

Machine Learning Assessment of Synoptic and Hydroclimatic Influence on Persistent Wildfire Events Across Brazilian Biomes (2003–2024)

Louise da Fonseca Aguiar¹, Priscila Esposte Coutinho¹, Livia Sancho², Vitor Luiz Victalino Galves¹, Aimée Guida³, Marcio Cataldi¹

¹ Postgraduate Program in Biosystems Engineering, Federal Fluminense University, 156 Passo da Pátria St., 24210-240 - Niterói, Brazil

² COPPE, Universidade Federal do Rio de Janeiro, 2030 Horácio Macedo Av., 21941-594 - Rio de Janeiro, Brazil

³ Geoscience Graduate Program, Faculty of Geology, Rio de Janeiro State University, 524 São Francisco Xavier St., 20550-900 - Rio de Janeiro, RJ, Brazil

louisedaguiar@gmail.com; priscila_esposte@id.uff.br; liviasancho@gmail.com; vitor_luiz@id.uff.br; aimeeguida@gmail.com; mcataldi@id.uff.br.

Abstract Brazil encompasses diverse climate regimes and biomes that are differently affected by climate change. Among the emerging challenges, fire activity has become a key environmental and socioeconomic concern due to its impact on carbon emissions, biodiversity integrity, and human health. Within this context, this study investigates the influence of different components on the persistence and propagation of wildfire events across Brazilian biomes from 2003 to 2024. Wildfire hotspots data were sourced from the satellite detection system of the Brazilian National Institute for Spatial Research (INPE). Persistent fire events were identified using a spatio-temporal clustering algorithm, where individual detections were considered connected when occurred within a 1 km spatial radius with a temporal continuity of at least 24 hours, defining a fire event. A machine learning forecasting framework based on XGBoost and Random Forest was developed to estimate the probability of wildfire persistence and propagation from the detection time, with synoptic-scale system indices representing the South Atlantic Convergence Zone (SACZ), Standard Precipitation Index (SPI) and the Extreme Heatwave Index (XHWI) as predictors to estimate the likelihood that a detected fire would persist or propagate on the following day (t+1). The propagation was defined by the emergence of a new hotspot buffer at the boundary of the original buffer, as long as no previous fire detection existed in that location. This procedure was iteratively evaluated on a daily basis, allowing the identification of successive propagation buffers originating from both the initial and previously propagated fire locations. Separate models were trained for each Brazilian biome, accounting for regional differences in fire behavior, and model performance was assessed using five-fold cross-validation, with 80% of the data used for training and 20% for validation in each iteration. This framework represents a first step toward predicting the short-term evolution of active wildfire events in Brazil.

Keywords: atmospheric blocking; SACZ; fire propagation; synoptic climatology; cross-validation; XGBoost; SPI; heatwaves.