

Appendix 4. Methods and results for city-specific urbanization gradient analysis.

Methods

In order to test whether city size impacted the urbanization gradient within a city, we used each eBird checklist as a data point and fitted individual Generalized Additive Models within each city, where urbanness was a smooth term. There were a potential of 1,581 models for each response variable. Models which failed to converge --- for a variety of reasons --- were excluded from further analyses. This left us with 1,166 cities for species richness, and 1,133 cities for abundance, Shannon diversity and phylogenetic diversity. For each model that converged, we extracted the smoothed term for urbanness (e.g., Figure A1 below). But because some cities are inherently more urban than others the level of urbanness was scaled and centered, as was the smoothed biodiversity response, thus making cities comparable in relation to biodiversity response to an urbanization gradient within that city. We then calculated the average slope of the smooth term by calculating the change in y (i.e., biodiversity response) divided by the change in x (i.e., urbanization level). This approach, however, left us with a bimodal distribution in slopes (Figure A2 below) because for many GAMs, the smooth term was best explained by a linear predictor with either a slope of positive 1 or negative 1. We then modelled these slopes against city size and our previously mentioned predictor variables, using a variety of exploratory data analysis approaches, reliant on GAMs. First, we modelled only the 1 and -1 values as a binomial model, and modelled the other data which were normally distributed after removing the 1 and -1 values with a normal gaussian distribution. Second, we modelled the relationship using a binomial model where everything > 0.5 received a 1 and everything < -0.5 received a 0. We found high correlation between city area and population (Figure A3 below) so we excluded population, but included population density of a city in these models (Figure A4 below). Lastly, we investigated the correlation of these slopes for each city with the residual species richness of each city.

Results

When looking at urbanization gradients within a city, there were varying responses (Figure A1; Figure A2), and this was consistent among species richness, abundance, Shannon diversity, and phylogenetic diversity. For abundance models, 51% of cities showed a negative response to urbanization while 49% showed a positive response (i.e., positive versus negative slope of their smoothed term). Similarly, for phylogenetic diversity 50% of cities showed a positive response and 50% of cities showed a negative response. Shannon diversity also showed a 50%/50% split between positive and negative responses. And for species richness, 53% of cities showed a positive response and 47% a negative response. However, this variation in biodiversity response to an urbanization gradient was not explained by city size (Figure A5) and we found no evidence ($p > 0.05$) for this variability in relation to our predictor variables (see full model results in Table A1). We also found no relationship between slope of response to urbanization and the residual species richness in a city (Figure A6). Lastly, when investigating the parametric terms (i.e., urbanness, water, trees, and evi) at a local-level sampling unit, we found no evidence that any predictor was generally more important than another and significant variation existed among cities (Figure A7). Although, the mean proportional effect size for water was slightly positive while the mean proportional effect size for urbanness was slightly negative (Figure A7).

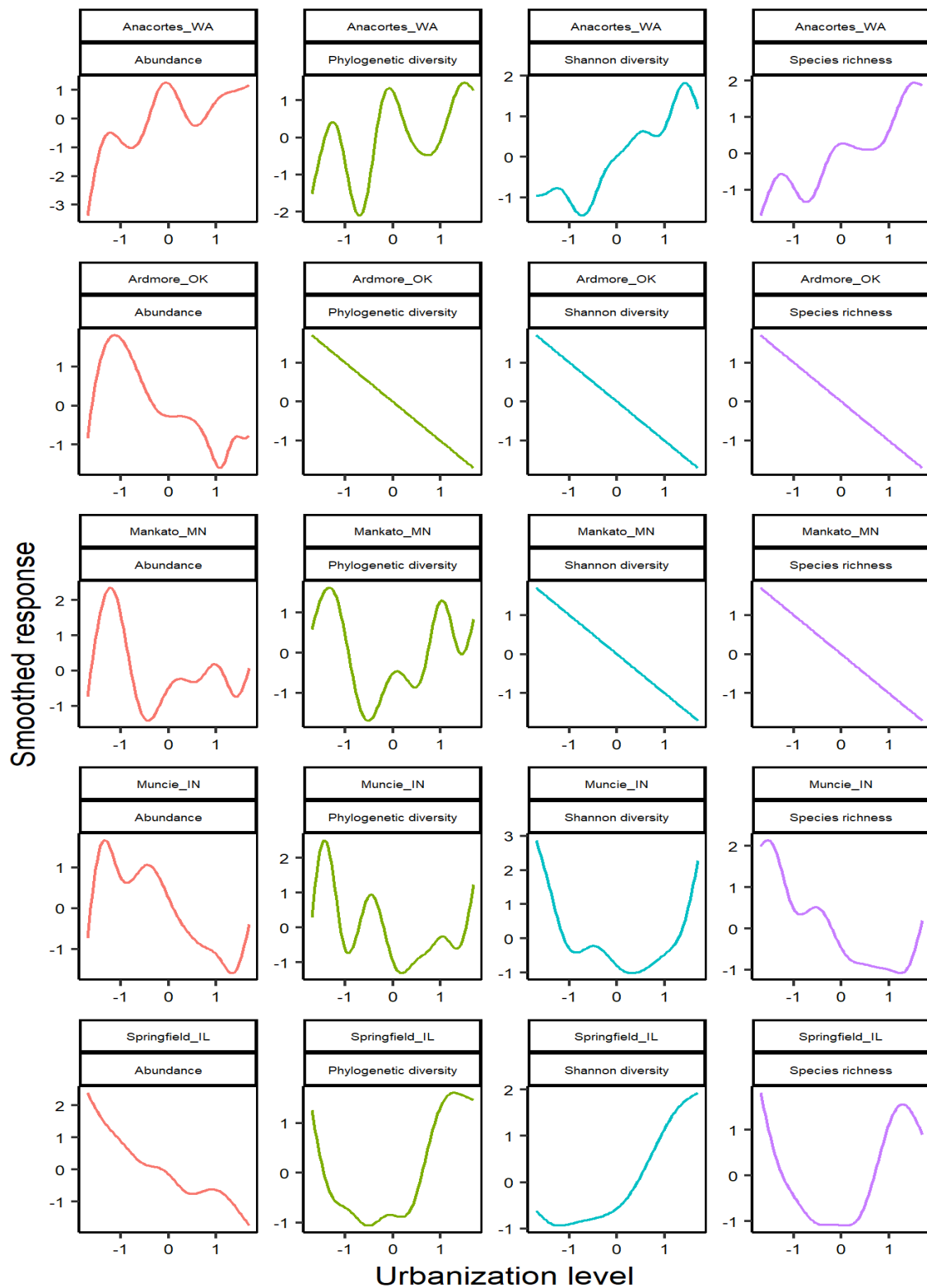


Figure A1. Smoothed responses of the urbanization gradient within 5 cities in the USA, demonstrating how we assessed the urbanization gradients within a city. The urbanization levels were scaled and centered so that slopes calculated for each city were comparable.

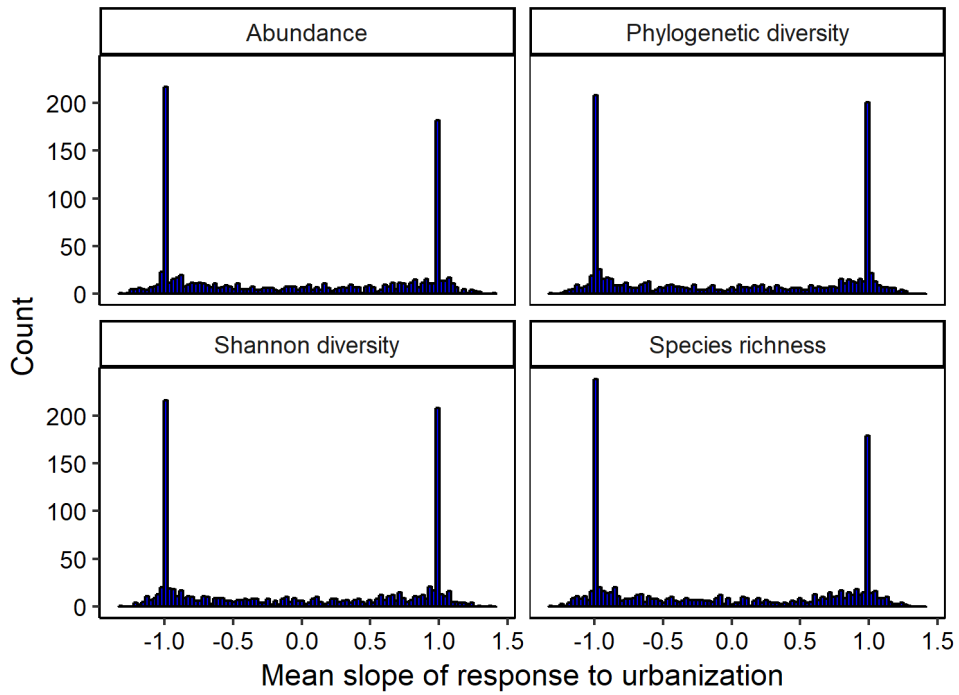


Figure A2. Histogram showing the mean slope for each city, stratified by the different analyses, of the response to urbanization. The slope was calculated from the scaled and centered response to urbanness from fitted GAMs (Figure 1).

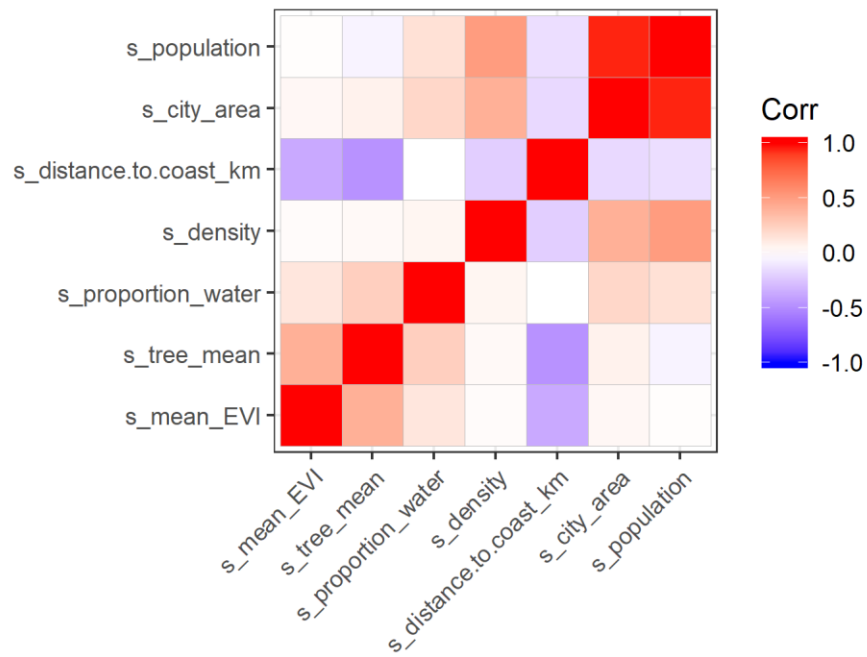


Figure A3. Correlation plot of predictor variables used in the slope of response to urbanization analysis, but population of a city was then excluded.

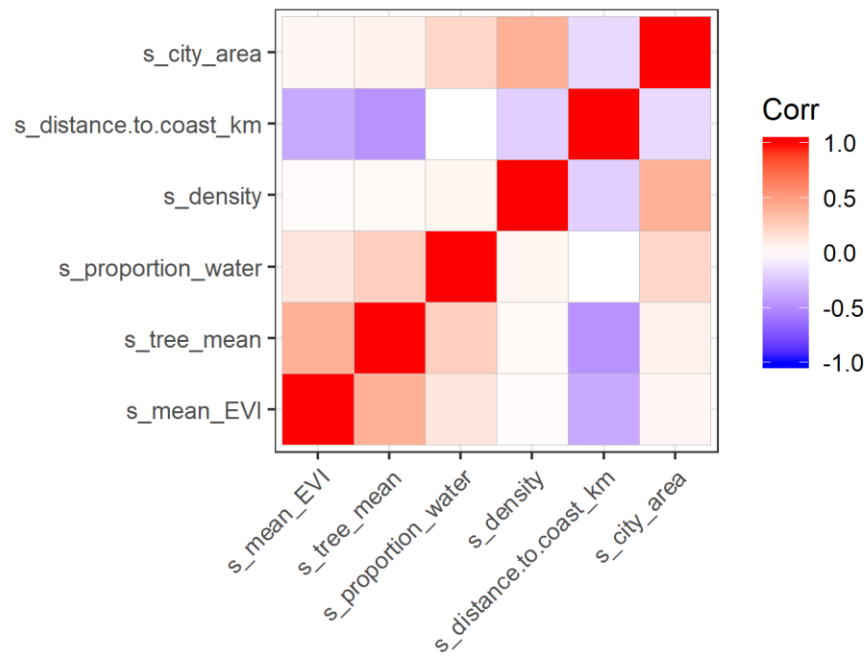


Figure A4. Correlation plot of predictor variables used in the slope of response to urbanization analysis, after excluding population.

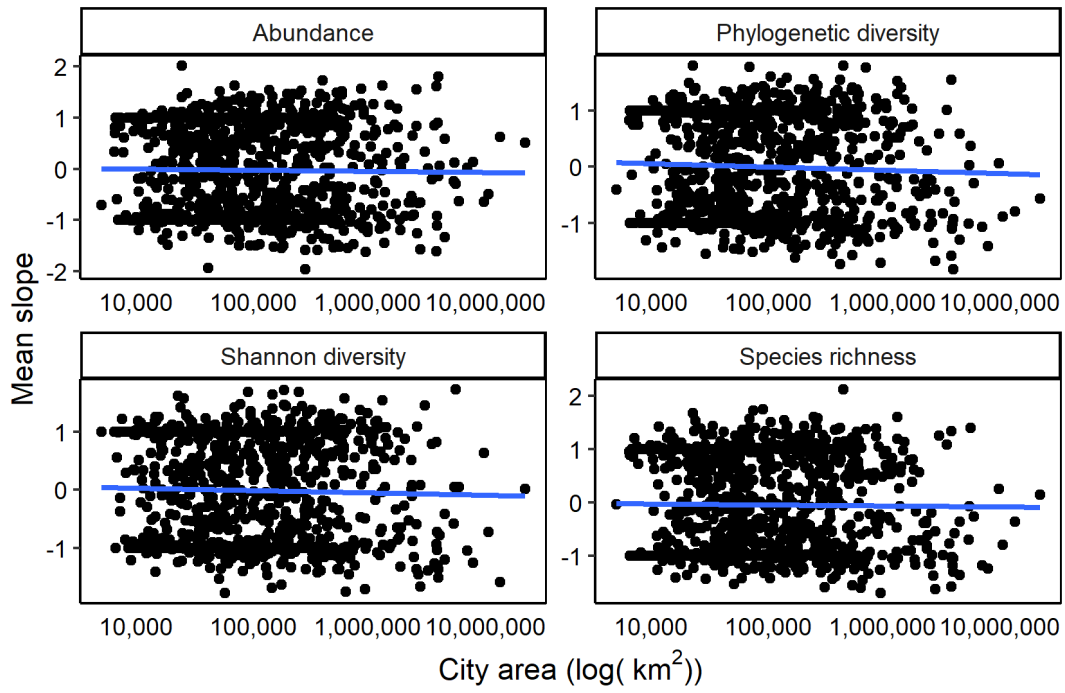


Figure A5. The relationship between city size and the mean slope of response to urbanization.

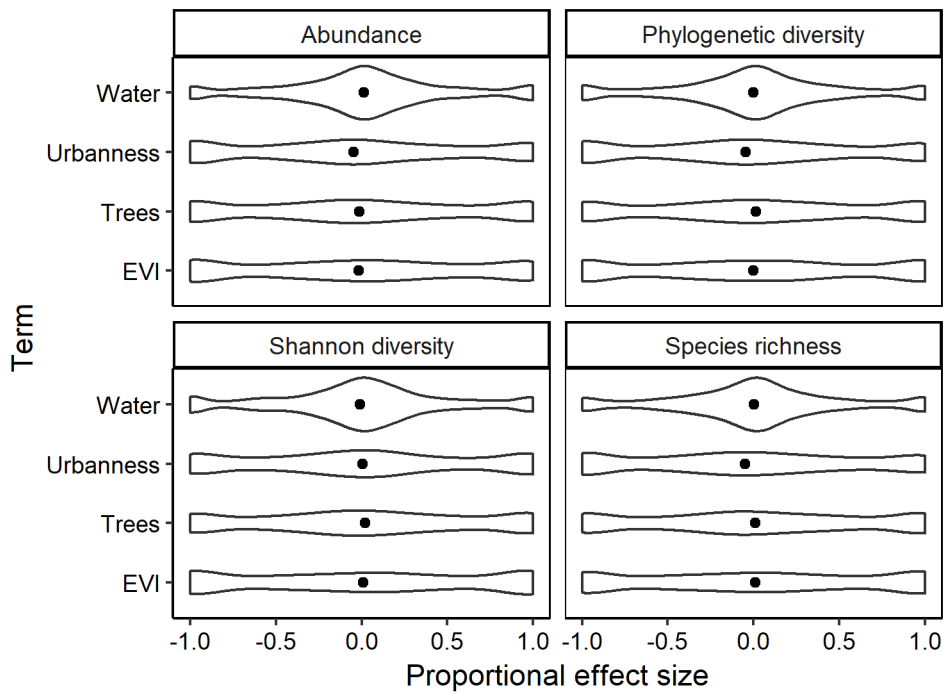


Figure A6. The proportional effect size for each city, calculated by dividing all effect sizes for each model by the largest effect size particular to that model. The black dot represents the mean among all city-specific models.

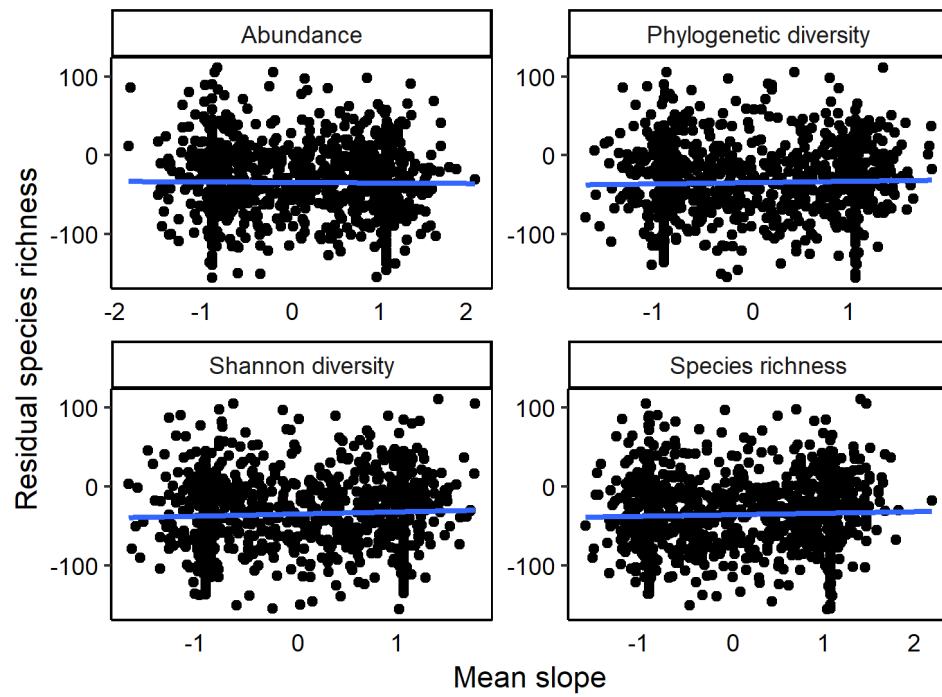


Figure A7. The relationship between the slope of response to urbanization, based on each of the four response variables, and the residual species richness in that city.

Table A1. Results of models which investigated the slope of urbanization response within cities as the response variable with the aforementioned predictor variables. These results are for the full models, and other modelling approaches (e.g., splitting to binomial response distributions) were also investigated and confirmed these results.

Species richness results:

```
Family: gaussian
Link function: identity

Formula:
mean_slope.all_s ~ s_distance.to.coast_km + s_mean_EVI + s_tree_mean +
  s_proportion_water + s_city_area + s(lat, lng) + s_density

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.049936  0.026623  -1.876   0.0609 .
s_distance.to.coast_km -0.202373  0.081784  -2.474   0.0135 *
s_mean_EVI    -0.064816  0.059901  -1.082   0.2795
s_tree_mean   -0.134515  0.090536  -1.486   0.1376
s_proportion_water -0.066801  0.059216  -1.128   0.2595
s_city_area   -0.010689  0.067449  -0.158   0.8741
s_density      0.002086  0.066989   0.031   0.9752
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df      F p-value
s(lat,lng)    2      2  1.721  0.179

R-sq.(adj) =  0.0016  Deviance explained = 0.846%
GCV = 0.45719  Scale est. = 0.45366    n = 1165
```

Abundance results:

```
Family: gaussian
Link function: identity

Formula:
mean_slope.all_s ~ s_distance.to.coast_km + s_mean_EVI + s_tree_mean +
  s_proportion_water + s_city_area + s(lat, lng) + s_density

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.03136  0.02697  -1.163   0.245
s_distance.to.coast_km  0.02690  0.14780   0.182   0.856
s_mean_EVI    -0.05082  0.06488  -0.783   0.434
s_tree_mean   -0.02813  0.09447  -0.298   0.766
s_proportion_water -0.02797  0.06261  -0.447   0.655
s_city_area   -0.01265  0.06975  -0.181   0.856
s_density     -0.03029  0.07157  -0.423   0.672

Approximate significance of smooth terms:
              edf Ref.df      F p-value
s(lat,lng)  6.843   9.66  0.515  0.851

R-sq.(adj) =  0.000753  Deviance explained = 1.21%
GCV = 0.37345  Scale est. = 0.36889    n = 1133
```

Shannon diversity results:

```
Family: gaussian
Link function: identity

Formula:
mean_slope.all_s ~ s_distance.to.coast_km + s_mean_EVI + s_tree_mean +
  s_proportion_water + s_city_area + s(lat, lng)

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.019839   0.026815  -0.740   0.460
s_distance.to.coast_km -0.127148   0.183966  -0.691   0.490
s_mean_EVI     -0.003434   0.066743  -0.051   0.959
s_tree_mean    -0.096772   0.094884  -1.020   0.308
s_proportion_water -0.044386   0.063522  -0.699   0.485
s_city_area    -0.004001   0.060390  -0.066   0.947

Approximate significance of smooth terms:
              edf Ref.df      F p-value
s(lat,lng)  9.749  13.56  0.882   0.577

R-sq.(adj) =  0.0115   Deviance explained = 2.44%
GCV = 0.38975   Scale est. = 0.38433   n = 1133
```

Phylogenetic diversity results:

```
Family: gaussian
Link function: identity

Formula:
mean_slope.all_s ~ s_distance.to.coast_km + s_mean_EVI + s_tree_mean +
  s_proportion_water + s_city_area + s(lat, lng) + s_density

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.0036232   0.0271062  -0.134   0.894
s_distance.to.coast_km -0.1069375   0.0992775  -1.077   0.282
s_mean_EVI     -0.0214886   0.0625410  -0.344   0.731
s_tree_mean    -0.1036956   0.0925148  -1.121   0.263
s_proportion_water -0.0709483   0.0611661  -1.160   0.246
s_city_area    -0.0412579   0.0689967  -0.598   0.550
s_density      0.0003296   0.0688048   0.005   0.996

Approximate significance of smooth terms:
              edf Ref.df      F p-value
s(lat,lng)  3.1   4.035  0.794   0.542

R-sq.(adj) =  3.85e-05   Deviance explained = 0.808%
GCV = 0.45718   Scale est. = 0.45311   n = 1133
```