Bayesian Spatio-Temporal Modeling of Fire Spots in the Amazonia 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 Amazonia biome over a 12-year period (2011–2022). Bayesian analysis allows for the integration of diverse data sources, handling uncertainty and variability inherent to complex ecological systems. By incorporating meteorological variables and land-use transitions, the report identifies key drivers of wildfire activity and provides robust insights into the patterns and mechanisms underlying fire dynamics.

Introduction

The Amazonia biome, home to the world's largest rainforest, is a key ecosystem facing growing threats from wildfires, fueld by climate change and human-driven land-use changes. Understanding the drivers of these fires is essential for effective management and conservation. Bayesian data analysis, with its ability to handle uncertainty and integrate diverse sources of information, offers a framework for studying the complex spatio-temporal dynamics of wildfires. 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 Amazonia biome over a 12-year period, with a focus on their spatio-temporal patterns and driving factors.

Understanding the data

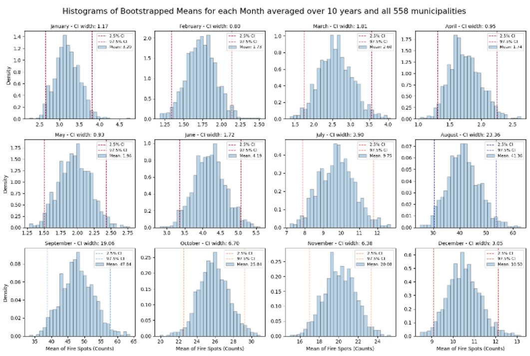
The dataset used integrates geographic, meteorological, and land-use transition data for 588 municipalities, all contained within the Amazônia biome, across 120 months of observations. We utilized the finalized and cleaned dataset prepared by Pimentel, Bulhões, and Rodrigues 2024.

Fire Spot Data: Collected by the Brazilian National Institute for Space Research (INPE) using satellite images (AQUA M-T), spanning from 2011 to 2022, and detecting approximately 2.2 million fire spots. 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, associating them with municipalities and assigning days without fire as zero.

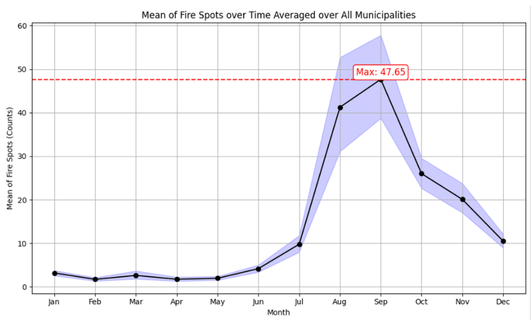
Meteorological Data: Provided by the Brazilian National Institute of Meteorology (INMET) with hourly records from 2012 to 2021, covering precipitation, air temperature, humidity, and wind speed.

The station count increased from 468 to 588 over 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.



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



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

Figure 1: We analyzed the mean of the number of fire spots to get an understanding in which months the most fire spots occurred. The by comparison rather large confidence intervals in the months August and September led us to believe that especially in these months certain municipalities were affected more strongly by wildfires than others

Correlations

With this large amount of covariate parameters it is necessary to first filter for correlations in our data. The weather data we used in this analysis is not the only available data, but for simplicity, we left out the less correlated weather data. But even in the 5 covariates we chose we get lower correlation with our Firespots and also strong correlations with other covariates. A step to a simplified model could therefore be to neglect those in the fitting process.

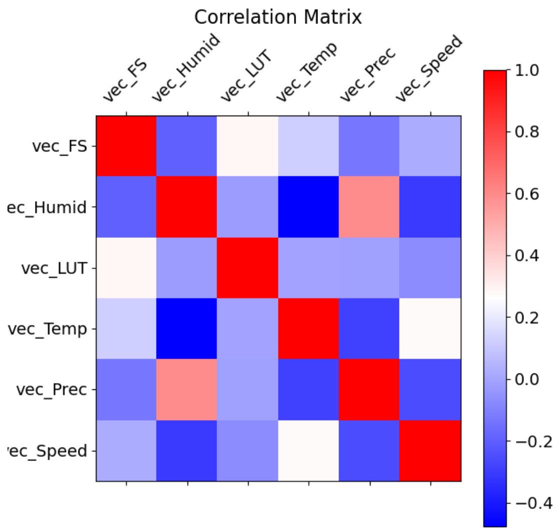


Figure 2: The figure depicts the correlations between our covariate parameters. We used the Pearson number for this analysis over the whole dataset.

We were not sure about the exact nature of our correlations and saw some indication, that there are non-linear correlations.

We therefore also applied the Kendall Tau and the Spearman Number, but were not able to confirm important deviations overall.

Fisher Information

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

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

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

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

Here $\vec{X}_n$ refers to one row of data.
For the whole dataset wee 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

Modeling Framework

To estimate the expected number of fire spots in Amazonia, we employ a Bayesian spatio-temporal generalized linear mixed model. The observed number of fire spots $Y_{i,t}$ for municipality i at time t follows 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. The log of the relative risk is expressed as a linear combination of covariates:

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

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 ϕ_i captures spatial variability, and α and δ_i represent global and local temporal trends, respectively. The parameters are assigned priors as follows:

$$\beta \sim \text{Normal}(\mu_\beta, \Sigma_\beta), \quad \alpha \sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2).$$

Posterior distributions of the parameters are estimated using Markov Chain Monte Carlo (MCMC) sampling. For each municipality i, the expected number of fire spots is predicted as:

$$\mu_{i,t}^{(k)} = \exp\Bigg(\ln(o_{i,t}) + \beta_0^{(k)} + \sum_{j=1}^{p-1} \beta_j^{(k)} x_{i,t,j} + \phi_i^{(k)} + (\alpha^{(k)} + \delta_i^{(k)}) \cdot \frac{t - \bar{t}}{T} \Bigg), \tag{3}$$

and the posterior mean is obtained as the average of the MCMC samples. The relative risk associated with a unit change in covariate $x_{i,t,j}$ is computed as:

$$RR(x_{i,t,j}; \xi) = e^{\beta_j \cdot \xi},$$

where ξ typically corresponds to the standard deviation of the covariate. Model fit is assessed through residual analysis, with Pearson residuals calculated as:

$$r_{i,t}^* = \frac{y_{i,t} - \mu_{i,t}}{\mu_{i,t}}.$$

Spatially aggregated residuals are used to evaluate the model's overall performance.

Results and Discussion

Descriptive Analysis

The Amazonia biome exhibited significant patterns in fire activity, with notable peaks observed in specific years and seasonal trends.

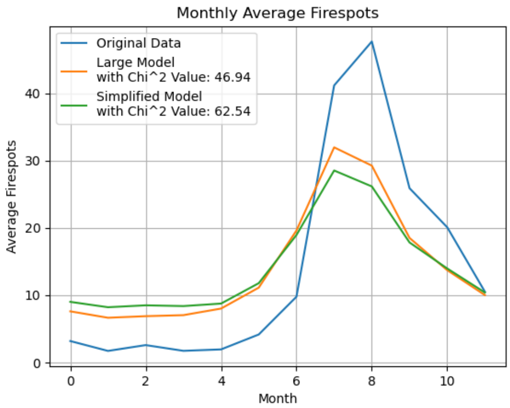


Figure 4: Monthly average Firespots showing peaks between August and November.

Modeling Results

Large Model

Key meteorological factors contributing to the fire activity were identified with the computation and sampling of the large spatial-temporal model including the weight matrix.

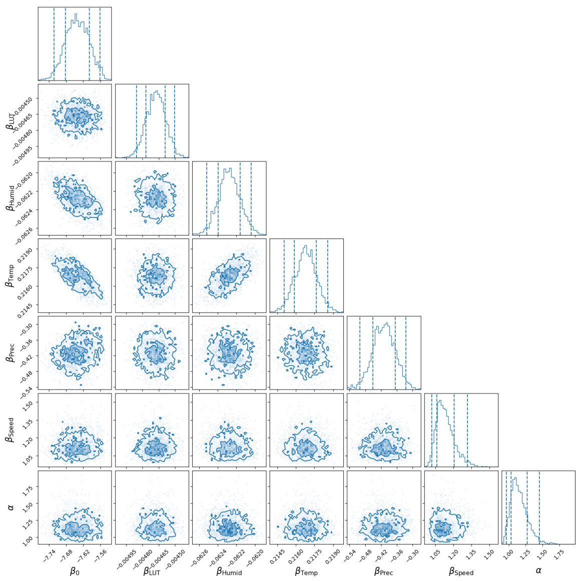


Figure 5: Corner plots of the large model parameters using the sampling of the large model 25

Parameter	Mean	95% CI
β_0	-7.6412	[-7.7224, -7.5607]
β_{LUT}	-0.0047	[-0.0049, -0.0045]
β_{HUMID}	-0.0623	[-0.0625, -0.0620]
β_{TEMP}	0.2168	[0.2150, 0.2186]
α	-0.4099	[-0.4997, -0.3254]
τ_{int}	1.1337	[1.0193, 1.3190]
τ_{slo}	1.1558	[0.9631, 1.4552]
ρ_{int}	0.0396	[0.0014, 0.1228]
ρ_{slo}	0.1134	[0.0114, 0.3066]

Table 1: Summary of model parameter estimates and their respective credible intervals from sampling (25)

Simplified Model

For later interpretation and comparison, we implemented a simplified model which does not use the weighted matrix and only depends on the β and α parameters reducing the effective parameter space from 1121 to 5.

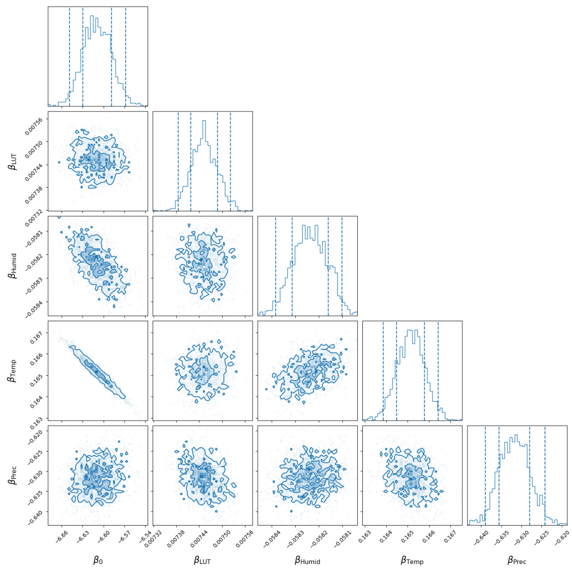


Figure 6: Corner plots of the simplified model parameters (Sampling: 26).

Parameter	Mean	95% CI
β_0	-6.6091	[-6.6490, -6.5683]
β_{LUT}	0.0075	[0.0074, 0.0075]
β_{HUMID}	-0.0582	[-0.0584, -0.0581]
β_{TEMP}	0.1651	[0.1639, 0.1664]
α	-0.6317	[-0.6390, -0.6242]

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

Interpretation of Findings

Therefore, 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 WAIC (Widely Applicable Information Criterion) we get:

The following table summarizes the WAIC results for the large and simplified models. These results offer insight into their predictive performance and complexity.

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

Suggestions for extending the analysis, such as incorporating additional variables or improving model complexity, are discussed here. Placeholder for your proposed next steps.

- Explore predictive models for real-time risk assessment.

- Integrate additional ecological variables, such as vegetation indices.

- Investigate the impact of long-term policy changes on fire dynamics.

Conclusion

By focusing on the Amazonia biome, this study provides critical insights into the drivers of wildfire activity in one of the world's most vital ecosystems. The application of Bayesian spatio-temporal modeling enhances our understanding of these dynamics and offers a foundation for improved management strategies.

Contribution Statement

The workflow for this project was divided into two main components:
1. Understand the data and Model the data/build a likelihood.
2. Model implementation, sampling the posterior and goodness of fit.

Luca Titze and Niklas Viebig focused on the first part while Jan Fritz and Victor Windhab concentrated on the second. All team members contributed collaboratively to writing and refining the report.

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 Sample Mean of the Incremental Intercept Parameter ϕ and the Incremental Slope Parameter δ

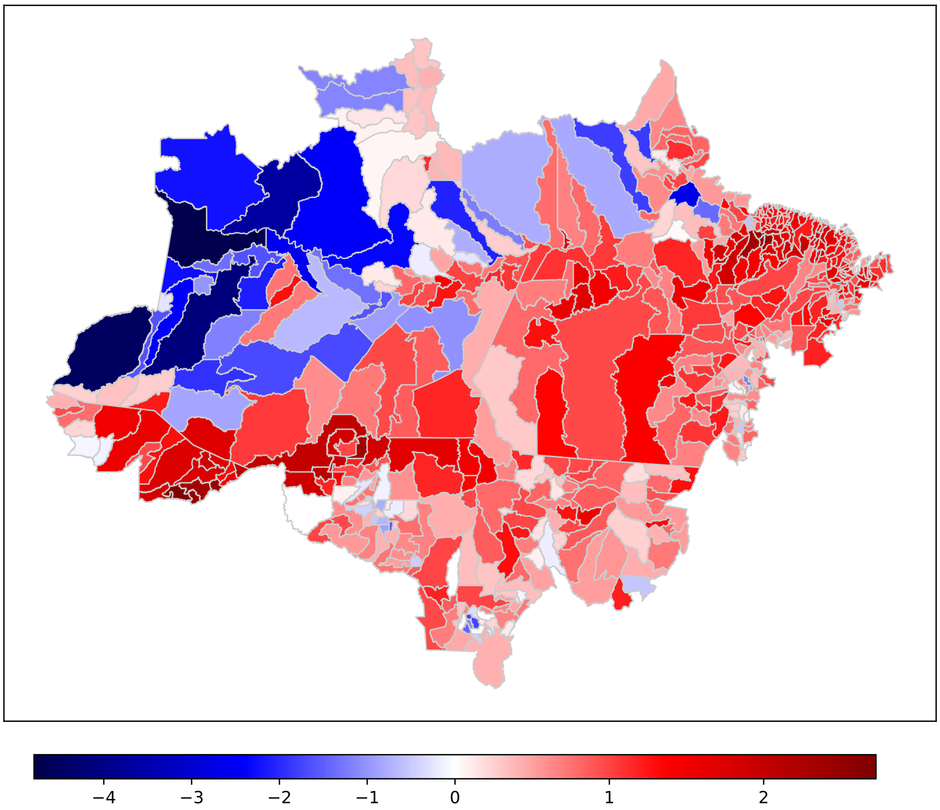


Figure 8: The sample mean of the the incremental intercept parameter ϕ.

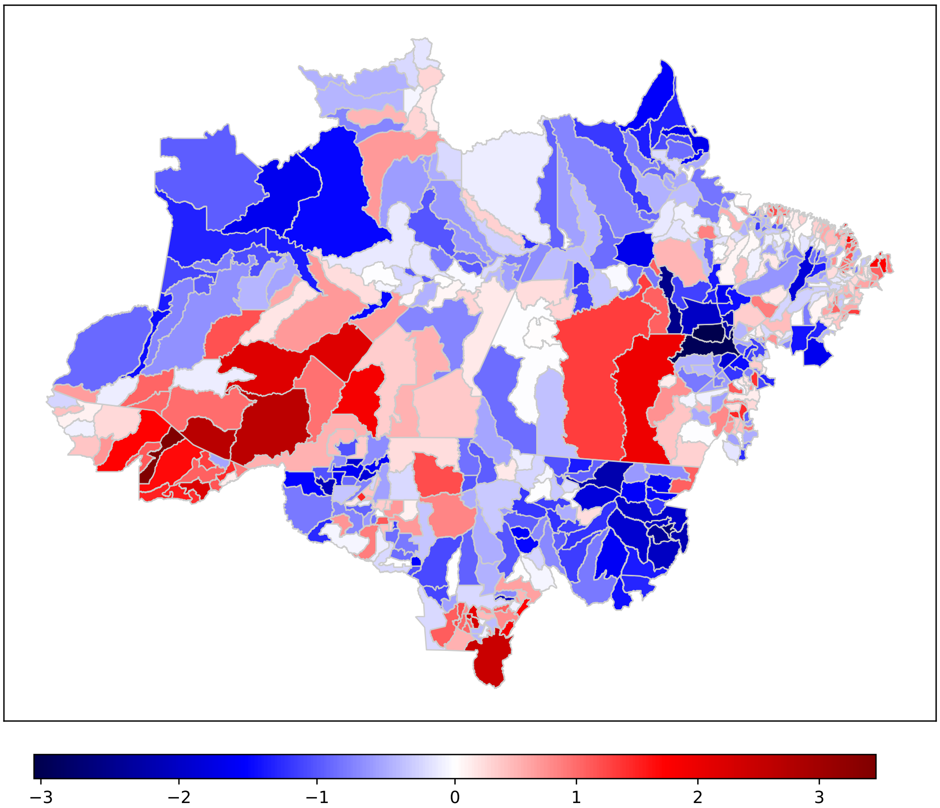


Figure 9: The sample mean of the the incremental intercept parameter δ.

A.2 Correlations

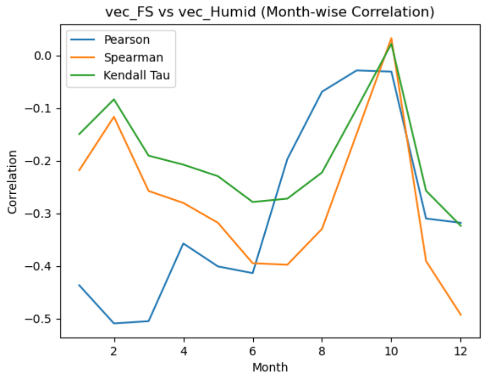


Figure 10: Correlation between `vec_FS` and `vec_Humid`.

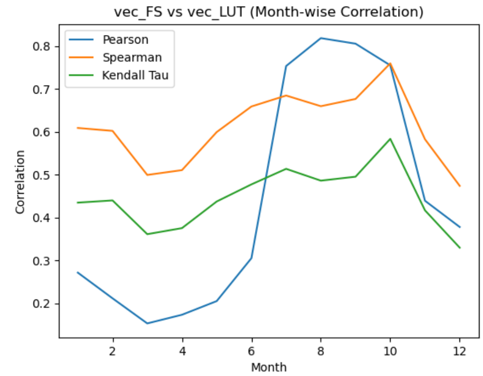


Figure 11: Correlation between `vec_FS` and `vec_LUT`.

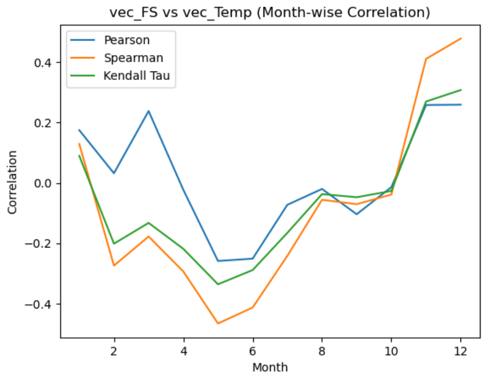


Figure 12: Correlation between `vec_FS` and `vec_Temp`.

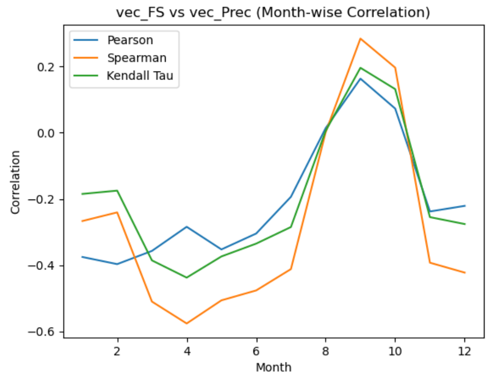


Figure 13: Correlation between `vec_FS` and `vec_Prec`.

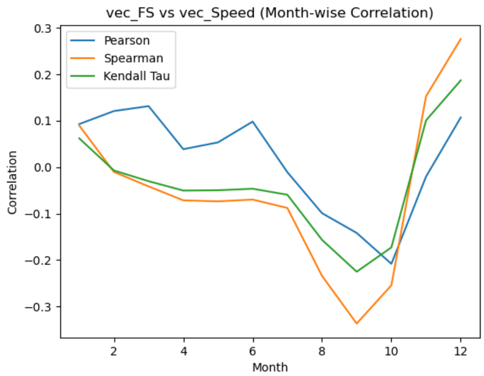


Figure 14: Correlation between `vec_FS` and `vec_Speed`.

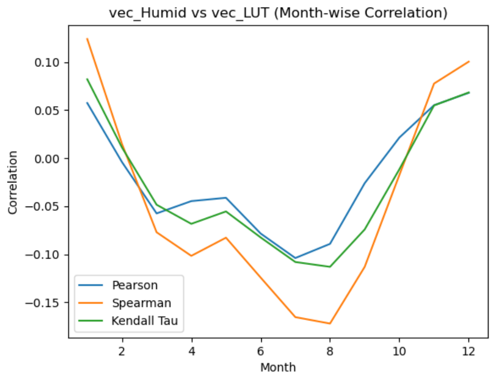


Figure 15: Correlation between `vec_Humid` and `vec_LUT`.

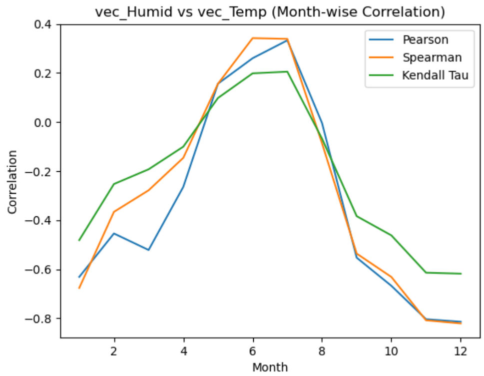


Figure 16: Correlation between `vec_Humid` and `vec_Temp`.

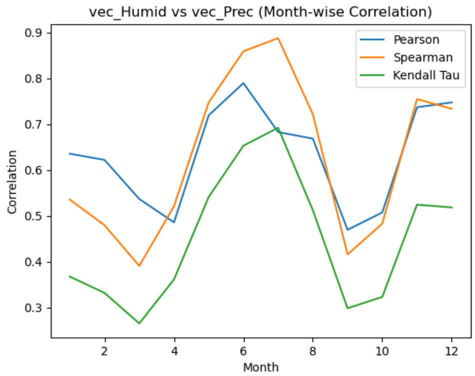


Figure 17: Correlation between `vec_Humid` and `vec_Prec`.

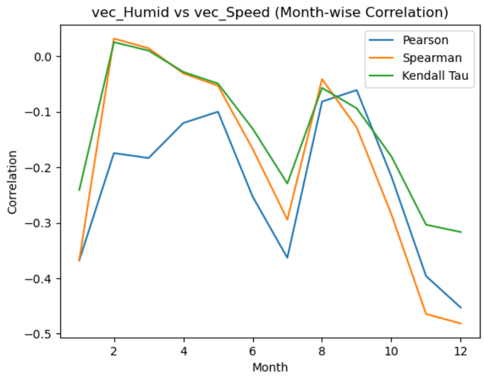


Figure 18: Correlation between `vec_Humid` and `vec_Speed`.

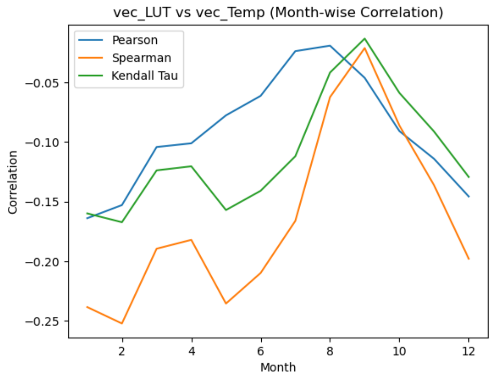


Figure 19: Correlation between `vec_LUT` and `vec_Temp`.

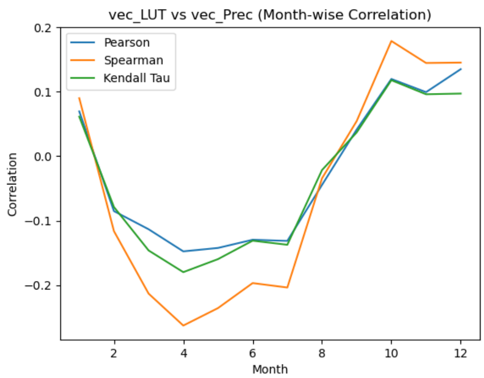


Figure 20: Correlation between `vec_LUT` and `vec_Prec`.

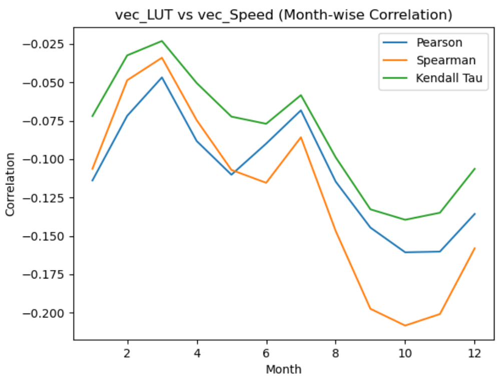


Figure 21: Correlation between `vec_LUT` and `vec_Speed`.

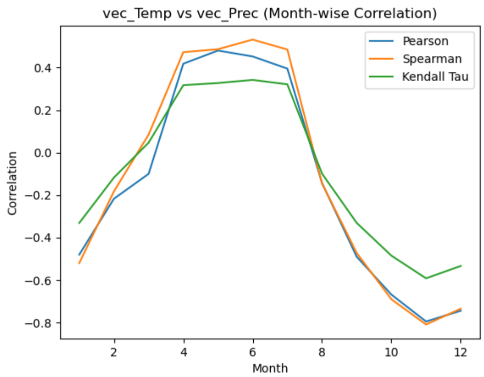


Figure 22: Correlation between `vec_Temp` and `vec_Prec`.

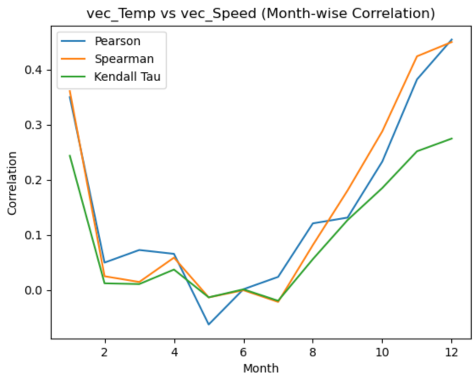


Figure 23: Correlation between `vec_Temp` and `vec_Speed`.

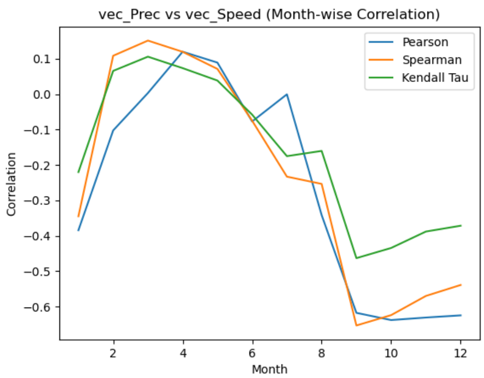


Figure 24: Correlation between `vec_Prec` and `vec_Speed`.

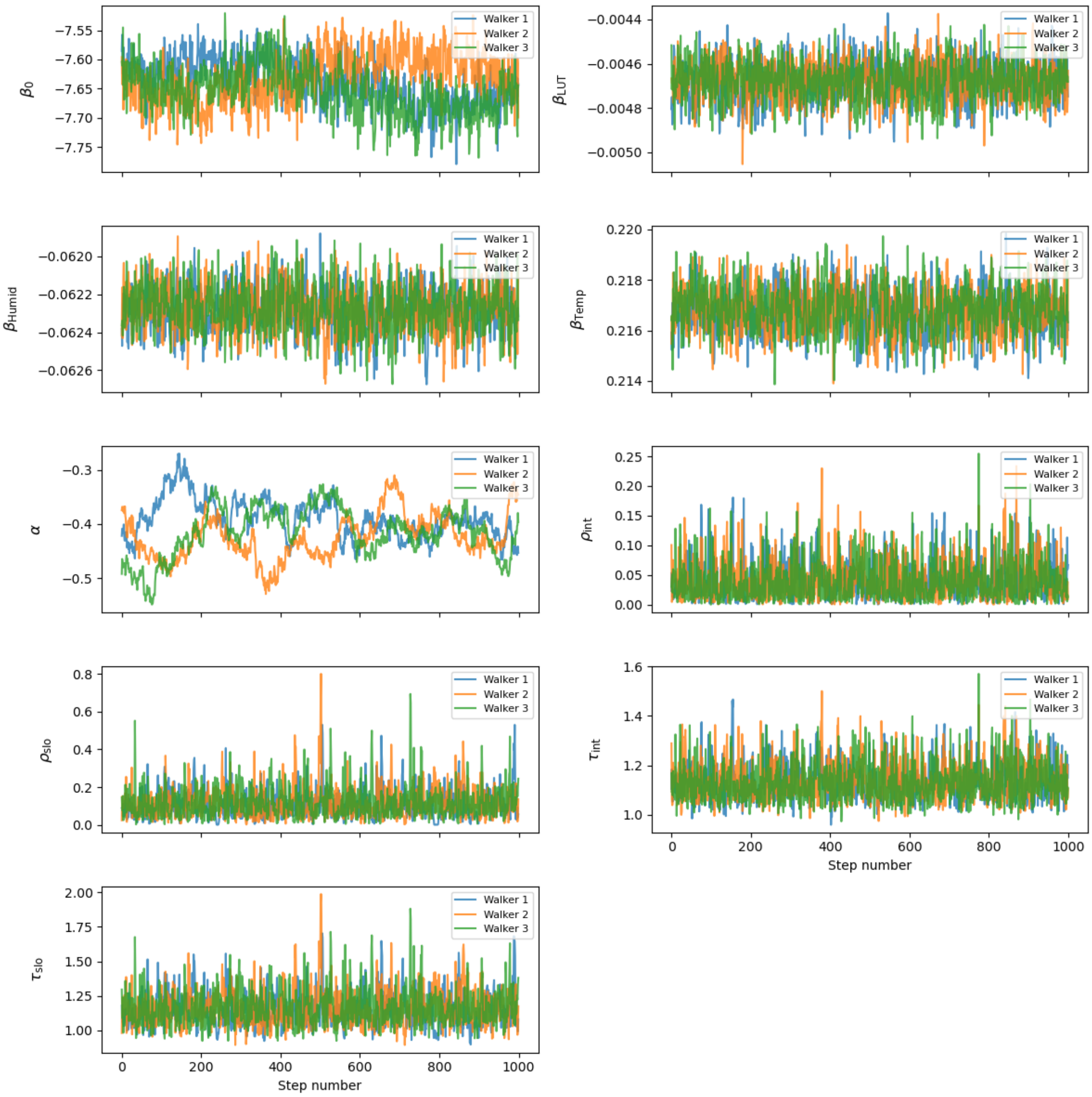


Figure 25: Sampling of large model.

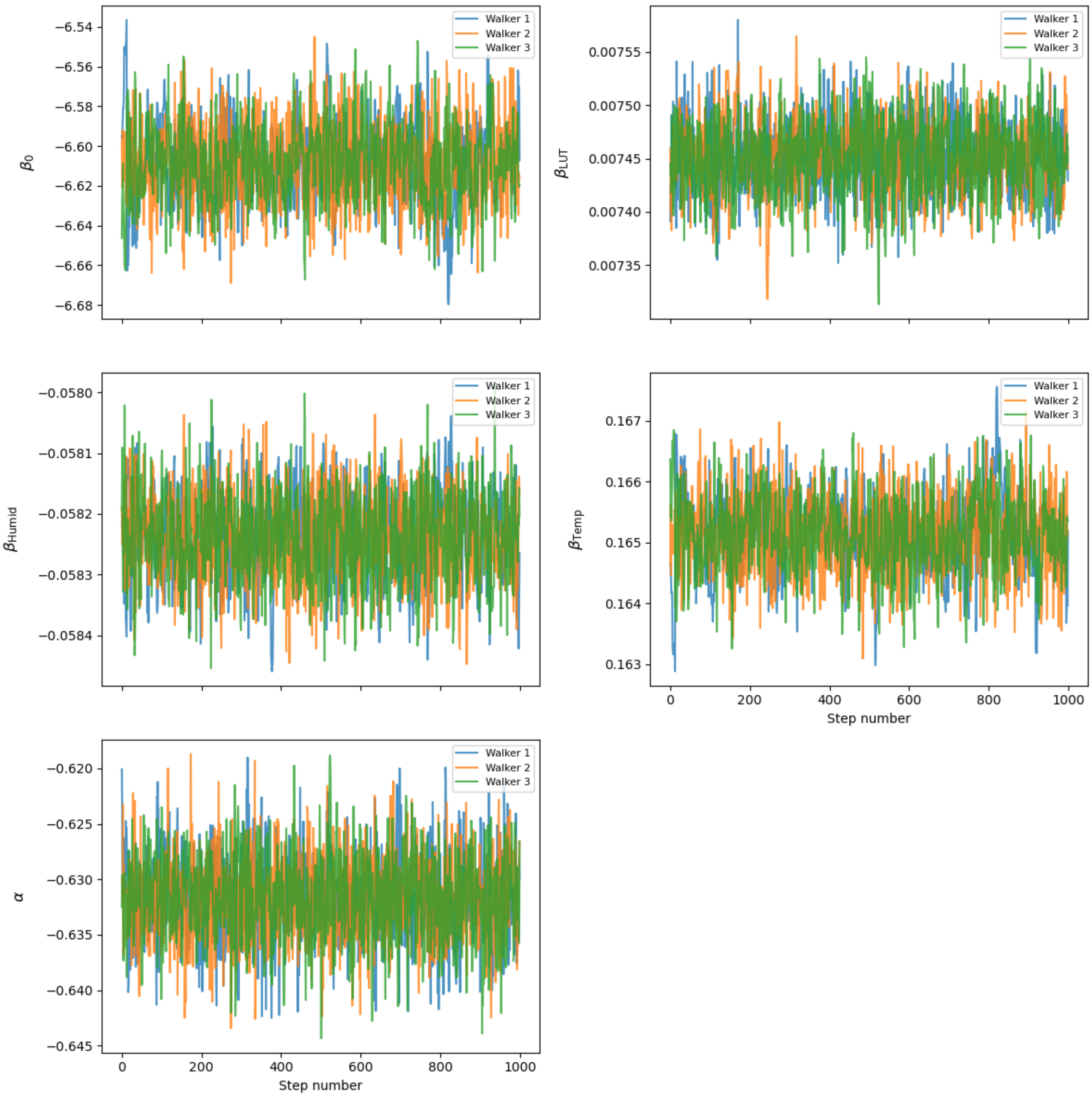


Figure 26: Sampling of simplified model.