, and blue brightness. Pairwise spatial correlations between products shown in lower left corner of the map quantify agreement. Segmenting each continuous fraction map at >25% and >50% thresholds produces binary maps. Rank size distributions of contiguous areas of agricultural land cover for each product have similar slopes over 4 orders of magnitude in size. Power law fits to each rank size distribution yield slopes near −1. Size cutoffs estimate lower bound of power law behavior. Small cutoffs for the IASA-IFPRI and MODIS products indicate that the power law fits all but the smallest segments, while the larger cutoffs in the HYDE and Earthstat products result from quantization of smaller segments due to their coarser 10 × 10 km resolution. Slopes near −1 indicate that areas of agricultural land cover diminish in size at the same rate they increase in number. The implication is that the total area of agricultural land cover is nearly uniform across a wide range of segment sizes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
This figure is adapted from Small and Sousa (2016).

Several remote sensing studies have used power laws to describe fire size distributions. For instance, Hantson et al. (2015) studied the global distribution of fire sizes using a 2° grid and found that a power law model successfully fit 93% of grid cells with significant fire activity. Similarly, Malamud et al. (2005) studied wildfires across the conterminous United States and found robust power law fits in 18 different ecoregions. Kumar et al. (2011) use these established power law relationships to estimate fire biomass and radiative energy.

Studies of a wide range of other phenomena in the natural sciences also find power law behavior. Desert vegetation in the Kalahari was found by Scanlon et al. (2007) to follow a robust power law distribution across 1.5 orders of magnitude. Horizontal cloud sizes were found by Wood and Field (2011) to be well represented by power laws from sizes ranging from 0.1 km to at least 1500 km. Earthquakes (famously described by Gutenberg and Richter, 1956), wind profiles (as characterized by Hsu et al., 1994), and landslide area (characterized by Guzzetti et al., 2002) provide but a few of the myriad other cases in which power laws have been used to describe Earth processes. Interested readers may find more detailed descriptions of similar processes in references such as Turcotte (1997) and Sornette (2006).

The goal of this paper is to investigate the question of whether the globally observed scaling property of agricultural land cover holds over smaller areas and at spatial scales relevant to the questions of pollination, pathogen transmission, and ecology. Specifically, we seek to assess the robustness of the global scaling relationship at the decameter spatial scale for a set of diverse agricultural landscapes spanning 5 continents. To our knowledge, the investigation of heavy-tailed size distributions of contiguous patches of agriculture has not yet been performed in the literature. We ultimately seek to answer the question: Do the size distributions of agricultural landscapes already observed at

global scales maintain similarity to true power laws at spatial scales resolving individual fields? This question has direct relevance to several agricultural applications because of the implications for spatial connectivity of agricultural networks at scales where interventions are feasible.

2. Background

2.1. Rank-size plots and heavy-tailed distributions

Some processes in nature tend to produce objects or events that cluster around one characteristic size, with large deviations from this value being relatively infrequent. However, other processes produce objects or events that can take on a wide range of sizes – sometimes varying over several orders magnitude. When viewed as realizations of random variables, distributions which can take on a wide range of values are said to be *heavy tailed*. In a heavy tailed distribution, concepts from Gaussian statistics such as mean and standard deviation have little utility since the random variable deviates highly from that of a Gaussian (i.e., extreme events are much more common than predicted by a Gaussian distribution). Several types of heavy-tailed distribution which have been invoked by different authors to describe natural phenomena include the Weibull (e.g. wind speed, Seguro and Lambert, 2000), log-normal (e.g. distribution of chemical concentrations, Limpert et al., 2001) and power law (e.g. city size, Auerbach, 1913; Lotka, 1941; Zipf, 1942) distributions.

In the case of phenomena characterized by heavy tailed distributions, rank-size plots can be an intuitive tool for displaying both the magnitude and frequency of observations. Because such processes often span several orders of magnitude, such plots are typically displayed on logarithmic axes. Such a visualization scheme can be desirable because of its conceptual simplicity and minimum of assumptions about the form of the data. In the case where a rank-size plot is linear on logarithmic axes, the power law distribution is often considered a likely candidate for the underlying process. A power law distribution is defined by a constant factor and an exponent. If a set of features is distributed according to a power law, the slope of the rank-size plot in log-log space is related to the power law exponent α by the following expression from Li (2002):

$$\text{slope} = -\frac{1}{\alpha - 1}$$

Nonparametric statistical methods have been developed to determine the power law of best fit, the portion of the distribution most likely to be power law, and confidence intervals using Monte Carlo and the Kolmogorov-Smirnoff goodness-of-fit statistic. For an excellent description of these tools and their application to a wide range of datasets, see Clauset et al. (2009). When observations are binned logarithmically, a rank-size distribution with a slope of -1 (corresponding to a power law exponent of -2) corresponds to a uniform distribution across scales (Small et al., 2011).

Importantly, linearity of the rank-size plot alone does not rule out the possibility of other similar heavy tailed distributions describing the data equally well – or even better (Clauset et al., 2009). For this reason, in this paper we only use power law fitting as a convenient way to quantify the degree of linearity and slope of the rank-size plots. We remain noncommittal about the ultimate form of the underlying probability density function and suggest more rigorous analysis as a direction for future work on this topic.

2.2. Scale-free networks and constrained networks

The most basic pieces of networks are *nodes* and *links*. Nodes are connected to each other by links. Depending on the network, some nodes may be linked to many other nodes, some may be linked to only a few, and some nodes may not linked to any other nodes at all. Each set of interconnected nodes is called a *component*. Within each component, all nodes are connected to each other either directly or indirectly (i.e. through other nodes within the same component). No node within one component can be linked to a node within another component. A *network* is a set of components. In many networks, all nodes are linked to each other (directly or indirectly) to form a single "giant" component (Newman, 2010). Other networks have many components.

In a network, each node has a certain number of links. The frequency distribution of the number of links per node is called the degree distribution of the network. In some networks, the degree distribution can be well-characterized by a power law. These networks are called scale-free networks. For these networks, when the distribution of degree sizes versus rank (ordinal number: 1 = largest, 2 = second largest, 3 = third largest, …) is plotted on logarithmic axes, the result is linear. The slope of this line can vary substantially for different networks (Barabási and Albert, 1999; Clauset et al., 2009). The wide range of degrees necessary for a power law distribution is possible in some cases because many networks have no limit (or some very large limit) to the number of links that each node can have. Networks are already used in the field of landscape ecology under the term graph theory (Cantwell and Forman, 1993; Gardner et al., 1992; McIntyre et al., 2014; Urban and Keitt, 2001). For a general review of network theory, see Albert and Barabási (2002) and Newman (2010).

In this paper we treat landscapes as networks of land cover. The spatial domain of interest determines the total possible spatial extent of the network it contains. As the network grows within the rectangular grid of the domain, each pixel is treated as a potential node. In this paper, a pixel becomes a node of a spatial network if it satisfies a single criterion: subpixel vegetation abundance above the threshold of analysis. We consider two pixels to be directly linked if they are spatially adjacent to each other. For this reason, nodes in land cover networks on a regular rectangular grid (as is the case in this study) have a maximum number of direct links (Steinwendner, 2002). Because we use the Queen's case for connectivity (all immediate neighbors including diagonals), this number is 8. In this case, the parameter of interest is not the degree distribution but the component size distribution, as the sizes of each component (spatially contiguous patch of agricultural land) can possess a wide range of values. The rank-size plots used in this paper show the distribution of component sizes in a single network. We refer to the particular type of spatial network defined in this way as a *bounded spatial network* (Small and Sousa, 2015).

Small and Sousa (2016) show that four land cover products which seek to map agricultural land use at the global scale exhibit empirical component size distributions characterized by linearity in logarithmic space and slope of -1, despite differences in spatial patterns (Fig. 1, from Small and Sousa (2016)). This result holds across a wide range of analysis thresholds (described in more detail below). This suggests that agriculture may be well characterized as a scale-free spatial network on the global scale. Other types of land cover products have also been found to exhibit similar properties on the global scale (Small and Sousa, 2015).

Scale-free networks have been shown to result from two simple conditions: network growth and preferential attachment (Barabási and Albert, 1999). Preferential attachment is sometimes described as "rich get richer"– i.e., new nodes to attach more frequently to existing nodes with greater numbers of links, or to components with a greater number of nodes, than to their less connected counterparts. The networks we consider fill space on a surface. This generates a mechanism for preferential attachment because the surface has finite area and larger components naturally have larger perimeters to which new nodes can link. If new nodes are generated randomly in space, components with larger perimeters will exhibit preferential attachment – without the need for a situation-specific mechanism for preference. To the extent that components with larger sizes (i.e. areas) also have larger perimeters, a mechanism for preferential attachment is inherent to

bounded spatial networks on a surface. For more detailed background and mechanism, see Small and Sousa (2015).

3. Data & methods

To quantify the scaling properties of different agricultural landscapes, we choose images that are dominated by agricultural land cover and then use the following procedure. Beginning with raw Landsat data, we first calibrate from DN to radiance to exoatmospheric reflectance. We then estimate vegetation fraction (F_v) at each pixel using the standardized global endmembers from Small and Milesi (2013), generating a continuous field of sub-pixel vegetation abundance.

Sentinel data are processed by first resampling all 12 bands to 10 m resolution and then unmixing into Substrate, Vegetation, and Dark components. Subpixel vegetation fraction from this unmixing is then used for subsequent network analysis. Sentinel-2 spectral unmixing is performed using local SVD endmembers since global Sentinel-2 endmembers are not currently available. As noted previously (Small and Milesi, 2013), local substrate EMs can differ substantially from the global EMs and produce systematic differences in fraction estimates between global and local EMs. These differences are most prominent in substrate fraction estimates.

We then segment the F_v images at several different fraction thresholds with the ENVI segmentation algorithm, using the Queen's case of 8 neighbors including diagonals. The ENVI segmentation algorithm finds spatially contiguous groups of pixels which all obey the rule used for segmentation. In this case, the rule used is F_v above a given threshold. All spatially adjacent pixels with F_v above this threshold are labeled with the same segment number. We use the Queen's case as our metric for spatial adjacency in order to provide the most liberal estimate of connectivity. We use a minimum segment size of 8100 m^2 (9 Landsat pixels or 81 Sentinel-2 pixels) to account for spatial autocorrelation of the input imagery and avoid large numbers of spurious detections. Allowing smaller segments in Sentinel imagery has the effect of resolving the characteristic field size of the landscape and is discussed in (Small & Sousa, forthcoming). The segmentation algorithm produces a map of segments corresponding to spatially contiguous patches of vegetation (for each threshold). The agricultural network is thus composed of all pixels in the image for which the following two conditions are true: (1) F_v of the pixel is above the given threshold and (2) the pixel is spatially adjacent (directly or indirectly) to at least 8 other pixels (Landsat) or 80 other pixels (Sentinel) which also have F_v above the given threshold.

Next, we calculate the total area of each segment (for each threshold). The resulting segment size maps (for each threshold) provide both the size distributions and a depiction of the spatial network structure. Segment areas are then sorted into a descending list and plotted against ordinal number (i.e. rank) on logarithmic axes.

A wide range of thresholds was applied in each case and results were compared. Fig. 2 shows the typical progression of a rank-size distribution at full resolution for an example region in northern California. Images of the spatial structure of the network are shown for several different thresholds, with inset size distributions. Segment sizes are color coded on both the image and the rank-size plot. At a threshold of 100% subpixel vegetation abundance, all pixels fall below the threshold and there is no network. As the threshold is lowered, more pixels are included in the network and form components (contiguous patches). At this phase the components correspond to individual fields or groups of closely spaced fields with high F_v. Continuing to lower the threshold eventually results in connection of more and more components into larger contiguous areas as pixels with lower F_v are added to the network. Eventually enough pixels become part of the network that components begin connecting to form much larger components. Eventually the larger components superconnect and form one massive unit. If the threshold continues to be lowered to negligible F_v, the entire spatial domain of the image becomes part of the network. For more detail on the general methodology used to segment continuous fields, see Small and Sousa (2015) and Small et al. (2011).

Disruption of agricultural networks was performed by sequential erosion using a morphological operator. For each iteration of the analysis, segment area maps were converted into binary maps indicating

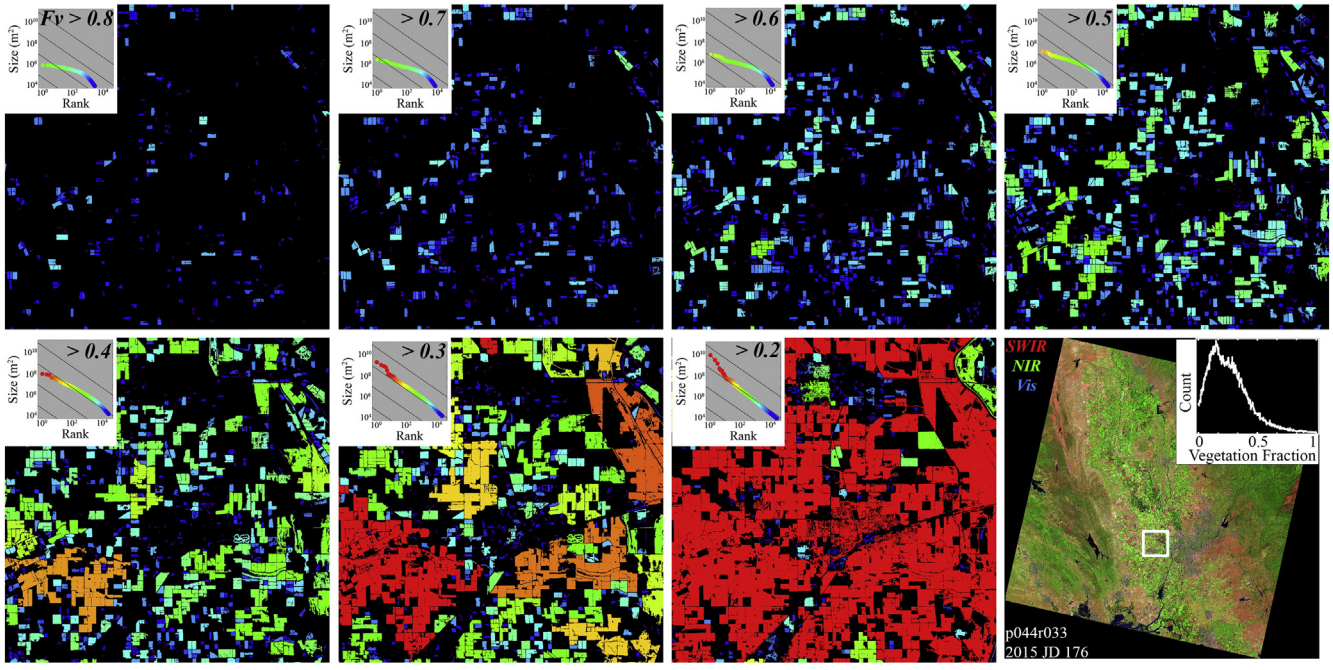

Fig. 2. Illustration of network progression with threshold. This 1000 × 1000 pixel subscene of Landsat vegetation fraction was thresholded at successive values (upper left of each image). Insets show size distributions for each threshold. At threshold of Fv > 0.8, only a few fields emerge as part of the network and the size distribution is nearlyflat. As the threshold decreases, more fields are included and adjacent fields connect. The size distribution steepens and loses its curvature. This continues until the size distribution becomes straight (near Fv > 0.3 in this case). After this point, further lowering of the threshold results in a majority of segments superconnecting into a small number of very large segments. This progression is typical of the other vegetation fraction images in this paper.

presence or absence of agriculture. These binary maps were then convolved with a 3 × 3 pixel Gaussian filter. Any pixel with a full set of 8 agricultural neighbors was unchanged, but any pixel with one or more non-agricultural neighbor decreased in value. A threshold of 1 was then applied and segment areas were recaluclated. This produced the effect of removing every pixel in the image on the boundary of the network. The output of one erosional step was then used as the input for the next step.

Power law exponents were fit using the statistically robust algorithm described by Clauset et al. (2009) and converted to slopes of the size distributions using the relation given in Eq. (1). Power law fits are also characterized by size cutoffs describing how far the power law properties plausibly extend down the lower tail of the distribution. Cutoffs were determined using the same algorithm by choosing the minimum of the Kolmogorov-Smirnoff (KS) statistic for sets of points extending sequentially farther into the lower tail of the distribution. Significance was estimated using a Monte Carlo approach to generate 1000 synthetic datasets and calculating the KS goodness-of-fit for each. Using this approach, large p values represent plausible power law fits. We use the suggestion of Clauset et al. (2009) in presenting significant power law fits as those with $p > 0.1$, which is a stricter test than accepting as plausible distributions with $p > 0.05$.

While significant p values indicate that a power law distribution cannot be ruled out, they do not decisively show it to be a better fit than other heavy tailed distributions such as log-normal. Comparison of power law versus log-normal and Weibull distributions was performed by computing the Likelihood Ratio for each fit to each distribution. In all 27 cases, the power law gave a better fit (i.e. yielded a higher Likelihood value) than either the log-normal or Weibull distributions. However, despite outperforming both of the other two heavy tailed distributions, we remain non-committal about the true form of the underlying distributions being a strict power law. More information would be required to substantiate this claim than is available at the present time, and rigorous statistical testing of the power law hypothesis could be a useful avenue for future work. In this paper, we simply use the power law as a convenient metric of linearity of the rank-size distributions in logarithmic space.

The data used in this study were (1) Landsat TM/ETM+/OLI scenes selected from diverse agricultural regions across 5 continents, (2) Sentinel-2 scenes selected from 3 agricultural regions in Europe, and (3) one

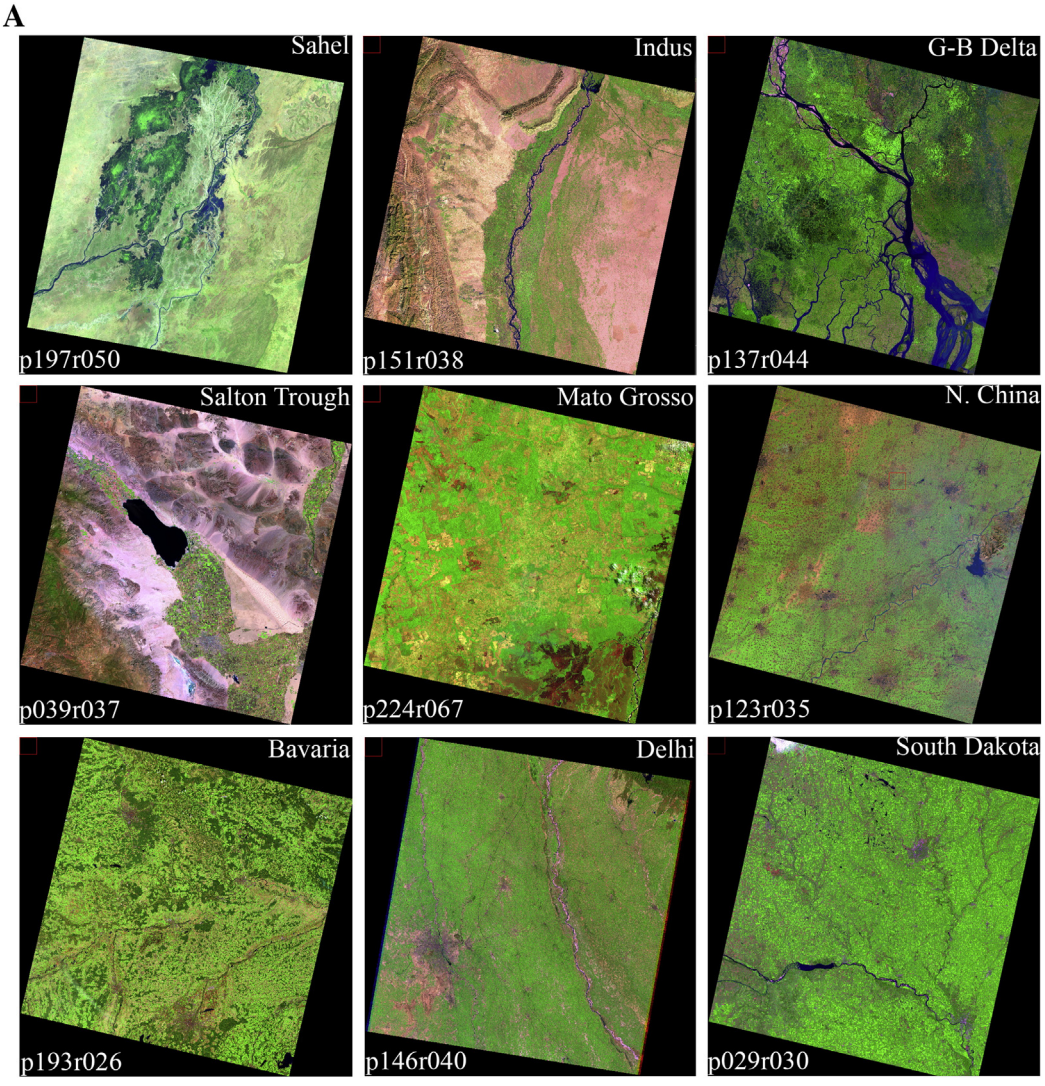

Fig. 3. A. Agricultural landscapes used for scaling analysis. Scenes were chosen to represent a diverse set of landscapes characterized by agricultural extensification and intensification. A range of field sizes, competing land uses, climate zones, and land management practices is depicted. B. Rank-Size distributions for vegetation fraction from the 9 Landsat scenes shown in panel a. Inset shows vegetation fraction histogram for each Landsat scene, with arrows indicating the segmentation thresholds. Rank-Size distributions for each scene illustrate the sensitivity of the network structure to threshold. Distributions of vegetation fraction are different for each landscape but most scenes have linear rank size distributions with slopes near −1 and giant components forming as thresholds approach the median vegetation fraction. C. Slope of Rank-Size distribution versus threshold for nine Landsat scenes. Local landscape properties vary from scene to scene, resulting in a wide range of vegetation fraction distributions. These distributions control the progression of slope of the size distributions.

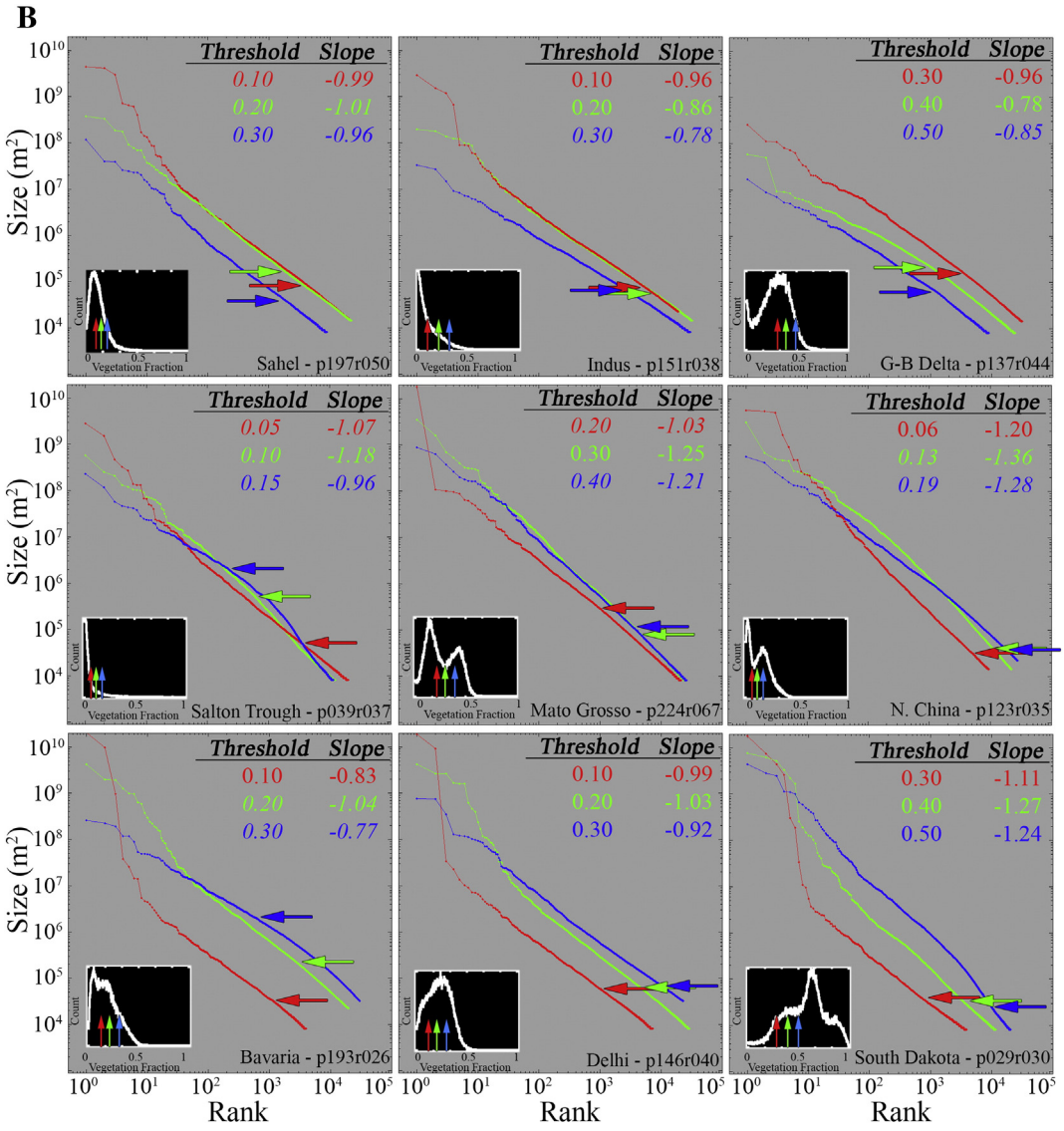

Fig. 3 (*continued*).

IKONOS scene of an intensively cultivated region in Anhui, China. All Landsat scenes were acquired from the USGS Earth Resources Observation and Science Center (www.glovis.usgs.gov). All Sentinel scenes were acquired from the ESA SciHub web portal (www.scihub.copernicus.eu). For Landsat and IKONOS data, we use UTM equal area projections at the native resolution of the sensor. For Sentinel-2 data, we resample all 12 bands to the 10 m resolution native to the Sentinel-2 VNIR sensor in UTM equal area projection. Landsat scenes in this analysis are referred to by their WRS-2 path and row identifiers: i.e. scene p029r030 corresponds to Path 29, Row 30 (South Dakota).

The scenes were chosen to represent a diverse set of landscapes dominated by extensive agriculture, spanning a range of field sizes, climate zones, phenologies, and land management practices. A wide range of crops are represented, including regions dominated by one or two grains (e.g. rice and/or wheat) as well as regions producing a balance of both commodity and specialty crops. The 9 Landsat scenes used in this study were selected quasi-randomly from the Landsat archive to meet the criteria of: lack of cloud cover, diversity of agricultural practices, and range of hydrologic and climatic milieu. They were selected quasi-randomly in time to represent a range of stages of the phenologic cycle. We do not claim that these 9 scenes fully sample the global distribution, but rather suggest that their results represent a diverse set of potential endmembers of the global distribution of agricultural landscapes.

Fig. 3A shows false color composites of the 9 Landsat scenes used for this analysis. Spatial configuration of agriculture across scenes varies widely from nearly wall-to-wall coverage (e.g. South Dakota and North China) to regions strongly limited in spatial extent by irrigation (e.g. Salton Trough and Indus). A range in extent of sectioning of the landscape by roads and rivers is apparent. Field size varies widely both across scenes and within scenes. Scenes were chosen at varying stages of the annual cycle, from soon after planting to maximum greenness. All scenes contain some non-agricultural vegetation ranging from tropical forest to desert shrubs – but all are dominated by agriculture. The Bavaria scene contains several forest patches, but all are managed forests so are effectively part of the agriculture/silviculture mosaic.

The spatial extent and abundance of non-agricultural vegetation varies from scene to scene. While the presence of some non-cultivated vegetation violates the assumption made in the analysis that networks of vegetation fraction strictly represent networks of agricultural activity, we have attempted to choose regions dominated by extensive cropland.

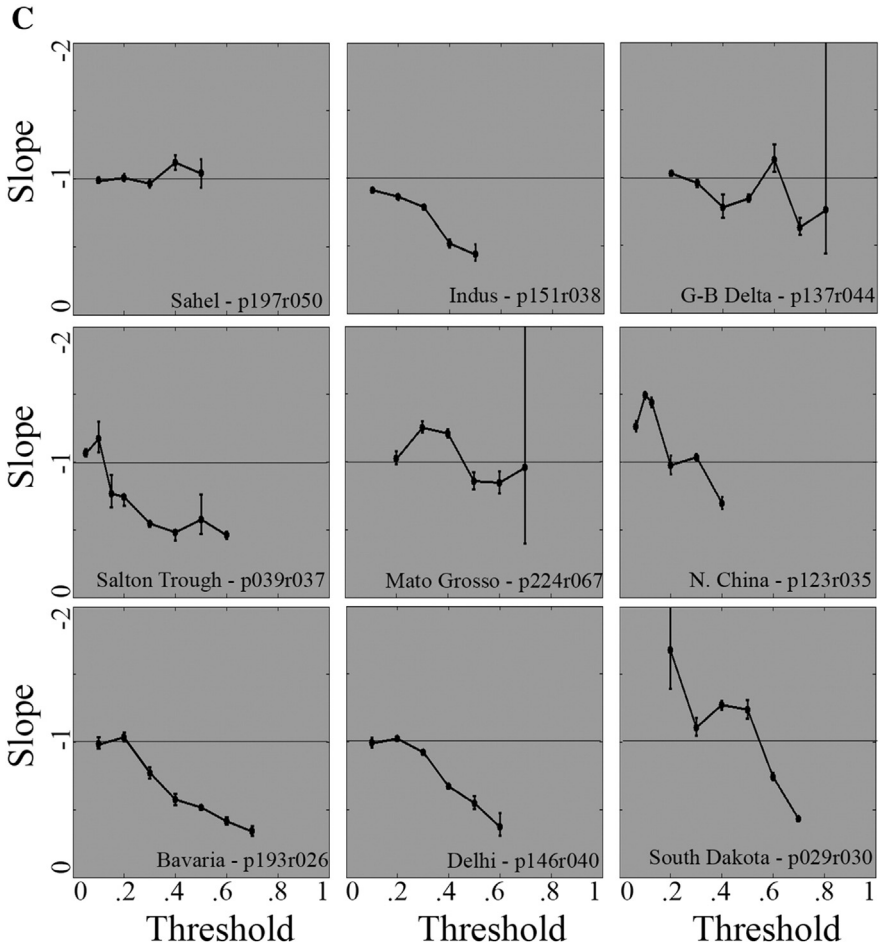

Fig. 3 (*continued*).

We also suggest that, for some applications such as species migration and pollination, vegetation networks may be closer to the phenomenon of interest than strict definitions of cropland. Further, while considerable uncertainty exists as to the definition of cropland in global agriculture maps (reviewed in Small and Sousa, 2016), subpixel vegetation abundance represents a physically meaningful quantity which can be directly compared across widely varying landscapes. While using F_v as a general proxy for agriculture would not be valid in many landscapes, we hold that its properties of simplicity and consistency justify its use in the examples chosen in the context of this analysis.

4. Analysis

4.1. Landsat

Fig. 3B shows Rank-Size plots for 3 different thresholds for each of the 9 Landsat scenes from Fig. 3A. F_v distribution for each scene is inset with the three thresholds indicated using vertical arrows. Histograms vary widely from scene to scene in central tendency, dispersion, and number of modes, reflecting differences between the landscapes described above. Thresholds are adjusted accordingly from scene to scene to capture similar positions in the distribution. Horizontal arrows on the rank-size plots indicate the cutoff for power law fit that maximized the goodness-of-fit criterion. Italicized thresholds and slopes have p values > 0.1, indicating a statistically plausible power law fit. The statistical significance of the fit is not critical for the purposes of this analysis because we use the power law exponent as a tool to quantify the slope of the rank-size plot, not as an assertion of the generating process itself. We include the goodness-of-fit result for the benefit of readers inclined to favor the power law mechanism.

Fig. 3C shows Rank-Size slope estimates for several thresholds for each of the 9 Landsat scenes. Error bars indicate 95% confidence. As the threshold is successively lowered, rank-size slopes generally increase toward more negative values. This corresponds to an increase in overall network size and in the size of individual components, consistent with the network growth mechanism proposed in Small and Sousa (2015). Prominent exceptions to this rule correspond to cases of severe non-Gaussianity of the vegetation histogram, e.g. bimodality in the Mato Grosso and North China Landsat scenes and a broad, asymmetric shoulder in the South Dakota scene. Slopes near −1 indicate that segments decrease in size at roughly the same rate that they increase in frequency. Slopes pass through a value of −1 for 8 of the 9 scenes considered here. The two scenes with slopes consistently shallower than −1, Indus and Salton Trough, are characterized by exponential-like F_v histograms with a mode of $F_v \approx 0$.

4.2. Sentinel-2

Fig. 4 shows network structure intercomparisons between Landsat and Sentinel-2 for three 30 km × 30 km agricultural landscapes in Europe. In Fig. 4A, we examine an agricultural basin in Abbruzzo, Italy imaged on successive days in December 2015. The segment size image of the agricultural network (middle row) reveals a range of spatial

patterns of contiguous photosynthetic agriculture, from isolated small fields to clusters of closely spaced fields with separators which are not resolved by either 30 m Landsat or 10 m Sentinel-2 sensors. However, close visual inspection reveals several cases where fields grouped together in Landsat imagery are broken apart in Sentinel imagery – and vice versa. This is possible because increasing spatial resolution can have (at least) two processes working in opposite directions: ability to resolve narrow connectors which do not emerge above threshold in coarse resolution imagery (enhancing connectivity) and ability to resolve narrow separators which are presented in coarse resolution imagery as a mixed pixel above threshold (reducing connectivity). Which of these processes dominates varies based on the local geometry of the segment at play.

In Fig. 4B, we examine a subset of the agriculture/silviculture mosaic in Bavaria used in the global analysis of Fig. 3. In this case, we examine the agricultural network over a range of nearly 25 years and 17 days offset in the phenological cycle. As a result, the overall greenness of the landscape is notably different, although the spatial arrangement of fields is

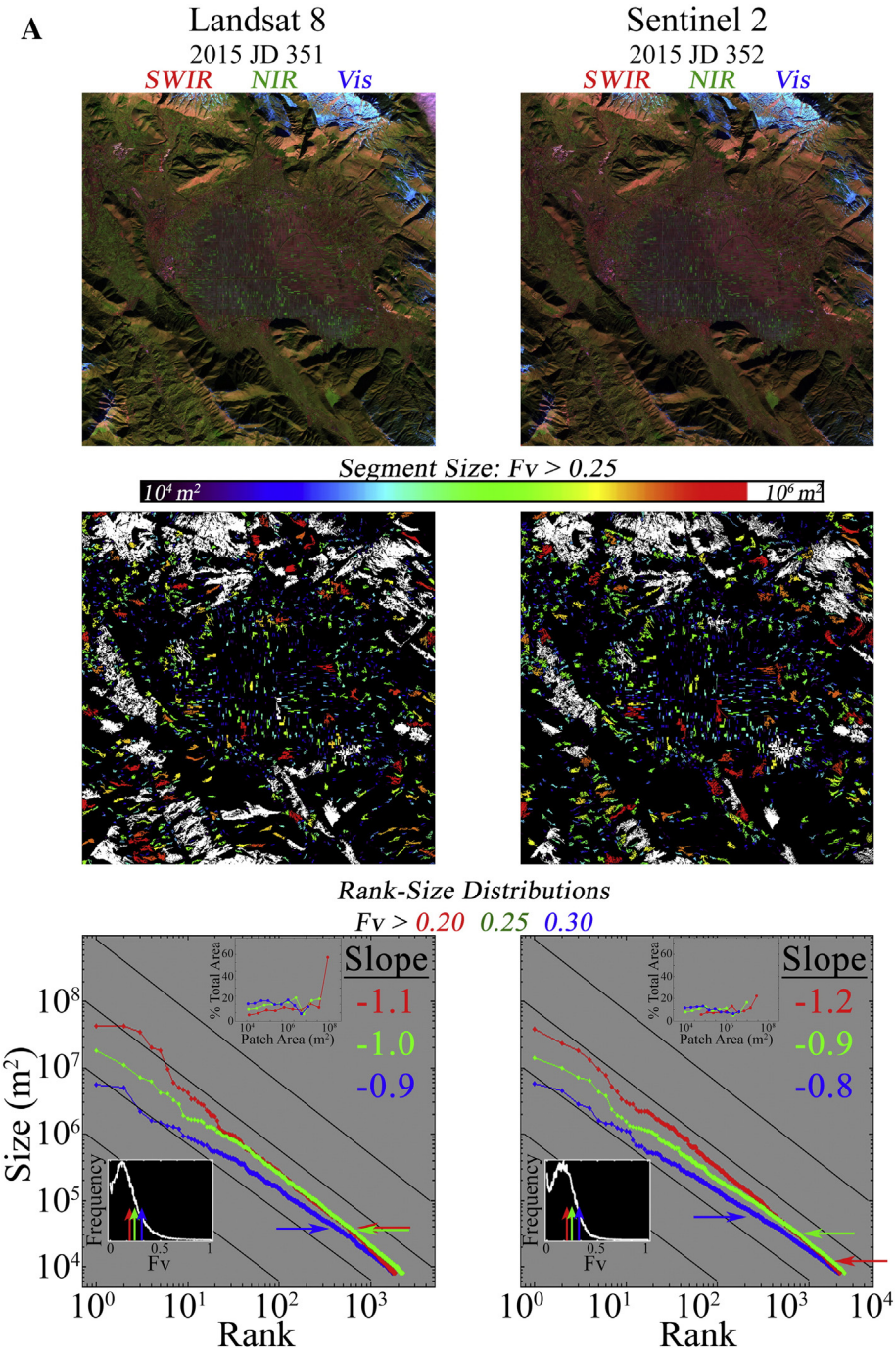

Fig. 4. A. Comparison of Landsat- and Sentinel-derived networks from a 30 km × 30 km agricultural region in Abbruzzo, Italy. 1:1 lines shown on the rank-size curves in black. Inset histograms show vegetation fraction (lower left) and total area by segment size (upper right) distributions. B. Comparison of Landsat- and Sentinel-derived networks from a 30 km × 30 km agricultural region in Bavaria, Germany. 1:1 lines shown on the rank-size curves in black. Inset histograms show vegetation fraction (lower left) and total area by segment size (upper right) distributions. C. Comparison of Landsat- and Sentinel-derived networks for a 30 km × 30 km agricultural region in Centre-Val de Loire, France. 1:1 lines shown on the rank-size curves in black. Inset histograms show vegetation fraction (lower left) and total area by segment size (upper right) distributions.

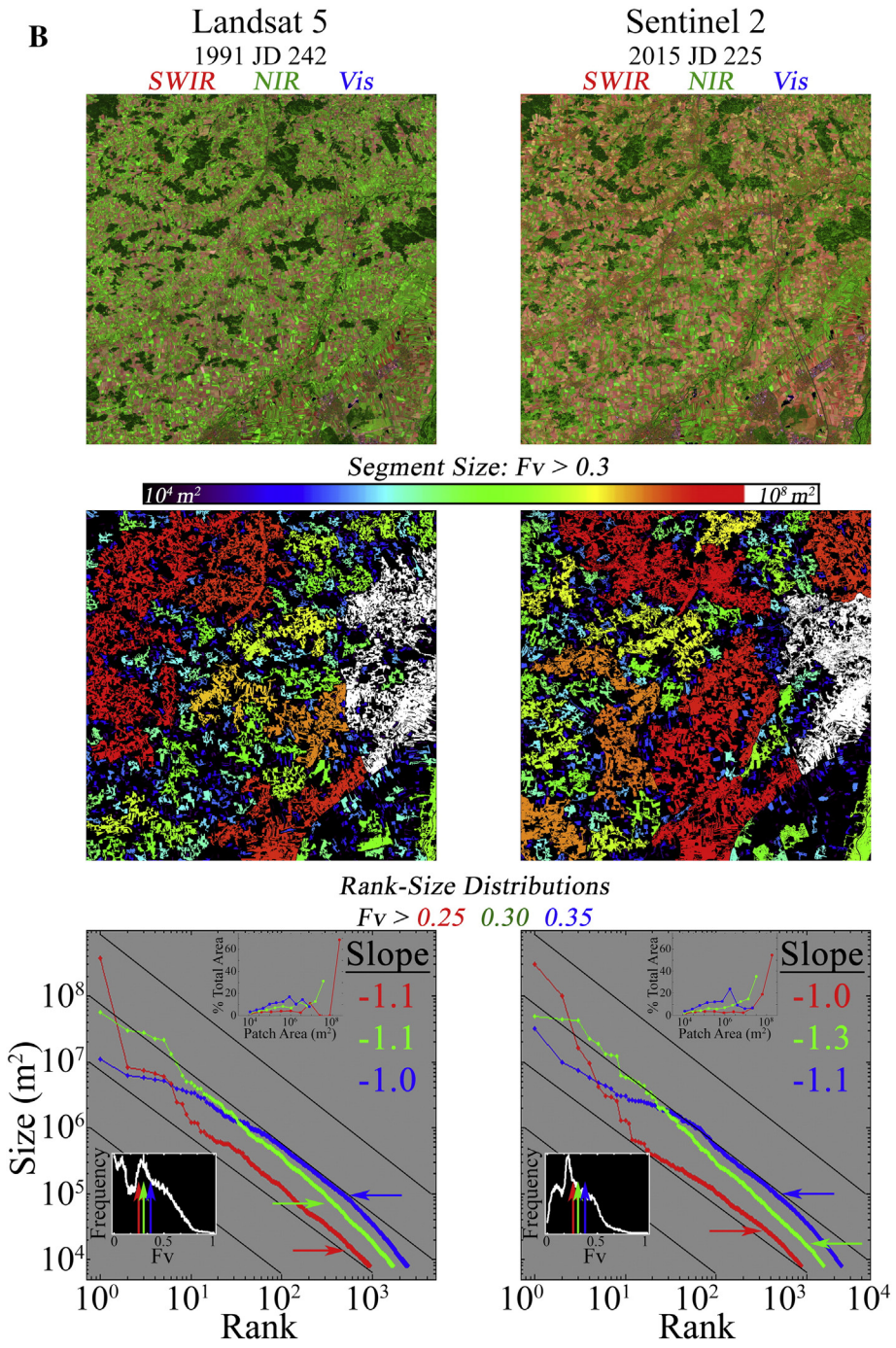

Fig. 4 (*continued*).

generally stable. The segment size images reveal the breakup and connection of segments as a result of interannual, phenological, and resolution-based differences in the images. In spite of these differences, the rank-size distributions of segment areas remain remarkably consistent.

Fig. 4C shows a region in Centre-Val de Loire, France 1 year and 14 days apart. In both cases, the agricultural network is dominated by large segments, identifiable both by visual inspection of the segment area images and by rank-size slopes steeper than −1. Loci of closely spaced fields which dominate the landscape appear to be generally stable in their position, but the connectivity between them varies. While the spatial positions of the largest components shift due to this variation in connectivity, the rank-size plots again remain remarkably stable.

4.3. IKONOS

Fig. 5 shows the procedure of successive thresholding when repeated for a ~39 km^2 IKONOS image in Anhui, China. The 4-band image was unmixed into SVD fractions using local endmembers. Successive thresholding was then applied to the F_v image. Segment area maps for four representative thresholds demonstrating the progression of the network are shown in the top 4 panels. The progression of the IKONOS size distributions with changing threshold (bottom right panel) is

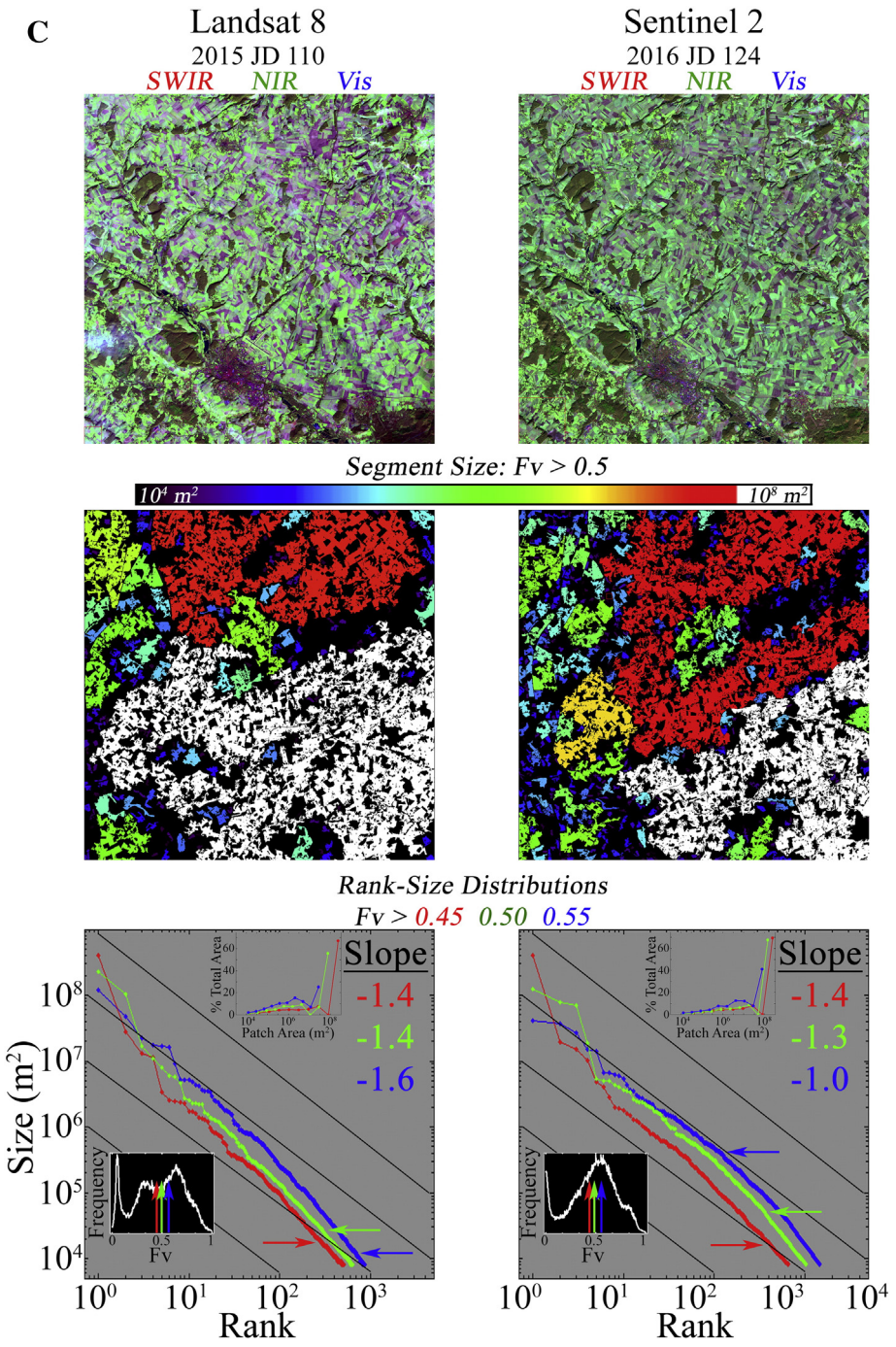

Fig. 4 (*continued*).

similar to that of the Landsat scene shown in Fig. 2. At high thresholds IKONOS size distributions have high curvature and shallow slopes. The slope of the size distribution steepens as the threshold is reduced and the lower-tail power law cutoff gradually moves up the distribution. Curvature is even more pronounced than for Landsat at this phase. Once a threshold near 0.3 is reached, however, the size distribution loses most of its curvature and becomes linear. The slope of the size distribution crosses −1 at this point and the lower-tail cutoff rapidly moves deep into the lower tail of the distribution. As the threshold is decreased below this level, the network superconnects into a few giant components. The total number of segments (i.e. maximum rank) begins to decrease and the bottom of size distribution moves to the left. These properties are all similar to those observed for the Landsat and Sentinel scenes in Figs. 3 and 4.

4.4. Practical example – disruption by node removal

Fig. 6 shows how two agricultural networks can respond differently to disruption by sequentially reducing the area of each component. In each iteration of this process, all segments in the image are simultaneously reduced in size by removing one pixel width from around the boundary. We refer to this type of disruption as “erosion”. This process has the potential to remove segments from the network by shrinking them below the 9 pixel threshold. It also has the potential to break a

small number of large segments into a large number of smaller segments (i.e. making little pieces out of big pieces) by separating dense intra-segment clusters which are only connected by narrow "bridges".

We disrupt two agricultural networks in this way: one in the Salton Trough (p039r037) and one in South Dakota (p029r030). The upper tails of the rank-size distributions are shown in detail for successive numbers of erosional steps. The Salton Trough network (top) maintains the structure of its rank-size distribution through 7 erosional iterations, while the largest segments in the South Dakota network (bottom) rapidly dissociate into components with area approximately 2 orders of magnitude smaller, resulting in a drastic shallowing of the slope of the size distribution. This is a consequence of the differences in spatial structure and fractal dimension of the two networks.

5. Discussion

Considerable range exists in the slope and curvature of the size distributions shown in Fig. 3B – but the similarities are much more surprising than the differences. Indeed, we find it remarkable that there is any similarity at all given the diversity of landscapes (Fig. 3A) and of vegetation abundance distributions (histogram insets of Fig. 3B) from which they are derived. While it is clear that none of the 9 size distributions here exactly resembles the global size distribution in Fig. 1, it is similarly clear that none of the 9 landscapes used in this study comes close to fully sampling the diversity or scope of agriculture at global scales. Furthermore, because the differences between size distributions emerge from the differences in landscapes, these differences can be diagnostic in characterizing the variability in spatial distributions of agriculture across widely variable landscapes. From a network perspective, a diversity of size distributions implies a diversity of network structures.

Several potential explanations exist for the differences in rank-size distributions shown in Fig. 3B. Some Landsat scenes, such as the Salton Trough and Indus scenes, feature heavily irrigated agricultural landscapes in which cropland is tightly clustered around the hydrologic distribution network. This clustering impacts the rank size distribution and corresponding power law fit. All scenes feature some variable amount of human settlement, and some scenes such as North China, Delhi, and South Dakota feature spatially extensive conurbations which visibly disrupt the agricultural landscape. The extent to which these populated areas influence the spatial pattern of the agricultural land in the scene impacts the rank-size distribution of agricultural land.

Furthermore, the scenes range widely in levels of agricultural development, from smallholder farms (e.g. G-B Delta and Delhi), to industrial scale production (e.g. Salton Trough and South Dakota). While our analysis captures clusters of agriculture rather than individual fields, the distribution of field sizes contributes to the size of these clusters and thus the rank-size distribution. Major non-agricultural curvilinear and rectilinear features (e.g. rivers and roads) also cross-cut all of the scenes, providing a plethora of subscene background geometries which break apart some contiguous segments and encourage others to grow together. Finally, a wide range of climate zones from tropical (G-B Delta and Mato Grosso) to hyperarid (Salton Trough and Indus) impact the

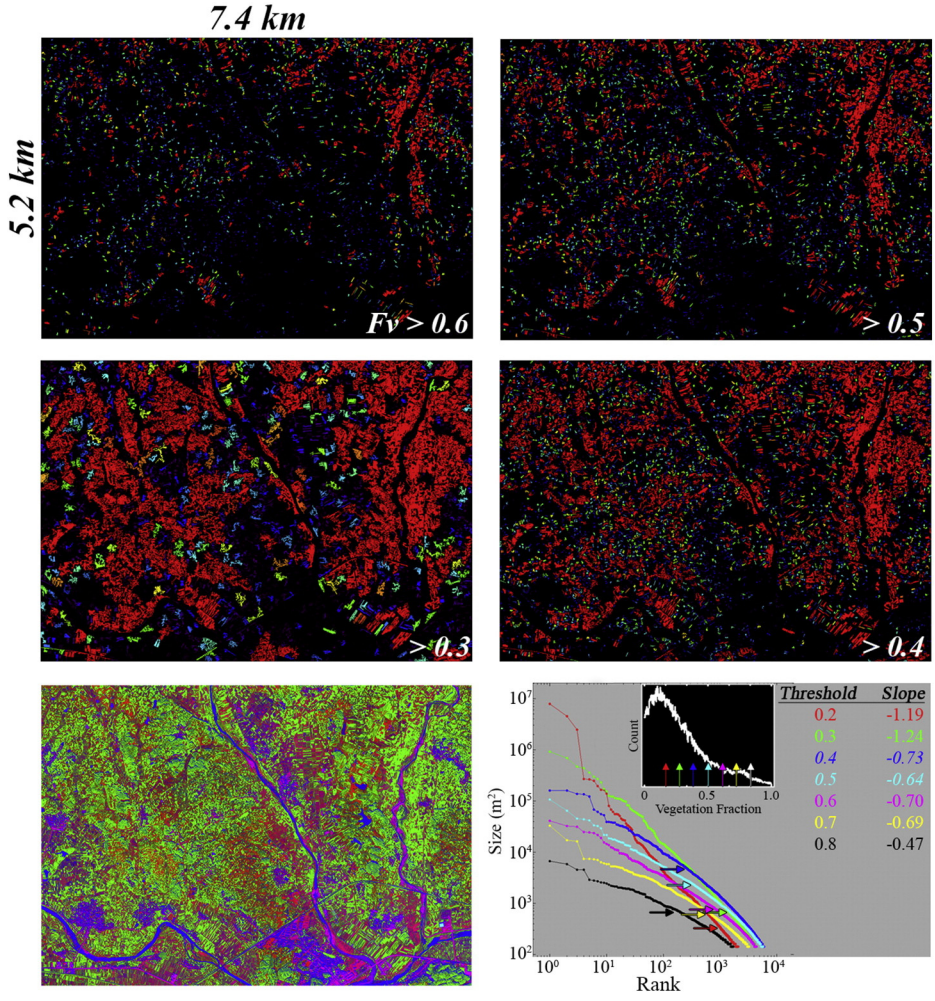

Fig. 5. Successive thresholding of vegetation fraction for a 39 km^2 IKONOS image of Anhui, China. Rank-Size plots show a similar succession to those from Landsat in Fig. 2.

background landscape mosaic within which the contiguous agricultural land for each scene is embedded. A detailed investigation of the way in which these (and other) factors generate the differences in rank-size distributions could provide a rich subject for future work.

Some of the size distributions in Fig. 3B cannot plausibly be described as power laws. Some exhibit power law behavior that truncates in the middle of the distribution. Others show statistically plausible power law behavior extending deep into the lower tail of the distribution. We suggest that the important characteristic of the size distributions is not presence or absence of statistically defensible power law behavior, but rather that every distribution shown here is similarly heavy tailed. Every size distribution shows many more small patches than large patches, and nearly all distributions show that when ordered by area patches become smaller and more frequent at similar rates – implying the total area sum of patches at any size is nearly equal to the total sum at any other size. This property corresponds to a slope near −1 on the plots in Fig. 3B. Further, Fig. 3C shows that many of the distributions vary with threshold in a predictable way: starting at high threshold (right side of the plots), the size distribution increases in slope as the threshold drops and the components grow (moving right to left on the plot) until reaching linearity near −1. At this point, a giant component emerges and dominates the network.

As the threshold is dropped even further, more and more of the remaining patches become connected into the giant component, reducing the total number of segments until every pixel in the entire domain is superconnected. The variations in progression of network structure with threshold are related to the fraction distributions, but the gross structure described above occurs in a consistent way across a wide range of conditions. A similar progression is also shown for the IKONOS image (Fig. 5) over a much smaller spatial domain. Similar progressions have even been observed in some random spatial networks and a general mechanism for the process been proposed (Small and Sousa, 2015). Despite this observed commonality, some of the distributions shown here vary with changing threshold in a more complex way than described above. This discrepancy often corresponds to severe non-Gaussianity in the F_v histogram. Detailed analysis of this complexity will be the subject of further study.

Landsat – Sentinel-2 intercomparisons provide an opportunity to make note of several important limitations of spatial network analysis for agricultural landscapes. A primary challenge, long recognized, is in the definition of "agriculture" as observed by remote sensing. In this study we use vegetation fraction because it is a physically meaningful quantity which has been shown to be consistent across sensors and scalable across spatial resolution. However, it cannot distinguish between anthropogenically-driven vegetation (i.e. agriculture, including silviculture) and natural vegetation. When scenes are pre-selected to be dominated by agriculture, as they have been in this study, this problem is minimized – but not completely eliminated. Inspection of the hillslopes around the agriculturally-dominated caldera in the Abbruzzo scene provides one example of this scenario.

Phenology provides another challenge which can be highly complex in agricultural landscapes. Network analysis of any single image provides a single snapshot of that landscape in time. As is apparent from Fig. 4B and C, temporal variability on the order of weeks can substantially alter the spatial connectivity of an agricultural landscape. Complete network-based analyses intended for practical applications must account for the phenology of the landscapes which they observe, as spatially-dependent processes are often temporally-dependent as well. As quantified by single-image vegetation fraction, landscapes possess a temporal progression of connectivity structures – of which the power law phase may be short-lived. We find it remarkable that so many agricultural networks, even when undersampled so extensively in both time and space, still display this unique structure which is so similar to that of the global network.

Analysis of two seemingly similar agricultural landscapes by network erosion shown in Fig. 6 demonstrates one potential application of the concepts presented in this paper. In one case (Salton Trough), power law behavior with slope near −1 is persistent even after removal

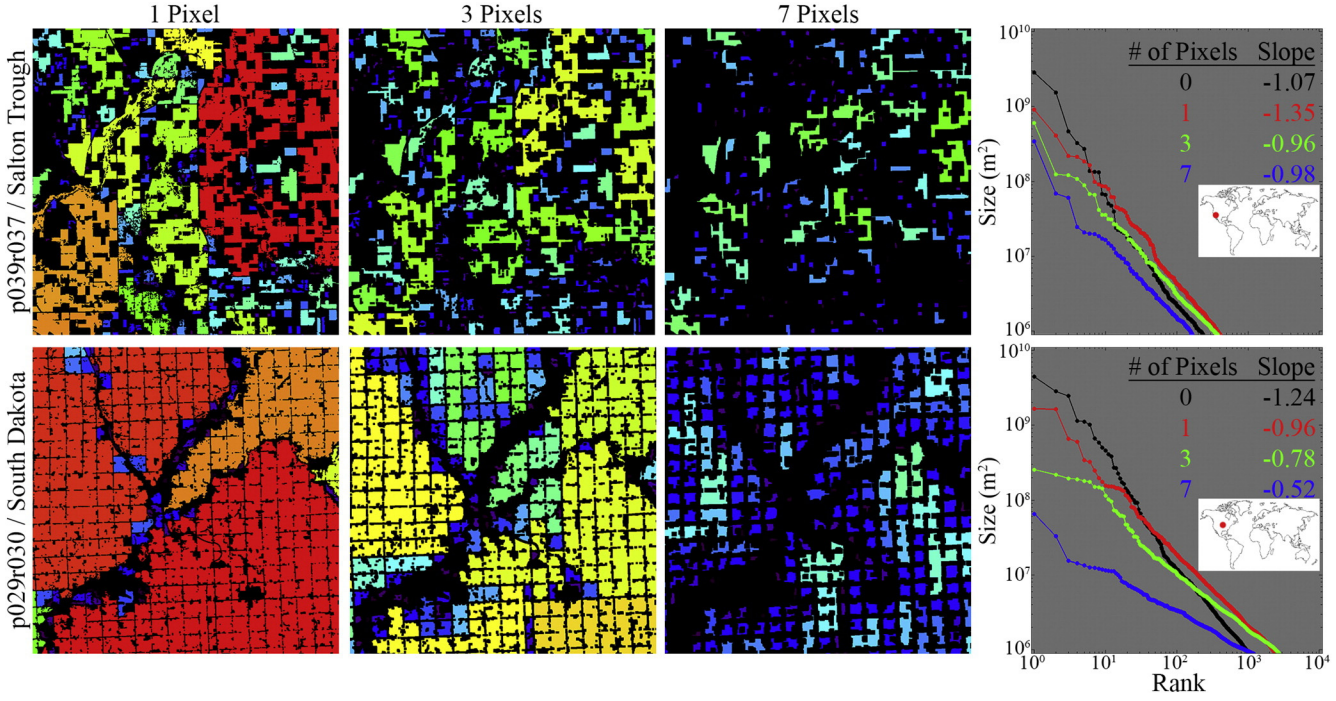

Fig. 6. Disruption of agricultural networks by erosion. Images show a 1000 × 1000 pixel subset of segment area maps derived from full Landsat scenes. In each step, all boundary pixels are removed from the network, akin to removing rings of an onion. Regular rectilinear patterns correspond to real features of the landscape (i.e. roads) which often serve to guide the erosion process. As pixels are removed, the upper tail of the size distribution may maintain a slope near −1 (top) or flatten considerably (bottom), depending on the spatial structure of the network. Networks with size distributions which maintain their slope in the face of erosional perturbations may be more robust to disruption. Whether this form of network stability is desirable or undesirable depends on the application.

of many pixels and considerable reduction of the total size of the network. In another case (South Dakota), the power law behavior of the network is much less robust. Removal of only a few pixels drastically reduces the sizes of the largest components (by a factor of ~100), rapidly breaking apart the largest segments of the network into much smaller disconnected components. This is clearly a result of the sectioning of the landscape by the regular grid of the road network. One could imagine a landscape which is more sensitive to small perturbations as being more easily disruptable – either a dangerous characteristic (as in the case of pollinator pathways) or a desirable one (as in the case of quarantining disease outbreaks). Understanding the robustness of the structure of an agricultural network to disruption could provide application-specific insight into practical methods for disrupting (or preventing disruption of) connectivity across an agricultural landscape.

Another possible application, not shown in this analysis, is to use multitemporal observations to constrain the growth and attenuation of agricultural networks in a landscape throughout the complete phenological cycle. As the agricultural mosaic evolves through time, different crops are planted, green up, senesce, and are harvested at different times of year. Taken together, the combination of the spatial distribution of these crops and their corresponding phenology time series govern the complete spatiotemporal agricultural network of a landscape. The diagnostic property of an agricultural landscape may be not just the network as observed at any one time but rather the robustness of the network properties throughout the course of the year. For instance, effective pollination may require an agricultural network to remain in a particularly interconnected state for a certain length of time. Crops may be particularly susceptible to disease outbreaks at one particular time of year. Native species may be more sensitive to disruptions of habitat in migration season than at other times of year. Furthermore, network adaptation to catastrophic environmental stresses such as drought or widespread disease outbreaks may be easily characterized. Finally, multitemporal network studies – like all of the analyses performed in this paper – have the added benefit of being easily performed nearly anywhere on Earth using simple methodologies and freely available remotely sensed observations.

Acknowledgements

Work done by D. Sousa was conducted with Government support under FA9550-11-C-0028 and awarded by the Department of Defense, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. D. Sousa thanks his father, F.J. Sousa II, for teaching him that the only wisdom we can hope to acquire is the wisdom of humility. CS was funded by the NASA Interdisciplinary Science program (grants NNX12AM89G & NNN13D876T) and by the NASA Socioeconomic Data and Applications Center (SEDAC) (contract NNG13HQ04C). The authors thank Aaron Clauset and colleagues for providing the power law fitting code and methodology. The authors also thank three anonymous reviewers.

References

Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47.
Auerbach, F., 1913. Das Gesetz der Bevölkerungskonzentration.
Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.
Barthélemy, M., 2011. Spatial networks. Phys. Rep. 499, 1–101.
Cantwell, M.D., Forman, R.T., 1993. Landscape graphs: ecological modeling with graph theory to detect configurations common to diverse landscapes. Landsc. Ecol. 8, 239–255.
Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. SIAM Rev. 51, 661–703.
Diekötter, T., Billeter, R., Crist, T.O., 2008. Effects of landscape connectivity on the spatial distribution of insect diversity in agricultural mosaic landscapes. Basic and Applied Ecology 9, 298–307.
Dixo, M., Metzger, J.P., Morgante, J.S., Zamudio, K.R., 2009. Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. Biol. Conserv. 142, 1560–1569.
Gardner, R.H., Turner, M.G., Dale, V.H., O'Neill, R.V., 1992. A Percolation Model of Ecological Flows. Springer.
Gilligan, C.A., 2008. Sustainable agriculture and plant diseases: an epidemiological perspective. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 741–759.
Gutenberg, B., Richter, C.F., 1956. Magnitude and energy of earthquakes. Ann. Geophys. 9, 1–15.
Guzzetti, F., Malamud, B.D., Turcotte, D.L., Reichenbach, P., 2002. Power-law correlations of landslide areas in central Italy. Earth Planet. Sci. Lett. 195, 169–183.
Hantson, S., Pueyo, S., Chuvieco, E., 2015. Global fire size distribution is driven by human impact and climate. Glob. Ecol. Biogeogr. 24, 77–86.
Hsu, S.A., Meindl, E.A., Gilhousen, D.B., 1994. Determining the power-law wind-profile exponent under near-neutral stability conditions at sea. J. Appl. Meteorol. 33, 757–765.
Kumar, S., Roy, D., Boschetti, L., Kremens, R., 2011. Exploiting the power law distribution properties of satellite fire radiative power retrievals: a method to estimate fire radiative energy and biomass burned from sparse satellite observations. J. Geophys. Res. Atmos. 116.
Li, W., 2002. Zipf's law everywhere. Glottometrics 5, 14–21.
Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: keys and clues. Bioscience 51, 341–352.
Lotka, A.J., 1941. The law of urban concentration. Science 94, 164 (New York, NY).
Luoto, M., Rekolainen, S., Aakkula, J., Pykälä, J., 2003. Loss of plant species richness and habitat connectivity in grasslands associated with agricultural change in Finland. AMBIO: A Journal of the Human Environment 32, 447–452.
Malamud, B.D., Millington, J.D., Perry, G.L., 2005. Characterizing wildfire regimes in the United States. Proc. Natl. Acad. Sci. U. S. A. 102, 4694–4699.
McIntyre, N.E., Wright, C.K., Swain, S., Hayhoe, K., Liu, G., Schwartz, F.W., Henebry, G.M., 2014. Climate forcing of wetland landscape connectivity in the Great Plains. Front. Ecol. Environ. 12, 59–64.
Newman, M., 2010. Networks: An Introduction. Oxford University Press Scholarship Online, Oxford, England.
Ricketts, T.H., Williams, N.M., Mayfield, M.M., 2006. Connectivity and ecosystem services: crop pollination in agricultural landscapes. Connectivity Conservation. Campridge Univeristy Press, Cambridge.
Scanlon, T.M., Caylor, K.K., Levin, S.A., Rodriguez-Iturbe, I., 2007. Positive feedbacks promote power-law clustering of Kalahari vegetation. Nature 449, 209–212.
Seguro, J., Lambert, T., 2000. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. J. Wind Eng. Ind. Aerodyn. 85, 75–84.
Small, C., Milesi, C., 2013. Multi-scale standardized spectral mixture models. Remote Sens. Environ. 136, 442–454.
Small, C., Sousa, D., 2015. Spatial Scaling of Land Cover Networks. ArXiv.
Small, C., Sousa, D., 2016. Humans on Earth; global extents of anthropogenic land cover from remote sensing. Anthropocene (Invited Manuscript, Submitted).
Small, C., Elvidge, C.D., Balk, D., Montgomery, M., 2011. Spatial scaling of stable night lights. Remote Sens. Environ. 115, 269–280.
Sousa, D., Small, C., 2016. Spatial Resolution and Characteristic Scale in Spatial Networks (Forthcoming manuscript).
Sornette, D., 2006. Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools. Springer Science & Business Media.
Steinwendner, J., 2002. Graph-theoretic issues in remote sensing and landscape ecology. EnviroInfo 1, 546–552.
Turcotte, D.L., 1997. Fractals and Chaos in Geology and Geophysics. Cambridge University Press.
Urban, D., Keitt, T., 2001. Landscape connectivity: a graph-theoretic perspective. Ecology 82, 1205–1218.
Wood, R., Field, P.R., 2011. The distribution of cloud horizontal sizes. J. Clim. 24, 4800–4816.
Zipf, G.K., 1942. The unity of nature, least-action, and natural social science. Sociometry 48–62.