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

Shu Wei Chou-Chen1,✉, Adolfo Quesada-Román2, and Daniela Vargas-Sanabria3

Abstract

This study investigates the spatial distribution and environmental drivers of wildfire occurrence in a humid tropical protected area in Southern Pacific Costa Rica using geomorphometric analysis. We applied multiple statistical models, including Poisson, Negative Binomial, Zero-Inflated Poisson (ZIP), and Zero-Inflated Negative Binomial (ZINB), to evaluate their performance in fitting wildfire data. Model diagnostics, including Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Validated Global Deviance (VGD), indicated that the ZIP\_red\_final model provided the best fit. Covariates such as land surface temperature (LST), slope, topographic wetness index (TWI), and land use (savanna areas) were significant predictors of wildfire counts, while vegetation exposure (WE) showed limited significance. Spatial predictions revealed higher wildfire probabilities in specific areas, enabling the identification of regions with more than 50% likelihood of wildfire occurrence. These findings enhance understanding of wildfire dynamics in tropical ecosystems and offer insights for targeted wildfire prevention and management strategies in protected areas.

1 Escuela de Estadística, Universidad de Costa Rica  
2 Physical Geography Lab, School of Geography, University of Costa Rica  
3 Laboratorio de Investigacion e Innovación Tecnológica. Universidad Estatal a Distancia, Costa Rica.

✉ Correspondence: [Shu Wei Chou-Chen <shuwei.chou@ucr.ac.cr>](mailto:shuwei.chou@ucr.ac.cr)

# 1. Introduction

…

# 2. Methods

## 2.1 Data

Data description…

The explanatory variables, FA and TRI, are logarithmically transformed as to address their asymmetry and to avoid undefined values for . Subsequently, all covariates (DAH, logFA, LST, logLSF, SLOPE, logTRI, TWI, WE, and WExpo) are scaled using z-scores to enable the comparison of their standardized contributions.

## 2.2 Statistical model

The goal is to model a count variable (INC), which ranges from 0 to 6 and has a high percentage of zeros (). Therefore, we considered count models, such as Poisson and Negative Binomial regression, as well as the zero-inflated count regression models. There are different ways in R (R Core Team 2023) that implement zero-inflated count models throughout different packages (Zeileis, Kleiber, and Jackman 2008). We performed all the statistical modeling by using a flexible statistical framework called Generalized Additive Models for Location, Scale, and Shape (GAMLSS), since all these models are special cases within this model class. All the statistical analysis in this paper was conducted using the gamlss package (R. A. Rigby and D. M. Stasinopoulos 2005). Although this model allows for non-linear effects from the covariates, we retain the model with linear covariates for simplicity.

Let be the INC of location . The count regression models described above are represented by either one-, two-, or three-parameter GAMLSS models:

where 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, and are the regression coefficients of the covariate , measured at location , for each parameter function .

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 .

For the zero-inflated models, a finite mixture distribution of a point mass at zero and a count distribution is specified as follows:

where represents the probability of observing , and is the probability function of a count distribution (either Poisson or Negative Binomial). In this way, zero-inflated models add an additional parameter to the GAMLSS framework.”

That is, the Poisson Zero-Inflated model (ZIP) includes two parameters, with and . Finally, the Zero-Inflated Negative Binomial model (ZINB) fits three parameters, with , , and . Note that the second parameter of ZIP and the third parameter of ZINB model the probability that a given location has zero wildfires, similar to a logistic regression.

Additionally, due to the large number of data points (60,484 locations), we conducted the analysis using a training-testing approach (R. A. Rigby and D. M. Stasinopoulos 2005). In other words, we divided the data into two subsets: a training set comprising (42,213 locations), used for model development and selection, and a testing set comprising (18,271 locations), utilized for out-of-sample validation.

The analysis strategy and the selection of the final model are as follows. First, we considered the models’ mean : PO, NB, ZIP, and ZINB, with full specifications of the location function, while holding other parameters constant. Next, by applying the stepwise algorithm in both directions, we obtained reduced models (PO-red, NB-red, ZIP-red, and ZINB-red). Since the zero-inflated models yielded better results, in the third step, we applied the stepwise algorithm again to add to ZIP-red and both and to ZINB-red, resulting in the final models, ZIP-red-final and ZINB-red-final.”

Finally, the best model is selected based on several metrics and an assessment of model adequacy. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), along with generalized pseudo- (Cox-Snell and Cragg-Uhler), are computed for the training set, while the validation global deviance (VGD) is calculated for the testing set (out-of-sample validation).

# 3. Results

## 3.1 Descriptive analysis

Creo que vale la pena incluir una serie de análisis descriptivos.

## 3.2 Statistical model

[Table 1](#tbl-ICmodels) presents the goodness-of-fit measures for all fitted models. AIC and BIC evaluate the goodness of fit for the training set, while VGD assesses the performance of the fitted models in predicting the out-of-sample data (validation data). All three criteria indicate that the best models are ZIP\_red\_final and ZINB\_red\_final. However, the estimation of the ZINB\_red\_final model is difficult to converge, and diagnostics suggest that this model is not appropriate. Conversely, the generalized pseudo- (Cox-Snell and Cragg-Uhler) indicates that PO\_red yields the highest values among all models, but the assumptions of this model are not satisfied.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | |

The model quality of the ZIP\_red\_final model is assessed using normalized quantile (NQ) residual plots ([Figure 1](#fig-ZIPdiagnostics)) and the worm plot ([Figure 2](#fig-ZIPworm)), which is a detrended form of quantile residuals, similar to a Q-Q plot. Overall, this demonstrates the best performance, except for the quantile residuals against the fitted values, which show slight asymmetry across different fitted values.

|  |
| --- |
| Figure 1: Normalized quantile residuals analysis plots of the final fitted model (ZIP\_red\_final). |

|  |
| --- |
| Figure 2: Worm Plot |

After checking the model adequacy, [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 location We observe that all these variables are statistically significant, except for logTRI and WE in the (Mu) equation. DAH, LST, SLOPE, TWI, and WExpo negatively affect INC, while logLSF and the presence of sabana positively affect INC. On the other hand, the presence of sabana and LST positively influence the probability of not observing INC (), whereas WE, TWI, and WExpo negatively influence this probability.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | |

Finally, [Figure 3](#fig-ZIPmusigma) presents the fitted and , which represent INC mean of each location and the probability of no INC, respectively. Those figures are difficult to assess due to the fact that [Figure 3 (a)](#fig-ZIPmu) shows higher fitted INC mean for northern parts of the region. However, if we take into account that the model fits a location with , meaning this location has greater probability to not observe INC (), we can filter out those locations, and only consider those location with , that is, those places will have more than 50% of chance to have more than zero INCs, then plot the fitted ([Figure 4 (b)](#fig-ZIPmu_sigma)). We can observe that the places with higher INC count mean are similar to the observed INC (fig-INC).

Finally, [Figure 3](#fig-ZIPmusigma) presents the fitted and , which represent the mean INC count for each location and the probability of no INC, respectively. These figures can be challenging to interpret, as [Figure 3 (a)](#fig-ZIPmu) shows a higher fitted INC mean in the northern parts of the region, which have high probability to not observe INC. However, if we consider that the model fits a location with , indicating a greater probability of no INC, we can filter out those locations and focus only on those with . This means that these locations have more than a 50% chance of experiencing at least one INC. We can then plot the fitted ([Figure 4 (b)](#fig-ZIPmu_sigma)). Notably, the areas with higher mean INC counts are similar to the observed INC ([Figure 4 (a)](#fig-INC)).

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

Figure 3: 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 4: Observed wildfire counts and the model predicted fire count mean, conditional on a probability greater than 0.5 of occurrence of wildfire.

# 4. Conclusiones

texto…

# 5. Referencias

Alvarado, G. E., and A. Quesada-Román. 2024. “Anthropogenic Geomorphology of Costa Rica.” In *Landscapes and Landforms of Costa Rica*, 259–79. Cham: Springer International Publishing.

Arroyo-Solórzano, M., and A. Quesada-Román. 2024. “Morphotectonic Regions of Costa Rica: A Review and Updated Classification.” In *Landscapes and Landforms of Costa Rica*, 35–48.

Artavia-Rodríguez, L. G. 2011. “LAS SABANAS HÚMEDAS DEL SUR DE COSTA RICA. UNA CARACTERIZACIÓN FITOGEOGRÁFICA.” *Revista Geográfica de América Central* 2 (47E).

Artavia-Rodríguez, L. G., and G. Avalos. 2020. “Historia Natural Del Ecosistema de Sabana Del Valle Del Río General, Costa Rica.” *Revista Geográfica*, no. 161: 11–28.

Artavia-Rodríguez, L. G., R. Masis-Campos, and G. Ávalos-Rodríguez. 2024. “Análisis Espacio-Temporal Del Régimen de Fuego En Las Sabanas Húmedas Del Pacífico Sur de Costa Rica Mediante Datos MODIS 2001-2022.” *Revista de Geografía Norte Grande*, no. 87.

Camacho, M. E., A. Quesada-Román, R. Mata, and A. Alvarado. 2020. “Soil-Geomorphology Relationships of Alluvial Fans in Costa Rica.” *Geoderma Regional* 21: e00258.

Camacho-Umaña, M. E., A. Quesada-Román, M. Villatoro-Sánchez, B. Alemán-Montes, R. Mata, C. Henríquez-Henríquez, and A. Alvarado. 2024a. “Agricultural Landscapes of Costa Rica.” In *Landscapes and Landforms of Costa Rica*, 221–58. Cham: Springer International Publishing.

———. 2024b. “Highly Weathered Soil Landscapes of Costa Rica.” In *Landscapes and Landforms of Costa Rica*, 431–51. Cham: Springer International Publishing.

Granados-Bolaños, S., N. Surian, C. Birkel, G. E. Alvarado, A. Quesada-Román, J. P. Galve, and A. Bonilla-Mata. 2024. “Fluvial Landscapes of Costa Rica: An Overview of Dynamic Rivers.” In *Landscapes and Landforms of Costa Rica*, 91–111.

Kariminejad, N., M. Jafari, F. Domazetović, and A. Quesada-Román. 2024. “An Overview of the Importance of DEM Resolution in Soil Erosion Assessment.” *Papers in Applied Geography*, 1–10.

Mejía-Agüero, O., and A. Quesada-Román. 2024. “Geoheritage Assessment for the Geoconservation and Geotourism Promotion of in the Altamira Sector of the La Amistad International Park, Costa Rica.” *Geoheritage* 16 (4): 95.

Pérez-Umaña, D., and A. Quesada-Román. 2024. “Geoheritage, Geoconservation, and Geotourism Promotion in Costa Rica.” In *Landscapes and Landforms of Costa Rica*, 281–93. Cham: Springer International Publishing.

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.

Zeileis, Achim, Christian Kleiber, and Simon Jackman. 2008. “Regression Models for Count Data in r.” *Journal of Statistical Software* 27 (8): 1–25. <https://doi.org/10.18637/jss.v027.i08>.