Geomorphometric analysis of wildfire occurrence in a humid tropical protected area: a case study in Southern Pacific Costa Rica

Primary Results

Shu Wei Chou Chen

Table of contents

[1. Introduction 1](#_Toc182056739)

[2. Methods 1](#_Toc182056740)

[3. Results 2](#_Toc182056741)

[4. Conclusiones 6](#_Toc182056742)

[5. Referencias 6](#_Toc182056743)

# 1. Introduction

…

# 2. Methods

* The explanatory variables, FA and TRI, are transformed logarithmically to address their asymmetry, specifically as . Subsequently, all covariates (DAH, logFA, LST, logLSF, SLOPE, logTRI, TWI, WE, and WExpo) are scaled using z-scores to allow for comparison of their standardized contributions.
* In total, 60484 centroids are divided into two datasets: a training set comprising 70% (42213 centroids) and a testing set comprising 30% (18271 centroids) to validate the model fit.
* The goal is to model a count variable (INC), which ranges from 0 to 6 and has a high percentage of zeros (). Therefore, zero-inflated count regression models are considered. These models are special cases within a flexible statistical framework called Generalized Additive Models for Location Scale and Shape (GAMLSS), and the statistical analysis was conducted in R (R Core Team 2023) with the package gamlss (R. A. Rigby and D. M. Stasinopoulos 2005).
* Essentially, the count regression models described above are represented by three-parameter GAMLSS models:

The response variable INC is distributed as a two-parameter or three-parameter distribution : the location or mean (), the scale (), and a parameter related to the skewness of the distribution (), as well as the link functions ( for ) for each parameter.

Regarding to the specification of , the Poisson regression model (PO) includes only one parameter with link function . Then, Negative Binomial model (NB) fits two parameters and , with link functions and . Similarly, the zero-inflated Poisson (ZIP) also includes two parameters but with , . Finally, the zero-inflated Negative Binomial model (ZINB) fits three parameters with , and . Note that the second paramter of ZIP and the third parameter of ZINB models the probability of a certain centroid has zero count of wildfires, similar to a logistic regresion.

Due to the complexity of the models, we suppose that all covariates affect in a linear way and perform the following procedure to define model terms:

First, we considered the models’ mean : PO, NB, ZIP, and ZINB with full specification of location function, and let other parameters as constant. Then, by applying the stepwise algorithm using both directions, reduced models are obtained (PO-red, NB-red, ZIP-red, and ZIMB-red). Since zero-inflated models have better results, in the third steps, we performed stepwise algorithm to add to ZIP-red and and to ZINB-red, called ZIP-red-final and ZINB-red-final.

# 3. Results

[Table 1](#tbl-ICmodels) shows the goodness of fit measures for the fitted models. Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) evaluate the goodness of fit for the training set, while Validated Globa Deviance (VGD) measures how well the fitted models perform for predicting the out-of-sample data (validation data). All three criterion shows that the best model is ZIP\_red\_final. On the other hand, note that the generalized pseudo r-squared (CoxSmell and Cragg Uhler) show that PO\_red are higher among all models, but the assumptions of this model are not satisfied. On other other hand, all assumptions of ZIP\_red\_final are satisfied and present the best measures (**aquí falta presentar los diagnósticos del modelo y pregunto si para artículos en su área, esto se presenta en el cuerpo del artículo o generalmente no lo toman con tanta importancia y podría ponerlos en anexo**).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Goodness of fit measures for the fitted models.   | Model | AIC | BIC | VGD | R2-CoxSnell | R2-Cragg Uhler | | --- | --- | --- | --- | --- | --- | | PO | 79,265.66 | 79,360.82 | 34,836.93 | 0.2723 | 0.3065 | | NBII | 70,882.70 | 70,986.50 | 31,089.78 | 0.1526 | 0.1812 | | ZIP | 73,309.74 | 73,413.55 | 32,133.96 | 0.0883 | 0.1052 | | ZINBI | 76,092.71 | 76,205.17 | 33,380.19 | 0.0219 | 0.0261 | | PO\_red | 79,269.74 | 79,356.24 | 34,842.41 | 0.2722 | 0.3063 | | NBII\_red | 70,881.29 | 70,976.45 | 31,090.24 | 0.1526 | 0.1812 | | ZIP\_red | 73,308.00 | 73,403.16 | 32,134.62 | 0.0883 | 0.1052 | | ZINBI\_red | 72,903.31 | 72,981.16 | 31,869.63 | 0.0929 | 0.1108 | | ZIP\_red\_final | 69,754.09 | 69,892.50 | 30,790.78 | 0.1621 | 0.1931 | | ZINBI\_red\_final | 69,532.61 | 69,688.32 | 30,566.35 | 0.1629 | 0.1942 | |

[Table 2](#tbl-ZIPmodels) describes the model estimates. The equation for Mu fits the wildfire counts, while the equation for Sigma models the probability of observing zero wildfires in a given centroid. We observe that all these variables are statistically significant, except for WE in the Mu equation.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Parameter estimates of the fitted models.   | Covariate | Estimate | Std. Error | t value | Pr(>|t|) | | --- | --- | --- | --- | --- | | Mu |  |  |  |  | | (Intercept) | 0.1652 | 0.0134 | 12.3008 | 0.0000 | | DAH | -0.0276 | 0.0077 | -3.5788 | 0.0003 | | LST | -0.4053 | 0.0242 | -16.7190 | 0.0000 | | logLSF | 0.3944 | 0.0241 | 16.3706 | 0.0000 | | SLOPE | -0.3849 | 0.0260 | -14.7907 | 0.0000 | | logTRI | -0.0012 | 0.0298 | -0.0413 | 0.9671 | | TWI | -0.4616 | 0.0239 | -19.3145 | 0.0000 | | WE | -0.0120 | 0.0131 | -0.9122 | 0.3617 | | WExpo | -0.1993 | 0.0175 | -11.3659 | 0.0000 | | sabanaTRUE | 0.4713 | 0.0181 | 26.0327 | 0.0000 | | Sigma |  |  |  |  | | (Intercept) | 0.8404 | 0.0206 | 40.7412 | 0.0000 | | sabanaTRUE | -2.4844 | 0.0658 | -37.7473 | 0.0000 | | LST | -0.3194 | 0.0305 | -10.4579 | 0.0000 | | WE | 0.1946 | 0.0211 | 9.2396 | 0.0000 | | TWI | 0.2514 | 0.0248 | 10.1444 | 0.0000 | | WExpo | 0.2546 | 0.0310 | 8.2237 | 0.0000 | |

Finall, [Figure 1](#fig-ZIPmusigma) presents the fitted and , which represent wildfire mean of each centroid and the probability of no wildfire, respectively. Those figures are difficult to assess due to the fact that [Figure 1 (a)](#fig-ZIPmu) shows higher fitted wildfire mean for northern parts of the region. However, if we take into account that the model fits a centroid with , meaning this centroid has greater probability to not have wildfire, we can filter out those centroids, and only consider those centroid with , that is, those places will have more than 50% of chance to have more than zero wildfires, then plot the fitted ([Figure 2 (b)](#fig-ZIPmu_sigma)). We can observe that the places with higher wildfire count mean are similar to the observed INC (fig-INC).

Finally, [Figure 1](#fig-ZIPmusigma) presents the fitted and , which represent the mean wildfire count for each centroid and the probability of no wildfire, respectively. These figures can be challenging to interpret, as [Figure 1 (a)](#fig-ZIPmu) shows a higher fitted wildfire mean in the northern parts of the region, which have high probability to not observe wildfire. However, if we consider that the model fits a centroid with , indicating a greater probability of no wildfire, we can filter out those centroids and focus only on those with . This means that these locations have more than a 50% chance of experiencing at least one wildfire. We can then plot the fitted ([Figure 2 (b)](#fig-ZIPmu_sigma)). Notably, the areas with higher mean wildfire counts are similar to the observed INC ([Figure 2 (a)](#fig-INC)).

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | (a) ZIP mean. | | |  | | --- | | (b) ZIP sigma. | |

Figure 1: The predicted and of the fitted ZIP model.

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | (a) Wildfire counts. | | |  | | --- | | (b) Predicted fire count mean, conditional on a probability greater than 0.5 of fire. | |

Figure 2: Observed wildfire counts and the model predicted fire count mean, conditional on a probability greater than 0.5 of occurrence of wildfire.

# 4. Conclusiones

…

# 5. Referencias

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

R. A. Rigby, and D. M. Stasinopoulos. 2005. “Generalized Additive Models for Location, Scale and Shape,(with Discussion).” *Applied Statistics* 54.3: 507–54.