**Droughts and controlled rivers: how Belo Monte dam has affected the food security of Amazonian riverine communities**

Lopes, PFM1,2\*; Cousido-Rocha, M3; Silva, MRO1,4; Carneiro, CC5; Pezzuti, JCB2,6; Martins, EG7; de Paula, EMS8; Begossi, A2,9†; Pennino, MG1,10

**Supplementary Material 1**

**This document brings additional information on the methods used and on the results of these models.**

We modeled ‘fish presence’ using a beta distribution (Eq. 1), which is well-suited for modeling proportion response variables. To avoid values of 0 and 1, we transformed the average presence variable:

$Y\_{i}^{\*}=(Y\_{i}$$\left(N-1\right)+0.5)/N)\~Beta(μ\_{i}^{Y}$*,* $ϕ^{Y})$(equation 1),

where $Y\_{i}$ is the average ‘fish presence’, $Y\_{i}^{\*}$ is the transformed average ‘fish presence’ response variable, $μ\_{i}^{Y}$ and $ϕ^{Y}$represent the mean and dispersion of response variable, and *N* is the sample size. Hereafter, for simplicity, we will refer to $Y\_{i}^{\*}$ as the proportion of ‘fish presence’.

We applied a Gamma distribution (Eq. 2) to model 'fish consumption', as it is suitable for strictly positive responses (never zero or negative):

$Z\_{i}\~Gamma($$μ\_{i}^{Z}$*,* $ϕ^{z}$) (equation 2),

where $Z\_{i}$is the average ‘fish consumption’; $μ\_{i}^{Z}$and $ϕ^{z}$ represent the mean and dispersion of average fish consumption conditional-to-presence. In both cases, the predictive variables used to explain the dependent variables (‘fish presence’ and ‘fish consumption’) were the same (Eq. 3, 4, 5 and 6), with the inclusion of ‘richness’ in the second model (Eq. 4 and 6). Two initial models were tested (below), followed by alternative formulations with the inclusion/removal of variables (as shown in the original Table 2 in the manuscript). These two models are considered separately due to the correlation between the explanatory variables river level rate and affected region.

Model including river level rates:

$$logit\left(μ\_{i}^{Y}\right)=α^{Y}+f\_{1}\left(time\_{i}\right)+ f\_{2}\left(season\_{i}\right)+ f\_{3}\left(level\_{i}\right) (equation 3) $$

$$logit\left(μ\_{i}^{Z}\right)=α^{Z}+g\_{1}\left(time\_{i}\right)+ g\_{2}\left(season\_{i}\right)+g\_{3}\left(level\_{i}\right)+g\_{4}\left(richness\_{i}\right) (equation 4) $$

Model including affected region:

$$logit\left(μ\_{i}^{Y}\right)=β\_{0}^{Y}+f\_{1}\left(time\_{i}\right)+ f\_{2}\left(season\_{i}\right)+ \sum\_{j=1}^{C}β\_{j}^{Y}D\_{ij} (equation 5)$$

$$logit\left(μ\_{i}^{Z}\right)=β\_{0}^{Z}+g\_{1}\left(time\_{i}\right)+ g\_{2}\left(season\_{i}\right)+g\_{4}\left(richness\_{i}\right) + \sum\_{j=1}^{C}β\_{j}^{Z}D\_{ij} (equation 6)$$

The intercepts of each response variable, 'fish presence' and 'fish consumption,' are represented by $α^{Y}$ and $α^{Z}$, respectively. The coefficients associated with the reference level of the affected region variable, which is the de-watered reach, are denoted by $β\_{0}^{Y}$ and $β\_{0}^{Z}$ for ‘fish presence’ and ‘fish consumption’, respectively. The term $\sum\_{j=1}^{C}β\_{j}^{Y}D\_{ij}$ (or $\sum\_{j=1}^{C}β\_{j}^{Z}D\_{ij}$) represents the parametric factor effect of deviations from the reference region value for the remaining affected regions. Here, $β\_{j}^{Y}(or β\_{j}^{Z})$ represents the deviation for region *j* in relation to the de-watered reach value and $D\_{ij}$ is a dummy variable that equals 1 if the *i*-th observation corresponds to region *j* and zero otherwise. The value *C=4* represents the total number of categories of the region factor excluding the reference one. Note that the four regions are sorted as follows: downstream – rural, downstream – urban, reservoir – rural, reservoir – urban. The effects of river level rate, measured in cm, are modeled as a second-order random walk (RW2) smooth function denoted by$ f\_{3}\left(level\_{i}\right)$and $g\_{3}\left(level\_{i}\right)$. Additionally, $g\_{4}\left(richness\_{i}\right)$ represents the RW2 effect of fish richness, measured as the number of species identified by their popular names, in meals.

The RW2 is a flexible model that overcomes the assumption of linear relation between the response variable and the covariables allowing more flexible relationships. In particular, the random walk 1 (RW1) formulation established that the covariable value at time instant *t* minus the value of the covariable in time *t-1* (termed first order increments) follows a gaussian distribution, therefore modeling a trend without linear assumptions. The RW2 generalizes this formulation assuming that the second order increments follow a gaussian distribution producing smoother curves.

The $f\_{1}\left(time\_{i}\right)$ and $g\_{1}\left(time\_{i}\right)$ correspond to a continuous increasing time covariable (the number of days since day 0, i.e., values from 0 to 3380) created using the *julian* R function from the date (R Development Team, 2022). The categorical date covariable was transformed into a continuous variable, $time$. This approach was chosen to estimate the effect of $time$ on the response variables, ‘fish presence’ and ‘fish consumption’, through a smooth function to detect temporal trends, instead of having a coefficient associated with each of the different dates. An analogous seasonal effect, $f\_{2}\left(season\_{i}\right) $and $g\_{2}\left(season\_{i}\right)$*,* was created using the *yday* function of the ‘lubridate’ R package (Grolemund, Wickham, 2011) and fitting cyclic RW2 effects in the models.

 It should be noted that we have also investigated the presence of other forms of dependency in our data, particularly focusing on temporal dependence related to the weekly effect. To achieve this, we constructed time series for each day of the week, incorporating the respective fish consumption values. Following this, we evaluated the dependency within each series and the inter-series dependency using autocorrelation and cross-correlation measures, employing the "acf" and "ccf" functions in R. The results suggest the absence of weekly dependency patterns in our data, which is why the final analyses do not include it.

*Additional modeling results*

Table S3. Estimates of the fixed effects in the beta and gamma predictors. SD = Standard deviation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean | SD | 2.5% percentile | 50% percentile | 97.5% percentile |
| **Beta model** $β\_{0}^{Y}$ | **-0.71** | **0.07** | **-0.84** | **-0.71** | **-0.57** |
| Downstream rural | 0.55 | 0.10 | 0.35 | 0.55 | 0.74 |
| Downstream urban | 0.02 | 0.09 | -0.15 | 0.02 | 0.19 |
| Reservoir rural  | 0.57 | 0.09 | 0.40 | 0.57 | 0.75 |
| Reservoir urban | -0.45 | 0.08 | -0.61 | -0.45 | -0.28 |
| **Gamma model**$β\_{0}^{Z}$ | **5.12** | **0.05** | **5.03** | **5.21** | **5.12** |
| Downstream rural | -0.26 | 0.05 | -0.35 | -0.26 | -0.17 |
| Downstream urban | -0.17 | 0.04 | -0.26 | -0.17 | -0.09 |
| Reservoir rural  | 0.34 | 0.04 | 0.25 | 0.33 | 0.42 |
| Reservoir urban | -0.21 | 0.04 | 0.30 | -0.21 | -0.13 |

Table S4: Estimates of the precision of the random effects in the beta and gamma predictors. The precision is defined as the variance inverse. SD = Standard deviation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | mean | sd | 2.5% percentile | 50% percentile | 97.5% percentile |
| Precision parameter for the Gamma observations | 3.42e+00 | 1.17e-01 | 3.19e+00 | 3.42e+00 | 3.65e+00 |
| Precision parameter for the beta observations | 1.18e+00 | 3.30e-02 | 1.11e+00 | 1.18e+00 | 1.24e+00 |
| Precision of $f\_{1}\left(time\_{i}\right)$ | 9.83e+04 | 5.78e+04 | 3.27e+04 | 8.38e+04 | 2.50e+05 |
| Precision of$f\_{2}\left(season\_{i}\right)$ | 8.65e+01 | 2.70e+02 | 2.32e+00 | 2.81e+01 | 5.37e+02 |
| Precision of$ g\_{1}\left(time\_{i}\right)$ | 9.82e+04  | 5.76e+04 |  3.27e+04 | 8.37e+04 | 2.50e+05 |
| Precision of $g\_{4}\left(richness\_{i}\right)$  | 9.80e-02 | 7.80e-02 | 2.50e-02 | 7.50e-02 | 3.06e-01 |
| Precision of $g\_{2}\left(season\_{i}\right)$ | 2.74e+03  | 6.07e+03 | 1.64e+02 | 1.19e+03  | 1.51e+04 |



Figure S4. Estimates $β\_{j}^{Y}$ and $β\_{j}^{Z} $coefficients associated with the deviation of each of the affected region (*j=1, …,4*) respect to the reference one (de-watered reach) are displayed in beta ('fish presence') and gamma ('fish consumption') predictors.

Literature cited

Grolemund G, Wickham H (2011) Dates and Times Made Easy with lubridate. *Journal of Statistical Software* 40: 1–25.