

# Bayesian Spatio-Temporal Modeling of Fire Spots in the Amazônia Biome

Jan Fritz, Niklas Viebig, Luca Titze, Victor Windhab

Autumn Semester 2024

## Abstract

This project employs Bayesian statistical methods to investigate the spatio-temporal dynamics of fire-spots in the Amazônia biome over a 10-year period (January 1, 2012, to December 31, 2021). Bayesian analysis allows for the integration of diverse data sources, handling uncertainty and variability inherent to complex ecological systems. This report was motivated by Pimentel, Bulhões, and Rodrigues 2024 which used a spatio-temporal approach consisting of a hierarchical high dimensional model. We successfully reproduced their model. To address the computational expense caused by the model's high dimensionality, we simplified it by omitting the auto-correlated term, reducing dimensionality by 99.56%. Our analysis revealed that temperature is the primary driver of wildfires among the considered covariates, while humidity serves as the most significant mitigating factor.

## Introduction

The Amazônia biome<sup>1</sup>, home to the world's largest rain-forest, is a key ecosystem facing growing threats from wildfires, fueled by climate change and human-driven land-use changes. Understanding the causes and factors influencing these fires is crucial for the effective conservation of this vital biome. Bayesian data analysis, with its capacity to manage uncertainty and incorporate diverse information sources, provides an ideal framework for examining the various contributing variables. Recent work, such as Pimentel, Bulhões, and Rodrigues 2024, highlights the use of Bayesian spatio-temporal models in identifying key variables, including meteorological conditions and land-use transitions. Building on this foundation, our project aims to understand fire-spot occurrences in the Amazônia biome over a 10-year period, with a focus on their spatio-temporal patterns and driving factors.

## Understanding the data

The dataset used integrates geographic, meteorological, and land-use transition data for 588 municipalities, all contained within the Amazônia biome across 120 months of observations. We utilized the finalized and cleaned dataset prepared by Pimentel, Bulhões, and Rodrigues 2024.

**Fire-spot Data:** Collected by the Brazilian National Institute for Space Research (INPE) using satellite images (AQUA M-T), spanning from 2012 to 2022, and detecting approximately 2.2 million fire-spots (FS). Data includes fire locations, dates, biomes, municipalities, and fire risk. Challenges such as missing values and sensor coarseness were addressed by daily summation of fire-spots, associ-

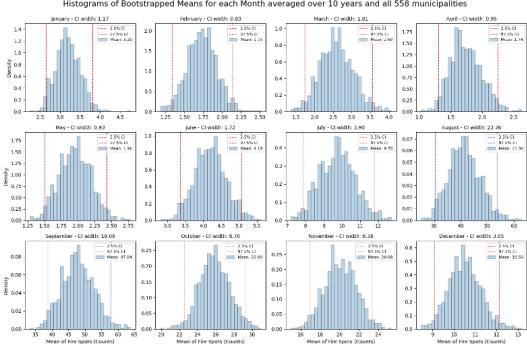
ating them with municipalities and assigning days without fire as zero.

**Meteorological Data:** Provided by the Brazilian National Institute of Meteorology (INMET) with hourly records from 2012 to 2021, covering precipitation (PREC), air temperature (TEMP), humidity (HUMID), and wind speed (SPEED). The station count increased from 468 to 588 during the period. Missing data issues were mitigated by calculating monthly averages, distance-weighted extrapolation from nearby stations, and imputing values using the exponential weighted moving average method.

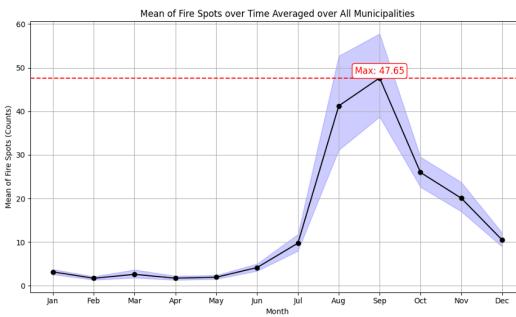
**Land Use Data:** Sourced from the Brazilian Annual Land Use and Land Cover Mapping Project (MapBiomas), tracking annual land-use transition (LUT) from 2011 to 2021, categorized into six classes. The data was interpolated to generate monthly values, assuming uniform distribution.

Figure 1 illustrates the seasonal trends in fire-spot occurrences averaged over the ten-year observational period. The data shows a significant increase in fire-spot activity during the months of August and September. The credible intervals during these months are noticeably wider, indicating greater variability in fire-spot occurrences across different municipalities. We can see that certain areas are more heavily impacted by wildfires during this season, emphasizing the spatial distribution of fire activity across the biome. This highlights the importance of employing a spatio-temporal modeling approach to get deeper insights into and predict these patterns across both regions and seasons.

<sup>1</sup>Pimentel, Bulhões, and Rodrigues 2024 define the Amazônia biome as being roughly comprised of municipalities within the Amazônia Legal: Amazonas, Acre, Rondônia, Pará, Amapá, Mato Grosso, and Roraima.



(a) Bootstrapped means of the count of fire-spots over the ten year observation .



(b) The graph shows the mean of the means of the bootstrapped count of fire-spots over the ten year period.

Figure 1: Analyzed monthly fire-spot averages with larger credible intervals in August and September, suggesting certain municipalities were more severely affected by wildfires during these months.

## Correlations

With the large number of covariate parameters, it is important to first filter for correlations in our data. The weather data used in this analysis is not the only available data, but for simplicity, we restricted our analysis to the five covariates: fire-spot count, humidity, land-use transition, temperature, precipitation, and wind speed, as mentioned in the paper Pimentel, Bulhões, and Rodrigues 2024. Among these covariates, we observed lower correlations with the count of fire-spots and stronger correlations between some covariates, as shown in Figure 2. For example, temperature showed a relatively strong negative correlation with humidity, while humidity showed a positive correlation with precipitation. A step toward simplifying the model could therefore involve excluding parameters like precipitation and wind speed in the fitting process, as their weaker relationships with fire-spot count suggest limited predictive value.

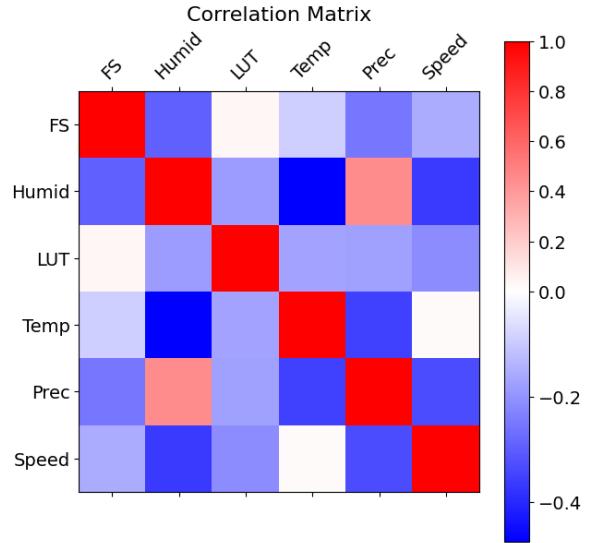


Figure 2: The Figure depicts the correlations between our covariate parameters. We used the Pearson number for this analysis over the whole dataset.

We also applied the Kendall Tau and the Spearman Number, as we saw indications for non-linear correlations but were not able to confirm important deviations overall, as seen in Figure 3.

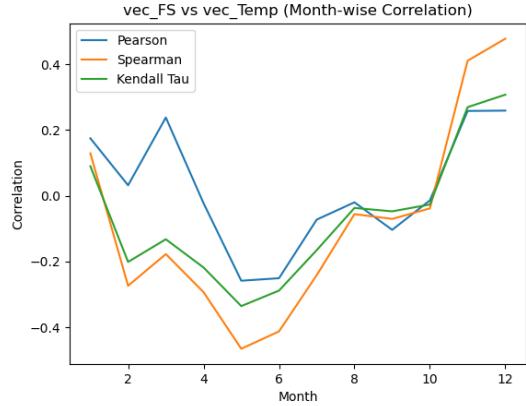


Figure 3: The correlation between fire-spots and temperature in an average year.

## Fisher Information

The likelihood of our data is too difficult to compute analytically. We thus needed to use the following estimator of the fisher information:

$$F_{ij} = \frac{\partial^2 \log(p(x|\vec{\theta}))}{\partial \theta_i \partial \theta_j}. \quad (1)$$

In our case the likelihood is  $p(X|\vec{\theta}) = \prod_{n=1}^{66960} \frac{\mu_n^{X_n^0} e^{-\mu_n}}{X^0!}$  with  $\mu_n = o_n e^{\vec{\theta}^T X_n}$  which leads

to the fisher information matrix

$$F_{ij} = \sum_{n=1}^{66960} o_n \vec{X}_n \vec{X}_n^T e^{\vec{\theta}^T \cdot \vec{X}}. \quad (2)$$

Here  $\vec{X}_n$  refers to one row of data containing the values for a certain municipality in a certain month. For the whole dataset we get very high values ( $\sim 10^5$ ), which could be a predictor for the low standard deviation in our model and as seen in the referenced paper.

## Methods

To estimate the expected number of fire-spots in Amazônia for each municipality for each month, we apply a Bayesian spatio-temporal generalized linear mixed model. This allows us to consider both spatial and temporal variations, while also incorporating explanatory variables.

For the observed number of fire-spots  $Y_{i,t}$  for municipality  $i$  at time  $t$  we assume a Poisson distribution:

$$Y_{i,t} \mid \mu_{i,t} \sim \text{Poisson}(\mu_{i,t}), \quad \mu_{i,t} = o_{i,t} \lambda_{i,t},$$

where  $o_{i,t}$  denotes the offset (e.g., municipality area) and  $\lambda_{i,t}$  represents the fire risk relative to the offset, such that  $\lambda_{i,t}$  is the expected number of fire-spots. This is motivated by the fact that the to be modeled data is counting data which are not bounded by an upper limit, excluding a binomial distribution from consideration.

The relative risk  $\lambda_{i,t}$  is expressed as a linear combination of known and unknown covariates  $\mathbf{x}_{i,t} = (1, x_{1;i,t}, x_{2;i,t}, \dots)$  and  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots)$ :

$$\ln(\lambda_{i,t}) = \mathbf{x}_{i,t}^\top \boldsymbol{\beta} + \psi_{i,t},$$

where  $\mathbf{x}_{i,t}$  includes covariates such as land-use transitions ( $LUT_{i,t}$ ), temperature ( $TEMP_{i,t}$ ), and humidity ( $HUMID_{i,t}$ ), while  $\psi_{i,t}$  accounts for spatio-temporal random effects. These effects are modeled with a linear time trend:

$$\psi_{i,t} = \phi_i + (\alpha + \delta_i) \cdot \frac{t - \bar{t}}{T},$$

where  $\phi_i$  captures spatial variability from the baseline risk  $\beta_0$ , and  $\delta_i$  the variations from the local temporal trend  $\alpha$ , per municipality.

The parameters are assigned priors as follows:

$$\begin{aligned} \boldsymbol{\beta} &\sim \text{Normal}(\mu_\beta, \Sigma_\beta), \\ \alpha &\sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2) \\ \phi &\sim \text{CAR}(W, \rho_{\text{int}}, \tau_{\text{int}}^2) \\ \delta &\sim \text{CAR}(W, \rho_{\text{slo}}, \tau_{\text{slo}}^2) \\ \rho_{\text{int}} &\sim \text{Uniform}(0, 1), \\ \rho_{\text{slo}} &\sim \text{Uniform}(0, 1), \\ \tau_{\text{int}}^2 &\sim \text{Inverse-Gamma}(a, b) \\ \tau_{\text{slo}}^2 &\sim \text{Inverse-Gamma}(a, b) \end{aligned}$$

where the deviations  $\phi$  and  $\delta$  follow a CAR prior (see A.1), which allows the modeling of spatial dependencies and neighborhood structure of the municipalities. Since the CAR prior relies on parameters itself, it makes the model hierarchical. For the uninformed priors we chose:  $\mu_\beta = 0$ ,  $\Sigma_\beta = 10^5$ ,  $\mu_\alpha = 0$ ,  $\sigma_\alpha^2 = 1000$ ,  $a = 1$  and  $b = 0.01$ .

The structure of the models allows to identify drivers and reduces of the fire-spot risk, since for some increase  $\xi$  in a covariate  $x_i$  the relative risk is

$$\begin{aligned} RR(x_i; \xi) &= \frac{\text{FS risk with } x_i + \xi}{\text{FS risk with } x_{i,t,j}} \\ &= \exp(\beta_j \xi). \end{aligned}$$

a positive  $\beta$  would make the corresponding covariant a driver of fire-spot risk and a negative  $\beta$  a reducer.

Due to the high dimensionality of this model (1125) we decided to reduce this model's dimensionality by not considering the deviations  $\phi$  and  $\delta$ , leading not only to a model with dimensionality 5 but also making the new model non-hierarchical, reducing the overall complexity of the new model considerably. We will refer to the model with higher dimensionality as 'large model' and the one with reduced dimensionality as 'simplified model'.

To sample the posterior of the two models a NUTS sampler was used due to the high dimensionality of the large model. The implementation of python package *PyMC* was used with a tune size of 500 samples and sample size of 1000 for 4 walkers each.

To get a quantitative comparison of the two models we compute the WAIC factor for each (see A.2).

# Results and Discussion

## Descriptive Analysis

The Amazônia biome experiences significant wild-fire activity, with a peak in August, as shown in Figure 4. Both the large and simplified models successfully captured this temporal trend, with both peaking slightly earlier in July at an average count of approximately 30 fire-spots. To analyze the spatial aspect of our models, we averaged the fire-spot counts over the 120 months, as illustrated in Figure 26. Notably, both models closely align with the observed averages and effectively capture the substantial differences between municipalities. Remarkably, the simplified model, despite excluding the auto-correlation component of the large model, is still capable of modeling significant differences between neighboring municipalities.

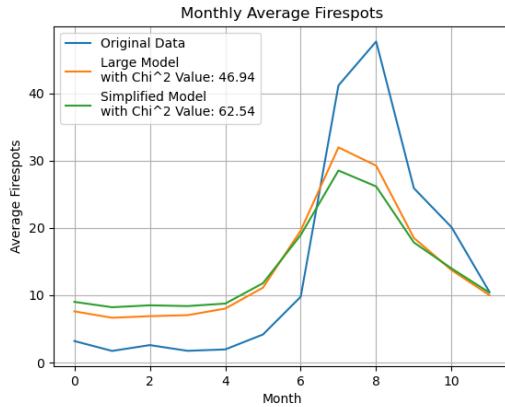


Figure 4: Monthly average fire-spots showing the general seasonal trend.

## Modeling Results

### Large Model

Key meteorological factors contributing to the fire activity can be identified from the sampling results of the large model's posterior, see Table 1 and Figure 1. We find that temperature ( $\beta_{TEMP} = 0.2168$ ) increases the fire-spot risk, while humidity ( $\beta_{HUMID} = -0.0623$ ) reduces it and land-use transition has no significant effect. This is in acceptance with our analysis of the data, see Figure 2.

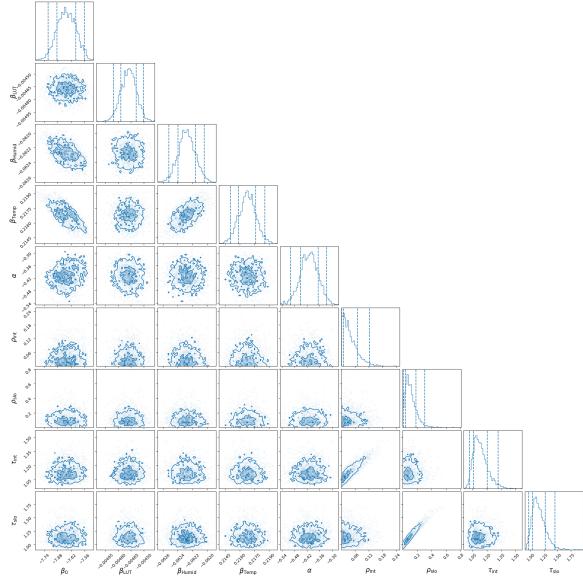


Figure 5: Corner plots of the large model parameters using the sampling of the large model 24

Parameter	Mean	95% CI
$\beta_0$	-7.6412	[-7.7224, -7.5607]
$\beta_{LUT}$	-0.0047	[-0.0049, -0.0045]
$\beta_{HUMID}$	-0.0623	[-0.0625, -0.0620]
$\beta_{TEMP}$	0.2168	[0.2150, 0.2186]
$\alpha$	-0.4099	[-0.4997, -0.3254]
$\tau_{int}$	1.1337	[1.0193, 1.3190]
$\tau_{slo}$	1.1558	[0.9631, 1.4552]
$\rho_{int}$	0.0396	[0.0014, 0.1228]
$\rho_{slo}$	0.1134	[0.0114, 0.3066]

Table 1: Summary of model parameter estimates and their respective credible intervals from sampling (24)

### Simplified Model

Even though we have simplified our model significantly we still find agreeing values for the correlations as seen in the corner plot 6 and Table 2.

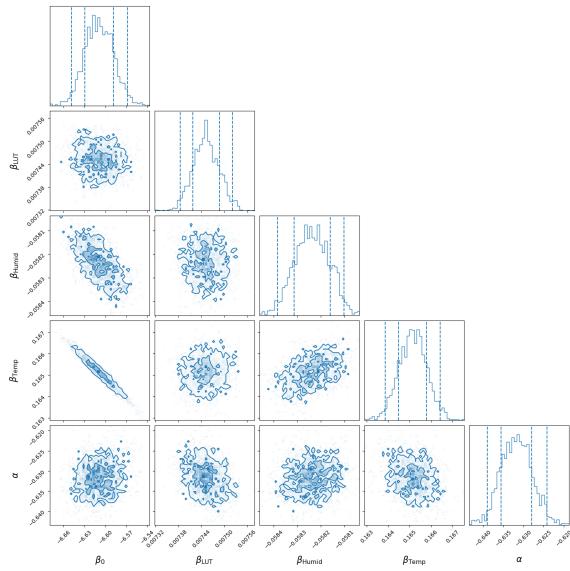


Figure 6: Corner plots of the simplified model parameters (Sampling: 25).

Parameter	Mean	95% CI
$\beta_0$	-6.6091	[-6.6490, -6.5683]
$\beta_{LUT}$	0.0075	[0.0074, 0.0075]
$\beta_{HUMID}$	-0.0582	[-0.0584, -0.0581]
$\beta_{TEMP}$	0.1651	[0.1639, 0.1664]
$\alpha$	-0.6317	[-0.6390, -0.6242]

Table 2: Summary of model parameter estimates and their respective credible intervals.

## Interpretation of Findings

We were able to build the large hierarchical spatial-temporal model used in the referenced paper and implement our own Bayesian statistical methods to find mean values similar to the literature and their variance comparable using the 95% Credible intervals from our sampling. In addition, we were able to compare the model with a simplified one. Using the WAIC (Widely Applicable Information Criterion), we generated the results summarized in Table 3, which compares the predictive performance and complexity of the large and simplified models.

Metric	L Model	S Model
elpd_waic	-805 858.52	-1 208 581.91
p_waic	48 417.23	1 807.82

Table 3: WAIC results for the Large Model and Simplified Model.

The large model has a less negative elpd\_waic (-805858.52) compared to the simplified model (-1208581.91). Comparison with the effective number of parameters (p\_waic = 48417.23 for the

large model and p\_waic = 1807.82 for the simplified model) suggests that the simplified model yields better results for the amount of complexity it contains.

## Future Directions

For further analysis, we could explore other predictive models for risk assessment. Integrating additional variables, such as vegetation indices that could improve the model. In addition, extending the analysis onto the spread of fire-spots using the meteorological data as well as vegetation information of the municipalities could be deemed valuable. Lastly, a long-term focus on policy changes in municipalities could investigate the fire dynamics. Especially in the context of climate change, this analysis highlights that higher temperatures significantly drive wildfire activity, making it crucial to consider temperature trends and their interaction with policy changes when addressing fire dynamics.

## Conclusion

By focusing on the Amazônia biome, this study provides critical insights into the drivers of wildfire activity in one of the world’s most vital ecosystems. By employing Bayesian spatio-temporal modeling, we were able to analyze the complex relationship of meteorological and land-use factors over a 10-year period. Our findings indicate that temperature is the primary driver of wildfire activity, while humidity acts as the most significant mitigating factor. The application of both a large hierarchical model and a simplified model allowed us to balance predictive accuracy with computational efficiency. The simplified model, despite its reduced complexity, produced results that closely aligned with the large model, making it a practical choice for further analyses. Future research should explore the integration of additional variables, such as vegetation indices, and consider the spread dynamics of fire-spots. Moreover, a long-term focus on policy changes and their impact on fire dynamics could provide deeper insights into managing wildfire risks.

## Contribution Statement

The workflow for this project was divided into two main components:

- Understanding the data and building a likelihood model.
- Implementing the model, sampling the posterior, and assessing goodness of fit.

Luca Titze and Niklas Viebig focused on the first component, while Jan Fritz and Victor Windhab concentrated on the second. All team members collaborated on writing and refining the report and the complementary presentation. All code can be found on GitHub: Bayesian SpatioTemporal Modeling.

## References

Pimentel, Jonatha Sousa, Rodrigo S. Bulhões, and Paulo Canas Rodrigues (2024). “Bayesian spatio-temporal modeling of the Brazilian fire spots between 2011 and 2022”. In: *Scientific Reports* 14, p. 21616. DOI: 10.1038/s41598-024-70082-6. URL: <https://www.nature.com/articles/s41598-024-70082-6>.

## A Appendix

### A.1 CAR Prior

The Conditional Auto-Regressive (CAR) for the two deviations  $\phi_{i,t}$  and  $\delta_{i,t}$  is formulated the following:

$$\begin{aligned}\phi_i | \phi_{-i}, W &\sim \text{Normal} \left( \frac{\rho_{\text{int}} \sum_{i' \neq i} w_{i,i'} \phi_{i'}}{\rho_{\text{int}} \sum_{i' \neq i} w_{i,i'} + 1 - \rho_{\text{int}}}, \frac{\tau_{\text{int}}^2}{\rho_{\text{int}} \sum_{i' \neq i} w_{i,i'} + 1 - \rho_{\text{int}}} \right), \\ \delta_i | \delta_{-i}, W &\sim \text{Normal} \left( \frac{\rho_{\text{slo}} \sum_{i' \neq i} w_{i,i'} \delta_{i'}}{\rho_{\text{slo}} \sum_{i' \neq i} w_{i,i'} + 1 - \rho_{\text{slo}}}, \frac{\tau_{\text{slo}}^2}{\rho_{\text{slo}} \sum_{i' \neq i} w_{i,i'} + 1 - \rho_{\text{slo}}} \right).\end{aligned}$$

where  $W$  is the neighborhood matrix, capturing the adjacency of municipalities: municipalities  $i$  and  $j$  are adjacent to each other if  $W_{i,j} = 1$ , and not if  $W_{i,j} = 0$ . The parameters  $\rho_{\text{int}}$ ,  $\rho_{\text{slo}}$ ,  $\tau_{\text{int}}$  and  $\tau_{\text{slo}}$  follow the priors

$$\begin{aligned}\rho_{\text{int}} &\sim \text{Uniform}(0, 1), & \rho_{\text{slo}} &\sim \text{Uniform}(0, 1), \\ \tau_{\text{int}}^2 &\sim \text{Inverse-Gamma}(a, b), & \tau_{\text{slo}}^2 &\sim \text{Inverse-Gamma}(a, b).\end{aligned}$$

where  $a = 1$  and  $b = 0.01$ .

### A.2 WAIC Factor

The WAIC factor is calculated as

$$\text{WAIC} = -2 \left( \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i | \theta^{(s)}) \right) - \sum_{i=1}^n \text{Var}_{s=1}^S (\log p(y_i | \theta^{(s)})) \right),$$

where:

- $n$  is the number of observations,
- $S$  is the number of posterior samples,
- $p(y_i | \theta^{(s)})$  is the likelihood of observation  $y_i$  given the  $s$ -th posterior sample  $\theta^{(s)}$ ,
- $\text{Var}_{s=1}^S (\log p(y_i | \theta^{(s)}))$  is the posterior variance of the log-likelihood for  $y_i$ .

and evaluates how well the model predicts the data, considering the models complexity.

### A.3 Sample Mean of the Incremental Intercept Parameter $\phi$ and the Incremental Slope Parameter $\delta$

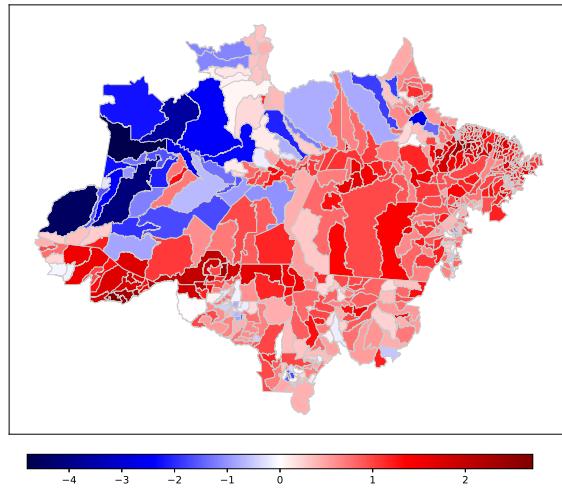


Figure 7: The sample mean of the incremental intercept parameter  $\phi$ .

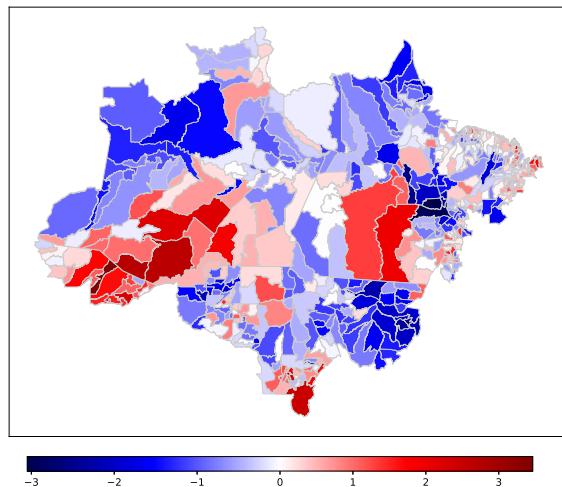


Figure 8: The sample mean of the the incremental intercept parameter  $\delta$ .

## A.4 Correlations

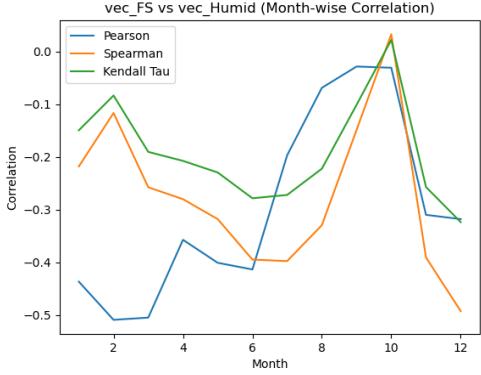


Figure 9: Correlation between `vec_FS` and `vec_Humid`.

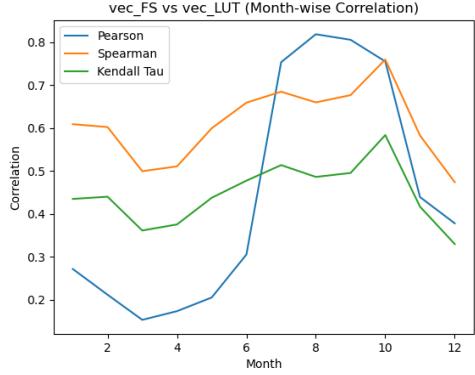


Figure 10: Correlation between `vec_FS` and `vec_LUT`.

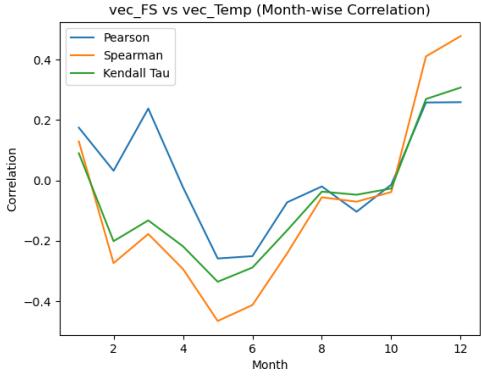


Figure 11: Correlation between `vec_FS` and `vec_Temp`.

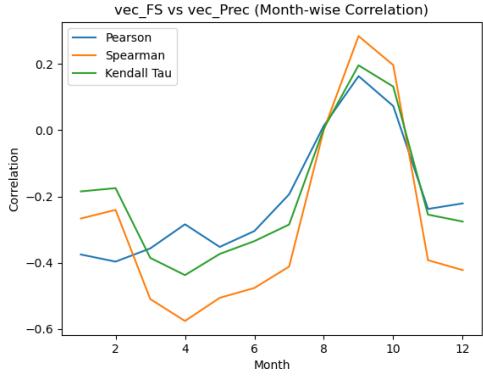


Figure 12: Correlation between `vec_FS` and `vec_Prec`.

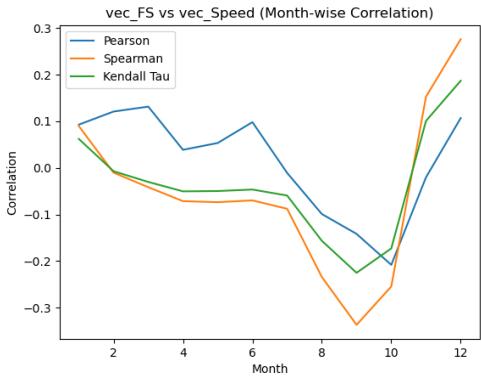


Figure 13: Correlation between `vec_FS` and `vec_Speed`.

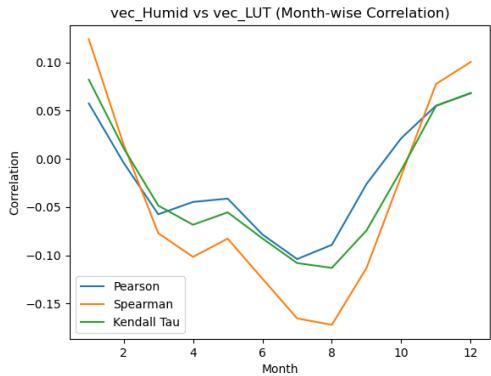


Figure 14: Correlation between `vec_Humid` and `vec_LUT`.

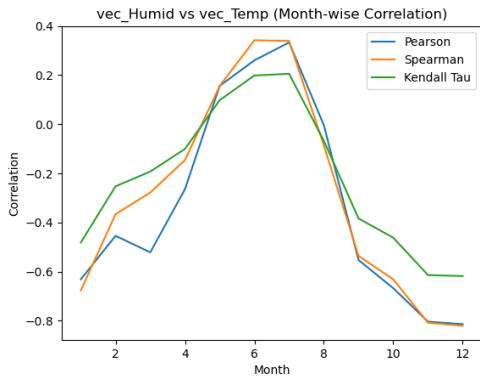


Figure 15: Correlation between `vec_Humid` and `vec_Temp`.

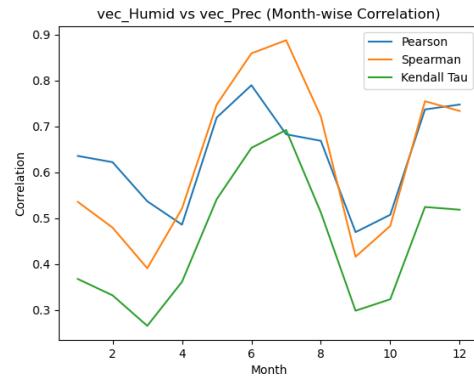


Figure 16: Correlation between `vec_Humid` and `vec_Prec`.

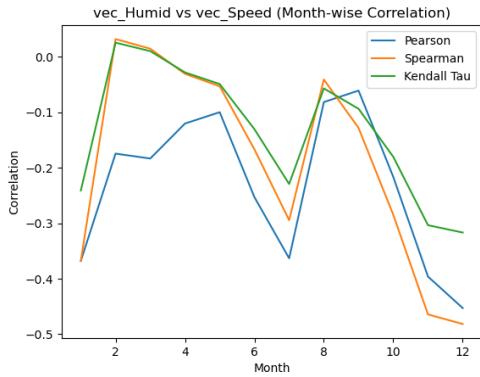


Figure 17: Correlation between `vec_Humid` and `vec_Speed`.

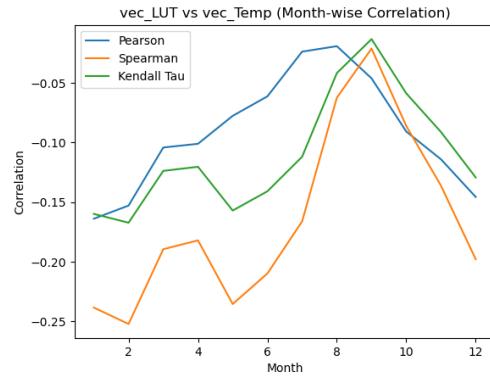


Figure 18: Correlation between `vec_LUT` and `vec_Temp`.

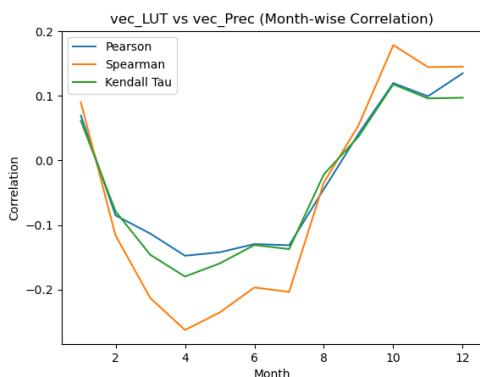


Figure 19: Correlation between `vec_LUT` and `vec_Prec`.

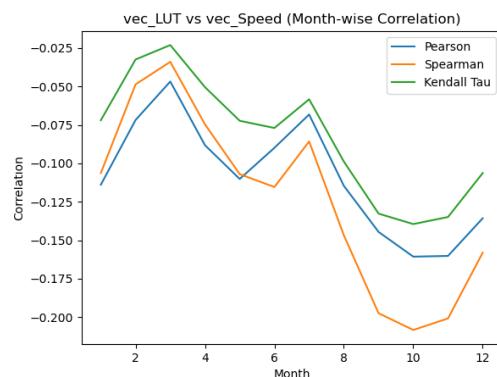


Figure 20: Correlation between `vec_LUT` and `vec_Speed`.

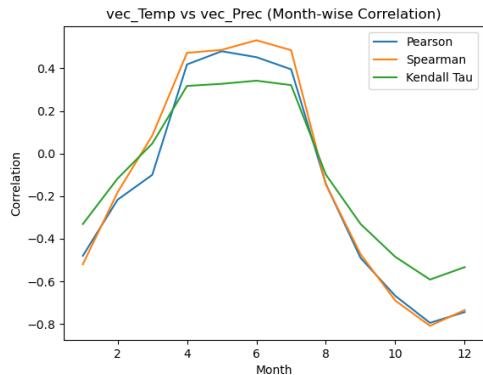


Figure 21: Correlation between `vec_Temp` and `vec_Prec`.

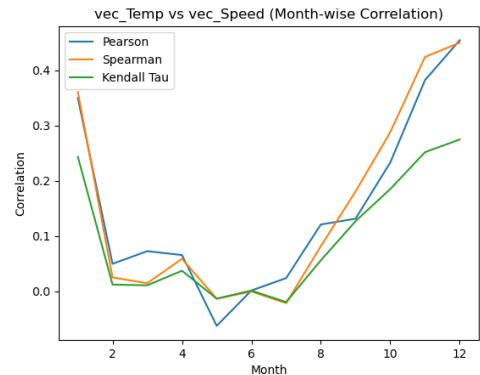


Figure 22: Correlation between `vec_Temp` and `vec_Speed`.

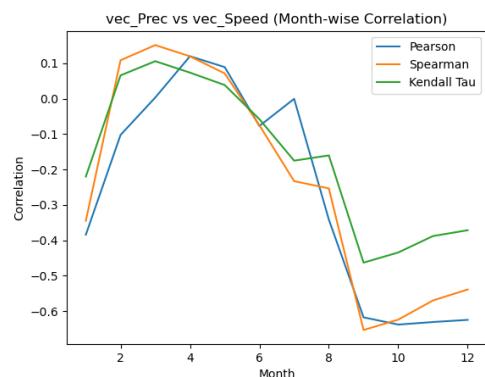


Figure 23: Correlation between `vec_Prec` and `vec_Speed`.

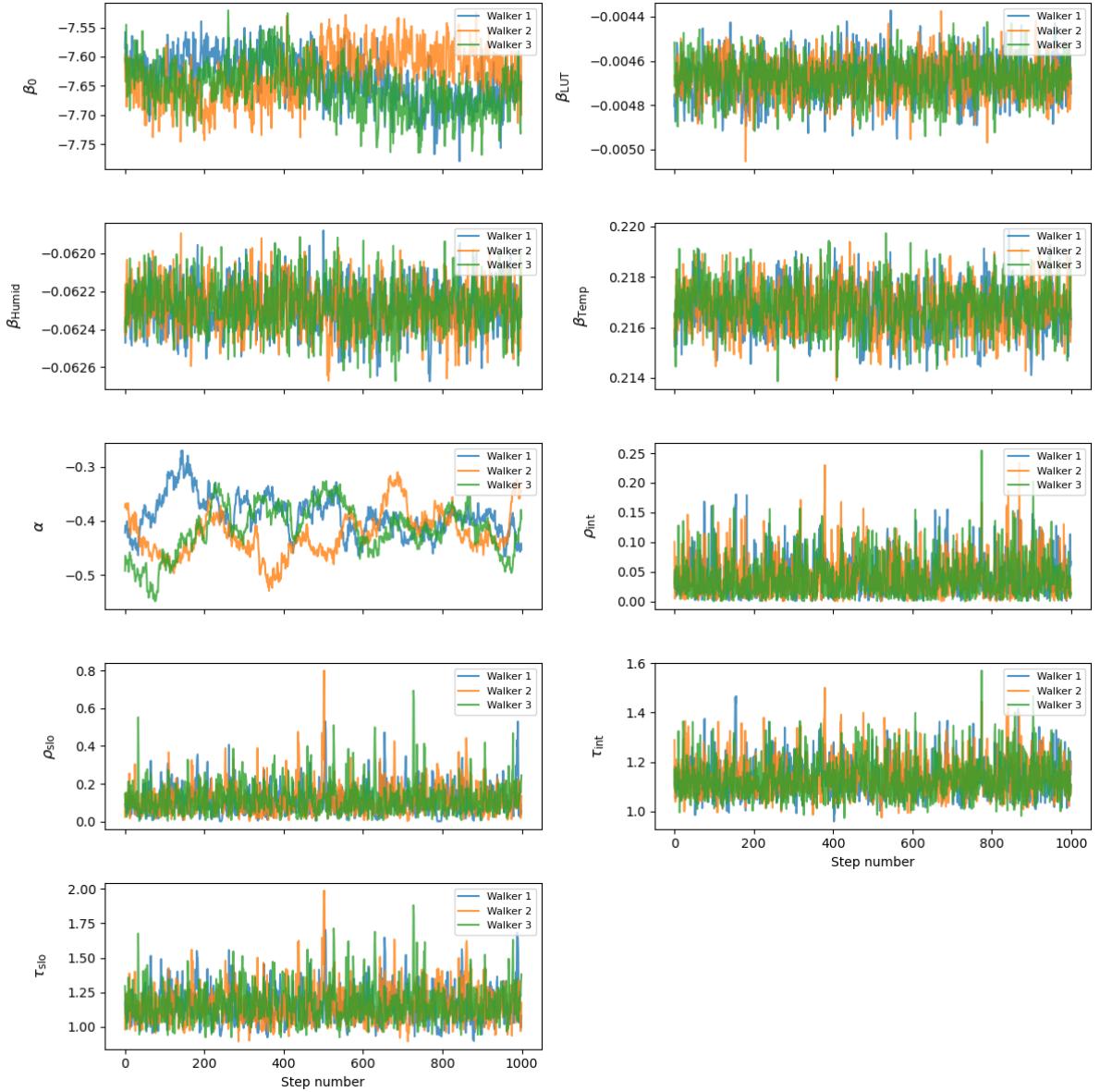


Figure 24: Sampling of large model.

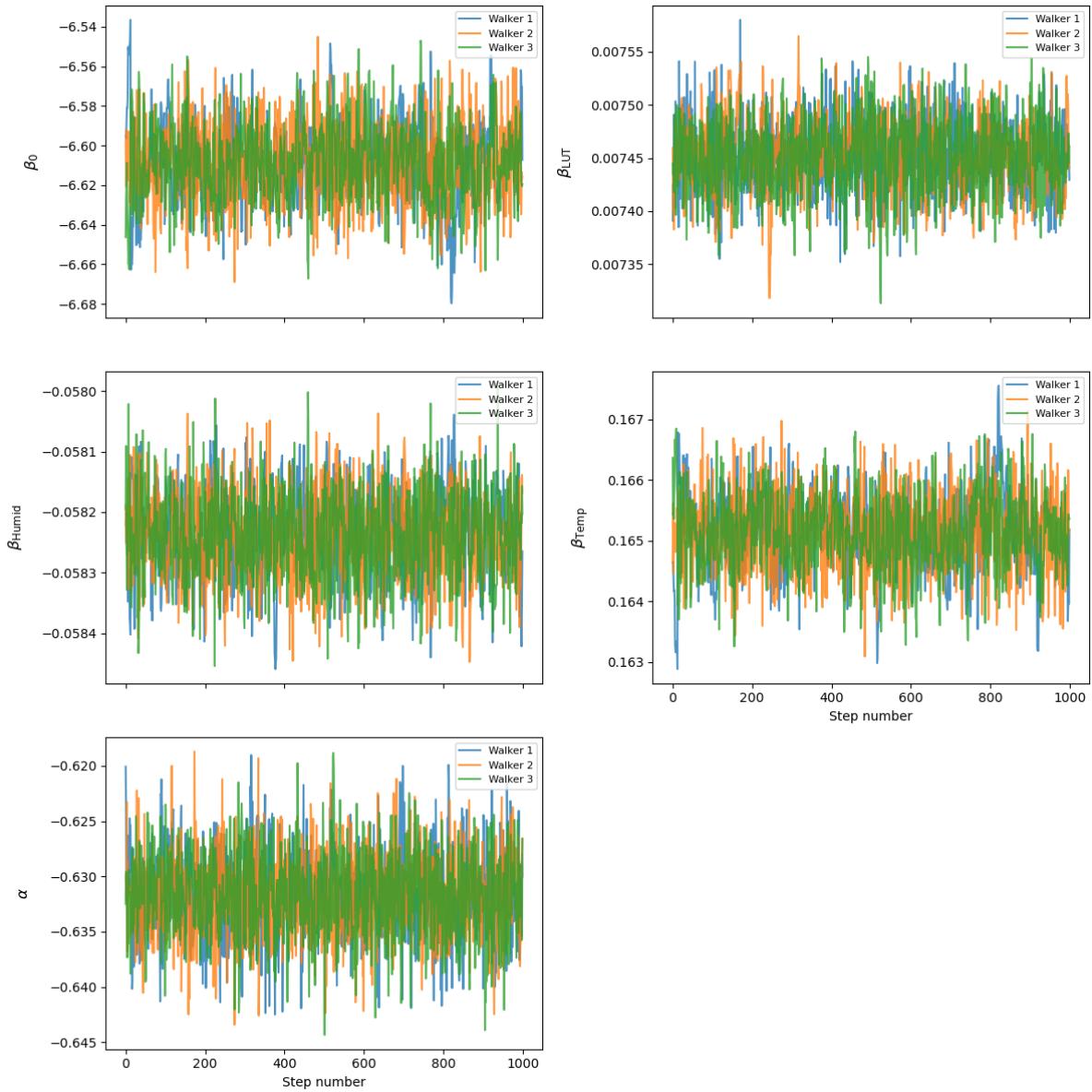
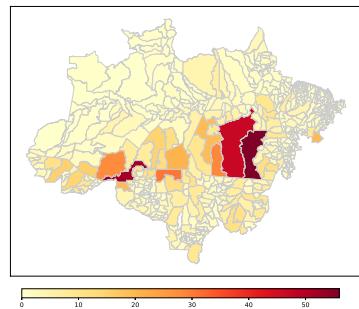
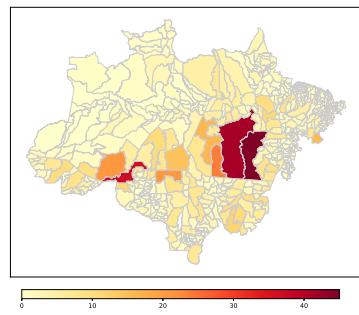


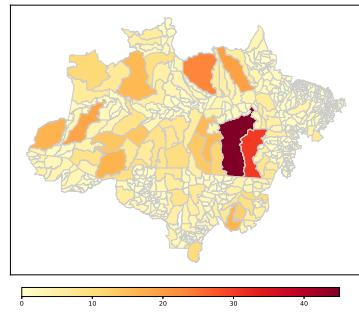
Figure 25: Sampling of simplified model.



(a) Observed data



(b) Large model



(c) Simplified model

Figure 26: 120-month average of the fire-spot risk for the two models and the observed data for every municipality