Spatio-temporal modelling of the Brazilian wildfires: The influence of human and meteorological variables

Paulo Canas Rodrigues[1,2]

(Joint work with Rodrigo Bulhões and Jonatha Pimentel)

[1]Department of Statistics, Federal University of Bahia, Brazil

[2]Statistical Learning Laboratory (SaLLy)

DaSSWeb
Data Science and Statistics Webinar

UFBA Universidade Federal da Bahia

Departamento de Estatística IME - UFBA

Statistical Learning Laboratory SaLLy

Paulo Canas Rodrigues

Email: paulocanas@gmail.com
Web: www.paulocanas.org
Lab: www.SaLLy.ufba.br

- BSc in Mathematics (FCT-UNL, Portugal; 2003), MSc in Statistics (IST-UTL, Portugal; 2007), PhD in Statistics (FCT-UNL, Portugal; 2012), Habilitation in Mathematics, specialization in Statistics and Stochastic Processes (IST-UL, Portugal; 2019).
- Published more than 90 scientific papers in collaboration with 126 co-authors from 61 universities in 24 countries and delivered more than 130 invited talks and seminars.

- Among other activities, Paulo Canas Rodrigues is currently:

- Professor of the Department of Statistics, Federal University of Bahia, Brazil
- Head & Principal Investigator of the Statistical Learning Laboratory (SaLLy)

- Co-founder and Vice-Coordinator of the Specialization in Data Science and Big Data
- Co-Editor for Computational Statistics, Brazilian Journal of Biometrics, Biometrical Letters, and Statistics, Optimization and Information Computing

- President of the International Society for Business and Industrial Statistics (2023 –2025)
- Member of the Representative Council of the International Biometric Society
- Member of the Board of Directors of the Brazilian Statistical Association
- Co-founder and Past-Chair of the ISI Special Interest Group (SIG) on Data Science (2021 - 2023)
- President-Elect of the International Association for Statistical Computing (10/2023 – 10/2025; President between 10/2025 and 10/2027)
- Council Member (*ex-officio*) of the International Statistical Institute (07/2023 – 07/2025)

Outline

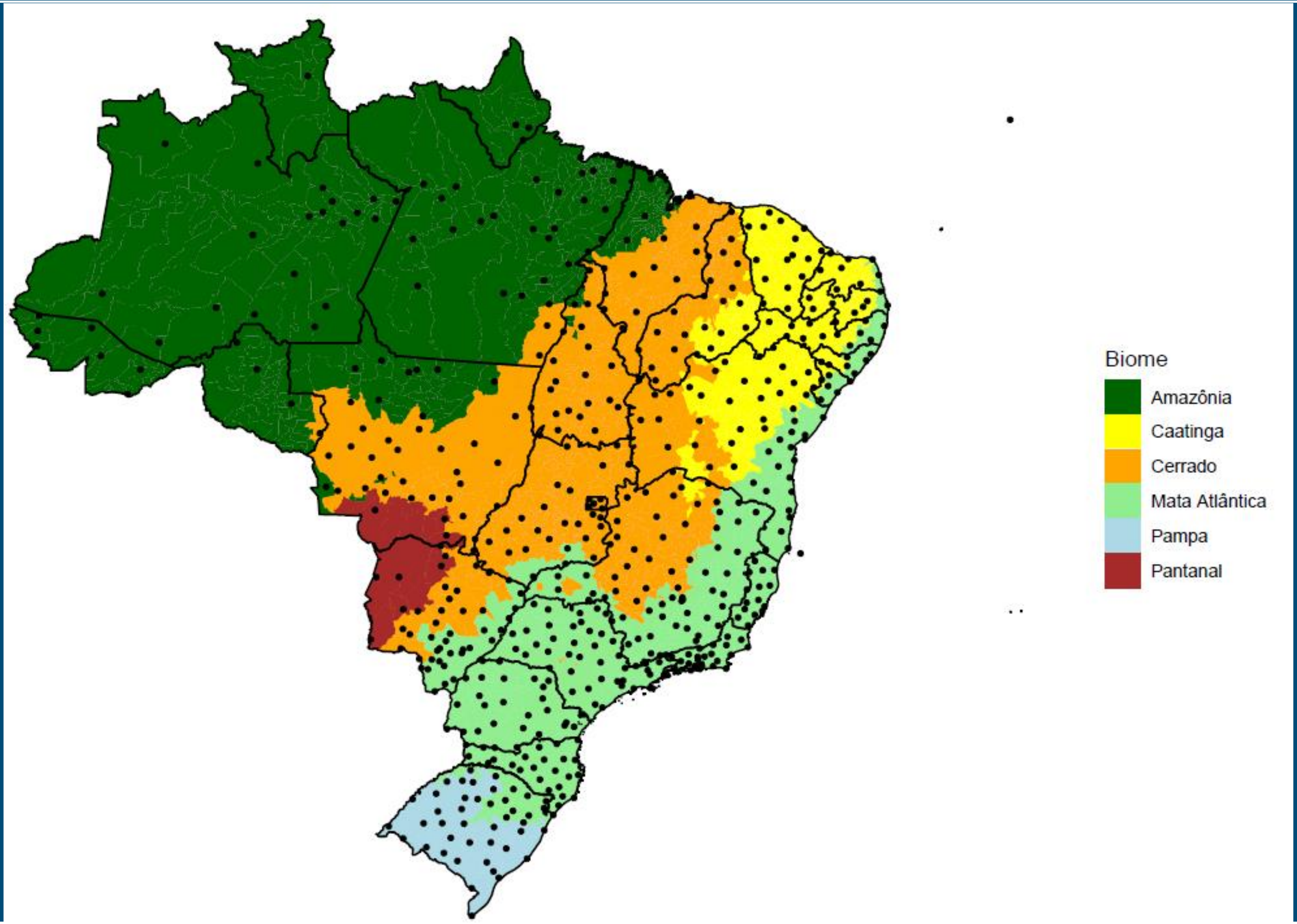
Introduction

- Wildfires are one of the most common natural disasters in many world regions and actively impact life quality.
- These events have become frequent due to climate change, local policies, and human behaviour.
- This study considers the historical data with the geographical locations of all the ``fire spots'' detected by the reference satellites covering the Brazilian territory between January 2011 and December 2022, comprising more than 2.2 million fire spots.
- This data was modelled with a spatio-temporal generalized linear model for areal unit data, whose inferences about its parameters are made in a Bayesian approach. We use meteorological variables (precipitation, air temperature, humidity, and wind speed) and a human variable (land-use transition and occupation) as covariates.
- The change in land use from the forest and green areas to farming significantly impacts the number of fire spots for all six Brazilian biomes.

Brazilian Biomes

- Brazil is the fifth country in the world in territorial extension.
- It is considered by many experts as the “country of megadiversity”, given that 15-20% of the known species in the world are found in its territory.
- Its fauna and flora are officially separated into six biomes: Amazônia, Caatinga, Cerrado, Mata Atlântica, Pampas, and Pantanal:
  - Amazônia: includes about 60% of the largest rainforest in the world, with extensive mineral reserves and 20% of the world's water availability.
  - Caatinga: a semi-arid climate, with great biological richness and unique species
  - Cerrado: recognized as the richest savanna in the world in terms of biodiversity, having remained unchanged until the 1950s when the federal capital was transferred to Brasília.
  - Mata Atlântica: located on the Brazilian coast, thus being the most threatened biome in the country, where only 27% of the original forest cover is still preserved.
  - Pampas: characterized by a rainy climate without a dry period and negative temperatures during the winter.
  - Pantanal: recognized as the planet's most extensive continuous floodplain.

Brazilian biomes and meteorological stations



The data

- This study used complex data from three different sources, considering a time interval of twelve years between January 1, 2011, and December 31, 2022. The data sources used were:

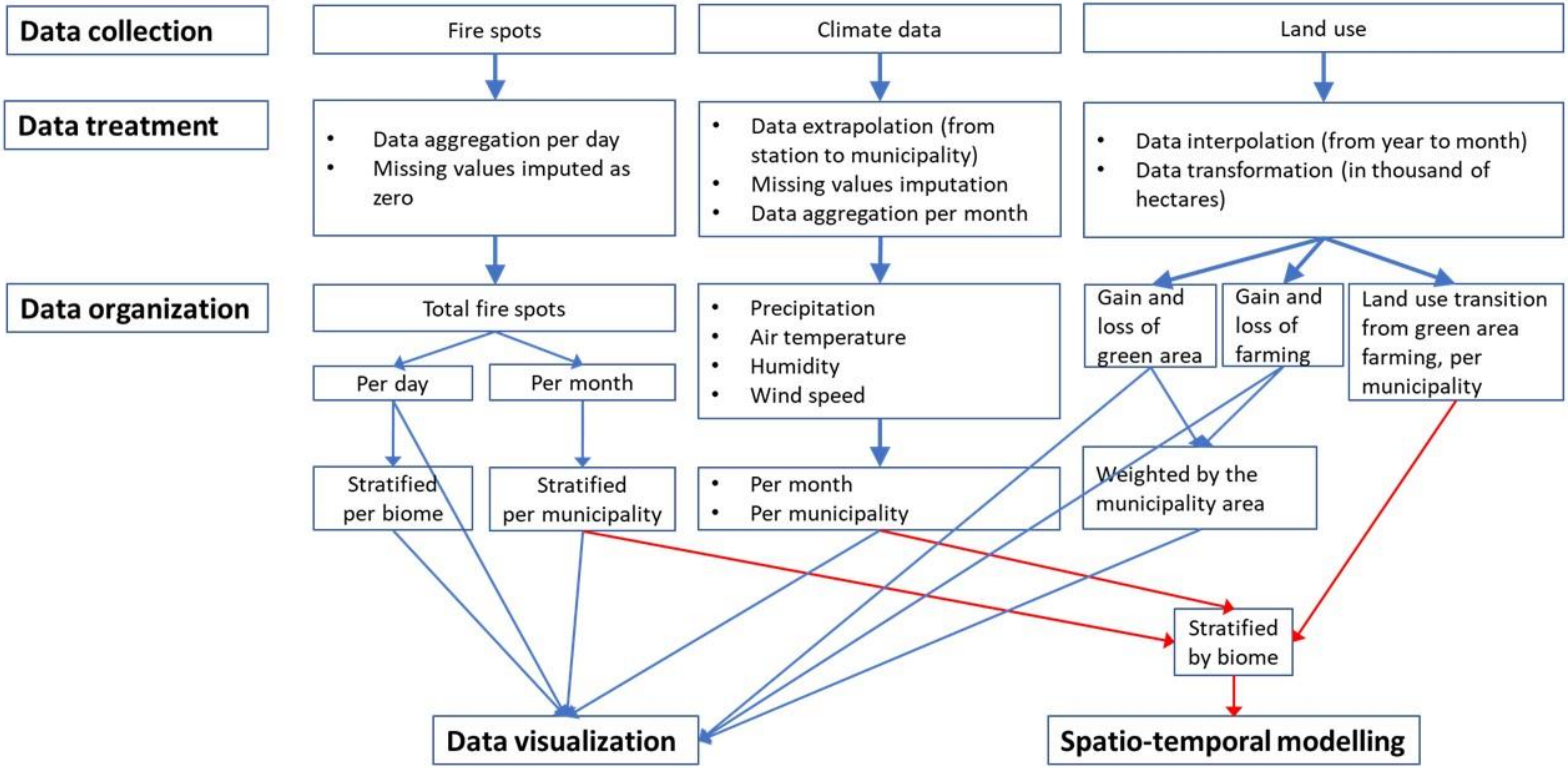
   i. Satellite images that resulted in a data set containing all fire spots in the whole Brazilian territory during the twelve years;

   ii. hourly climatic data from all available meteorological stations in Brazil during the ten years used for modelling, i.e., between January 1, 2012, and December 31, 2021; and

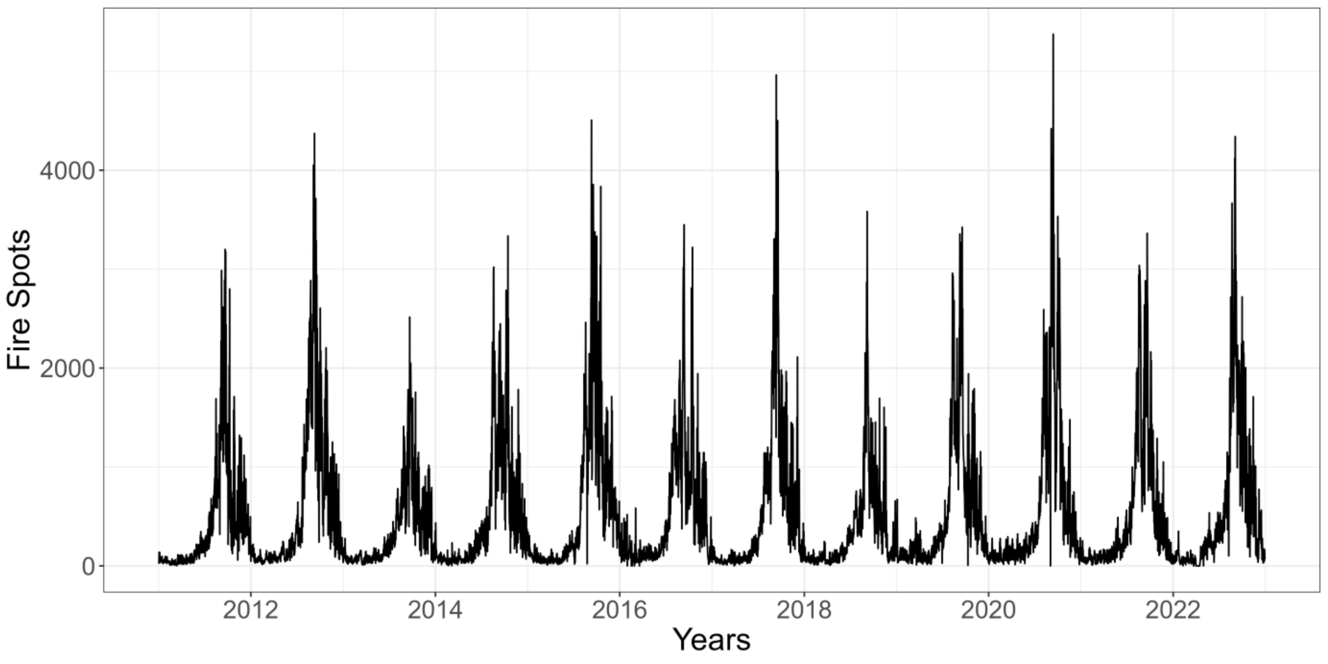
   iii. data related to land use and land-use transition during the ten used for modelling, i.e., between January 1, 2012, and December 31, 2021.

- The difference in data collection periods was due to data availability. Land use and land-use transition are made available with a large delay. The temporal and spatial descriptive and exploratory analysis of the number of fire spots considered 12 years (2011-2022) and the model building considered 10 years (2012-2021).

Flowchart of the methodology



Number of fire spots in Brazil

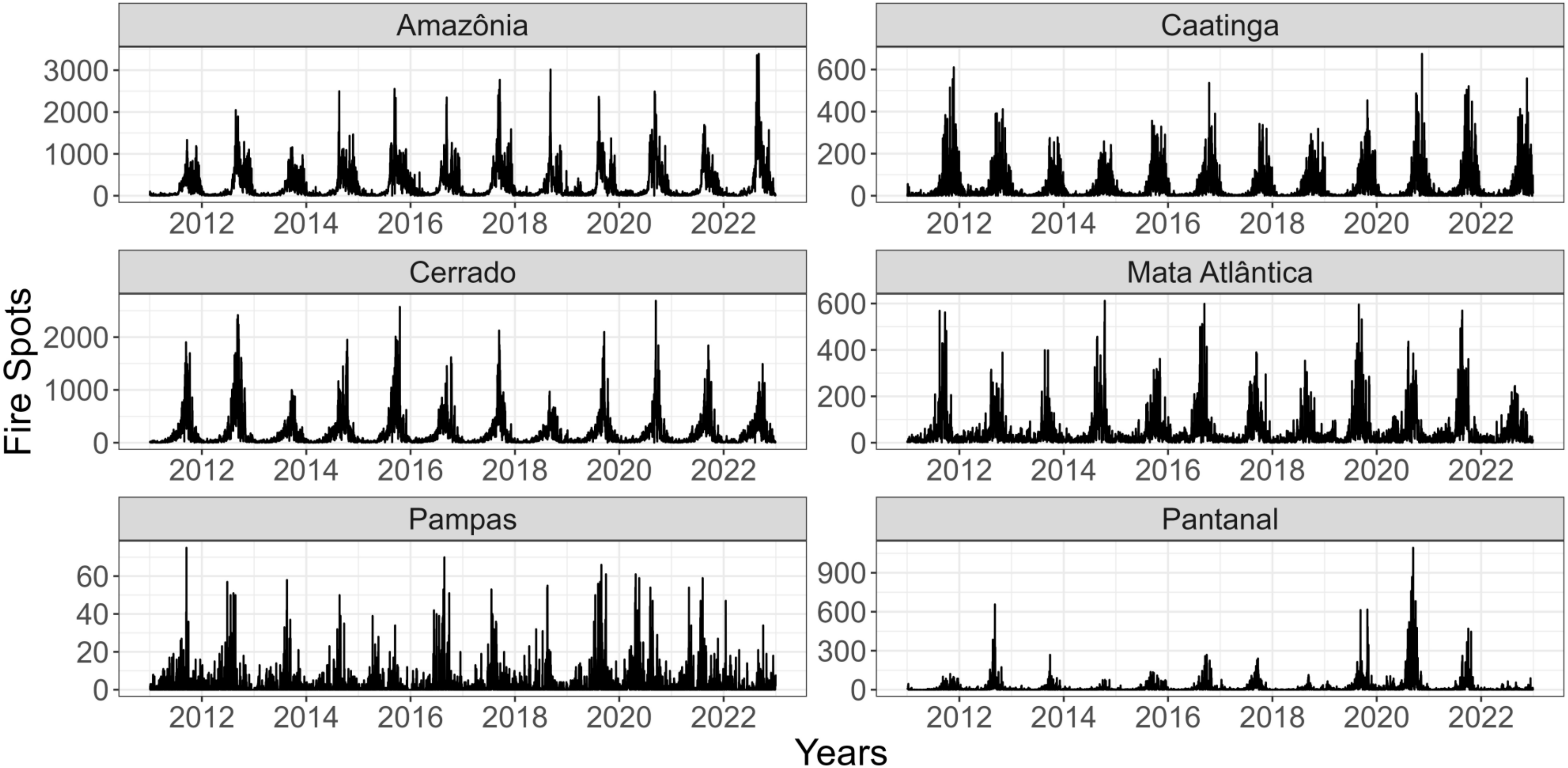


Daily number of fire spots in the Brazilian territory between January 1, 2011, and December 31, 2022

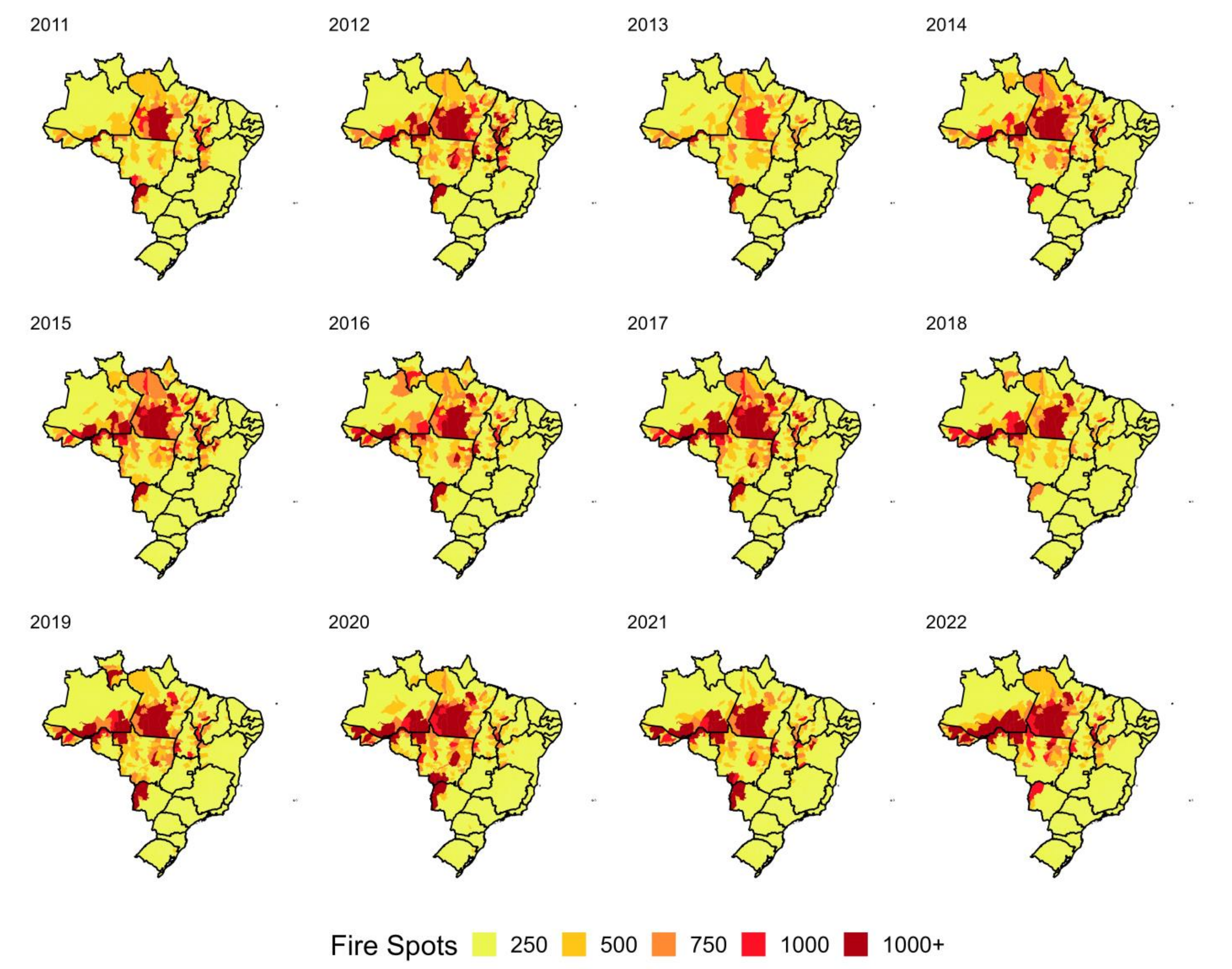
Total number of fire spots in the Brazilian territory per year, between January 1, 2011, and December 31, 2022, and its variation rate compared with the previous year.

Year	Fire Spots	Variation Rate (%)
2011	158,099	-
2012	217,238	37
2013	128,149	-41
2014	175,900	37
2015	216,782	23
2016	184,218	-15
2017	207,511	13
2018	132,872	-36
2019	197,632	49
2020	222,798	13
2021	184,081	-17
2022	200,763	9

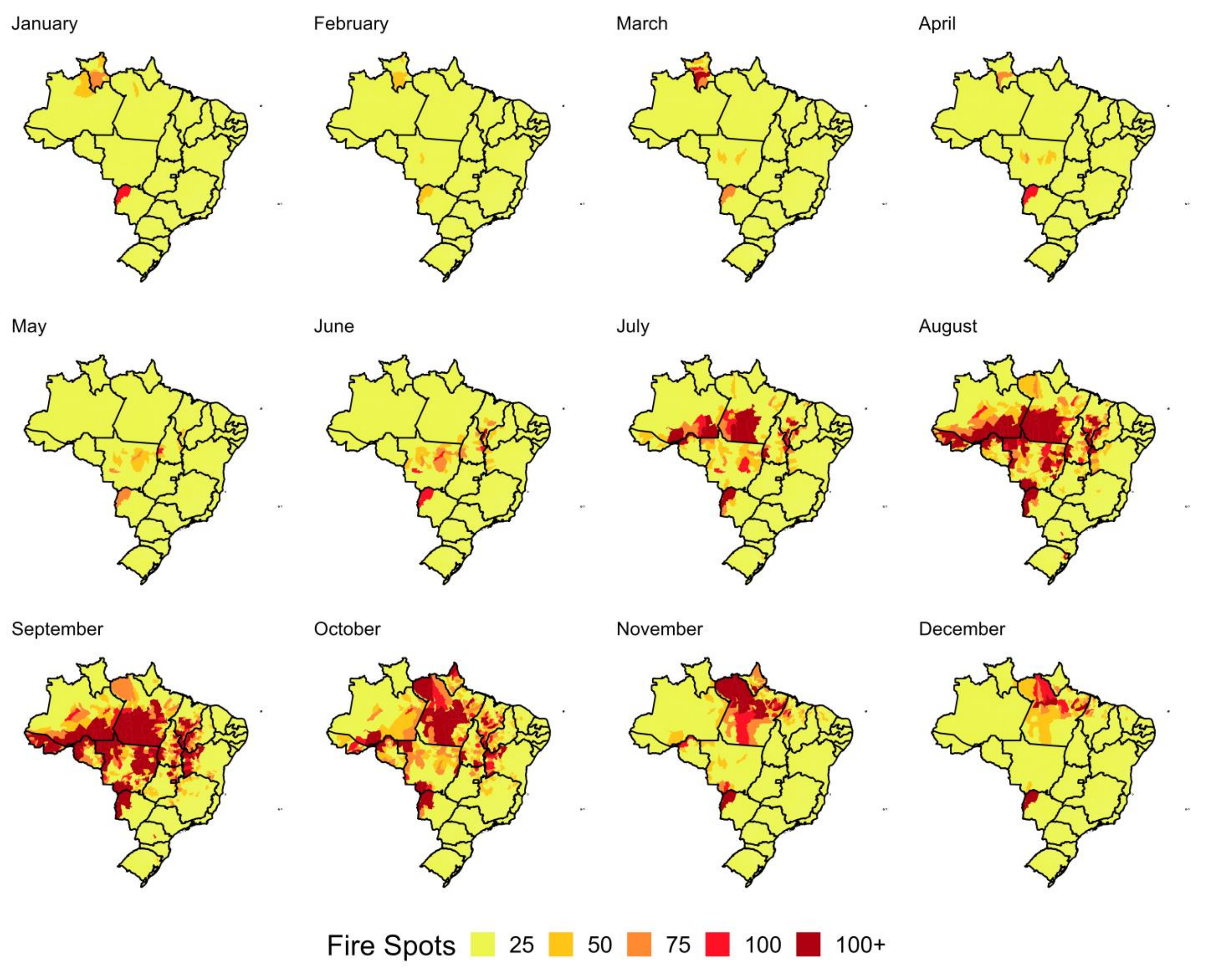
Daily number of fire spots in Brazil, per biome



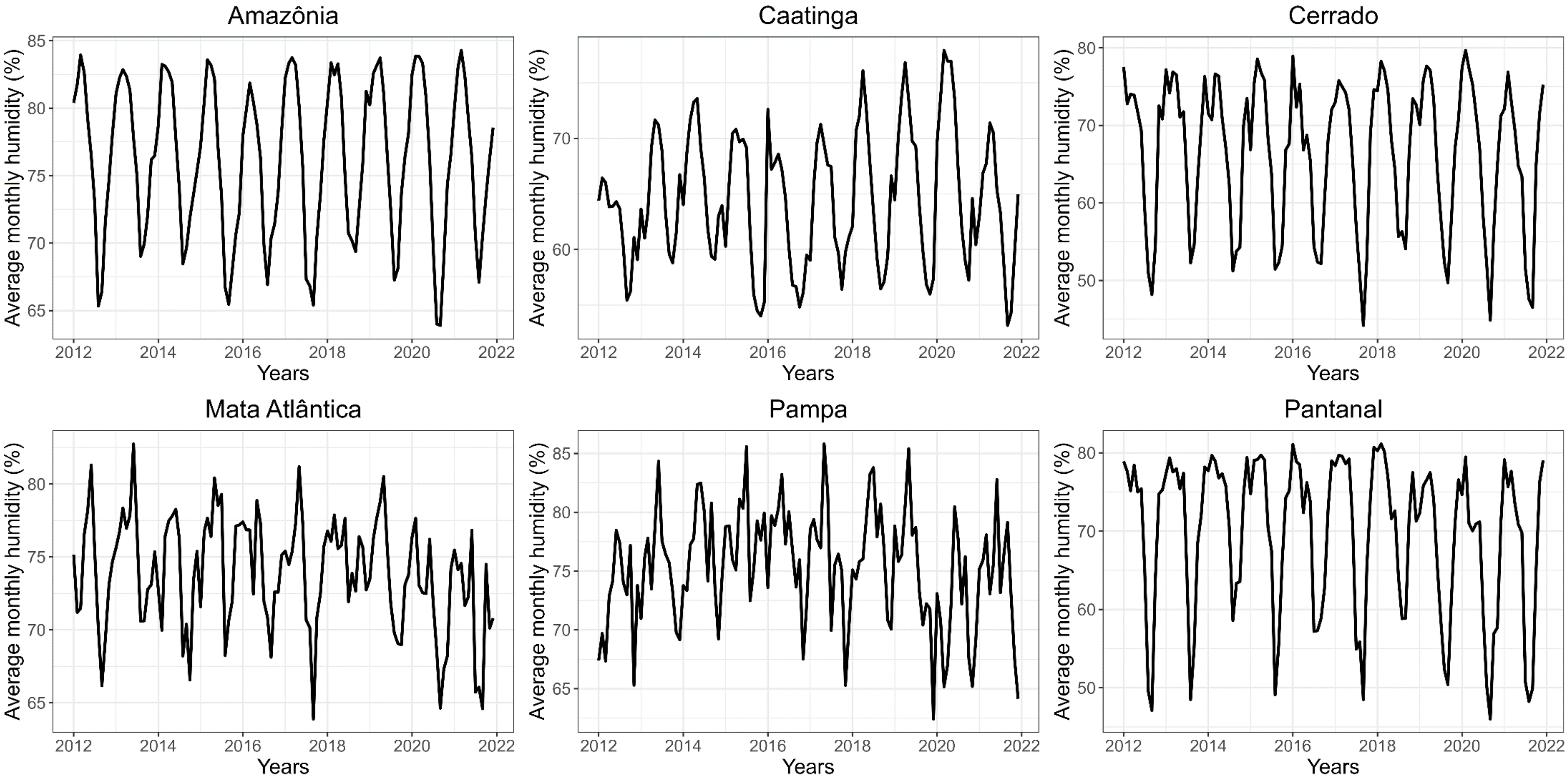
Fire spots per year per Brazilian municipality



Fire spots per month per Brazilian municipality

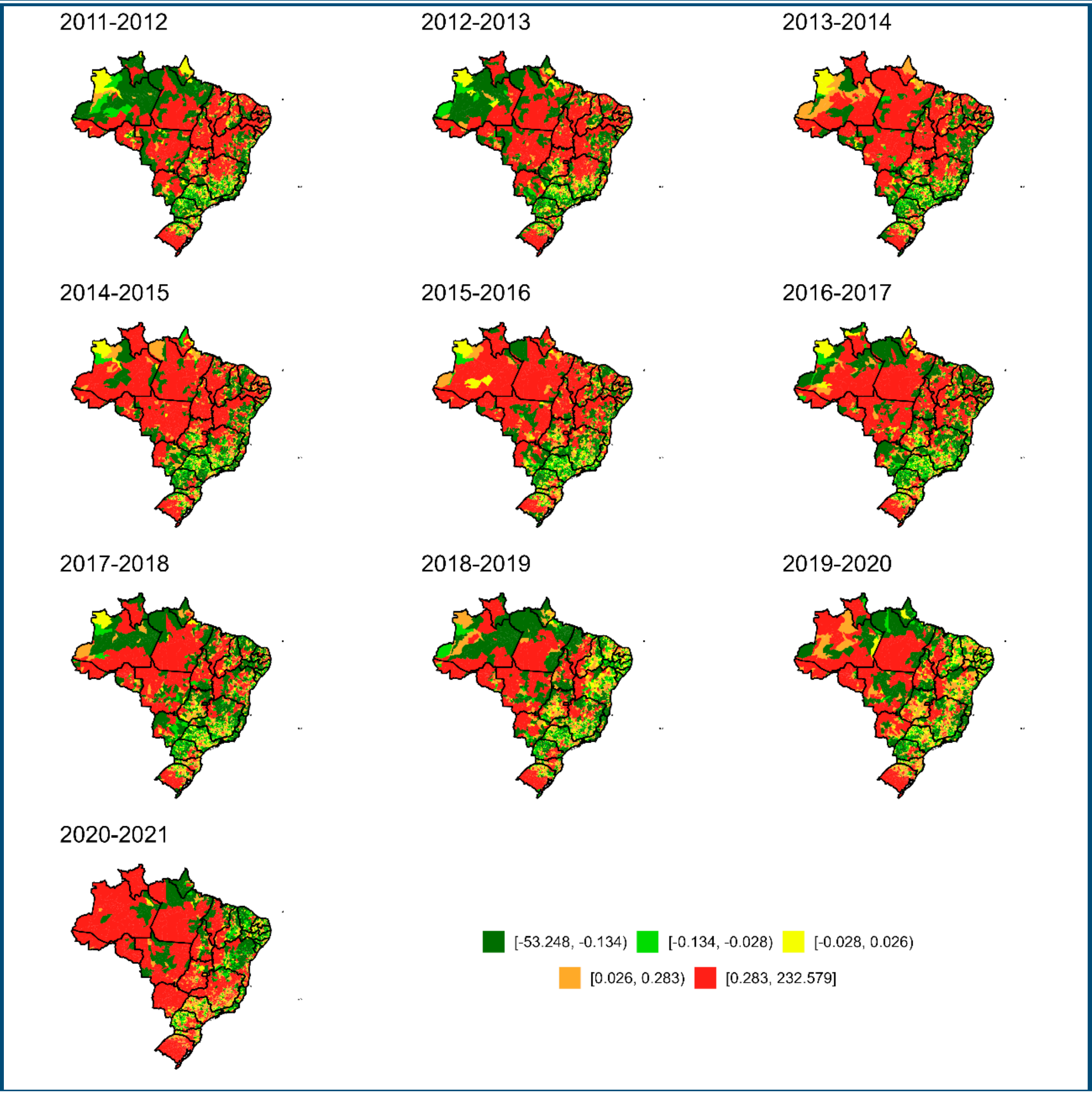


Covariates – Meteorological variables



Monthly average of the meteorological variable relative air humidity in each Brazilian biome between 2012 and 2022.

Covariates – Land-use transition



Evolution of land use and occupation by green areas (forests or non-forest and natural formations) between 2011 and 2021 for each Brazilian municipality.

The maps show the difference in the green area occupation between the years t and $t+1$, for $t = 2011, \ldots, 2020$, in thousands of hectares (overall numbers per municipality).

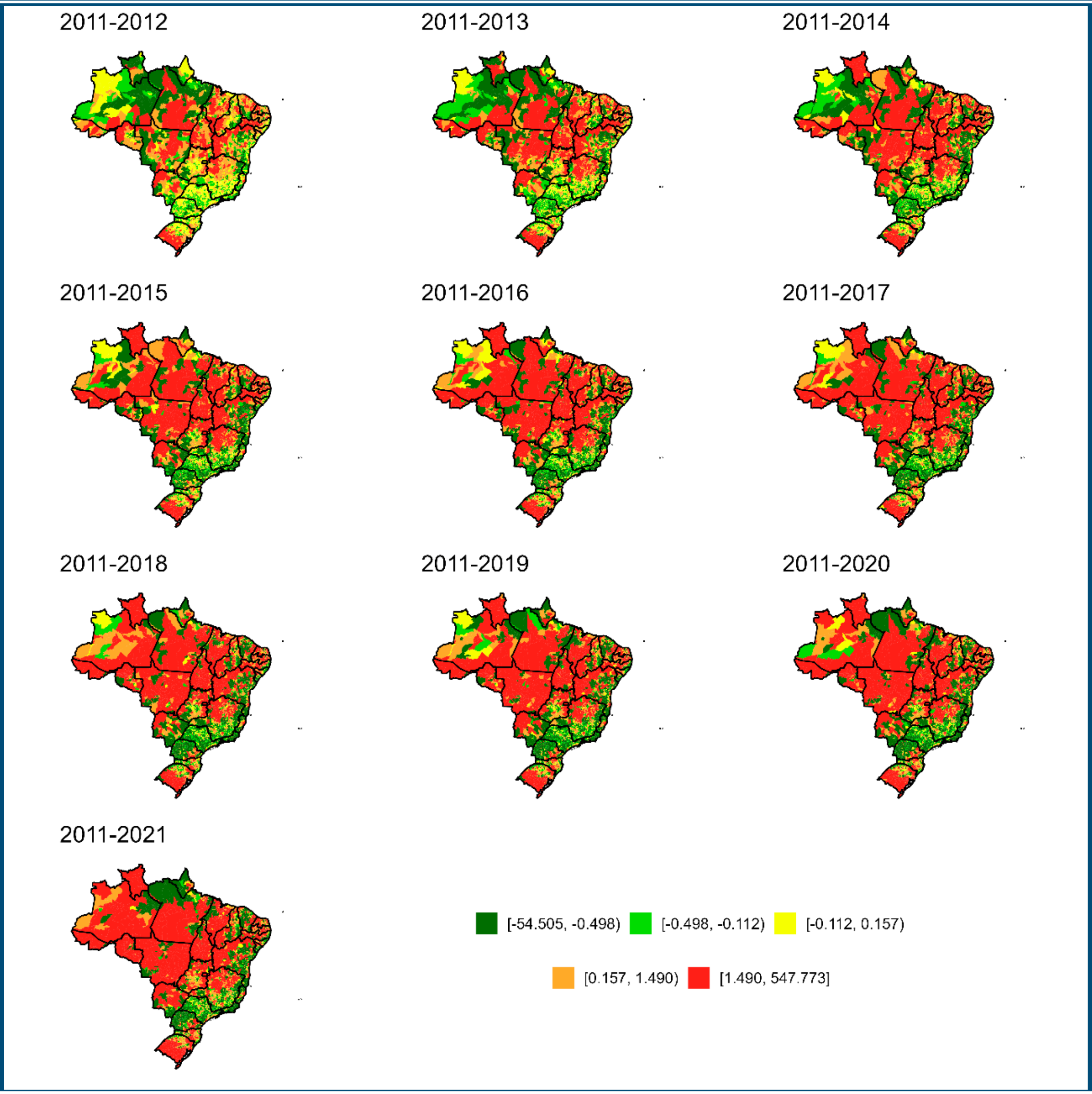
Positive/green values represent an increase in the green area in the given municipality from one year to the next, whereas negative/orange-red values indicate a decrease in the green area in the given municipality from one year to the next.

UFBA Universidade Federal da Bahia

Departamento de Estatística IME - UFBA

Statistical Learning Laboratory SaLLy

Covariates – Land-use transition



Evolution of land use and occupation by green areas (forests or non-forest and natural formations) between 2011 and 2021 for each Brazilian municipality, using 2011 as reference.

The maps show the difference in the green area occupation between the years 2011 and $t+1$, for $t = 2011, \ldots, 2020$, in thousands of hectares (overall numbers per municipality).

Positive/green values represent an increase in the green area in the given municipality from one year to the next, whereas negative/orange-red values indicate a decrease in the green area in the given municipality from one year to the next.

Spatio-temporal modelling

- Objective: to model the total number of fire spots, aggregated by municipality and month, by including explanatory variables such as precipitation, temperature, humidity, radiation, and land-use transition.
- As we deal with panel data (counts) that vary in time and space, we decided to work with a spatio-temporal generalized linear model for areal unit data, whose inferences about its parameters are made in a Bayesian framework.
- Models of this class help to fit areal unit data given in discrete periods while allowing the inclusion of explanatory variables.
- We consider the model proposed by
  - Lee, D., Rushworth, A., & Napier, G. (2018). Spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST package. Journal of Statistical Software, 84, 1-39.

Spatio-temporal generalized linear model

$$
\begin{aligned}
Y_{k,t} \mid \mu_{k,t} &\sim \text{Poisson}(\mu_{k,t}) \\
\ln \mu_{k,t} &= \mathbf{x}_{k,t}^{\intercal}\boldsymbol{\beta} + \psi_{k,t} \\
\boldsymbol{\beta} &\sim \text{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \\
\psi_{k,t} &= \beta_1 + \phi_k + (\alpha + \delta_k)\frac{t - \bar{t}}{N}
\end{aligned}
$$

$$
\begin{aligned}
\phi_k \mid \boldsymbol{\phi}_{-k}, \mathbf{W} &\sim \text{N}\left(\frac{\rho_{int}\sum_{j=1}^{K} w_{k,j}\phi_j}{\rho_{int}\sum_{j=1}^{K} w_{k,j} + 1 - \rho_{int}}, \frac{\tau_{int}^2}{\rho_{int}\sum_{j=1}^{K} w_{k,j} + 1 - \rho_{int}}\right) \\
\delta_k \mid \boldsymbol{\delta}_{-k}, \mathbf{W} &\sim \text{N}\left(\frac{\rho_{slo}\sum_{j=1}^{K} w_{k,j}\delta_j}{\rho_{slo}\sum_{j=1}^{K} w_{k,j} + 1 - \rho_{slo}}, \frac{\tau_{slo}^2}{\rho_{slo}\sum_{j=1}^{K} w_{k,j} + 1 - \rho_{slo}}\right) \\
\tau_{int}^2, \tau_{slo}^2 &\sim \text{Inverse-Gamma}(a, b) \\
\rho_{int}, \rho_{slo} &\sim \text{Uniform}(0, 1) \\
\alpha &\sim \text{N}(\mu_\alpha, \sigma_\alpha^2),
\end{aligned}
$$

- $\mu_{k,t}$ denotes the expectation of $Y_{k,t}$;
- $\psi_{k,t}$ is a latent component for municipality k and month t embracing one or more sets of spatiotemporally autocorrelated random effects;
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is a p-dimensional vector of covariate regression parameters;
- $\mathbf{W} = [w_{k,j}]$ is a binary neighborhood matrix $N \times N$, with $w_{k,j} = 0$ if $k = j$ (diagonal elements equal to zero), $w_{k,j} = 1$ if the municipalities k and j share a common border, and $w_{k,j} = 0$ if the municipalities k and j do not share a common border;
- The random effects $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)$ are modeled as spatially autocorrelated by the CAR prior, satisfying $\sum_{k=1}^{K} \phi_k = \sum_{k=1}^{K} \delta_k = 0$, with $\boldsymbol{\phi}_{-k}$ and $\boldsymbol{\delta}_{-k}$ denoting the vectors $\boldsymbol{\phi}$ and $\boldsymbol{\delta}$ without their corresponding k^{th} components, respectively;
- ρ_{int} and ρ_{slo} are two spatial dependence parameters.

The model gives the spatio-temporal pattern in the mean response with a spatially varying linear time trend. Municipality k has its own linear time trend, with a spatially varying intercept $\beta_1{}' + \varphi_k$ and a spatially varying slope $\alpha + \delta_k$. The hyperparameters are chosen to have non informative prior distributions.

Results – Spatio-temporal modelling

Model by biome	Mean	CI 95%	Geweke
Amazônia			
Intercept	6.5019	6.4914 to 6.5133	-1.2
Land-use transition	0.0358	0.0356 to 0.0359	-0.1
Humidity	-0.0596	-0.0598 to -0.0594	1.5
α	-0.0913	-0.0995 to -0.0828	-0.5
τ^2_{int}	0.5027	0.4370 to 0.5820	-1.0
τ^2_{slo}	1.6055	1.3820 to 1.9266	-0.2
ρ_{int}	0.0130	0.0004 to 0.0424	-1.6
ρ_{slo}	0.0186	0.0007 to 0.0588	-0.7
Caatinga			
Intercept	4.8904	4.8665 to 4.9104	-1.0
Land-use transition	0.1762	0.1745 to 0.1779	0.3
Humidity	-0.0873	-0.0877 to -0.0868	0.8
α	-0.4777	-0.4969 to -0.4607	1.8
τ^2_{int}	0.2188	0.1990 to 0.2461	2.3
τ^2_{slo}	2.4793	2.1110 to 2.9296	-0.5
ρ_{int}	0.0082	0.0002 to 0.0258	2.3
ρ_{slo}	0.0638	0.0251 to 0.1119	-0.1
Cerrado			
Intercept	5.3411	5.3289 to 5.3528	-1.1
Land-use transition	0.1449	0.1446 to 0.1453	0.2
Humidity	-0.0705	-0.0707 to -0.0702	0.8
α	0.1192	0.1086 to 0.1286	-0.1
τ^2_{int}	0.5735	0.4910 to 0.6755	-1.4
τ^2_{slo}	3.8550	3.3675 to 4.4606	0.1
ρ_{int}	0.0849	0.0449 to 0.1399	-1.2
ρ_{slo}	0.0967	0.0587 to 0.1404	0.7

Model by biome	Mean	CI 95%	Geweke
Mata Atlântica			
Intercept	3.6267	3.6016 to 3.6504	0.6
Land-use transition	0.1269	0.1244 to 0.1291	2.7
Humidity	-0.0650	-0.0654 to -0.0647	-0.9
α	-0.1250	-0.1442 to -0.1063	1.9
τ^2_{int}	0.2240	0.2045 to 0.2492	-1.9
τ^2_{slo}	2.2069	1.9724 to 2.5072	3.2
ρ_{int}	0.0150	0.0008 to 0.037	-1.9
ρ_{slo}	0.0681	0.0410 to 0.1117	1.6
Pampas			
Intercept	-0.6336	-0.6698 to -0.5933	-0.2
Land-use transition	0.1221	0.1195 to 0.1247	-0.5
Precipitation	-3.6167	-3.8325 to -3.4265	0.7
α	0.0856	0.0235 to 0.1502	0.2
τ^2_{int}	0.1083	0.0823 to 0.1436	0.0
τ^2_{slo}	0.9141	0.6831 to 1.2822	0.4
ρ_{int}	0.0286	0.0011 to 0.0947	-1.1
ρ_{slo}	0.0381	0.0015 to 0.1163	0.3
Pantanal			
Intercept	8.0972	8.0621 to 8.1294	-3.9
Land-use transition	0.0316	0.0313 to 0.0319	0.8
Humidity	-0.0808	-0.0813 to -0.0802	3.6
α	-1.2732	-1.3003 to -1.2447	1.9
τ^2_{int}	0.1171	0.0542 to 0.2531	0.1
τ^2_{slo}	0.8423	0.4029 to 1.5877	-1.5
ρ_{int}	0.2272	0.0078 to 0.7122	1.4
ρ_{slo}	0.4138	0.0297 to 0.8861	-0.5

Bayesian spatio-temporal models by biome: posterior mean, 95% credible interval and Geweke statistic.

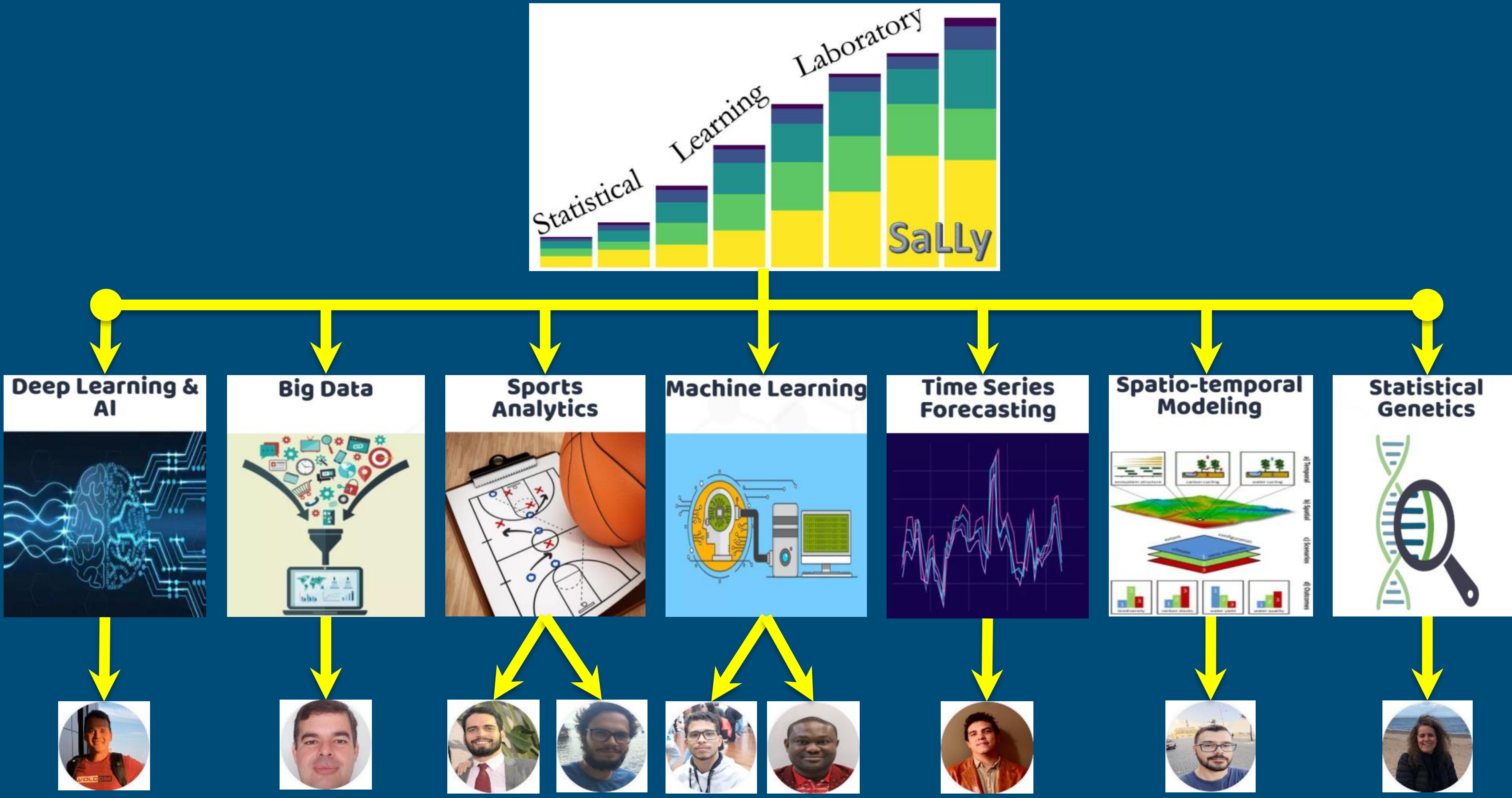
- The sign of the coefficient of the meteorological variable is negative for all models.
- The sign of the coefficient of the variable "land-use transition" is positive for the six models.

UFBA Universidade Federal da Bahia — Departamento de Estatística IME - UFBA — Statistical Learning Laboratory SaLLy

Concluding Remarks

- The number of fire spots in Brazil has been increasing in the last years.
- The most affected biomes are Amazônia, Cerrado, and Pantanal.
- The highest incidence of fire spots is in August, September, and October.
- Loss in land-use and occupation by green areas (forests or non-forest natural formations) in 2011 - 2021. Opposite trend for land-use and occupation by farming.
- Land-use transition (from green areas to farming) was statistically significant for all models/biomes with positive coefficients, representing a considerable increase in the transition from green areas to farming. A higher value for this variable represents a higher number of fire spots.
- The atmospheric variable (precipitation for Pampas and humidity for the other five biomes) was statistically significant for all models/biomes with negative coefficients, i.e., the variable is inversely correlated with the number of fire spots.

SaLLy – Statistical Learning Laboratory



Thank you for your attention!

Questions/Remarks/Suggestions?

E-mail: paulocanas@gmail.com

Web: www.paulocanas.org

Statistical learning laboratory: www.SaLLy.ufba.br

/company/88682389

/sally.laboratory

References

- Lee, D., Rushworth, A., & Napier, G. (2018). Spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST package. Journal of Statistical Software, 84, 1-39.
- Pimentel, J., Bulhões, R. and Rodrigues, P.C. (2023). Spatio-temporal modelling of the Brazilian wildfires: The influence of human and meteorological variables (submitted).