https://doi.org/10.1590/1982-0224-2021-0050

SUPPLEMENTARY MATERIAL S5

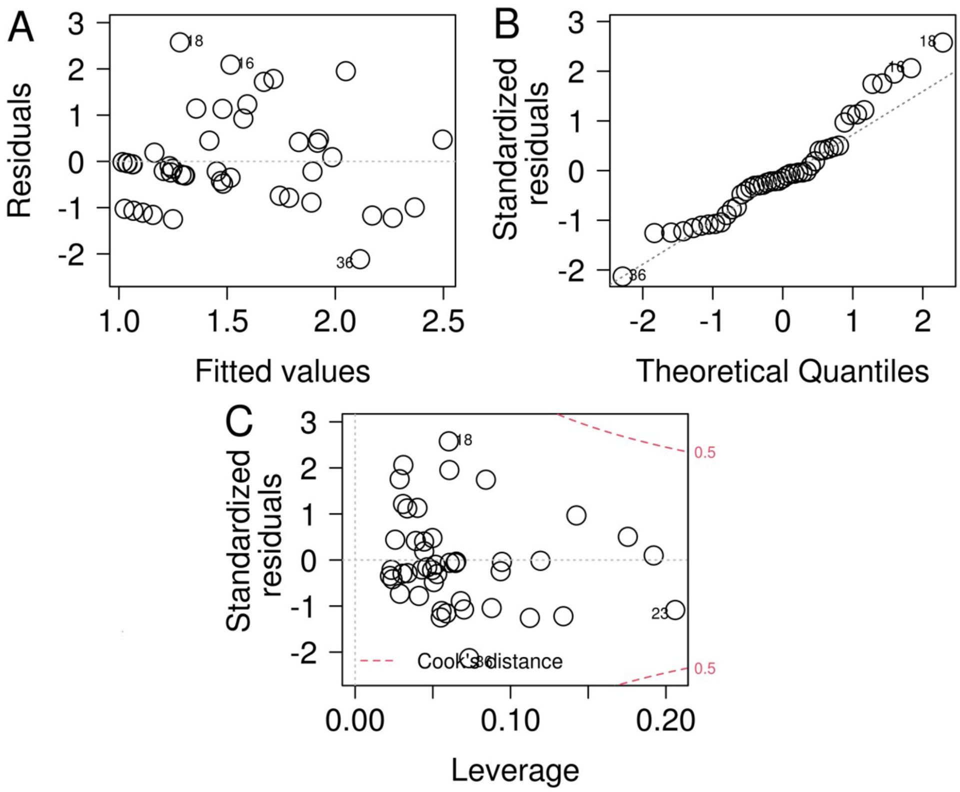

FIGURE S5 | Assessment of model assumptions and model quality of the linear model with rarefied species richness as the response variable. Assessment of homogeneity of variance (**A**), normality of residual distribution (**B**) and presence of influential observations (**C**).

We assessed the relationship between EH and urbanization with species diversity ($S_{rarefied}$) or local contribution to beta diversity (LCDB) with linear models. $S_{rarefied}$ or LCBD were the response variables and distance to median (a quantitative proxy to EH) and the proportion of impervious surface in a buffer of 500m (a quantitative proxy to urbanization) were the explanatory variables. We transformed distance to median by $log_e(x)$ to improve the linearity of the relationships.

For $S_{rarefied}$ as a response, we used an ordinary least squares model (OLS). Considering that LCBD values are bounded to 0 and 1, we used a beta-regression (Ferrari, Cribari-Neto, 2004; Douma, Weedon, 2019) to relate LCBD and both explanatory variables. For beta-regression we used a logit link function (Ferrari, Cribari-Neto, 2004; Cribari-Neto, Zeileis, 2010). We assessed the assumption of homogeneity of variance with dispersion plots with residuals and fitted values, and the presence of influential observations with the Cook's distance (Quinn, Keough, 2002; Zuur *et al*., 2010). We assessed for spatial autocorrelation in the residuals with Mantel correlogram and bubble plot of model residuals (Zuur *et al*., 2009; 2010; Legendre, Legendre, 2012).

We estimated Pearson correlation coefficient (*r*) between distance to median and proportion of impervious surface. We considered a correlation coefficient greater than |0.6| as indicative on multicollinearity (Zuur *et al*., 2010; Dormann *et al*., 2013). Multicollinearity was not an issue in all linear models because there was low correlation between the explanatory variables ($S_{rarefied}$ as response, r = –0.07, P = 0.627; LCBD as response, r = –0.08, P = 0.642).

We assessed the assumption of homogeneity of variance with dispersion plots with residuals and fitted values, the normality of residuals with quantile plots of standardized residuals with fitted values, and the presence of influential observations with a plot with standardized residuals in function of leverage and with Cook's distance thresholds (Quinn, Keough, 2002; Zuur *et al*., 2010). We assessed for spatial autocorrelation in the residuals with Mantel correlogram and bubble plot of model residuals (Zuur *et al*., 2009; 2010; Legendre, Legendre, 2012). For Mantel correlogram we first built a spatial connectedness matrix computing watercourse distance among sampled streams (Landeiro *et al*., 2011). Then, we computed Mantel correlation between the matrix of watercourse distance and an Euclidean distance matrix computed with model residuals. We established distance classes for the Mantel correlogram following Sturges (1926), computed significance of Mantel correlations with 999 permutations and applied Holm correction for multiple testing. We ran all analyses in the R software (R Core Team, 2020), with "vegan" (Oksanen *et al*., 2019), "adespatial" (Dray *et al*., 2020), "riverdist" (Tyers, 2020), "betareg" (Cribari-Neto, Zeileis, 2010) and "lmtest" (Zeleis, Hothorn, 2002). We adopted a 5% significance level for all analyses.

The distribution of residuals from the model with $S_{rarefied}$ as response variable against the fitted values did not show any systematic variation, which indicates that this model attended the assumption of homoscedasticity (Fig. **S5A**). The relation between standardized residuals and the expected value from a normal distribution also do not show substantial deviation from normality (Fig. **S5B**). Furthermore, the model with $S_{rarefied}$ as response variable did not present influential observations considering Cook's distance threshold (Fig. **S5C**). Finally, both Mantel correlogram (Fig. **S6A**) and bubble plot (Fig. **S6B**) indicates that spatial autocorrelation was not an issue in model residuals.

REFERENCES

- **Cribari-Neto F, Zeileis A.** Beta regression in R. J Stat Softw. 2010; 34(2):1–24. https://doi.org/10.18637/jss.v034.i02 https://doi.org/10.18637/jss.v034.i02
- **Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, *et al*.** Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 2013; 36(1):27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x
- **Douma JC, Weedon JT.** Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. Methods Ecol Evol. 2019; 10(9):1412–30. https://doi.org/10.1111/2041-210X.13234
- **Dray S, Bauman D, Blanchet G, Borcard D, Clappe S, Guenard G *et al*.** Adespatial: Multivariate Multiscale Spatial Analysis. Version 0.3-8 [Internet]. 2020. Available from: https://CRAN.R-project.org/package=adespatial
- **Ferrari S, Cribari-Neto F.** Beta regression for modelling rates and proportions. J Appl Stat. 2004; 31(7):799–815. https://doi.org/10.1080/0266476042000214501
- **Landeiro VL, Magnusson WE, Melo AS, Espírito-Santo HMV, Bini LM.** Spatial eigenfunction analyses in stream networks: do watercourse and overland distances produce different results? Freshwater Biology. 2011; 56:1184–92. https://doi.org/10.1111/j.1365-2427.2010.02563.x

- **Legendre P, Legendre L.** Numerical ecology. developments in environmental modelling, Vol. 24. 3rd ed. Amsterdam: Elsevier; 2012.
- **Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D *et al*.** Vegan: Community ecology package. Version 2.5-6 [Internet]. 2019. Available from: https://cran.r-project.org/web/packages/vegan/index.html
- **Quinn GP, Keough MJ.** Experimental design and data analysis for biologists. Cambridge: Cambridge University Press; 2002.
- **R Core Team.** R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org
- **Sturges HA.** The choice of a class interval. J Am Stat Assoc. 1926; 21:65–66.
- **Tyers M.** Riverdist: River Network Distance Computation and Applications. R package version 0.15.1. Available from: https://CRAN.R-project.org/package=riverdist; 2020.
- **Zeleis A, Hothorn T.** Diagnostic checking in regression relationships. R J. 2002; 2(3):7–10. https://CRAN.R-project.org/doc/Rnews
- **Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM.** Mixed effects models and extensions in Ecology with R. New York: Springer; 2009. https://doi.org/10.1007/978-0-387-87458-6
- **Zuur AF, Ieno EN, Elphick CS.** A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol. 2010; 1(1):3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

Neotropical Ichthyology

OPEN ACCESS

SBI
Official Journal of the
Sociedade Brasileira de Ictiologia

HOW TO CITE THIS ARTICLE

- **Ortega JCG, Bacani I, Dorado-Rodrigues TF, Strüssmann C, Fernandes IM, Morales J, Mateus L, Silva HP, Penha J.** Effects of urbanization and environmental heterogeneity on fish assemblages in small streams. Neotrop Ichthyol. 2021; 19(3):e210050. https://doi.org/10.1590/1982-0224-2021-0050