Characteristics of African Savanna Biomes for Determining Woody Cover

This companion file contains the supplemental information provided with (Sankaran et al., 2005).

Cite this data set as follows:

Sankaran M., N. P. Hanan, and R. J. Scholes. 2007. Characteristics of African Savanna Biomes for Determining Woody Cover. Data set. Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

Sankaran, M., N. P. Hanan, R. J. Scholes, et al. 2005. Determinants of woody cover in African savannas. NATURE 438 (7069): 846-849.

Supplementary methods and results

Details of soil analysis. Soil samples were obtained from 166 sites and analyzed under standardized laboratory conditions for texture, total N and P and N mineralization potential. Soils were typically sampled at 4-5 evenly spaced points (~25m apart) along a randomly placed transect at each site. At each point, soil samples (typically N=3, top 10 to 20cm) were collected a few feet apart from one another and bulked, yielding 4-5 replicates per site. Replicate samples were analyzed for soil C and N content by combustion with a LECO CHN analyzer (LECO Corp, St. Joseph, MI), and total soil P was determined by Inductively Coupled Plasma (ICP) spectrometry (Leman Labs, Hudson, MA) following acid digestion of soils with HNO_3-$HClO_4$. To provide an index of N availability and turnover in soils, potential N mineralization rates were quantified following laboratory incubations of soils under anaerobic conditions[1]. Five gram soil-samples were immersed in water-filled scintillation vials (1:4 w/v) for 7 days and then extracted with 25ml 2M KCl. Extracts were analyzed for NH_4-N and NO_3-N using an ALPKEM rapid flow analyzer (Alpkem, Clackamas, OR). An additional set of extractions were carried out on untreated 10g soil samples to determine initial amounts of inorganic N present. Potential N mineralization rates ($\mu g * g\ soil^{-1} * 7\ days^{-1}$) were determined as the difference in inorganic N between incubated and control soils. Finally, replicate soil samples from each site were pooled and soil texture determined on bulked soils using the Bouyoucos hydrometer method[2]. Including other published and unpublished data, estimates for soil % sand, %clay, %N, total P and mineralization potential were available for 175, 279,152, 155 and 146 sites, respectively.

Comparison of methods employed to estimate woody cover. Our final data set encompassed measures of woody cover derived using different sampling methodologies, including estimates based on densiometers (N=397), summed canopy areas of individuals in plots (N = 308), Bitterlich method (N=103), line intercepts (N=25), visual estimates (N=13), aerial photos/ satellite images (N=5) and presence/absence of woody canopies above fixed points (N = 3). A comparison of woody cover estimates based on the two most commonly employed methods in the dataset, densiometers and summed canopy areas (705 of 854 sites), indicate a close correspondence between the two (Fig S4; data are from 7 sites where both methods were employed to estimate woody cover and span a precipitation range from 409 to 970mm MAP), suggesting that the methods are comparable. Although data were not available to allow for a direct comparison of the other methods, visual inspection of the data revealed no consistent bias in woody cover estimates derived using any of the other methods. Further, our results remained qualitatively unchanged when the analysis was restricted to sites where cover was estimated using densiometers or by summing canopy areas, and hence, measurement methodologies do not appear to influence the results of our analysis. We have, therefore, presented results based on the larger dataset in the main text.

Analysis details for the piecewise linear model. The bent-cable form of the piecewise linear model allows the breakpoint between the pieces to be estimated and for a smooth, curved transition spanning the breakpoint that connects the two linear pieces. The width of the quadratic nonlinear curve connecting the two linear slopes can be used to indicate how sharp or gradual the transition is between the different linear changes. We examined changes in the loss function minimized by 'quantreg', checked that the proportion of residuals less than or equal to the estimate for a selected quantile $\tau \in \{0.90, 0.99\}$ was close to τ, and plotted estimated functions against the data to decide when 'quantreg' had converged to a good solution. Different starting values for parameters and tolerance values often were required before 'quantreg' would iterate to a reasonable solution.

We also used a b-spline approach to estimating piecewise linear quantile regression models[3, 4], which provided additional support for the piecewise linear form of response change. As this approach does not estimate an optimal break point (knots), we performed a manual search across an interval of values (MAP = 400 – 700 by increments of 25) suggested by the bent-cable form to find near optimal breakpoints. Quadratic and cubic versions of b-spline quantile regression models also were examined with up to 7 equally spaced breakpoints in MAP. While these nonlinear models always fit slightly better than the linear b-splines, they confirmed that the major trends in the 0.90 – 0.99 quantiles of tree cover could be estimated with two linear pieces: a strong positive linear trend below MAP = 425 for 0.90 – 0.98 quantiles and below MAP = 625 for 0.99 quantile, and a weak postive to nonzero trend above MAP = 425 - 625. All of these piecewise linear splines had coefficients that differed from zero ($P < 0.005$). With the b-spline form of the piecewise linear model it was possible to add other variables to the model and test for their effect with quantile rankscore tests. Nonzero ($P < 0.06$) effects of fire-return intervals (0.18), %clay (-0.25), %sand (0.15), N mineralization potential (-0.12), and herbivore biomass (0.003) were present for lower to mid quantiles (0.10 – 0.60) of tree cover when these variables were added individually to the piecewise linear model of MAP. None of these variables contributed ($P > 0.10$) to the 0.90 – 0.99 quantiles.

Details of linear quantile regression analysis. For sites receiving less than 650 mm MAP, we used linear quantile regression analysis to further investigate how fire regimes, herbivory and soil properties influenced the upper bound on woody cover that was evident in these sites. In all cases, 0.90 to 0.99 conditional quantiles were estimated using the modified version of the Barrodale and Roberts algorithm[3], with confidence intervals estimated based on inversion of rank score tests[5]. Sites for which fire return periods were reported as >*x* years were excluded from the analysis. For soil %clay, separate analyses were carried out for sites where measurements were from the upper regions of the soil profile (typically top 10 – 20 cm, N=131) and those from the top 100 cm (N=103). Because results did not differ, data were pooled (N=234) for the graph in Figure 2.

Regression tree analysis. This technique creates binary trees by recursively partitioning data into two sets at each step based on an explanatory variable that minimizes a measure of variation within the resultant subgroups[6]. Regression trees are invariant to monotonic changes in explanatory variables, do not assume relations are linear or fixed across the entire dataset, are adept at capturing non-additive behavior, and are easy to interpret[6]. Because partitions near the top of the regression tree reflect strong relationships between predictor and response variables while branches near the bottom of the tree are less precise[6-7], trees were pruned back to a size that best represents relationships generalizable outside the sample to the rest of the continent[6-7]. Tree size for pruning was determined based on 500, 10-fold cross-validations[7]. For each cross-validation, data were divided into 10 groups of similar size, 9 of which were used to create the tree while the remainder was used to determine the predictive accuracy for all possible tree sizes. This procedure was repeated 10 times so that all groups were used as cross-validation groups and the entire process repeated 500 times. The tree size which produced the lowest mean predictive error was chosen as the optimal size.

Our data set included estimates of rainfall and fire frequency for 854 sites, whereas soil sand content and herbivore biomass estimates were available for only 175 and 183 sites, respectively. Although regression trees are capable of handling missing data among predictor variables[6-7], an analysis based on the complete dataset would nevertheless be unduly weighted toward the rainfall/ fire data. For this reason, we have presented in the main text results from a regression tree analysis based on a subset of 161 sites for which all data on rainfall, fire return periods, soil sand content and herbivore biomass were available. However, the general conclusion of these results in terms of the primary importance of rainfall and fire in determining woody cover is unchanged by the inclusion of additional sites with missing data (Table S2). Only rainfall and fire return periods were retained as significant predictor variables in the pruned tree for analyses based on either the complete dataset (N=854), sites for which at least 3 of the 4 predictor variables were available (N=197) or all sites for which data on sand content (N=175) were available. For the analysis based on all sites for which herbivore biomass data were available (N=183), results were identical to those presented in Figure 3.

The ‘rpart’ routine also provides information on surrogate splitting variables at each node, i.e. variables that best agree with the original splitting variable and which can be used to assign cases to different branches when data on the original variable are missing[6-7]. For the tree shown in Fig. 4, sand content > 90.3% was identified as the surrogate criterion for the first split, reflecting the fact that drier sites in the dataset for the most part also tended to have high sand content. However, because some high MAP sites (>356 mm) were also characterized by high sand content (as reflected in the third split in Fig 3), there was a 13% disagreement between the surrogate and original criterion in site allocation to nodes. For split 2, sand content was again chosen as the surrogate variable, with sites having >13.9% sand assigned to the left branch (fire return < 10.5 yrs). In this case, the disagreement between the surrogate and original variables was 15%. No surrogates were identified for the third split.

Supplemental references

1. Keeney, D. R. in *Methods of Soil Analysis* (eds Page, A.L., Miller, R.H. & Keeney, D.R.) 711 – 733 (Amer. Soc. Agron., Madison, WI, 1982).
2. Bouyoucos, G. J. A recalibration of the hydrometer method for making mechanical analysis of soil. *Agron. Jour.* **43**, 434 - 438 (1951).
3. Koenker, R. W. & d’Orey, V. Computing regression quantiles. *Applied Statistics* **36**, 383 – 393 (1987).

4. Koenker, R., Ng, P. & Portnoy, S. Quantile smoothing splines. *Biometrka* **81,** 673-680 (1994).
5. Koenker, R. W. in *Asymptotic Statistics* (eds Mandl, P. & Huskova, M.) 349 – 359 (Springer-Verlag, New York, 1994).
6. Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. G. *Classification and Regression Trees* (Wadsworth International Group, Belmont, 1984).
7. De'ath, G. & Fabricus, K. E. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* **81**, 3178 – 3192 (2000).

Supplemental references for data sources

1. Abbadie, L., Gignoux, J., Le Roux, X. & Lepage, M. *Lamto: Structure, Functioning and Dynamics of a Savanna Ecosystem* (Springer Verlag, Ecological Studies 179, New York, 2005).
2. Aranibar, J. N. *et al*. Nitrogen cycling in the soil-plant system along a precipitation gradient in the Kalahari sands. *Global Change Biology* **10**, 359 – 373 (2004).
3. Caylor, K. K., Shugart, H. H., Dowty, P. R. & Smith, T. M. Tree spacing along the Kalahari transect in southern Africa. *Journal of Arid Environments* **54**, 281 – 296.
4. Diallo, O., Diouf, A., Hanan, N. P., Ndiaye, A. & Prevost, Y. AVHRR monitoring of savanna primary production in Senegal, West Africa: 1987-1988. *Int. J. Remote Sensing* **12**, 1259 – 1279 (1991).
5. Feral, C. J. W. *et al.* Carbon and nitrogen in the soil-plant system along rainfall and land-use gradients in southern Africa. *J. Arid Environ.* **54**, 327 – 343 (2003).
6. Isichei, A. O. & Muoghalu, J. I. The effects of tree canopy cover on soil fertility in a Nigerian savanna. *J. Trop. Ecol.* **8**, 329 – 338 (1992).
7. Patten, R. Pattern and process in an arid tropical ecosystem: The landscape and system properties of Ngisonyoka Turkana, North Kenya (PhD thesis, Colorado State University, 1991).
8. Scholes, R. J. & Walker, B. H. *An African savanna: Synthesis of the Nylsvley Study* (Cambridge University Press, Cambridge, 1993).
9. Scholes, R. J. *et al*. Trends in savanna structure and composition along an aridity gradient in the Kalahari. *J. Veg. Sci.* **13**, 419 – 428 (2002).

Supplementary tables

Table S1. Relationships between measures of soil properties for sites included in the regression tree analysis. Values are correlation coefficients. All correlations were significant at the α = 0.05 level.

	%Clay	%Sand	%N	Total P	N mineralization potential
%Clay	1.0				
%Sand	-0.93	1.0			
%N	0.72	-0.73	1.0		
Total P	0.53	-0.59	0.54	1.0	
N mineralization potential	0.4	-0.38	0.59	0.18	1.0

Table S2. Results of regression tree analysis for different subsets of the dataset.

Criterion for site inclusion	N	Variables retained in final tree	Tree size	R^2
All sites	854	MAP, fire return period	8 nodes	41.0%
At least 3 of 4 predictors	197	MAP, fire return period	4 nodes	45.9%
Sand content data available	175	MAP, fire return period	4 nodes	45.6%
Herbivore biomass data available	183	MAP, fire return period, sand content	4 nodes	45.2%

Supplementary Figure S1

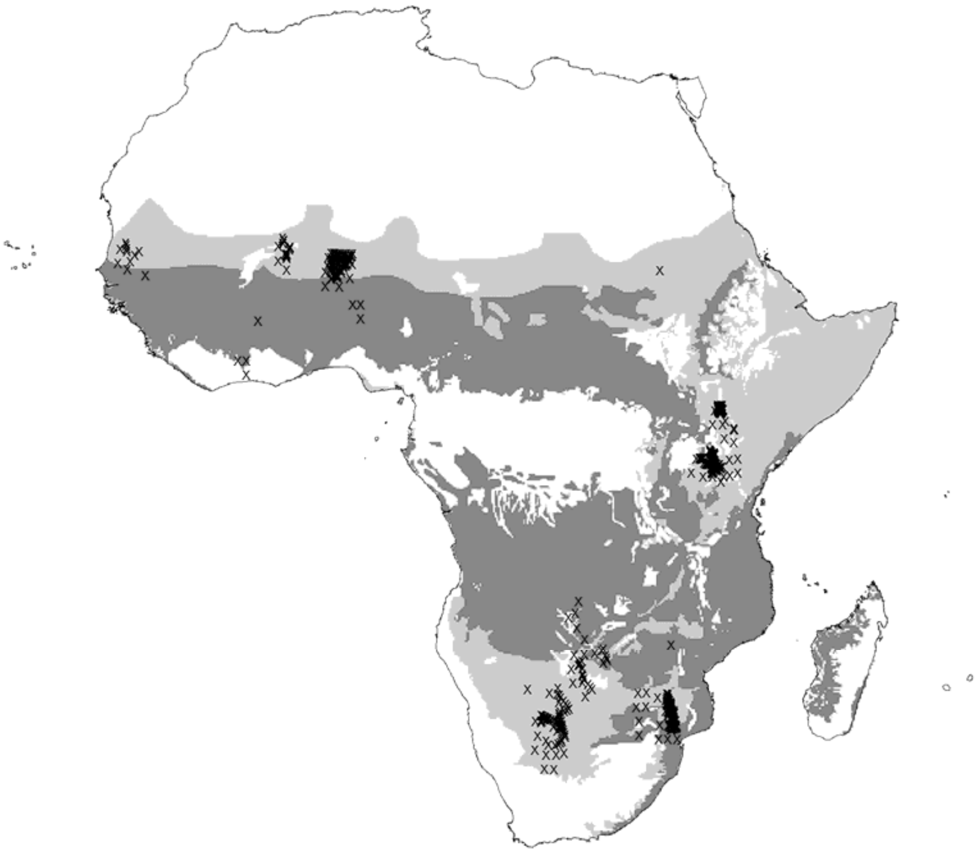

Figure S1. Map of Africa showing the location of sample sites (indicated by crosses) used for the analysis. Light gray areas represent fine-leaved, nutrient-rich savannas; darker areas are broad-leaved, nutrient-poor savannas. The underlying map was derived from White's vegetation map of Africa (ref. 30) by reclassifying woodland and wooded grassland map units into one or other of the two savanna classes according to the dominant tree species. For display purposes, sites geographically close to one another have been jittered to facilitate discrimination of points.

Supplementary Figure S2

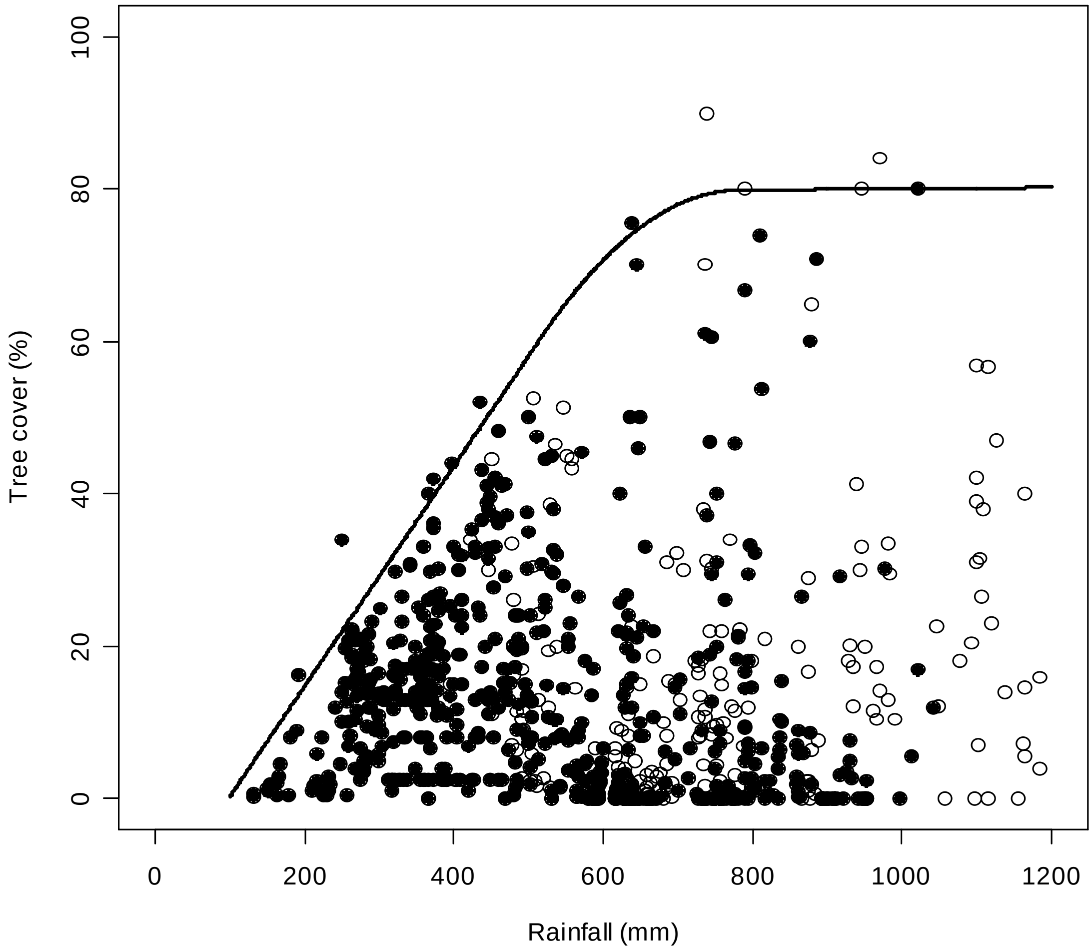

Figure S2. Woody cover as a function of MAP as related to the fire regimes of the different sites. Filled circles represent savannas with low fire frequencies (average fire return times of > 3 yrs), and open circles represent sites with high fire frequencies (fire return intervals of <= 3 years). Fires typically tend to be more frequent in sites with greater rainfall. For sites receiving <650mm MAP, the upper bound on tree cover exists despite low fire frequencies (some of the sites had not burned for > 50 years). Data are from 854 savanna sites across Africa.

Supplementary Figure S3

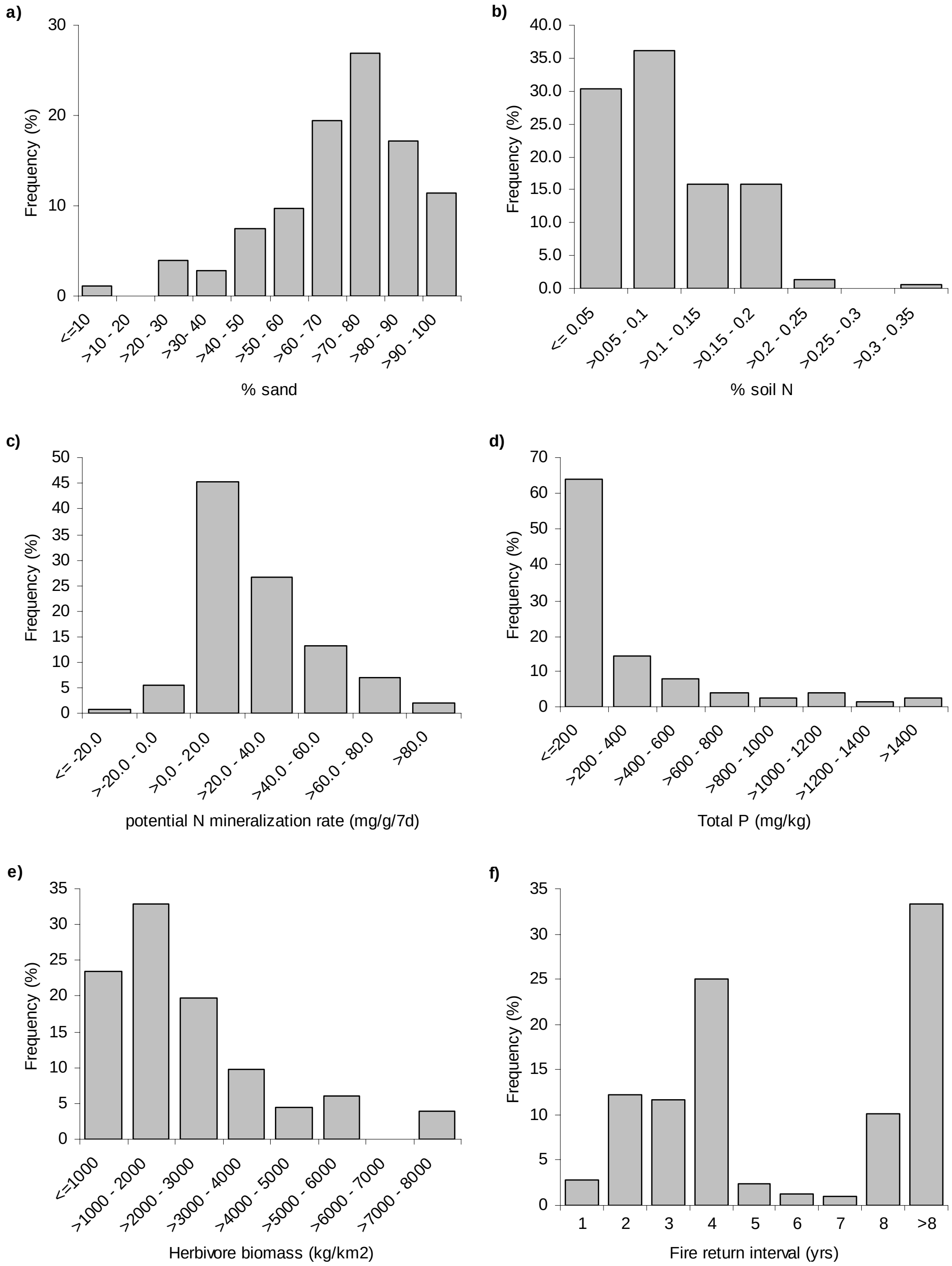

Figure S3. Frequency distributions for a) soil %sand (N = 175), b) soil %N (N = 162), c) N mineralization potential (μg.g soil^{-1}.7d^{-1}; N = 146), d) soil total P (mg.kg^{-1}; N = 155), e) herbivore biomass density (kg.km^{-2}; N = 183), and f) fire return periods (yrs; N = 854) across study sites.

Supplementary Figure S4

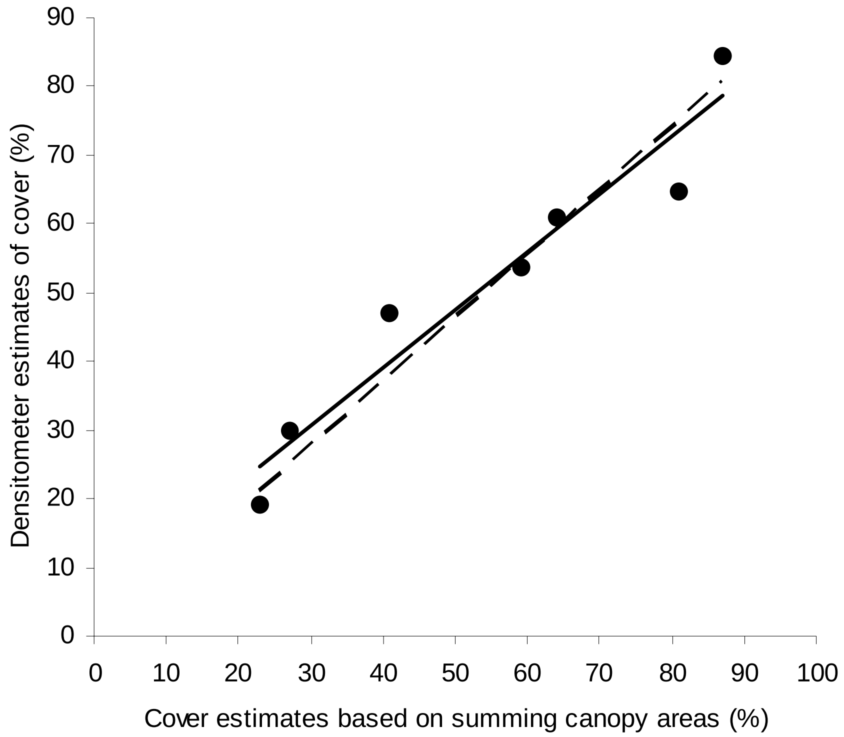

Figure S4. Comparison of woody cover estimates based on densiometer and summed canopy area methods. Data are from 7 sites where both methods were used to estimate woody cover and span a rainfall range from 409 – 970mm MAP. Of the 854 sites in the complete dataset, the densitometer method was used at 397 sites, while canopy areas were measured on belt transects or plots at 308 sites. The solid line represents a freely fitted regression ($y = 0.84x + 5.5$; $R^2 = 0.93$), while the dotted-line is for one that is constrained through the origin ($y = 0.93x$, $R^2 = 0.92$). The intercept for the freely fitted relationship was not different from zero ($P = 0.4$). In neither case was the slope of the regression line significantly different from 1.0 at the α=0.05 level.