

Supplementary Information

Empirical Analysis of Emergent Spatial Properties of Urban Landscapes

Methods

We tested the hypothesis that urban landscapes exhibit unique landscape patterns by quantifying landscape heterogeneity and connectivity across urban and non-urban areas using two metrics to characterize landscape structures. Landscape heterogeneity was measured using the Shannon diversity index (SHDI) of land cover, while connectivity was assessed using the aggregation index (AI) for both natural forested areas and urban development. We then contrasted the distributions of these landscape metrics across urban and non-urban landscapes within selected drainage basins intersecting 100 US metropolitan areas (refer to the sampling method). We compared urban and non-urban landscape metrics using Bayesian generalized linear mixed models with urban status as a fixed effect and metropolitan region as a random effect.

Sampling Design

This section describes the methods for the empirical analysis of the emergent network properties of urban landscapes. To quantify urban landscape structure, we utilized a hierarchical, stratified random sampling design spanning the drainage basin, metropolitan, and 1km-by-1km cell scales using high-resolution (30m) data. Following the methodology of Alberti and Wang (2022), we selected 100 metropolitan areas from a pool of 933 within the contiguous United States. Our selection strategy aimed for a spatially balanced sample, factoring in population, ecological, and socio-economic attributes of each metropolitan area. Each of the four census regions—Northeast, South, West, and Midwest—had their metropolitan areas ranked by population and stratified into 25 bins. We randomly selected a metropolitan area from each bin to represent its population tier, thus creating an initial sample set of 100 metropolitan areas (25 cities per region; refer to Figure 2). We then evaluated our sample set to ensure coverage across all ten Level I ecological regions of the contiguous United States (Omernik & Griffith, 2014). In cases where an eco-region was not represented in our initial sample set, we included a randomly selected metropolitan area from the underrepresented eco-region while removing one from the corresponding population stratum and census region. Our final metropolitan area selection thus represents all ecological regions and population tiers, evenly distributed among the four census regions.

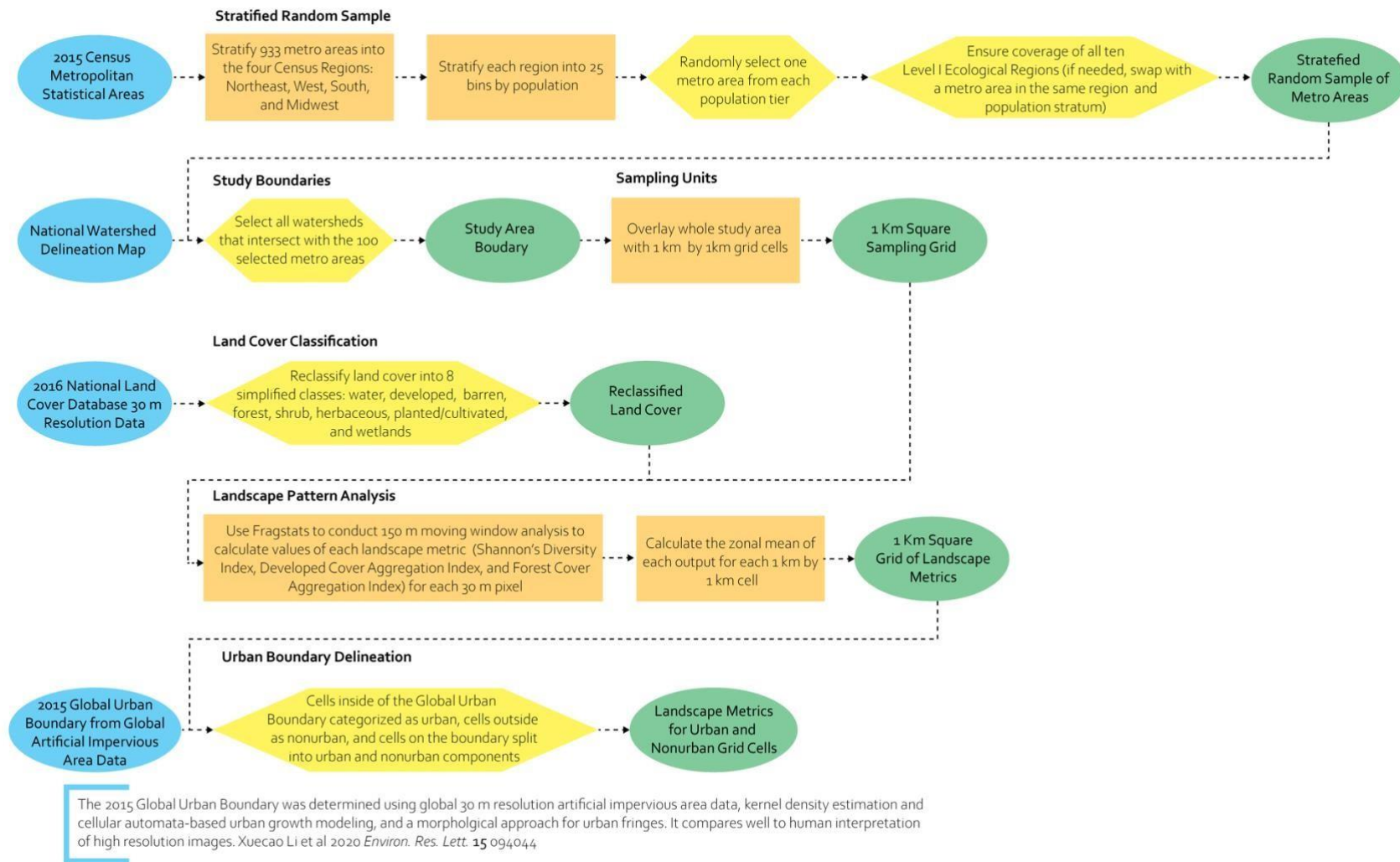


Figure 1. Flowchart of sampling design and boundary delineation.

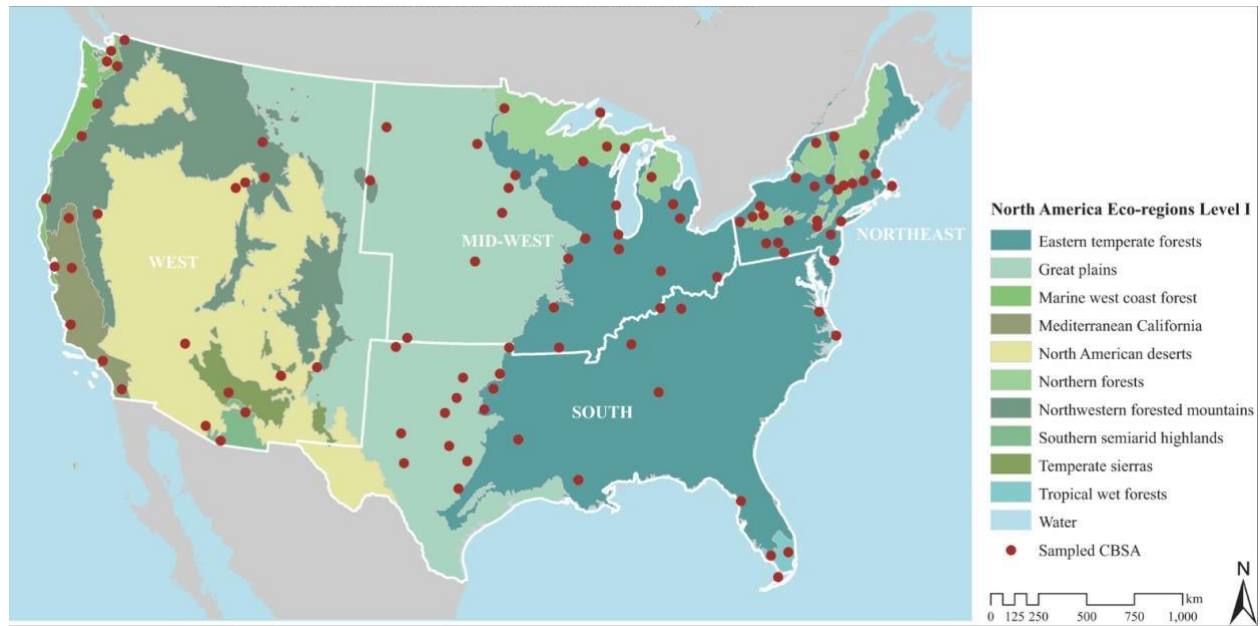


Figure 2. 100 metropolitan areas sampled by census and ecological regions, from Alberti and Wang (2022).

Urban Boundary Delineation

The study boundary was set by intersecting all 100 metropolitan areas with a national watershed delineation map. This created a study area that included all watersheds containing any portion of the 100 selected metro areas. We delineated the urban boundary according to the method described in Li et al. (2020). For the land cover data, we used the 2016 National Land Cover Database of the U.S. Geological Survey (NLCD) 30m resolution data and clipped it to the study boundaries (Dewitz and U.S.G.S. 2021). We reclassified the land cover data as in Alberti and Wang (2022).

Table 1. New land cover classifications based on original NLCD land cover classifications.

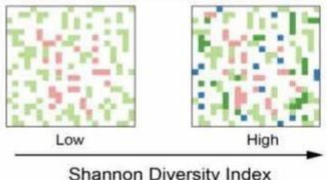
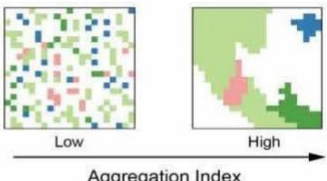
New Classification	Original Classification
Water	Open Water Perennial Ice/Snow
Developed	Low Intensity Medium Intensity High Intensity Open Space
Barren	Barren Land
Forest	Deciduous Forest Evergreen Forest Mixed Forest
Shrub	Shrub, Scrub
Herbaceous	Grassland/Herbaceous

Planted/Cultivated	Pasture/Hay Cultivated Crops
Wetlands	Woody Wetlands Emergent Herbaceous Wetlands

Landscape Metrics

We measure landscape heterogeneity using the Shannon diversity index (SHDI). Additionally, we measured both natural habitat connectivity (Cushman 2006; Shochat et al. 2006) and compactness of the urban structure using aggregation index (AI) of forest cover and developed land (Alberti 2008). Fragstats (4.2) was used to calculate Shannon's Diversity Index, Aggregation Index for forest cover, and Aggregation Index for developed cover using the same methods as in Alberti and Wang (2022): a 150m moving window was used to calculate the landscape metric values for each 30m pixel. Zonal means for each landscape metric were calculated for 1km x 1km grid cells (each associated with a unique cell ID), using the same methods as in Alberti and Wang (2022).

Table 2. Descriptions and equations for landscape metrics.

Dimensions	Metrics	Equation	Description
Heterogeneity	Landscape Shannon Diversity Index (SHDI)	$SHDI = - \sum_{i=1}^m (P_i * \ln P_i)$ $P_i = \text{percentage of the land occupied by class } i$	 Low High Shannon Diversity Index
Connectivity	Aggregation Index (AI) of Development and AI of Forest	$AI = \frac{g_{ii}}{\max \rightarrow g_{ii}}$ $g_{ii} = \text{number of like adjacencies between grid cells of class } i$ $\max \rightarrow g_{ii} = \text{max number of like adjacencies between grid cells of class } i$	 Low High Aggregation Index

The 1km x 1km grid cells were overlaid with the urban boundaries to identify each cell as urban or nonurban. Cells that were on the border were split into two features—one urban and one nonurban—that still shared the same cell ID, but with different areas, to allow for weighting. The landscape metrics were matched to their location (as urban or nonurban) and cell area (1 square km for most cells, but less for cells split by the urban boundary) by cell ID.

The distributions of the landscape metrics (Figure 3; Table 3) exhibited distinct distributional characteristics, including zero inflation for SHDI and bounded non-normal distributions for the aggregation indices. These distributional properties informed the selection of the Bayesian generalized linear mixed models used in the analysis.

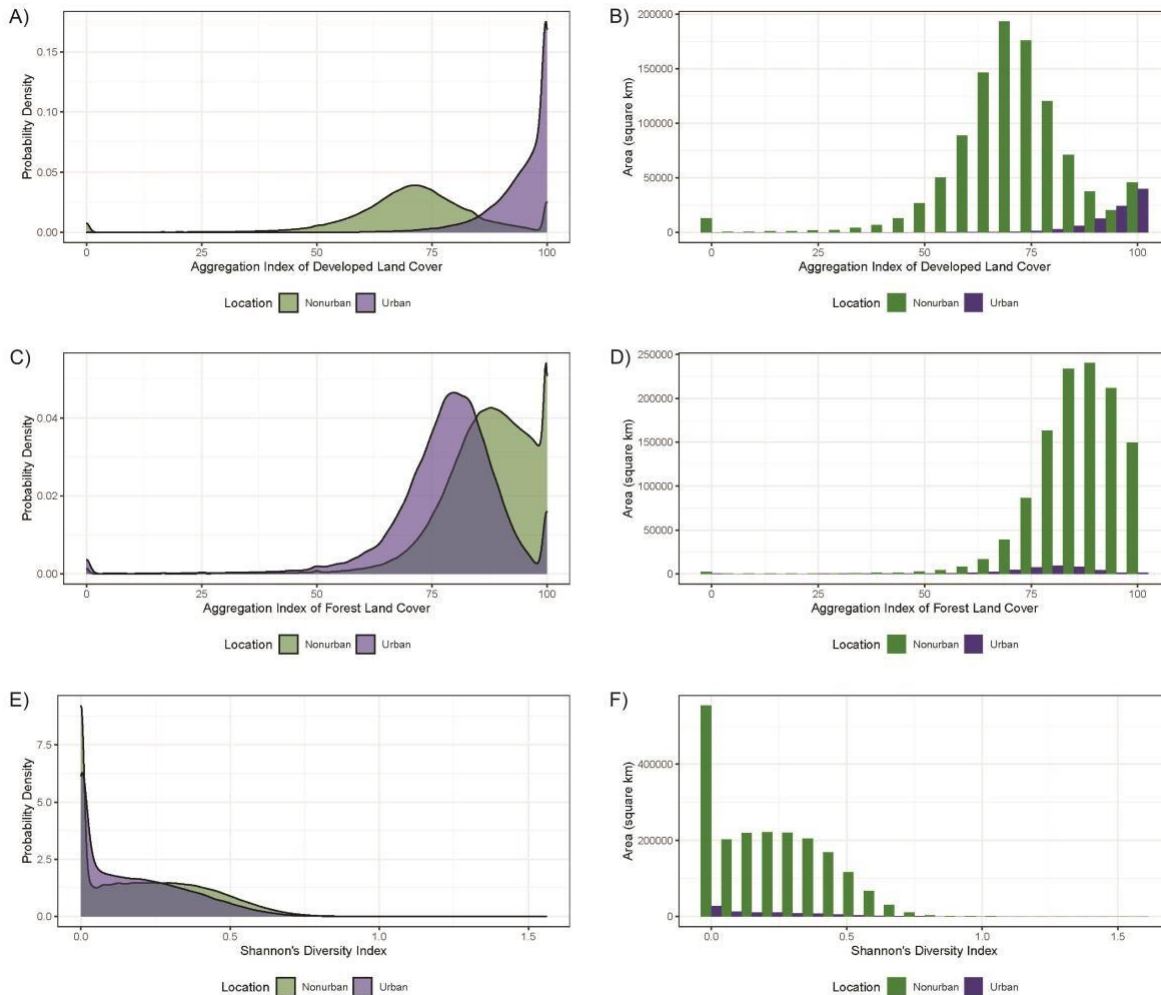


Figure 3. Distributions of landscape metrics in urban and non-urban areas.

Table 3. Descriptive statistics for landscape metrics in urban and non-urban areas.

	Shannon's Diversity Index		Forest Aggregation Index		Developed Aggregation Index	
	Urban	Nonurban	Urban	Nonurban	Urban	Nonurban
Mean	0.244	0.225	78.8	86.7	91.5	71.3
Median	0.228	0.201	80.1	87.8	93.8	71.6
Variance	0.0371	0.0392	140.	105	75.9	229

Standard Deviation	0.192	0.198	11.8	10.2	8.71	15.1
Sample Size	114,381	1,807,171	66,503	1,060,419	113,925	946,579
Weighted Mean*	0.187	0.223	78.1	86.8	94.6	70.9
Weighted Variance*	0.0300	0.0385	164	106	46.3	229
Weighted Standard Deviation*	0.173	0.196	12.8	10.3	6.80	15.1

*Weighted by grid cell area

Bayesian Analysis

The distributions of Shannon’s Diversity Index (SHDI), Agglomeration Index (AI) for forest cover, and AI for developed cover were compared between urban and non-urban areas using Bayesian statistics. We used generalized linear mixed models with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region. Appropriate models for each comparison were chosen based on the shapes and characteristics of the distributions. Models were run using the R packages *brms* and *cmdstanr* (Bürkner 2017; Gabry et al. 2023). Each model was run with 2,000 iterations, including 1,000 warmup.

Shannon’s Diversity Index:

The distributions of SHDI had a high number of zeroes, so they were modeled by a gamma hurdle model. The shape of the gamma distribution depends on two parameters—the shape parameter k and the scale parameter θ —and the hurdle parameter θ_2 represents the probability that a value is zero. The model summary is shown in Table 4, posterior summaries are shown in Table 5, posterior distributions and trace plots are shown in Figure 4, and posterior predictive checks are shown in Figure 5.

Table 4. Estimates and diagnostics for fixed effects and family-specific parameters for the gamma hurdle generalized linear mixed model of Shannon’s Diversity Index with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region.

Parameters		Estimate	95% CI	R-hat	Bulk ESS	Tail ESS
Population-Level	Intercept	-1.27	(-1.32, -1.21)	1.01	142	321
	Slope (Urban)	-0.06	(-0.07, -0.06)	1.00	3,553	3,375
	Shape	1.56	(1.56, 1.57)	1.00	10,432	9,223

Family-Specific	Hu	0.19	(0.19, 0.19)	1.00	4,941	7,195
-----------------	----	------	--------------	------	-------	-------

Table 5. Posterior summaries for fixed effects and family-specific parameters for the gamma hurdle generalized linear mixed model of Shannon's Diversity Index with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region.

Parameters		Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Population-Level	Intercept	-1.36	-1.29	-1.27	-1.27	-1.25	-1.18
	Slope (Urban)	-0.072	-0.065	-0.063	-0.063	-0.061	-0.052
Family-Specific	Shape	1.56	1.56	1.56	1.56	1.57	1.57
	Hu	0.193	0.194	0.194	0.194	0.194	0.195

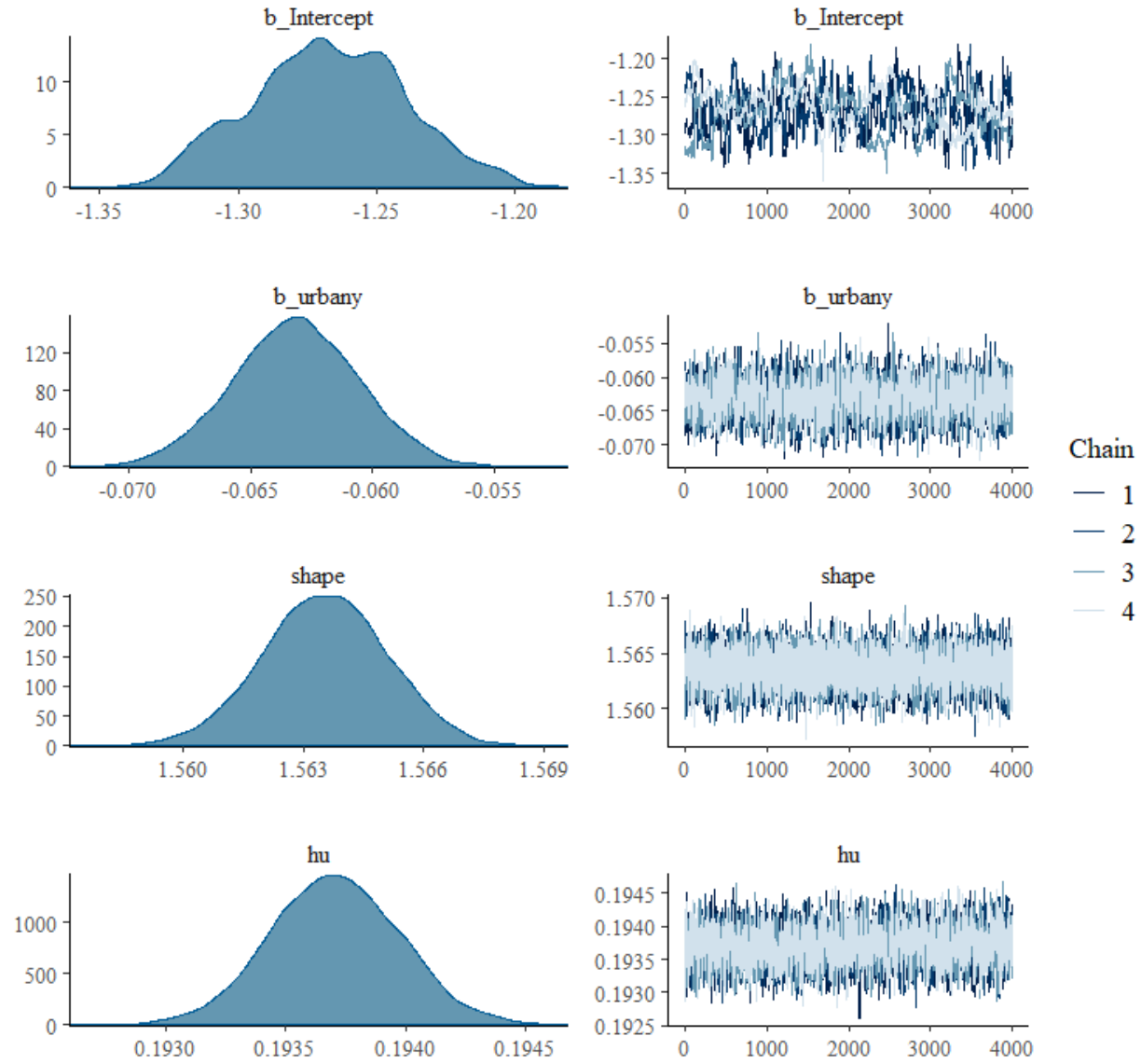


Figure 4. Posterior distributions and trace plots for fixed effects and family-specific parameters for the gamma hurdle generalized linear mixed model of Shannon's Diversity Index with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region.

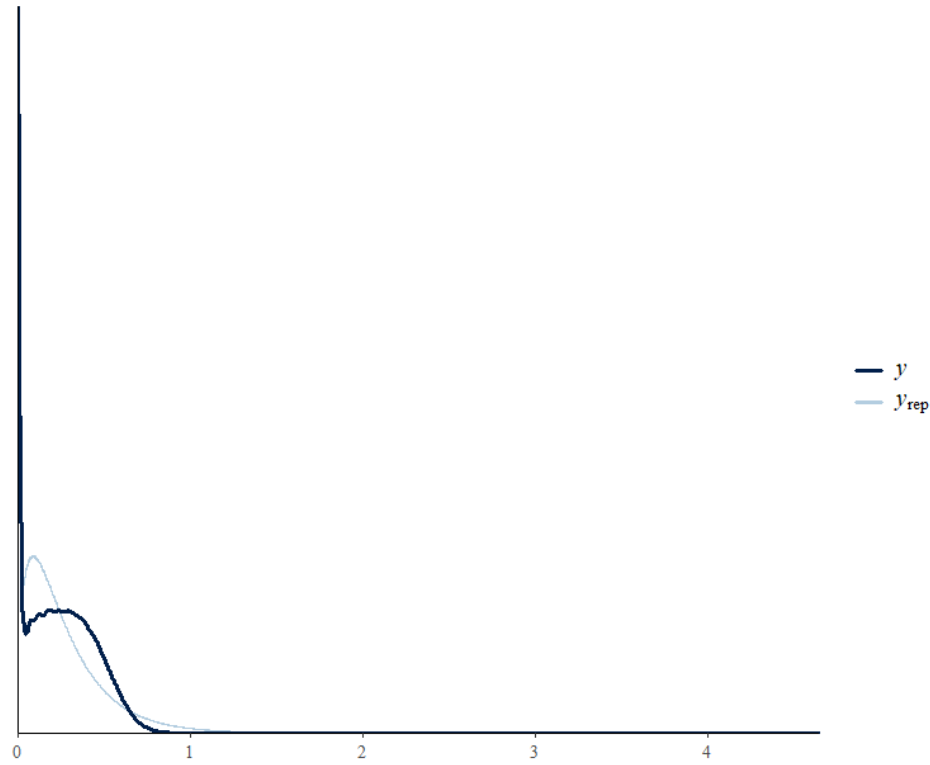


Figure 5. Density overlay plot comparing 100 draws from the posterior distribution of the gamma hurdle generalized linear mixed model of Shannon's Diversity Index (y_{rep}) to the input SHDI data (y).

Finally, we used the model to produce predictions for values of SHDI in urban and nonurban cells.

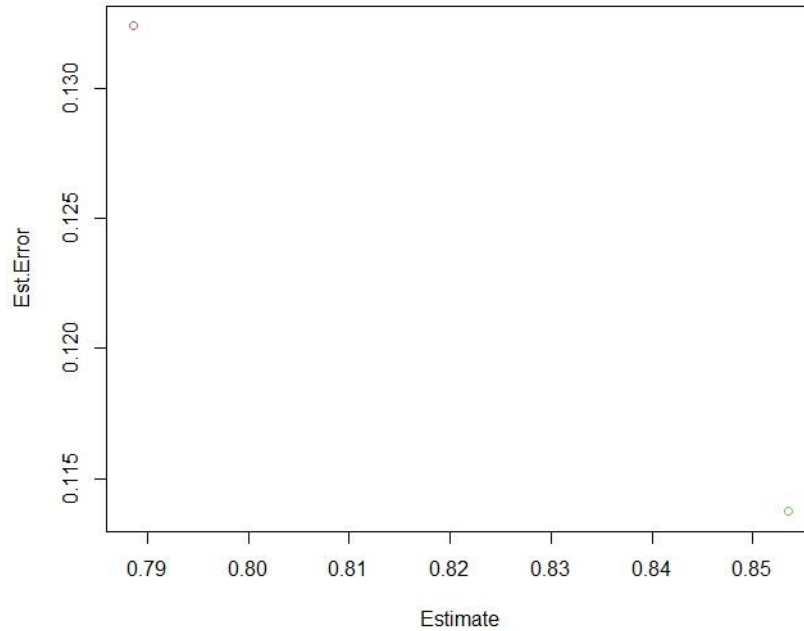


Figure 6. Predictions for values of SHDI in urban and nonurban cells. The urban prediction is represented in red and the nonurban prediction is represented in green.

Forest Cover Aggregation Index:

The distributions of Forest Cover Aggregation Index were converted to a 0-1 scale modeled using the beta distribution.

The beta distribution is determined by two parameters, *Shape 1* and *Shape 2*. In R, these are treated as μ and ϕ as follows:

$$\text{Shape 1} : a = \mu\phi$$

$$\text{Shape 2} : b = (1 - \mu)\phi$$

The model summary is shown in Table 6, posterior summaries are shown in Table 7, posterior distributions and trace plots are shown in Figure 7, and posterior predictive checks are shown in Figure 8.

Table 6. Estimates and diagnostics for fixed effects and family-specific parameters for the beta generalized linear mixed model of Forest Cover Aggregation Index with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region.

Parameters		Estimate	95% CI	R-hat	Bulk ESS	Tail ESS
Population-Level	Intercept	1.83	(1.77, 1.89)	1.03	103	182
	Slope (Urban)	-0.40	(-0.41, -0.39)	1.00	1,636	1,903
Family-Specific	Phi	6.49	(6.47, 6.50)	1.00	3,582	2,710

Table 7. Posterior summaries for fixed effects and family-specific parameters for the beta generalized linear mixed model of Forest Cover Aggregation Index with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region.

Parameters		Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Population-Level	Intercept	1.72	1.81	1.83	1.83	1.85	1.92
	Slope (Urban)	-0.412	-0.401	-0.399	-0.399	-0.397	-0.386
Family-Specific	Phi	6.46	6.48	6.49	6.49	6.49	6.52

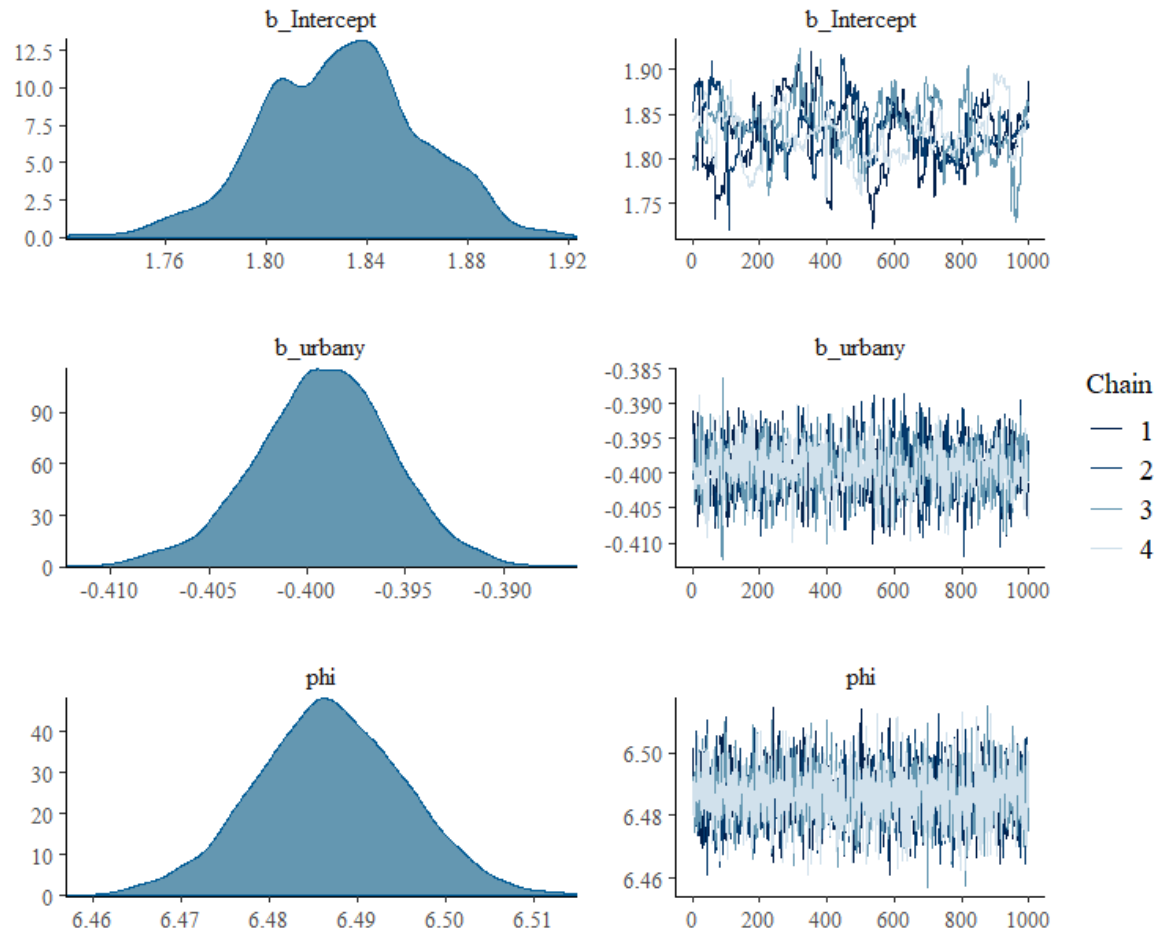


Figure 7. Posterior distributions and trace plots for fixed effects and family-specific parameters for the beta generalized linear mixed model of Forest Cover Aggregation Index with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region.

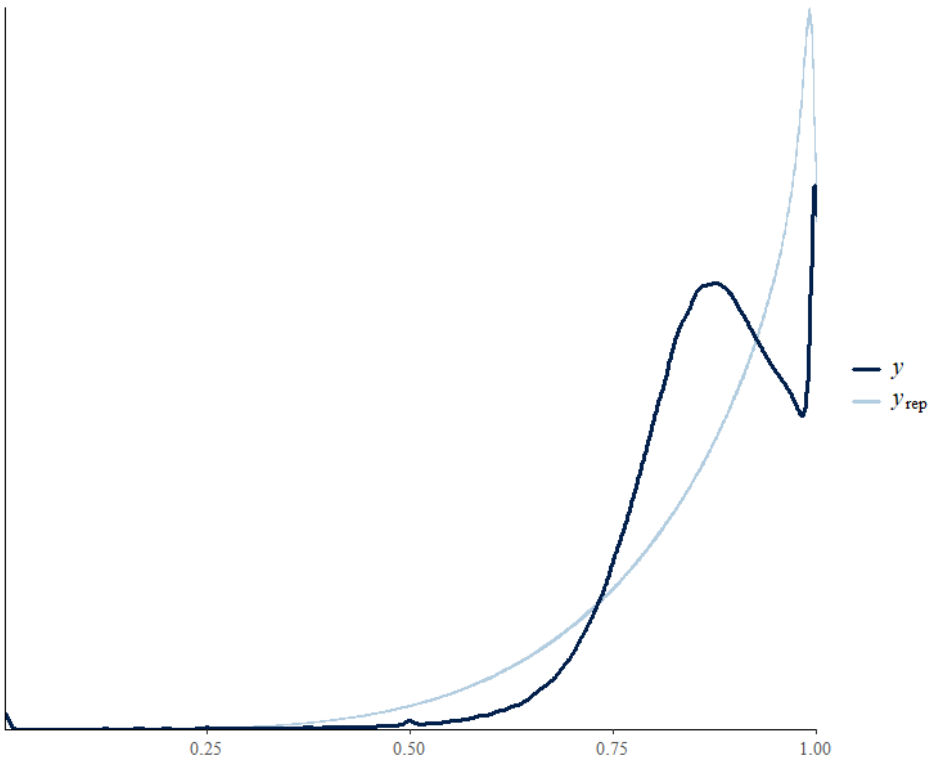


Figure 8. Density overlay plot comparing 100 draws from the posterior distribution of the beta generalized linear mixed model of Forest Cover Aggregation Index (y_{rep}) to the input Forest AI data (y).

Finally, we used the model to produce predictions for scaled values of Forest AI in urban and nonurban cells.

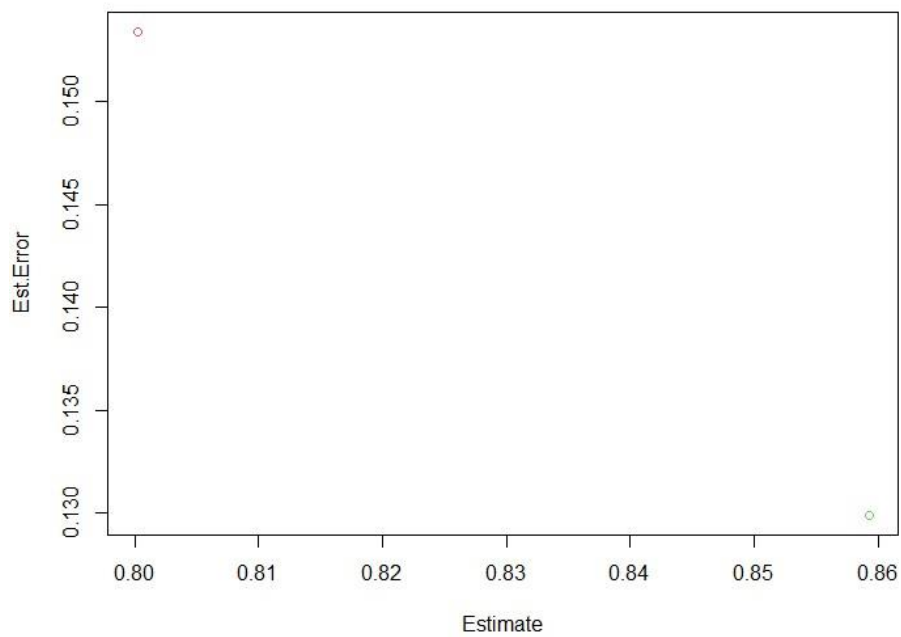


Figure 9. Predictions for scale values of Forest Aggregation Index (Forest AI / 100) in urban and nonurban cells. The urban prediction is represented in red and the nonurban prediction is represented in green.

Developed Cover Aggregation Index:

Developed Cover Aggregation Index was modeled using a beta model with the same methods as used for the Forest Cover Aggregation Index.

Table 8. Estimates and diagnostics for fixed effects and family-specific parameters for the beta generalized linear mixed model of Developed Cover Aggregation Index with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region.

Parameters		Estimate	95% CI	R-hat	Bulk ESS	Tail ESS
Population-Level	Intercept	0.59	(0.57, 0.62)	1.03	60	113
	Slope (Urban)	0.53	(0.52, 0.54)	1.00	5,311	2,920
Family-Specific	Phi	2.64	(2.64, 2.65)	1.00	7,132	2,776

Table 9. Posterior summaries for fixed effects and family-specific parameters for the beta generalized linear mixed model of Developed Cover Aggregation Index with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region.

Parameters		Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Population-Level	Intercept	0.554	0.584	0.594	0.593	0.602	0.631
	Slope (Urban)	0.518	0.528	0.530	0.530	0.532	0.542
Family-Specific	Phi	2.63	2.64	2.65	2.65	2.65	2.66

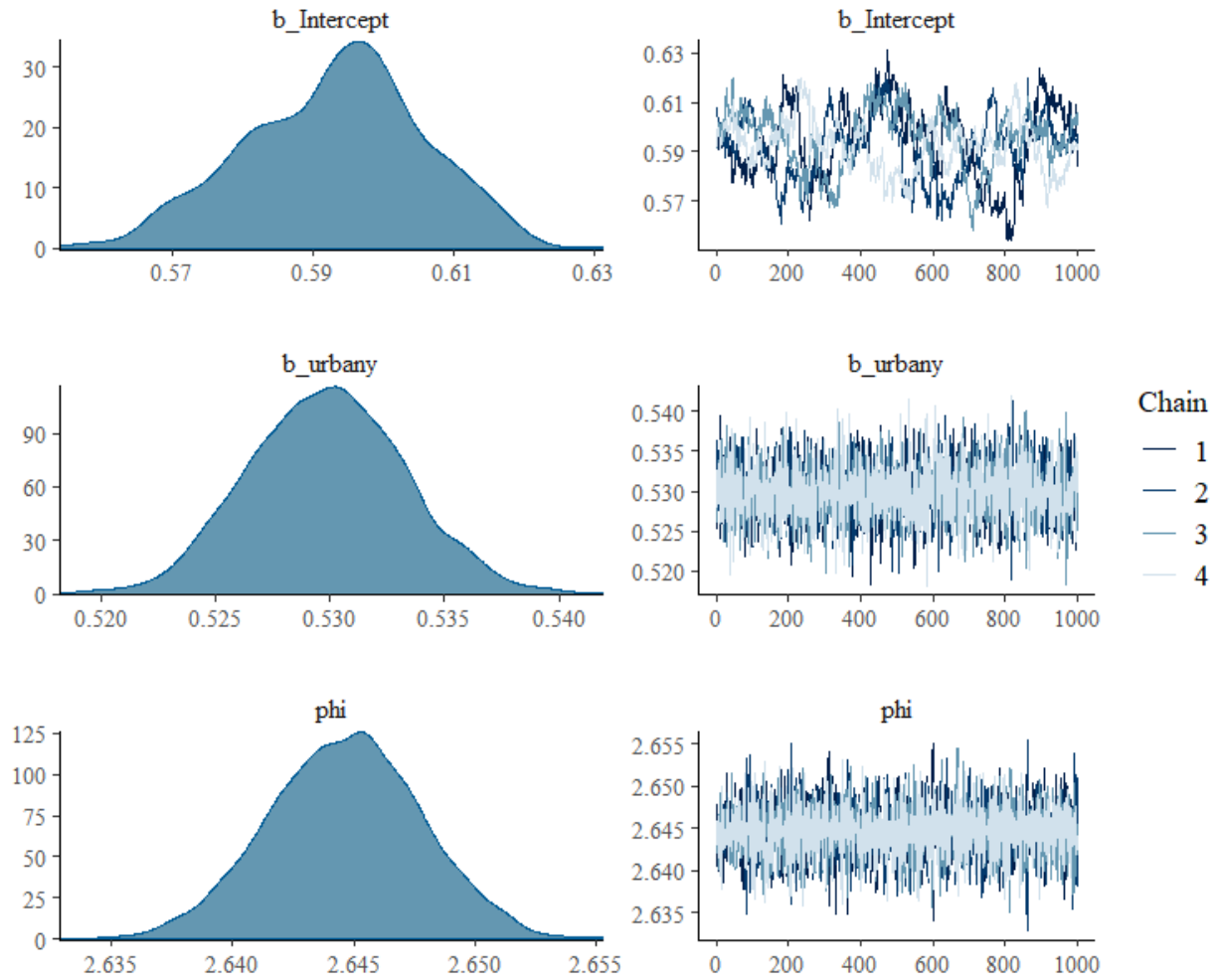


Figure 10. Posterior distributions and trace plots for fixed effects and family-specific parameters for the beta generalized linear mixed model of Developed Cover Aggregation Index with fixed effects for whether the cells were urban or nonurban and random effects for each metropolitan region.

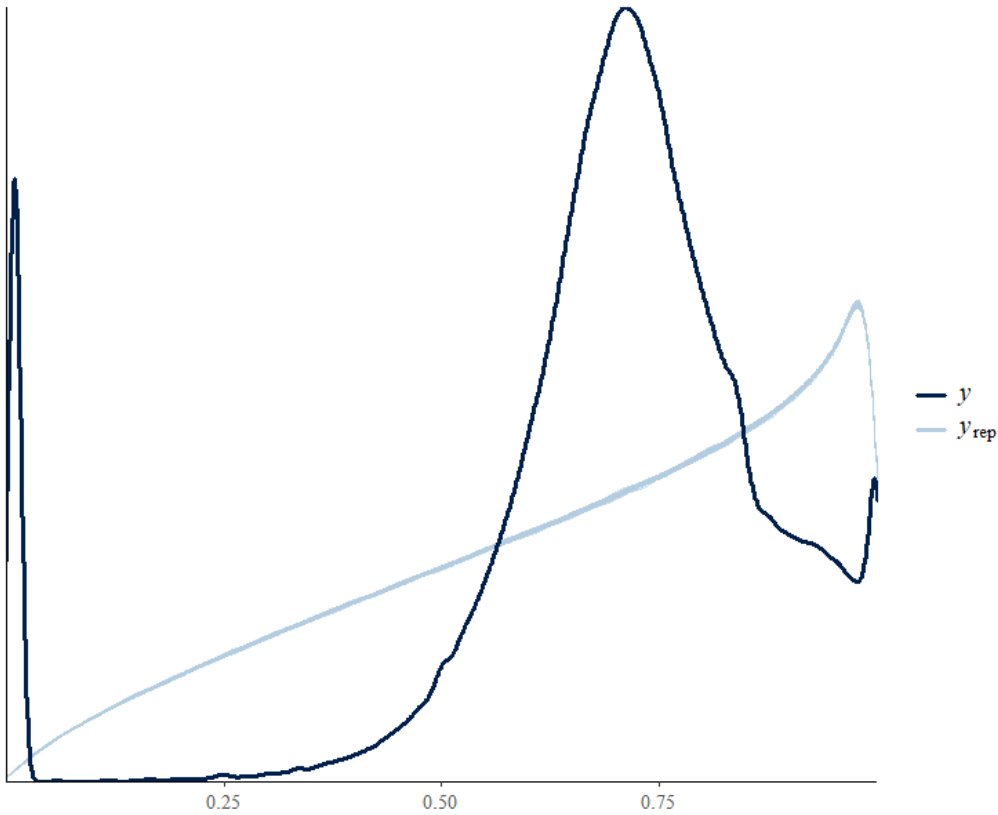


Figure 11. Density overlay plot comparing 100 draws from the posterior distribution of the beta generalized linear mixed model of Developed Cover Aggregation Index (y_{rep}) to the input Developed Cover Aggregation Index data (y).

Finally, we used the model to produce predictions for scaled values of Developed AI in urban and nonurban cells.

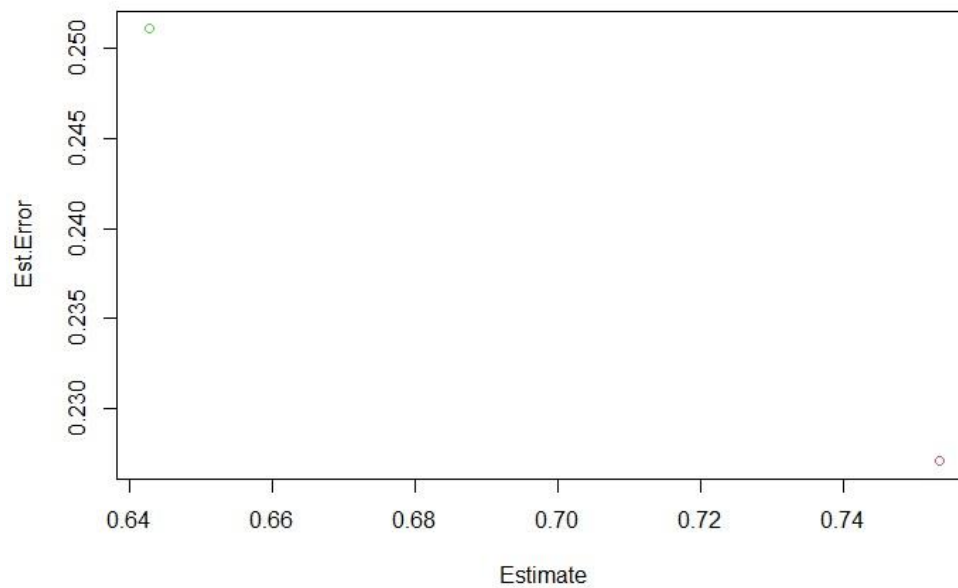


Figure 12. Predictions for scale values of Developed Aggregation Index (Developed AI / 100) in urban and nonurban cells. The urban prediction is represented in red and the nonurban prediction is represented in green.

Moran's I of Bayesian Residuals

We used Moran's I to test for spatial autocorrelation among model residuals. Residuals for each of the three Bayesian models were estimated based on 200 draws from the posterior distribution, and Moran's I was calculated across each full set of residuals using the package 'moranfast' (Cooper 2020). This package calculates the index based on inverse distance weights, which were calculated based on the Euclidean distances between the centroids of the grid cells. Results, shown in **Table 10**, indicate low levels of residual spatial autocorrelation in all three models.

Table 10. Moran's I of the residuals of all three Bayesian models of landscape metrics, estimated based on 200 draws from their posterior distributions.

Model	Observed	Expected	Standard Deviation	P Value
Shannon's Diversity Index – Gamma Hurdle	0.0771	-0.000000473	0.00000278	<0.0000001
Forest Cover Aggregation Index – Beta	0.0250	-0.000000791	0.00000504	<0.0000001
Developed Cover Aggregation Index – Beta	0.00845	-0.000000845	0.00000535	<0.0000001

References

- Alberti, M. (2008). *Advances in urban ecology: integrating humans and ecological processes in urban ecosystems*. Boston, MA: Springer.
- Alberti, M., and Wang, T. (2022). Detecting patterns of vertebrate biodiversity across the multidimensional urban landscape. *Ecology Letters*, 25(4), 1027-1045.
- Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Cooper, M. (2020). *moranfast: Conduct a Quick and Memory-Efficient Moran's I Test*. <https://github.com/mcooper/moranfast>.
- Cushman, S.A. (2006). Effects of habitat loss and fragmentation on amphibians: a review and prospectus. *Biological Conservation*, 128, 231–240.
- Dewitz, J., and U.S. Geological Survey. (2021). National Land Cover Database (NLCD) 2019 Products (ver. 3.0, February 2024): U.S. Geological Survey data release, <https://doi.org/10.5066/P9KZCM54>.
- Gabry, J., Češnovar, R., Johnson, A. (2023). *cmdstanr: R Interface to 'CmdStan'*. <https://mcstan.org/cmdstanr/>, <https://discourse.mc-stan.org>.
- Li, X., Gong, P., Zhou, Y., Wang, J., Bai, Y., Chen, B., Hu, T., Xiao, Y., Xu, B., & Yang, J. (2020). Mapping global urban boundaries from the global artificial impervious area (GAIA) data. *Environmental Research Letters*, 15, 094044.
- Omernik, J. M., and Griffith, G. E. (2014). Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environ Manage.*, 54(6), 1249-66. doi: 10.1007/s00267-014-0364-1.
- Shochat, E., Lerman, S.B., Anderies, J.M., Warren, P.S., Faeth, S.H. & Nilon, C.H. (2010). Invasion, competition, and biodiversity loss in urban ecosystems. *BioScience*, 60, 199–208.
- Zar, J. H. (2010). *Biostatistical Analysis*. 5th edn. Upper Saddle River: Prentice Hall.