

Spatial patterning among savanna trees in high-resolution, spatially extensive data

A. Carla Staver^{a,1}, Gregory P. Asner^b, Ignacio Rodriguez-Iturbe^{c,1}, Simon A. Levin^d, and Izak P.J. Smit^{e,f}

^aDepartment of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511; ^bCenter for Global Discovery and Conservation Science, Arizona State University, Tempe, AZ 85287; ^cDepartment of Ocean Engineering, Texas A & M University, College Station, TX 77843; ^dDepartment of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; ^eScientific Services, South African National Parks, Skukuza 1350, South Africa; and ^fCentre for African Ecology, School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg 2050, South Africa

Contributed by Ignacio Rodriguez-Iturbe, February 25, 2019 (sent for review November 13, 2018; reviewed by Dara Entekhabi and Ricardo Holdo)

In savannas, predicting how vegetation varies is a longstanding challenge. Spatial patterning in vegetation may structure that variability, mediated by spatial interactions, including competition and facilitation. Here, we use unique high-resolution, spatially extensive data of tree distributions in an African savanna, derived from airborne Light Detection and Ranging (LiDAR), to examine tree-clustering patterns. We show that tree cluster sizes were governed by power laws over two to three orders of magnitude in spatial scale and that the parameters on their distributions were invariant with respect to underlying environment. Concluding that some universal process governs spatial patterns in tree distributions may be premature. However, we can say that, although the tree layer may look unpredictable locally, at scales relevant to prediction in, e.g., global vegetation models, vegetation is instead strongly structured by regular statistical distributions.

savanna | spatial pattern | LiDAR | heterogeneity

Savannas are defined as having open tree canopies and a more-or-less continuous grass layer—a classification that differentiates savannas from closed-canopy forests (1) and perhaps even from pure grasslands (2), but which encompasses substantial variability in vegetation and ecosystem structure (3). Beyond this broad definition, variability in savanna vegetation structure has resisted mechanistic characterization and explanation, despite repeated analysis (4–8), perhaps because savanna vegetation is the product of complex interactions and feedbacks between bottom-up constraints from climate and soils and top-down control from fire and herbivory. As such, savannas pose a challenge for predictive frameworks: diverse biosphere models consistently underpredict savanna distributions or even fail to capture savanna vegetation entirely. A more fundamental understanding of how savanna vegetation varies, and its drivers, will be critical for improving the representation of this globally widespread biome that covers at least 40% of the global tropics.

There may be mundane reasons that savanna structural variability has been difficult to constrain. One possibility is that existing efforts have examined vegetation structure at scales that are minuscule compared with the processes that shape it (3). These may include rare events (9, 10), which are difficult to sample adequately, or the problem may be more trivial, since field plots—often as small as fractions of hectares—may fail to statistically sample tree populations sufficiently. By the same token, larger-scale analyses (5, 11) often leverage optical remote sensing data, which cover relatively short time scales and are often insufficient for characterizing savannas (12). Scaling up, with high-quality data, will clearly be necessary for deepening our mechanistic understanding of savanna tree-layer variability (3).

Savannas may also vary for reasons less mundane than those related to undersampling. Extensive theoretical work (13, 14) and more limited empirical observations (15) suggest that variation in savannas may be more organized, and less haphazard, than plot-based studies suggest. While locally patchy, savannas

can be highly regular. In special cases, patterns can repeat themselves at characteristic scales (16, 17); these patterns may result from the combination of local activation (e.g., facilitation via hydraulic redistribution; ref. 18) with longer-range inhibition (e.g., tree-tree competition; ref. 19). In other cases, or at least on other scales, patterns may be scale-free, best described by a power-law relationship between, e.g., the size of a patch of trees and its frequency (15, 20) of the form:

$$f(x) \propto x^{-\alpha},$$

where the frequency f of clusters of size x decreases with x via the exponent α (known as the power-law exponent). Short-range facilitation without inhibition (a type of Yule process) could be a possible mechanism (21), but power laws can emerge from diverse processes such that inferring process from pattern is impossible without extremely careful parameterization, and is difficult even then (21). Either way, distributions at larger scales speak to a degree of predictability—albeit not at the plot level—that predictive approaches to savanna ecosystems have so far mostly ignored.

These power-law distributions may describe widespread tree-layer variation in savannas and other arid ecosystems (15, 20). However, empirical work examining power laws in data (not just in savannas) has come under scrutiny for a lack of statistical rigor in how power laws are estimated (22, 23), broadly invalidating results based on fitting linear regressions to statistical distributions. Data are also improving dramatically in both extent and quality; for instance, airborne Light Detection and Ranging (LiDAR) techniques can now provide highly

Significance

Understanding how environmental variation shapes vegetation distributions is crucial for predicting biosphere responses to global change. Savannas represent a major challenge, and repeated efforts have failed to constrain variation in savanna vegetation structure to a useful degree. Here, we find that trees cluster in savannas and that their clustering is governed by power laws, with parameters and geometries that are invariant to environment. This explains why savannas have resisted mechanistic characterization and offers a possible tool for constraining variability in vegetation structure.

Author contributions: A.C.S., G.P.A., I.R.-I., S.A.L., and I.P.J.S. designed research; A.C.S. performed research; G.P.A. contributed new reagents/analytic tools; A.C.S. and G.P.A. analyzed data; and A.C.S., G.P.A., I.R.-I., S.A.L., and I.P.J.S. wrote the paper.

Reviewers: D.E., Massachusetts Institute of Technology; and R.H., University of Georgia.

The authors declare no conflict of interest.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email: carla.staver@yale.edu or irodriguez@ocn.tamu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1819391116/-DCSupplemental.

Published online May 13, 2019.

accurate data on vegetation occurrence and structure at the submeter scale over large landscapes. Methodologically, these data represent a tangible improvement over optical approaches for differentiating trees from the grasses in between; LiDAR also allows for evaluations of power law occurrence over a range of spatial scales (24) that is unprecedented in savannas. Comprehensive reevaluation of claims about the ubiquity of power-law distributions in the savanna tree layer is necessary and now possible.

New high-resolution LiDAR data also permit more nuanced examinations of tree clusters. Theory suggests that scale-free patterns in vegetation structure should be reflected not just in power law distributions of cluster sizes, but also in the geometry of those clusters (23). In a classic example, the length of the coastline of an island classically scales as a power of its area, via an exponent that defines its fractal dimension; however, because coastlines are tortuous (rough), these fractals do not scale as strictly 2D objects (with dimensionality of 2) but rather as something between a line and an area (25).

Here, we reexamine the statistical distributions and geometries of tree patch sizes in Kruger National Park in South Africa (Fig. 1), asking whether distributions are best described by power laws, and how their parameters and geometries depend on underlying environmental variation in climate and soils. To do this, we used high-resolution (56 cm) airborne LiDAR-based estimates of tree distributions (trees > 3.5 m height) across 10 large landscapes (>6,000 ha each) in Kruger, using robust distribution fitting techniques to estimate distribution parameters (see *Methods and Materials* for more detail). Kruger spans significant orthogonal ecological gradients in rainfall (300–750 mm) and soils (clay-rich versus sandy) (6), originating from differences in parent material, which also give rise to strongly contrasting topography and river structure, as well as contrasting fire and herbivory regimes.

Results and Discussion

We found that power laws were widespread, in some cases over two or even three orders of magnitude in tree clump size (Fig. 2A). Across all landscapes, the tree cluster size above which power laws adequately fit the fat tail of the distribution (defined as x_{\min}) is significantly related to independent, field-based estimates of the maximum canopy area of large trees (Fig. 2B; $R^2 = 0.630$, $n = 10$, $P = 0.0062$), suggesting that, intuitively, power-law tree clustering can only occur above the size of the individual tree. Both tree size ($R^2 = 0.622$, $n = 10$, $P = 0.0067$) and x_{\min} ($R^2 = 0.552$, $n = 10$, $P = 0.013$) increased significantly with rainfall (Fig. 3). Together, these findings substantially corroborate the idea that tree-clustering distributions are predictable, if not at small scales then at larger ones, and that universal statistical distributions (power laws) may be useful for describing the clustering of trees in savanna systems.

What is much more surprising is the unexpected consistency in the distribution of large tree clumps across landscapes in Kruger. Once we controlled for the smallest meaningful cluster size (x_{\min}), power law slopes were relatively consistent across landscapes, showing no systematic variation with respect to rainfall ($t_{10} = 0.13$, $P = 0.90$), landform ($t_{10} = 1.19$, $P = 0.30$; Fig. 2C), fire frequency ($t_{10} = -0.31$, $P = 0.77$), or mean distance to river ($t_{10} = -0.48$, $P = 0.66$), although distributions were impossible to estimate with confidence (Fig. 2C) on two landscapes with low tree cover (0.1% and 1.5%) and should not be considered conclusive with respect to geology. Excluding these, the mean power-law α across landscapes was 2.72 ± 0.16 (or equivalently 1.72 ± 0.16 for the power-law exponent of the inverse cumulative distribution function), showing dramatically less variation in tree cluster size distributions over gradients in rainfall—in high resolution, spatially extensive

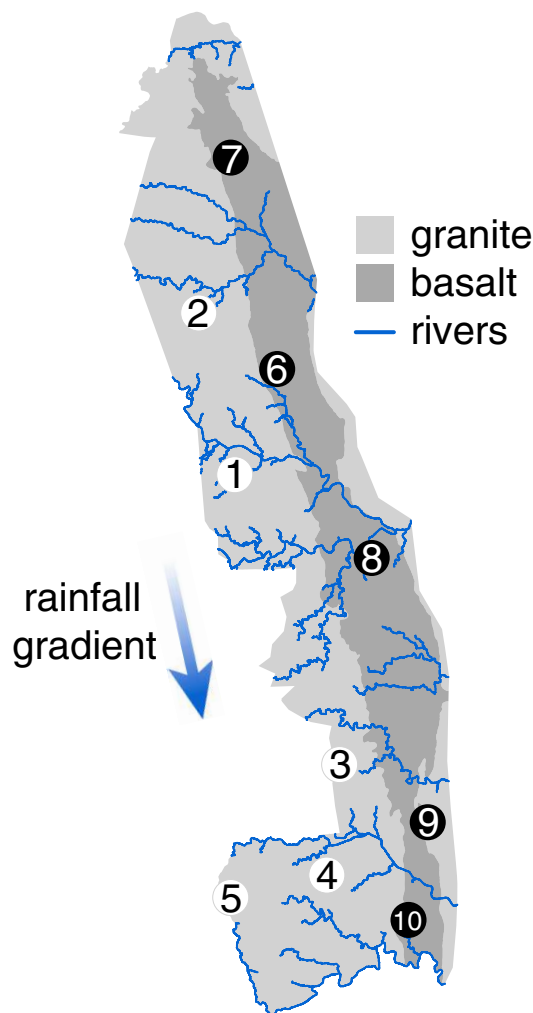


Fig. 1. Map of Kruger National Park, showing locations of large landscapes included in this study, parent material, rivers and streams, and the overall direction of the rainfall gradient. Large landscape positions are colored by parent material and then numbered by increasing rainfall (note the inversion of the rainfall gradient in the north of Kruger).

data—than previous studies have described (15, 20). Excluding clumps that were close to rivers from the analysis did not change estimates of α , which further suggests that rivers were not primarily responsible for observed tree clustering patterns (excluding clumps within 500 m, 1 km, and 5 km yield estimates of $\alpha = 2.71$ or 2.72).

Perimeter-area relationships were also predictable across landscapes, included those with lower tree cover (Fig. 4). Perimeter-area methods have been criticized as a method for characterizing scale-free distributions, because approaches are based on strong a priori assumptions about self-similarity, and are subject to computational issues and interactions with scales of analysis (23). Here, we found that patch geometry seemed to scale consistently within landscapes (Fig. 4A) in a way that did not depend on the scale of analysis [a frequent limitation of analyses of this type (23); Fig. 5]. Unlike patch-size distribution parameters, however, fractal dimension changed with rainfall (Fig. 4B), with fuller patches at higher rainfall and more linear ones at lower rainfall ($R^2 = 0.586$, $n = 10$, $P = 0.0099$; see Table 1). Again, changes were small but may reflect increases in tree size with increasing rainfall or may alternatively suggest trends in tree clustering across a broader rainfall gradient. Regardless, tree-cluster geometry may be more accessible at typical

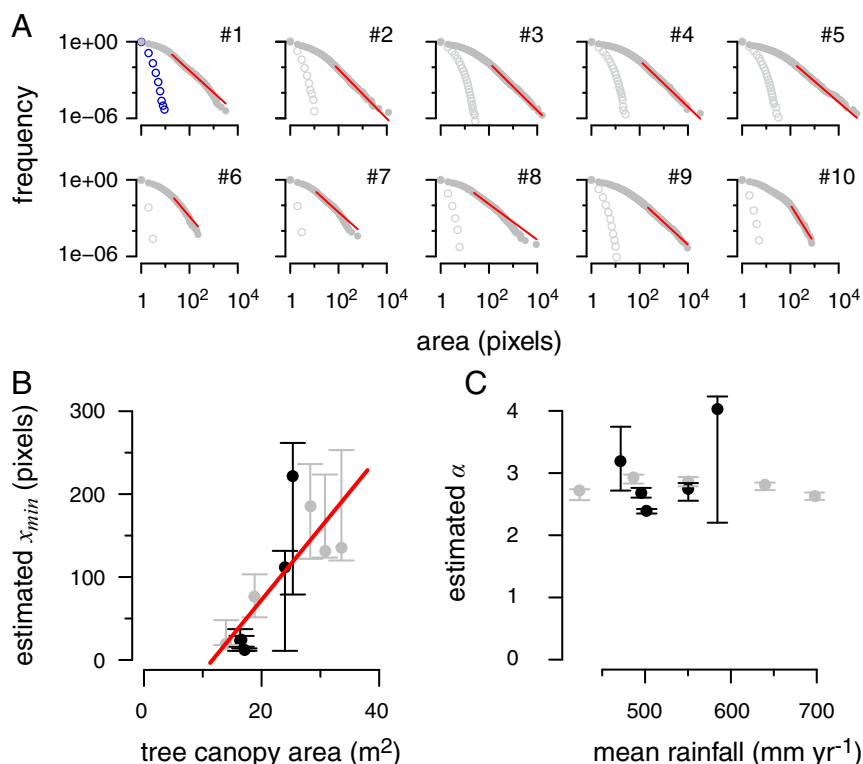


Fig. 2. Tree clump size distributions with estimated power law fits for all landscapes (A), response of power-law minimum clump size (x_{\min}) to independent estimates of individual tree canopy area (B), and response of power-law slope to mean annual rainfall (C). In A, red lines show power-law fits, gray dots show actual data, and open circles show properties of a randomized landscape with tree cover equal to that of each real landscape; in B, the red line denotes the best-fit estimate for the response ($R^2 = 0.630$, $n = 10$, $P = 0.0062$). In B and C, light gray points correspond to landscapes on granitic soils and dark ones to those on basaltic soils.

scales of analysis (23) and more sensitive to underlying environment than tree cluster size distributions at large scales (Fig. 2), which are not often available, or direct plot-based assessments of tree density at smaller scales, which yield poor predictive ability (3, 4, 6).

From this analysis, tree clusters and their distributions look more consistent across environments than previously thought. Theory gives us no real reason to suppose this should be the case; while generating power laws in tree cluster sizes is not difficult, most mechanistic models yield distributions with parameters that vary with respect to environment (15, 26). One possibility could be that tree cluster distributions are governed by some universality class; another could be that tree clustering is governed by some predictable but hitherto unappreciated aspect of tree physiology and its scaling properties. Either would represent a major breakthrough in our understanding of savanna vegetation heterogeneity.

Before such radical conclusions can be supported, however, a next line of future inquiry should be to broaden the scope of the current analysis to include more diverse environmental variation. Although Kruger encompasses broad variation in soils and a reasonable rainfall gradient (6), landscapes here range in tree cover (with height > 3.5 m) from near zero only up to ~13% and in rainfall from a minimum of 300 mm to 750 mm rainfall, and as such are not representative of the full range of possible savanna variation (7), especially among continents across which tree architecture may differ (27). This will require comprehensive large-scale, high-quality data, and may not be possible to the exceptional degree presented here, but establishing whether cluster size distributions are in fact invariant and accurately estimating their parameters will depend on a broader geographical approach.

What we can say with confidence is that, overall, the tree layer in savannas is far more predictable than smaller-scale plot analyses have suggested. While the tree layer looks unpredictable locally (3, 6, 8), it is instead strongly structured by regular statistical distributions at scales relevant to prediction. Savanna ecologists should aim for mechanistic understanding, both of invariant distributions (if indeed they are), and of parameters that vary smoothly with environment (here, individual tree size). Meanwhile, global models, which often operate assuming no spatial structure either within or among simulation units (28, 29), should consider how to leverage emergent distributions to improve predictions of vegetation distributions and carbon cycles in savannas.

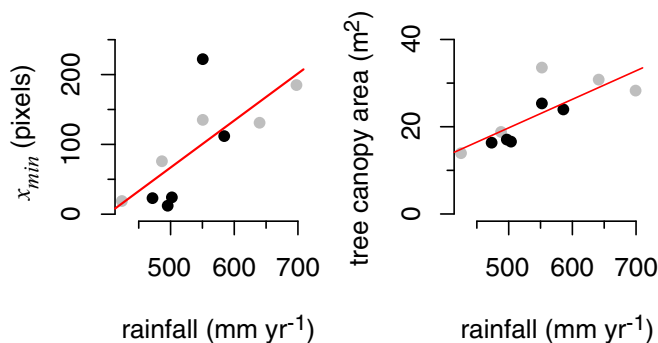


Fig. 3. The response of power-law minimum clump size (x_{\min}) (*Left*) and independently estimated tree-canopy area to rainfall (*Right*). Best-fit relationships are shown in red ($R^2 = 0.622$, $n = 10$, $P = 0.0067$ and $R^2 = 0.552$, $n = 10$, $P = 0.013$, respectively).

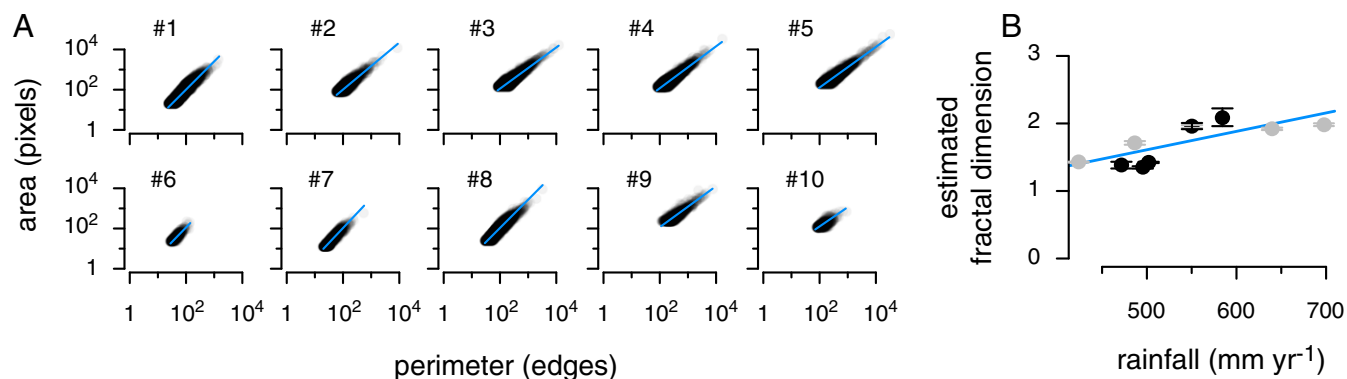


Fig. 4. Tree cluster area vs. tree cluster perimeter for all landscapes (A), and the response of resulting fractal dimension estimate to mean rainfall (B). Estimated fractal dimension increased with rainfall ($R^2 = 0.586$, $n = 10$, $P = 0.0099$).

Methods and Materials

Tree clusters were mapped across 10 landscapes in Kruger National Park, South Africa, in April 2012 using the Carnegie Airborne Observatory (CAO)-AToMS System (30). AToMS includes a waveform LiDAR scanner with integrated Global Positioning System-Inertial Measurement Unit (GPS-IMU), which returned 3D positioning and altitude data for the sensor onboard the aircraft, allowing for highly precise and accurate projection of laser ranging measurements on the ground. The GPS-IMU data were combined with the laser ranging data to determine the 3D location of each laser return. LiDAR data were collected from 2,000 m above ground level at 50 kHz laser pulse repetition rate, a 17° half-scan angle and 50% overlap between adjacent flight lines, resulting in LiDAR measurements with 56-cm laser spot spacing. These raw LiDAR returns were then used to derive top-of-canopy and ground digital elevation models (DEMs) (31). Woody canopy height was estimated as the difference between top-of-canopy and ground DEMs.

This woody canopy height map was then thresholded at 3.5 m to yield a presence/absence map of trees; the threshold of 3.5 m was chosen to conservatively accommodate LiDAR accuracy (which erodes strongly below 2 m in height) but also has ecological meaning in savannas, since fires and most herbivores have much lighter impacts on savanna trees above this height. These maps were analyzed using the package “raster” in R (version 3.2.2), with tree clusters identified via the “clump” algorithm with a Queen (Moore) neighborhood. The algorithm is more robust to gaps in tree canopies than the alternative Rook (von Neumann) neighborhood would be. Having identified clusters and estimated their perimeter and area, we then proceeded to fit power law distributions with the package “powerLaw” (22). This package fits the inverse cumulative distribution function $F(x)$, defined as:

$$F(x) = \left(\frac{x}{x_{\min}} \right)^{-(\alpha-1)}$$

rather than fitting either a statistical distribution to the frequency distribution or, worse, a linear model to binned data, which has been a common approach historically; this approach has been shown theoretically and via testing on distributions with known parameters to yield more robust estimates of power-law distribution parameters (22). This package also extracts an estimate for the minimum cluster size above which a fat-tailed distribution fits data (x_{\min}); note too that, here x and α correspond to those quantities defined in the main text. Values of α and x_{\min} were estimated by minimizing the Kolmogorov–Smirnov statistic. To evaluate sensitivity of parameter estimates to samples and sample sizes, we also bootstrapped (100x) parameter estimates to derive 95% confidence intervals.

Perimeter (P):area (A) relationships were evaluated as the linear relationship between log-transformed perimeter and log-transformed area, yielding an estimate of slope β and the overall relationship

$$A \propto P^\beta.$$

The fractal dimension of a set of self-similar clusters (25) is given by $2/\beta$. Lines were fit in the package “lmodel2” in R via model II major axis regression, which assumes the possibility of errors in observations in both variables, appropriate here because both were estimated via the same data collection

techniques. Only clusters above the x_{\min} estimated for each landscape were included for estimating this relationship.

To evaluate the effects of scale of analysis on estimates of power-law and P:A relationships (Fig. 5), we aggregated rasters from their original resolution by a factor of 1–8, taking average tree cover within each new pixel and defining presence when proportional tree cover was greater than or equal to 0.5. We found that no parameter estimate depended strongly on the scale of analysis, suggesting that results presented in the main text are robust.

Finally, we have used a dataset maintained by Kruger National Park management (the Veld Condition Assessment dataset; see ref. 6 for a complete analysis and description of these data) for ground-based estimates of individual tree canopy area, which were used to compare against estimates of the minimum cluster size above which power-law fits apply (i.e., are minimum cluster sizes related to tree size?) (Datasets S1 and S2). In 2008, vegetation structure data were collected at 457 sites throughout Kruger. At each site, tree size (including diameter, height, and maximum canopy diameter) and species were collected for all trees with height > 3 m within a radius of 5 m of eight points distributed evenly within a 50 m × 60 m plot. For the purposes of this analysis, we then selected the tree with the largest canopy diameter at each plot. To maximize data use, estimates of canopy diameter for each landscape were estimated as mean canopy diameter for all VCA plots within 50 km of the centroid of each large landscape (larger than the landscapes included here, but a better balance given sparse field sampling). Canopy area was

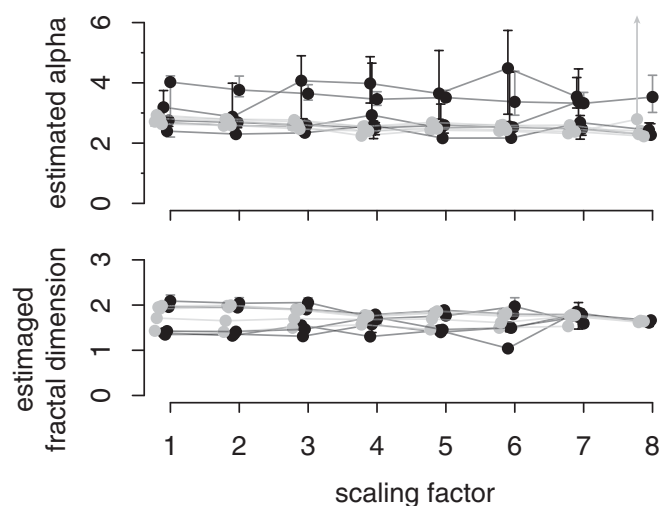


Fig. 5. Scaling effects on power law estimates and P:A relationships. A factor of one corresponds to original data resolution (56-cm pixels), a factor of two to a resolution that is twice that of the original (112-cm pixels), a factor of three to a resolution that is three times that of the original (168-cm pixels), and so on. Scaling has negligible effects on estimated parameters.

Landscape	Estimated x_{\min}	Estimated α	Fractal dimension
1	19	2.72	1.43
2	76	2.93	1.71
3	135	2.89	1.95
4	131	2.81	1.92
5	185	2.63	1.98
6	23	3.19	1.38
7	12	2.68	1.35
8	24	2.40	1.42
9	222	2.75	1.96
10	112	4.02	2.09

determined as $A = \pi r^2$. We are clearly overestimating canopy area, however, given that field measurements recorded only maximum canopy diameter; tree size trends with rainfall would thus benefit from further, careful direct examination.

ACKNOWLEDGMENTS. We thank Juan Bonachela and Sally Archibald for helpful discussion of this work. This collaboration was supported by a grant from the Andrew W. Mellon Foundation, National Science Foundation Division of Mathematical Sciences Grants 1615531 (to A.C.S.) and 1615585 (to S.A.L.), and National Science Foundation Macrosystems Biology Grant 1802453 (to A.C.S.). Airborne LiDAR data collection and processing were made possible by grants from the Andrew W. Mellon Foundation (to G.P.A.). The Carnegie Airborne Observatory has been made possible by grants and donations to G.P.A. from the Avatar Alliance Foundation, Margaret A. Cargill Foundation, David and Lucile Packard Foundation, Gordon and Betty Moore Foundation, Grantham Foundation for the Protection of the Environment, W. M. Keck Foundation, John D. and Catherine T. MacArthur Foundation, Mary Anne Nyburg Baker and G. Leonard Baker Jr., and William R. Hearst III.

1. Staver AC, Archibald S, Levin SA (2011) The global extent and determinants of savanna and forest as alternative biome states. *Science* 334:230–232.
2. Dantas VdL, Hirota M, Oliveira RS, Pausas JG (2016) Disturbance maintains alternative biome states. *Ecol Lett* 19:12–19.
3. Staver AC (2018) Prediction and scale in savanna ecosystems. *New Phytol* 219:52–57.
4. Sankaran M, et al. (2005) Determinants of woody cover in African savannas. *Nature* 438:846–849.
5. Sankaran M, Ratnam J, Hanan N (2008) Woody cover in African savannas: The role of resources, fire and herbivory. *Glob Ecol Biogeogr* 17:236–245.
6. Staver AC, Botha J, Hedin L (2017) Soils and fire jointly determine vegetation structure in an African savanna. *New Phytol* 216:1151–1160.
7. Lehmann CER, et al. (2014) Savanna vegetation–fire–climate relationships differ among continents. *Science* 343:548–552.
8. Colgan MS, Asner GP (2014) Coexistence and environmental filtering of species-specific biomass in an African savanna. *Ecology* 95:1579–1590.
9. Prins H, van der Jeugd HP (1993) Herbivore population crashes and woodland structure in East Africa. *J Ecol* 81:305–314.
10. Gillson L (2006) A “large infrequent disturbance” in an East African savanna. *Afr J Ecol* 44:458–467.
11. Bucini G, Hanan NP (2007) A continental-scale analysis of tree cover in African savannas. *Glob Ecol Biogeogr* 16:593–605.
12. Staver AC, Hansen MC (2015) Analysis of stable states in global savannas: Is the CART pulling the horse?—A comment. *Glob Ecol Biogeogr* 24:985–987.
13. Tarnita CE, et al. (2017) A theoretical foundation for multi-scale regular vegetation patterns. *Nature* 541:398–401.
14. Rieterk M, Dekker SC, de Ruiter PC, van de Koppel J (2004) Self-organized patchiness and catastrophic shifts in ecosystems. *Science* 305:1926–1929.
15. Scanlon TM, Caylor KK, Levin SA, Rodriguez-Iturbe I (2007) Positive feedbacks promote power-law clustering of Kalahari vegetation. *Nature* 449:209–212.
16. White LP (1970) Brousse tigrée patterns in southern Niger. *J Ecol* 58:549–553.
17. Bonachella JA, et al. (2015) Ecological feedbacks. Termit mound can increase the robustness of dryland ecosystems to climatic change. *Science* 347:651–655.
18. Scholz FG, et al. (2008) Biophysical and life-history determinants of hydraulic lift in Neotropical savanna trees. *Funct Ecol* 22:773–786.
19. Dohn J, Augustine DJ, Hanan NP, Ratnam J, Sankaran M (2017) Spatial vegetation patterns and neighborhood competition among woody plants in an East African savanna. *Ecology* 98:478–488.
20. Berdugo M, Kefi S, Soliveres S, Maestre FT (2017) Plant spatial patterns identify alternative ecosystem multifunctionality states in global drylands. *Nat Ecol Evol* 1:3.
21. Newman M (2005) Power laws, Pareto distributions and Zipf’s law. *Contemp Phys* 46:323–351.
22. Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51:661–703.
23. Halley JM, et al. (2004) Uses and abuses of fractal methodology in ecology. *Ecol Lett* 7:254–271.
24. Asner GP, et al. (2013) Forest canopy gap distributions in the southern Peruvian Amazon. *PLoS One* 8:e60875.
25. Sugihara G, M May R (1990) Applications of fractals in ecology. *Trends Ecol Evol* 5:79–86.
26. Li Q, Staver AC, Weinan E, Levin SA (2019) Spatial feedbacks and the dynamics of savanna and forest. *Theor Ecol*, 10.1007/s12080-019-0428-1.
27. Moncrieff GR, et al. (2014) Contrasting architecture of key African and Australian savanna tree taxa drives intercontinental structural divergence. *Glob Ecol Biogeogr* 23:1235–1244.
28. Fisher RA, et al. (2018) Vegetation demographics in Earth system models: A review of progress and priorities. *Glob Change Biol* 24:35–54.
29. Scheiter S, Langan L, Higgins SI (2013) Next-generation dynamic global vegetation models: Learning from community ecology. *New Phytol* 198:957–969.
30. Asner GP, et al. (2010) Carnegie airborne observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sens Environ* 124:454–465.
31. Asner GP, Levick SR (2012) Landscape-scale effects of herbivores on treefall in African savannas. *Ecol Lett* 15:1211–1217.