Supplemental Material

## S1

Precipitation. Historical series of precipitation obtained from the Serra de Navio weather station (id: 8052000), available in the virtual database of the National Water Agency/Agência Nacional de Águas (ANA-<http://www.snirh.gov.br/hidroweb/serieshistoricas>).

Figure S1. Monthly trends in precipitation during the sampling years. Boxplots show means and 95% confidence limits estimated via nonparametric bootstrap.



## S2 R code to calculate vulnerability

The below example uses a data frame “vulnerability\_values” that holds the values of the threats to be rescaled.

# packages

library(tidyverse)

library(scales)

# function to scale values 0-1, i.e. all threats have equal weight

range01 <- function(x, ...){(x-min(x, ...))/(max(x, ...)-min(x, ...))}c

# Now calculate vulnerability for each threat

vulnerability\_values <- vulnerability\_values |>

 mutate(vul\_casa = tanh(1/ifelse(dist\_casa\_km > 15, 15, dist\_casa\_km)),

 vul\_bar = range01(1/ifelse(dist\_barra\_km > 75, 75, dist\_barra\_km)),

 vul\_pg = scales::rescale(1/(dist\_pg\_km + 100), to = c(0,1)),

 vul\_f500 = range01(-1\*ifelse(forest21\_500 > 70, 70, forest21\_500)),

 vul\_f10km = range01(-1\*ifelse(forest21\_10000 > 70, 70, forest21\_10000)),

 vul\_500 = dsmt500,

 vul\_10km = dsmt10km,

 vul\_com = ifelse(rio == "Araguari",1,0)

 )

# Overall vulnerability score

vcols\_all <- c("vul\_casa","vul\_bar","vul\_pg","vul\_500","vul\_10km",

 "vul\_f500", "vul\_f10km", "vul\_com")

vcols\_pca <- c("vul\_casa","vul\_bar","vul\_pg","vul\_10km", "vul\_f500", "vul\_f10km", "vul\_com")

n\_vul = length(vcols\_pca)

vulnerability\_values <- vulnerability\_values |>

 mutate(vul\_sum = vul\_casa + vul\_bar + vul\_pg +

 vul\_10km + vul\_f500 + vul\_f10km, vul\_com) |>

 mutate(vul\_mean = vul\_sum / n\_vul)

# Calculate overall vulnerability index

#https://conservationbytes.com/2014/08/01/a-fairer-way-to-rank-conservation-and-ecology-journals-in-2014/

# calculate variance across columns

vulnerability\_values$vul\_var <- apply(vulnerability\_values[,vcols\_pca], 1, var)

vulnerability\_values$vul <- vulnerability\_values$vul\_sum \* (vulnerability\_values$vul\_sum^vulnerability\_values$vul\_var) # increase vulnerability when there is more variation

vulnerability\_values$vul\_index <- vulnerability\_values$vul / n\_vul

## S3 Hierarchical Clustering on Principle Components (HCPC)

Hierarchical Clustering on Principal Components (HCPC) is a versatile technique that can be used to identify clusters of data points in a variety of applications. HCPC is a well-established technique that is widely used in ecological and environmental vulnerability analysis (Tian *et al.* 2022, Wu *et al.* 2020, Alaniz *et al.* 2022). It combines two fundamental concepts: principal component analysis (PCA) and hierarchical clustering to efficiently explore and group data with complex structures. This method is particularly useful to discover underlying patterns and relationships in high-dimensional datasets. HCPC works by first applying PCA to the data to reduce the dimensionality and identify the most important underlying dimensions. Then, it applies hierarchical clustering to the principal components to identify clusters of data points that are similar in terms of these dimensions.

Principal Component Analysis is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while retaining the most important information (Vaughan and Ormerod 2005, Greenacre *et al.* 2022). In summary, PCA transforms a set of correlated variables into a smaller set of uncorrelated variables, called principal components. It does so by identifying the principal components, which are linear combinations of the original features that capture the maximum variance in the data (Greenacre *et al.* 2022). The principal components are ordered so that the first principal component explains the most variance in the data, and the subsequent principal components explain decreasing amounts of variance.

HCPC leverages PCA to reduce the dimensionality of the dataset, making it computationally more efficient and reducing the impact of noise or irrelevant features. Hierarchical clustering aims to organize data points into a dendrogram. Hierarchical clustering does not require a predetermined number of clusters. Instead, it arranges data points in a hierarchy where individual data points initially form distinct clusters, and these clusters are progressively merged into larger ones based on their similarity.

The R package FactoMineR provides an integrated and reproducible framework to implement diverse types of HCPC (Lê *et al.* 2008). The details are provided elsewhere (Lê *et al.* 2008), but in brief, this includes applying PCA to the dataset to reduce its dimensionality. Then hierarchical clustering is run on the reduced dataset. The result is a dendrogram that shows how data points are grouped at different levels of granularity. The dendrogram is then analyzed to determine the optimal number of clusters. Finally, data points are assigned to clusters based on the dendrogram analysis. We ran all analysis using default settings of the package functions. The results were then mapped to prioritize vulnerability and restoration actions.



Figure S3.1. Showing the distribution in principal coordinate space of vulnerability to threat variables. Biplot showing variables in relation to the first two principal components (Dimension 1 and Dimension 2). Arrow lengths are proportional to the associations between variables and principal components. Variable included were commercial use (“commercial”), distance to nearest dam, town, and house (“dam”, “town”, “house”), natural forest loss (“FL”) and cover (“FC”) both at 500 m and 10 km radii. See main text for definitions of vulnerability variables.

The HCPC analysis identified six clusters (Figure S3.2).



Figure S3.2. Hierarchical Clustering on Principle Components (HCPC) of nesting area vulnerability. Showing the distribution in principal coordinate space of (A) nesting areas and the variable with strongest correlation on the first PCA axis (Dimension 1) and (B) nesting areas and HCPC clusters identified. See main text for definitions of vulnerability variables.

References

Alaniz, A. J.; Smith-Ramírez, C.; Rendón-Funes, A.; Hidalgo-Corrotea, C.; Carvajal, M. A.; Vergara, P. M. ; Fuentes, N. 2022. Multiscale spatial analysis of headwater vulnerability in South-Central Chile reveals a high threat due to deforestation and climate change. *Science of The Total Environment* 849, 157930.

Greenacre, M.; Groenen, P. J. F.; Hastie, T.; D’Enza, A. I.; Markos, A. ; Tuzhilina, E. 2022. Principal component analysis. *Nature Reviews Methods Primers* 2, 100.

Lê, S.; Josse, J. ; Husson, F. 2008. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* 25, 1 - 18.

Tian, Y.; Wen, Z.; Zhang, X.; Cheng, M. ; Xu, M. 2022. Exploring a multisource-data framework for assessing ecological environment conditions in the Yellow River Basin, China. *Science of The Total Environment* 848, 157730.

Vaughan, I. P. ; Ormerod, S. J. 2005. METHODOLOGICAL INSIGHTS: Increasing the value of principal components analysis for simplifying ecological data: a case study with rivers and river birds. *Journal of Applied Ecology* 42, 487-497.

Wu, J.; Li, P.; Wang, D.; Ren, X. ; Wei, M. 2020. Statistical and multivariate statistical techniques to trace the sources and affecting factors of groundwater pollution in a rapidly growing city on the Chinese Loess Plateau. *Human and Ecological Risk Assessment: An International Journal* 26, 1603-1621.