Agricultural drought stress forecasting across different biomes in Brazil using machine learning

1. Abstract

Drought events across Brazil have become more common in recent years. The northeast, marked by the dry semi-arid Caatinga biome, experiences the most extreme drought events, which can have a large impact on economic productivity including disruptions to energy and transport infrastructure as well as agricultural losses. However, droughts are an issue across Brazil and can have significant effects on vegetation across the six biomes which make up the country. Drought stress, defined by a reduction in the vegetation health index (VHI) below the 40% threshold is indicative of significant crop yield reductions. Here we evaluate monthly forecasted reductions in VHI below the 40% threshold using machine learning. We then use the results to understand how climatological features influence drought stress in each biome. This is achieved through a combination of Empirical Orthogonal Function (EOF) analysis, Shapley plots, and correlation analysis of predicted spatial trends. We observe distinct behaviour in the Caatinga biome which is rooted in its unique regional climate patterns and physical geography. For other biomes, VHI, RZSM and precipitation are less strongly linked; however high inertia in VHI allows for high forecast performance. The findings strongly imply that drought forecasting models should treat Northeast Brazil as a distinct system, training models specifically for this region to improve the prediction accuracy of vegetation health and integrated drought indices.

2. Introduction

In recent years, drought episodes have become increasingly frequent across Brazil. The Northeast, marked by its semi-arid climate and the Caatinga biome, is the most severely affected by extreme droughts, which can significantly undermine economic productivity—leading to agricultural losses and disruptions in energy and transportation systems (Cunha et al., 2019). Drought stress impacts in Northeast Brazil have been reported since the sixteenth century (Marengo et al., 2017). For example, drought in 1997–1998 resulted in agricultural losses of 57% with the total economic damage estimated to be 5% of the regions GDP (Marengo et al., 2017). Agriculture in northeast Brazil is both important and vulnerable to drought impacts. The region has 7.8 M ha of agricultural lands with a share of 6–13% of the Brazilian production of soybean, maize, coffee, sugarcane, milk and beef. However, 95% of farmed land consists of rainfed agriculture, with a lack of irrigation increasing vulnerability to drought impacts (Marengo et al., 2022).

Many studies have been conducted on droughts in Northeast Brazil (Marengo et al., 2022, Cunha et al., 2019, Marengo et al., 2017, Lopes Ribeiro et al., 2021, Gallear et al., 2025). Previous drought monitoring studies estimated soil moisture trends in the region (Zeri et al., 2022). They found that soil moisture anomaly (SMA) and standardized precipitation index (SPI), showed a lagged correlation of 1 to 1.5 months in the annual scale with vegetation health index (VHI), suggesting that negative trends in SMA and SPI can be an early warning for yield losses (approximated by vegetation health) during the growing season. Gallear et al. (2025) built on this work by showing high correlations between soil moisture, SPEI and precipitation with a one-month time delay.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

However, drought impacts are not limited to the northeast region of Brazil. Most recently, severe drought in Amazonia saw precipitation deficits of the order of 50 to 100 mm/month and temperatures 3˚C greater than normal, leading to reduced evapotranspiration and soil moisture indicators (Marengo et al., 2024). In the west, drought has also impacted the Pantanal wetland biome. During the period of 2019 – 2020, a lack of precipitation caused by predominance of warm dry air masses from subtropical latitudes led to reduced river levels which subsequently affected transport systems (Marengo et al., 2021, Ferreira Barbosa et al., 2022). Droughts have also impacted South Brazil, causing US$ 3.5 billion in soybean yield losses from 1974 to 2019 (Miyamoto, 2024).

Here, we build on the work of Zeri et al. (2022) and Gallear et al. (2025) to determine how variations in climate across the six biomes of Brazil affect the probability of agricultural drought stress, and the propagation from meteorological variables to agricultural drought stress. Machine learning models are trained to forecast agricultural drought for each biome and the results are analysed to extract the importance of different predictors and study spatial-temporal variations. Further contextual analyses are performed using Principal Component Analysis (PCA, also referred to as Empirical Orthogonal Function or EOF analysis) and simple linear correlations to help interpret the results.

We seek to answer the following research questions:

1. How may monthly agricultural drought stress forecast model performance vary across Brazil's six biomes?
2. How does monthly propagation from meteorological variation to agricultural drought stress vary across Brazil
3. How does variation between biomes in Brazil affect probability of agricultural drought stress?

3. Methodology

3.1. Study regions

Figure 1 shows the biomes used to separate the data used in this study, the soybean harvested area in relation to biome boundaries and the cropping calendar with planting and harvesting periods of soybean for the different regions of Brazil.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

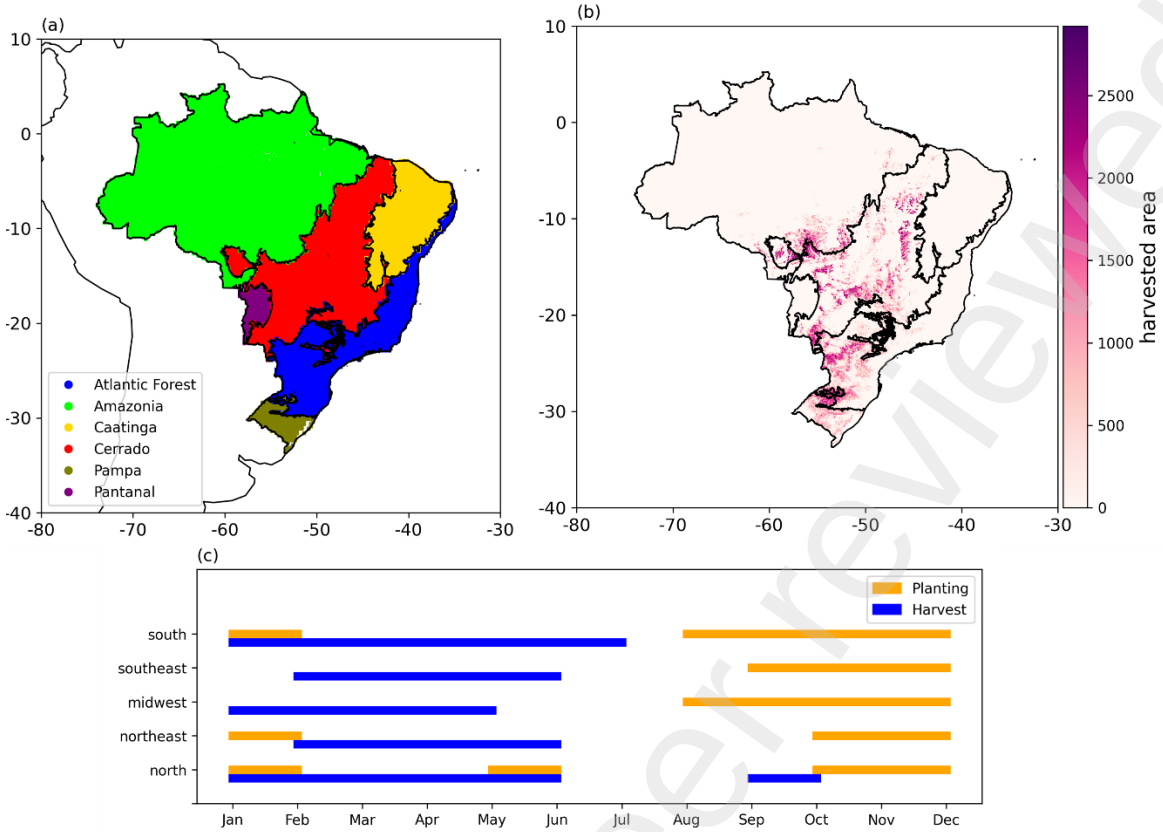


Figure 1. (a) *Each biome of Brazil: Atlantic Forest, Amazonia, Caatinga, Cerrado, Pampa and Pantanal. (b) Soybean harvested area, data from CROPGRIDS dataset (Tang et al., 2024). (c) Growing season calendar for regions of Brazil (BOPAR, 2025).*

The Caatinga biome averages an annual rainfall below 800 mm per year (Alvares et al., 2013). This, combined with high evapotranspiration rates (exceeding 2500 mm yr^{-1}) (Lopes Ribeiro et al., 2021) contributes to reduced water availability and a very limited storage capacity of rivers which are mostly intermittent (Lopes Ribeiro et al., 2021). Furthermore, Caatinga soils are generally shallow (0-50 cm) meaning limited soil moisture storage capacity (Cirilo, 2008). Annual precipitation in the Cerrado varies between 600–2200 mm, it is the second largest biome with wettest regions bordering the amazon and driest regions close to the Caatinga biome. Amazonia is the largest biome, mainly characterized by rainforest areas and an equatorial climate, with daily mean temperatures ranging from 22–28 ˚C. Torrential rains are distributed throughout the year (Lopes Ribeiro et al., 2021). The Pantanal is the smallest biome in Brazil; however, it is one of the world's largest wetlands (Marengo et al., 2021, Lopes Ribeiro et al., 2021). The Pantanal acts as a large reservoir, causing a lag of up to 5 months between inflows and outflows; summer rains determine the flooding seasons, being between November and March in the north and May and August in the south (Marengo et al., 2021). The Atlantic Forest is the biome which includes the major urban centres of São Paulo and Rio de Janeiro. The Atlantic Forest is extremely heterogenous in composition, spanning from 4˚–32˚ S and covers a wide range of climates and vegetation formations. Elevation ranges from sea level to 2900 m and there can be abrupt changes in rainfall, soil moisture, depth, and type as well as average air temperature (Tabarelli et al., 2005). The Pampa biome located in South Brazil has a wet subtropical climate, characterized by high precipitation throughout the year with hot summers and cold winters, and mostly consists of grasslands with some small patches of forest (Lopes Ribeiro et al., 2021, Overbeck et al., 2007).

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

Agricultural land is unevenly distributed across each biome with large areas of soybean farmland found in the Cerrado biome and the boundary between the Atlantic Forest and Pampa biomes. The boundary between Cerrado and the southern part of the Amazonia biome also shows large areas of expanding soybean producing land which has been shown to be of ecological concern (Song et al., 2021, Barona et al., 2010). Unlike the previous study by Gallear et al. (2025), the soybean growing area is not used to filter agrometeorological data to exclude grid cells without significant soybean farming activity. This decision was made to ensure inclusion of areas which may be used by small holder farmers in the analysis and to minimise bias of region selection towards large industrial scale farms. Although there are some changes in planting windows between the regions, October to December is the common planting window across all five regions in Figure 1 (c). For this reason, October to December are the most important months for risk to agricultural productivity, particularly because soybean exports are a significant driver of economic growth for the Brazilian economy (Cattelan and Dall'Agnol, 2018). Therefore, some analysis of model forecast performance should focus on this period of months.

3.2. Data

Data used in this study were either obtained from satellite products such as the Climate Hazards Center Infrared Precipitation with Station data (CHIRPS) (Funk et al., 2015) or from reanalysis data (such as ECMWF's ERA5 (Hersbach, 2023)) (Table 1). Table 1 shows the data sources with the abbreviations used henceforth throughout the text. The variables of interest are derived from the data sources described in Table 1 at the 0.25˚grid scale across all biomes shown in Figure 1. The variables were split into separate datasets according to biome with a separate model trained on each. Each grid cell in the dataset is a time series of months running from February 2003 to December 2021. 2 metre Temperature (t2m), potential evaporation (pev), and longwave radiation (longrad) data was obtained from the monthly averaged ERA5 reanalysis database. The reason for including these three variables in the model input data is to capture temperature effects which may result in drought conditions. T2m is defined as temperature two metres above the surface of the land, Potential evaporation (pev) is defined as a measure of the extent to which the near surface atmospheric conditions are conducive to evaporation, and mean surface downward longwave radiation flux (longrad) is included to capture potential heat effects. ERA5 is a reanalysis database which combines climate model data with observations using data assimilation to provide better estimates of meteorological variables at the grid scale (Hersbach, 2023).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

143

144

145

Table 1. *Variables used in this study with data sources and abbreviations used throughout the text.*

Variable (units)	Abbreviation	Source	Usage
Precipitation (mm/month)	Precip	CHIRPS	Feature
2 metre temperature (K)	T2M	ERA5	Feature
Potential evaporation (kg m^{-2})	pev	ERA5	Feature
Mean surface downward long-wave radiation flux (W m^{-2})	longrad	ERA5	Feature
Root Zone Soil Moisture (%)	RZSM	NASA GRACE	Feature
Vegetation health index (%)	VHI	NOAA STAR	Next month's value used to derive dependent variable
Standardized evapotranspiration -precipitation index (Unitless)	SPEI	Calculated from CHIRPS precipitation data and ERA5 data	Feature

146

147 ### 3.3. Vegetation health index (VHI)

148 The vegetation health index (VHI) is a satellite-based proxy for estimating vegetation health,
149 with values below 40% representing stress conditions. The VHI is a weighted sum of the
150 vegetation condition index (VCI) and the temperature condition index (TCI), determined
151 through the following formulae:

152 $$\mathrm{VCI} = \frac{100(\mathrm{NDVI} - \mathrm{NDVI_{min}})}{\mathrm{NDVI_{max}} - \mathrm{NDVI_{min}}} \#(1)$$

153 $$\mathrm{TCI} = \frac{100(\mathrm{BT_{max}} - \mathrm{BT})}{\mathrm{BT_{max}} - \mathrm{BT_{min}}} \#(2)$$

154 $$\mathrm{VHI} = \alpha\mathrm{VCI} + (1 - \alpha)\mathrm{TCI} \#(3)$$

155 where BT is the brightness temperature recorded from a thermal sensor, max/min
156 represents the maximum and minimum values of a variable over the study period and α is a
157 fixed coefficient used to determine the relative contribution of VCI and TCI to VHI. NDVI is
158 the normalized Difference Vegetation Index. The VHI data is obtained from the NOAA STAR
159 satellite-based vegetation health system based on the Advanced Very High Resolution
160 Radiometer (AVHRR) found on NOAA polar orbiting satellites (Kogan, 1997). The data is
161 upscaled from 0.036˚ to 0.25˚spatial resolution.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

3.4. Root Zone Soil Moisture (RZSM)

Root Zone Soil Moisture (RZSM) is used to determine the propagation of meteorological drought to soil moisture drought affecting plants (Gallear et al., 2025). Root zone refers to the top metre of soil, and RZSM is estimated using the NASA GRACE satellite (Li et al., 2019). The NASA GRACE satellite data are based on two satellites which record changes in the earth's gravity field caused by the redistribution of water. Similar to the previous study (Gallear et al., 2025), RZSM is averaged from weekly to monthly timescales to align with monthly vegetation health index data.

3.5. Total monthly Precipitation (Precip)

Precipitation data is obtained from the Climate Hazards Group Infrared precipitation with station data (CHIRPS) database (Funk et al., 2015). CHIRPS is a dataset which combines satellite data with in-situ measurements to provide a gridded dataset which has been used in many large-scale studies such as (Gallear et al., 2025, Lees et al., 2022, de Oliveira‐Júnior et al., 2021).

3.6. Standardized Precipitation-evapotranspiration Index (SPEI)

Standardised precipitation index (SPI) is the most widely adopted metric for monitoring meteorological drought (Svoboda and Fuchs, 2017). It's a measure of the precipitation relative to the climatology for that specific location and time of year. SPEI builds upon SPI by integrating potential evapotranspiration (PET) data to calculate a climatic water balance using the difference between precipitation and PET (Vicente-Serrano et al., 2010). In this study, daily PET is estimated using the Penman-Monteith method, as recommended by the United Nations' Food and Agriculture Organisation, and subtracted from the daily precipitation to calculate the climatic water balance (Allen et al., 1998). The climatic water balance is aggregated over running 30-, 60- and 90-day periods, with each then fitted to a log-logistic distribution to calculate daily standardised values, which were then used to calculate the SPEI (Vicente-Serrano et al., 2010). Daily SPEI values were finally averaged over monthly periods to get one, two- and three-month SPEI values used in this study. The base period used for this dataset to define the climatology is 1995–2014, reflecting the IPCC AR6 base period (IPCC, 2021). We used SPEI in this work rather than SPI due to stronger monthly correlations with VHI at the grid scale across Brazil, particularly in the Northeast Caatinga and central Midwest regions, and at shorter accumulation periods (Gallear et al., 2025).

3.7. Definition of drought stress

The accepted VHI threshold to signify serious drought stress on vegetation is 40%, this threshold has been applied globally (Kogan et al., 2013). This threshold stems from reductions in VCI and TCI being indicative of reductions in maize yield of more than 50% (Kogan, 1997). This value is therefore used to issue drought warnings and so can be interpreted as the threshold at which meteorological drought propagates to impact vegetation health to which we infer agricultural impacts such as yield losses (Kogan et al., 2013, Kogan, 1997, Gidey et al., 2018, Kloos et al., 2021, Dalezios et al., 2014).

This absolute definition of drought stress (rather than a relative value based on the distribution of VHI at each biome) leads to thresholds at different extremities of each dataset. Figure 2 shows the location of the threshold in relation to the monthly VHI distribution for each of the biomes. Most important to note is that the threshold of drought stress is closest to the median value of VHI for the Caatinga biome. This means that the ratio of drought to

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

non-drought cases is most equal in the Caatinga, meaning that drought occurs nearly half of all instances in this biome.

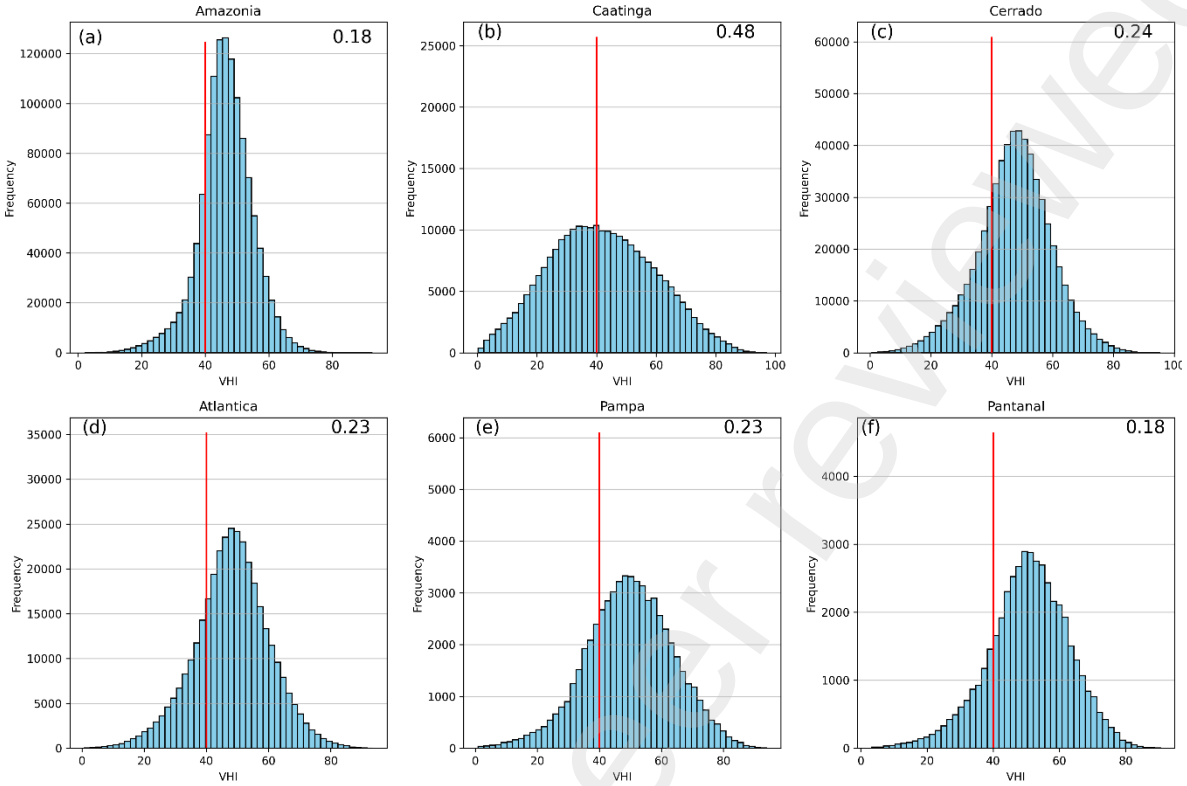


Figure 2. *Spatio-temporal distributions of monthly VHI for each biome across both time and space showing the threshold value (40%) as a red vertical line through each histogram (a-f). The ratio of drought to non-drought cases across the spatio-temporal distribution of VHI for each biome is shown in the top right corner of each panel with drought being defined by the values which fall below the 40% threshold.*

The position of the 40% threshold in relation to each biome results in differing proportions of drought cases in each biome. The drought to non-drought ratio is highest in Caatinga (0.48), with the second highest ratio (Cerrado) being half that of the Caatinga (0.24). Amazonia and Pantanal jointly have the lowest drought to non-drought ratios.

3.8. Machine Learning methods

Following Gallear et al. (2025), we used a random forest method to forecast the probability of drought stress for the subsequent month (defined as mean monthly VHI being below 40%). Random forest is an ensemble method first developed by (Breiman, 2001). The method partitions data into subsets based on conditions at each leaf node of the tree. Random forest constructs a specified number of trees then averages the result of each individual tree (Marsland, 2011, Breiman, 2001). There are several advantages to using tree-based methods over more complex ones such as neural networks, the main advantages being better interpretability, greater robustness to outliers, and simpler calibration (Delerce et al., 2016, Gallear, 2023).

Spatial and per biome model validation was undertaken using a leave-one-out approach in which each model is trained and tested successively with a leave-one-year-out approach, The test scores from each set are then used to produce the distribution of test scores shown in Figure 3. This is done to provide a better overall estimate of model performance and an

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

estimate of model stability across different training and testing datasets. Hyperparameters were left at default values to reduce risk of over-tuning and provide a more general estimate of model performance. A year-based train and test split prevents spatial information leakage from training to testing set. Forecast model performance is evaluated using several classification metrics. We use accuracy, recall, precision, F1 score as well as True positive rate and false positive rate. F1 score is the harmonic mean of precision and recall; this is especially important for the grid cell level evaluation which contains many imbalanced grid cells with much fewer drought instances than non-drought instances. These metrics are defined by the following formulae:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \#(4)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \#(5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \#(6)$$

$$F1 = \frac{2 \times \text{Precion} \times \text{Recall}}{\text{precision} + \text{Recall}} \#(7)$$

Where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

To determine the effect of changing the drought definition threshold on model performance, for each biome we produce Receiver Operating Characteristic (ROC) curves using a range of drought thresholds. The ROC is a standard technique for summarizing classifier performance across the trade-offs between true positive and false positive rates (Chawla, 2010). An ROC curve is a line plot of true positive rate versus false positive rate. The greater the area under the curve, the greater the model performance is relative to incorrectly forecasting droughts (i.e. The greater the ratio of true positives to false positives). If model results are plotted as a diagonal line across with a 1-1 ratio of true positives to false positives this would show the model performance is the same as random chance. ROC curve analysis was performed using training data from 2003 to 2018 with a testing period of 2018 to 2021.

SHAP (with associated Shapley values) is a feature attribution method used to determine the contribution of each feature to the model output. SHAP is a computational method which assigns proportional values to features depending on the influence of each feature on the prediction (Molnar, 2025). SHAP is computationally expensive and complex and so requires approximations and large computational resources (Chen et al., 2023). However, it is also a model agnostic method which, crucially, considers the influence of correlations between features (Rodríguez-Pérez and Bajorath, 2020). This is especially important for meteorological features such as rainfall and solar radiation which will likely be correlated.

3.9. Empirical Orthogonal Function (EOF) analysis

We use EOF analysis to describe how the spatio-temporal patterns of VHI, RZSM and precipitation may vary, providing indication for the reasons of the variation in forecast model performance. EOF analysis has been previously applied to climate and agronomic data to detect and correlate dominant patterns of variability (Challinor et al., 2003). EOF analysis (also referred to as Principal Component Analysis) reduces a larger set of variables to a new

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

smaller set of variables which are linear combinations of the original ones, chosen to represent the maximum possible fraction of the variability contained in the original data. EOF analysis has the potential for great insights into both the spatial and temporal variations in the data depending on the nature of the linear combinations which are most effective at compressing the data (Wilks, 2006).

4. Results

4.1. Classification metrics and model validation

The overall accuracy of the random forest models for each biome is between 0.8 and 1. Some biomes have a much higher proportion of non-drought cases than others (Figure 2). Therefore, to gain a full picture of model performance, recall and precision are also shown per biome and grid cell (Figure 3). Recall and precision are combined into the F1 score (Equation 7) and is shown with the other metrics in Figure 3. Results from Figure 3 show increased variability of model performance metrics in Caatinga, but high overall grid cell level metrics for all biomes across Brazil. High precision and recall mean that models are not significantly biased towards predicting false positives or false negatives. Greater variability in recall in the Caatinga biome indicates that in some years, more droughts are incorrectly forecasted than others. The maps also show that the central Caatinga has a lower recall score (between 0.6 and 0.4) than the rest of the biome and much of the Cerrado, this indicates that droughts are overestimated in this location. There is also a marked difference between model performance in the north of the Amazonia as compared to south Amazonia. Cross validation shows that median model performance for all metrics is between 0.8 and 1 showing high consistency in model performance across test years.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

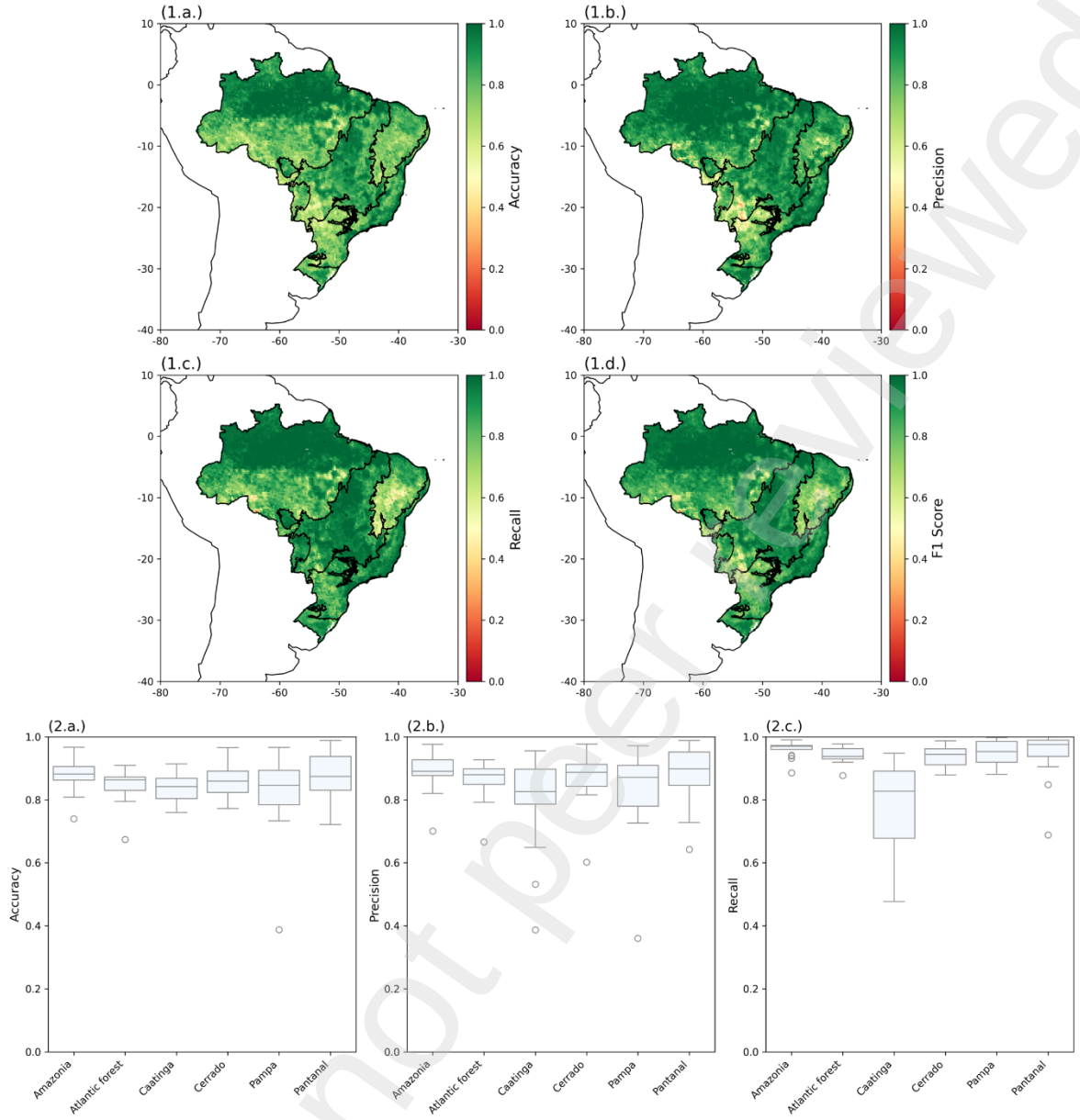


Figure 3. *Accuracy (1.a.) precision (1.b.), recall (1.c.) and f1 score (d) of the random forest forecast model for predicting stress condition in VHI data 1 month in advance. Test data is for the last 4 years of the time series (2018 – 2021). Accuracy (2.a.), precision (2.b.) and recall (2.c.) of random forest models using leave one out cross validation for each year (2003 – 2021)*

Figure 4 shows the monthly consistency of forecast accuracy, recall and precision, allowing us to determine the usefulness of forecasts throughout the soybean growing season and contrast this with the rest of the year. Typically, recall, accuracy and precision are consistent throughout the year regardless of month of forecast, though there are some variability and anomalous predictions. Dotted vertical lines which represent a typical planting window across Brazil do not contain significantly different values for model performance than the rest of the year in most cases (with October in Caatinga being the exception). Model performance during this planting window tends to be in the upper end of the annual range for Amazonia, Atlantic Forest, and Pampa, but is in the lower end for Caatinga.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

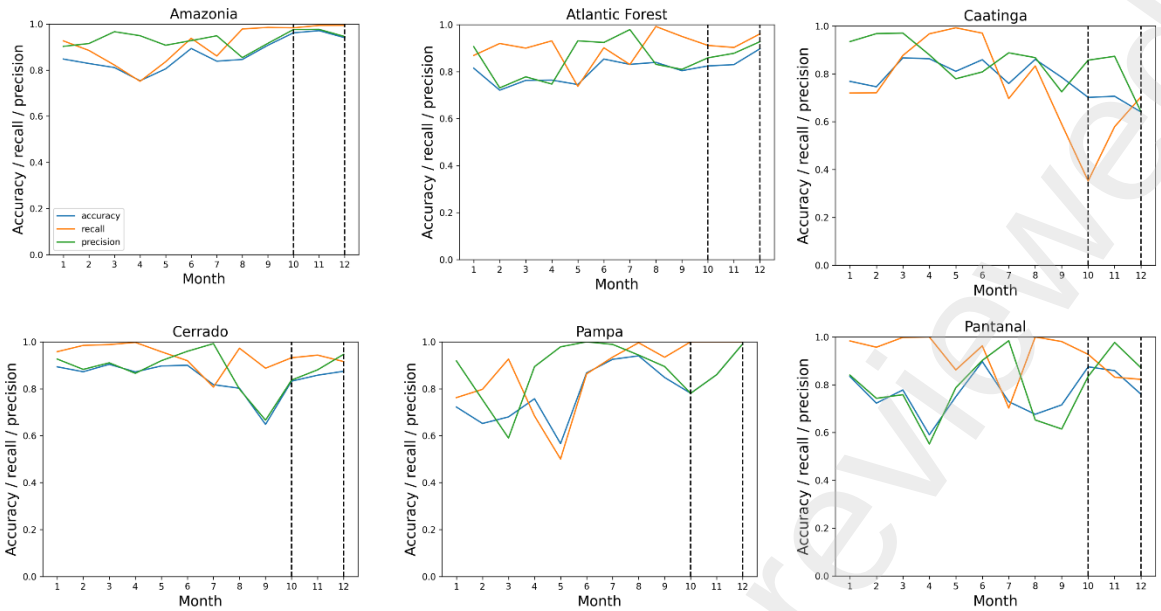


Figure 4. *Line plots of model accuracy, recall and precision across each month for the 4-year test period for the six biomes. Dotted lines show Start of October to December period which is typically the common period of months for Soybean planting across Brazil.*

4.2. ROC curve analysis

ROC curves are here used for determining the effect of changing the drought definition threshold on the trade-off between the true positive and false positive error rates. Therefore, we can examine to what extent more extreme definitions of drought can be forecasted using the same modelling set up for each biome. The greater true positive rate (shown in panel (a) of Figure 5, along with the larger area under the curve in panel (b) indicate both a higher true positive rate at lower thresholds and a greater true positive rate in relation to the number of false positives. panel (a) of Figure 5 shows how the drought threshold of 40% VHI has a much higher true positive rate for the Caatinga biome than other biomes (80%). Panel (b) shows how at the same time the ratio of true positives to false positives is also high. This means that in Caatinga, models can forecast more severe droughts at a greater true positive rate without incorrectly forecasting more droughts at the same time. Other biomes show very comparable results in terms of true positive rates at different thresholds, and the ratio of true positives to false positives at different thresholds. From this we can say that the Caatinga biome stands out as a biome which is unique in the ability of models to forecast severe droughts.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

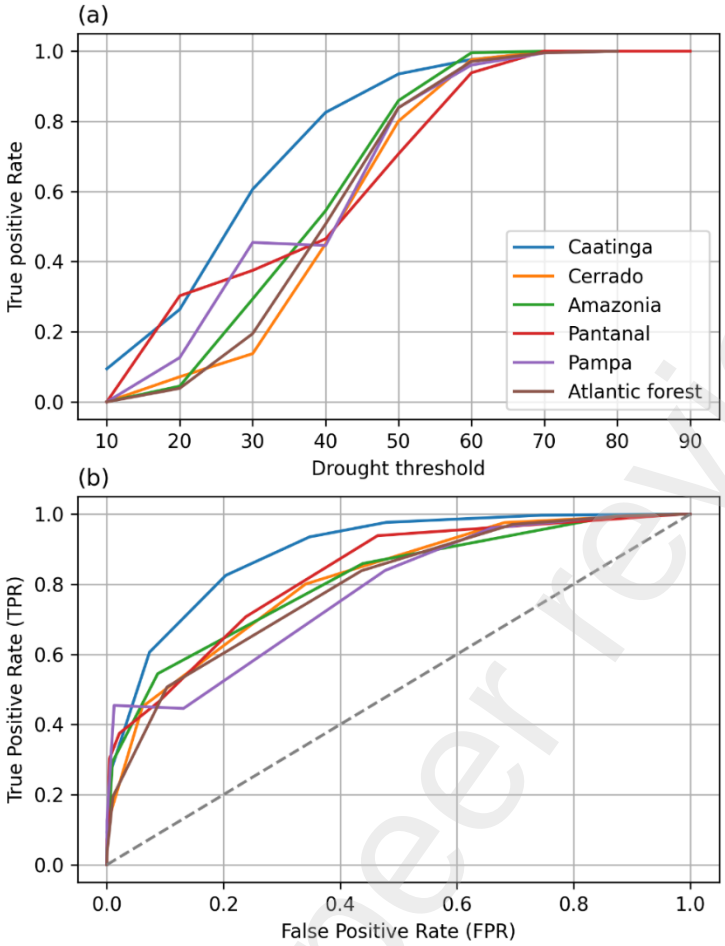


Figure 5. *(a) Rate of true positives against threshold VHI value for each biome, (b) True positive rate against False positive rate for each drought threshold value.*

4.3. Shapley values and spatial relationships

Shapley values explain how much each feature contributes to the variation in the modelled forecasts for each biome (Figure 6). Lagged VHI dominates the contribution to model forecasts across all biomes (ranging from 1.4 to 3 average relative contribution across biomes). High contribution from VHI indicates the slow inertia of VHI which changes gradually month to month. The relative contributions of other variables depend on which biome model forecasts are trained upon. But month of the year (month) and total monthly precipitation (precip) often make relatively high contributions to model forecasts across multiple biomes. SPEI often appears in the top four of the features when ordered by importance for model forecast. Which accumulation period is most important depends on the biome, with sometimes 90 days and sometimes 30 days being a higher contributing feature.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

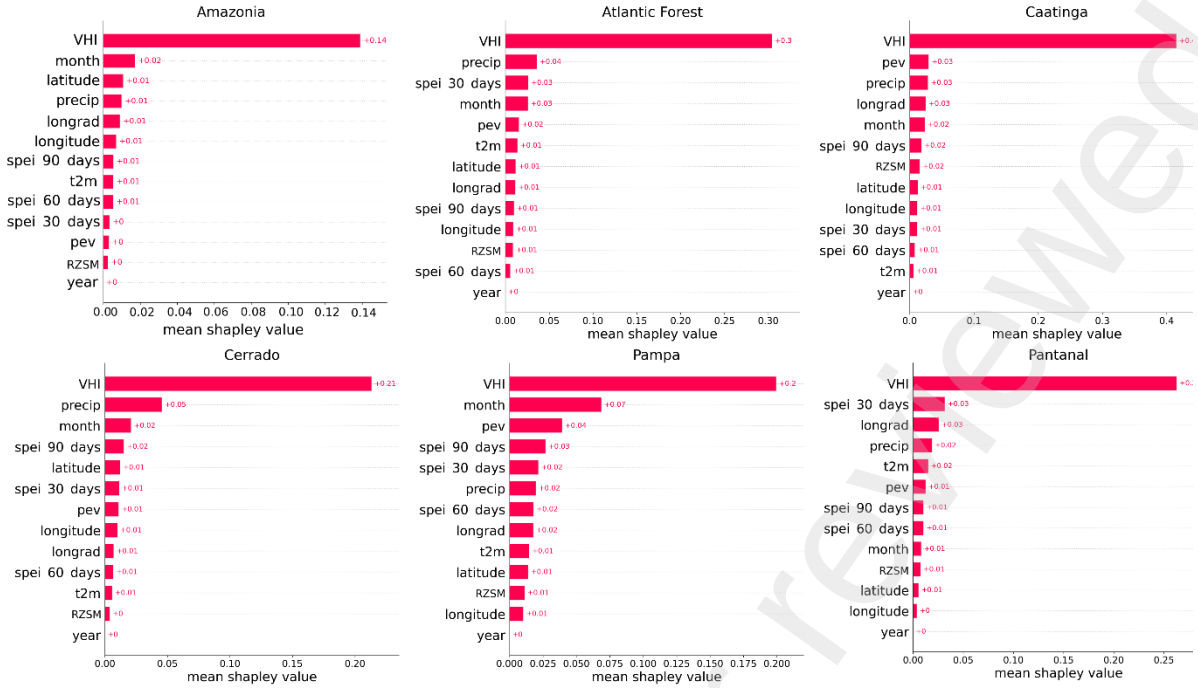


Figure 6. *Shapley plots for each of the six biomes in Brazil.*

Figure 7 shows scatter plots of mean drought probability against monthly total precipitation (a–f) and against RZSM (g–l). Each variable has been averaged over time to produce one value per grid cell. This allows for an understanding of the potential spatial relationships between drought and climate for each biome. Precipitation data indicates that the strongest spatial relationships between average rainfall and probability of drought are for the Caatinga and Cerrado biomes. The Amazonia biome shows a medium Pearsons correlation coefficient value (-0.526) however the relationship is not clear by eye. This may be because the correlation value is heavily influenced by the density of data points in this biome. Other biomes show much weaker relationships between these two variables. There is a correlation coefficient of -0.606 between RZSM and drought probability in Caatinga, however other biomes do not show strong correlations. There is a high degree of confidence in the r values reported for many of the correlations, this is because the number of data points used for the correlation is quite high.

The plots clearly show differences in the spatial variability in rainfall across biomes. In Pampa and Pantanal, the range of rainfall values throughout space is small, whereas in Cerrado, Caatinga and Amazonia, there are much larger ranges in the precipitation values. Moreover, these latter biomes also exhibit stronger correlations between average drought probability and average precipitation. Atlantic forest also has a larger range in precipitation values across space but shows little relationship between average rainfall and average drought probability. RZSM values show a similar pattern with lower spatial variability in Pampa and Pantanal, higher spatial variability in Cerrado, Caatinga and Amazonia, but only Caatinga shows a strong correlation between spatial RZSM values and average drought probability.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

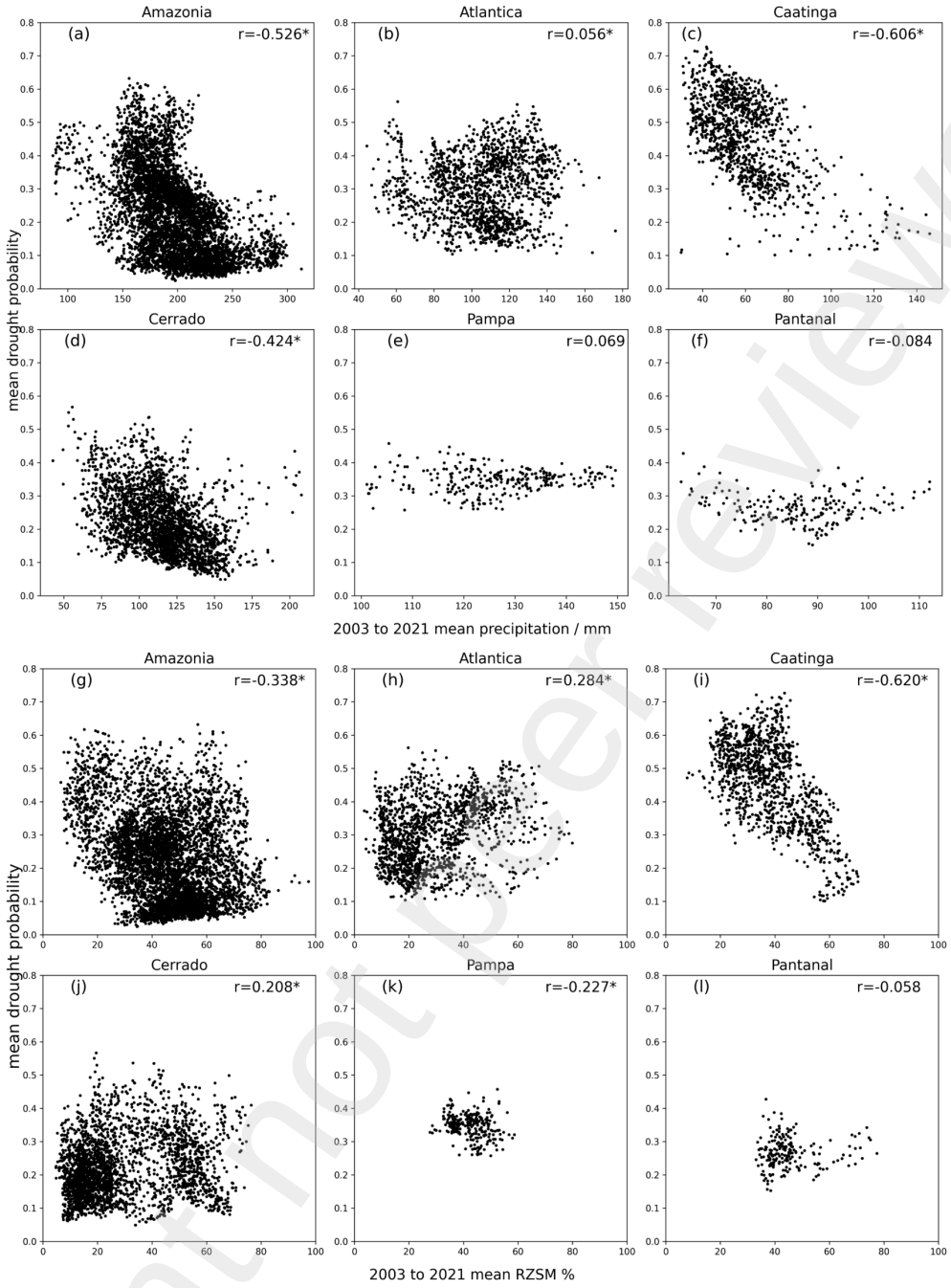


Figure 7. *Scatter plots of average precipitation (a-f) quantities and RZSM percentages (g-l) against probability of drought threshold exceedance per grid cell. Data is shown separately for the 6 biomes of Brazil. Starred r values represent greater than a 95% degree of confidence.*

357

358 4.4. Empirical Orthogonal Function (EOF) analysis

359 The EOF analysis provides insights into explaining the differences in drought stress
360 forecasts across biomes and how propagation of drought can have different spatial
361 dependencies. Figure 8 shows the results of the analysis for RZSM, precipitation and VHI.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

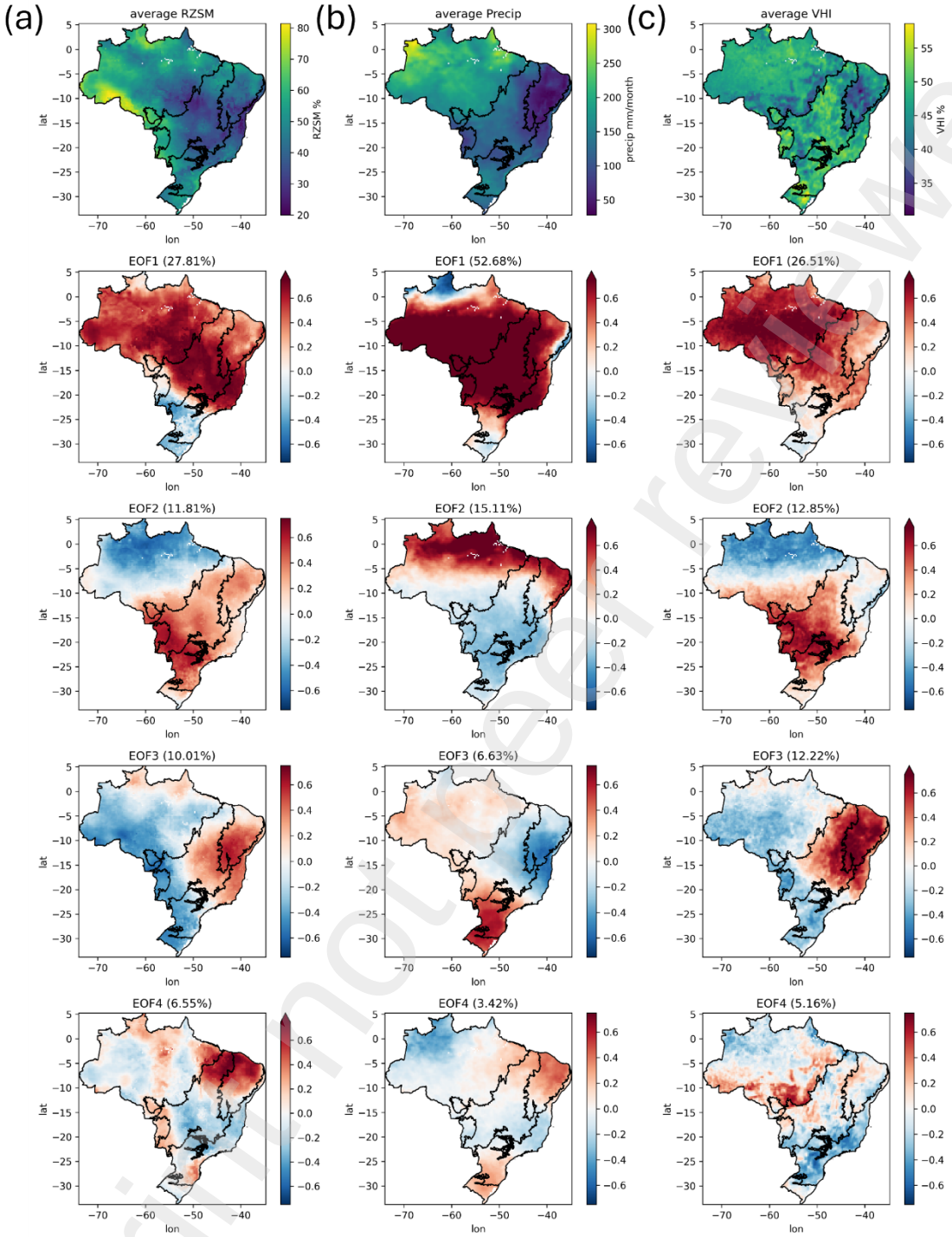


Figure 8. *Average and 4 principal components/EOFs for RZSM (a), Precip (b) and VHI (c). The first row shows the 2003-2021 average values, followed by EOFs 1–4 for each variable in subsequent rows. Labelled percentages refer to the percentage variation of the variable explained by heach EOF.*

Dominant patterns of variation among the three satellite product derived variables are captured by the first EOF modes, with decreasing percentages of variation captured with increasing EOF mode (Figure 8). EOF 1 for each variable captures a pattern with southern Brazil (most notably the Pampa biome) varying distinctly from the rest of the country. EOF 2 for each variable captures a clear distinction between north and south. RZSM EOF2 also

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

shows a pattern in the amazon distinct from the rest of Brazil. EOF 3 presents a distinction between northeast Brazil (where the Caatinga biome and parts of the Cerrado are located) from the rest of Brazil. RZSM, precipitation and VHI show very similar patterns for EOF 3 however there are some very minor differences in the pattern of VHI EOF 3 from the other 2 variables (namely a stronger positive mode in the northeast tip of the Caatinga biome). EOF 4 explains less proportion of the variation for each variable and less similarities are shown between the three variables for EOF 4. EOF 4 of RZSM and precipitation show similar patterns of variation in the northeast, this contrasts with EOF 4 of VHI which is quite different and shows some patterns of variation around the edges of the amazon rainforest in the south, and east of the Amazonia biome. Further EOFs after 4 would capture less and less variation of the data however these were not plotted or analysed as the importance of these EOF modes would be less.

5. Discussion

Overall, accuracy, precision, and recall showed high values (above 0.8) for all biomes. The Caatinga biome had a lower recall median of around 0.8, but with greater spatial variability, with lower values in the central and southern parts of the biome. High accuracy and precision but slightly larger variability in recall in the Caatinga biome indicates that the model over predicts drought in some cases. Results in this biome merit closer examination as it's also more susceptible to droughts. For future forecasting work it is important to note the (Tomasella et al., 2025) study which demonstrated changes in the aridity index towards more arid conditions in the Caatinga. Future work may seek to understand if longer term trends towards increased aridity may affect the performance of agricultural drought forecasts.

When comparing the differences between how each of the biomes affect agricultural drought stress probability at different thresholds, ROC curve analysis shows a true positive rate of 0.4 to 0.5 at the 40% threshold for all biomes excepted Caatinga, where this rate reaches 0.8. When the true positive rate is 0.4 this results in a very low false positive rate of below 0.1 for all biomes. For the Caatinga 0.8 true positive rate results in a false positive rate of around 0.2, and so errors are minimised. This indicates that the threshold of 40% is particularly useful for forecasting drought stress conditions for the next month in this biome. This result is most likely associated with the characteristics of the Caatinga species, which are well adapted to long periods without precipitation due to the short rainy season in most of the region (usually from February to May). Plants in this biome respond quickly to precipitation pulses and dry spells (Medeiros et al., 2022, Santos e Silva et al., 2024). Results from other biomes indicate lower drought stress probability at lower thresholds indicating that more severe droughts are not just less common but also harder to forecast.

Propagation of meteorological drought to agricultural drought is an important question which could be used for earlier warning of the impacts of drought on crop health depending on weather forecasts. Here, we show that the relationship between precipitation and agricultural drought stress probability varies per biome, with some biomes such as the Pantanal showing no correlation between rainfall and agricultural drought but others such as the Caatinga showing a much stronger relationship. Strong correlation between rainfall and vegetation health in the northeast is also shown in (Gallear et al., 2025). Both the results in this study and the previous, show that drought more easily propagates from meteorological to agricultural drought in the northeast, this should make it easier to forecast agricultural drought in this region at longer lead times.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

The Shapley plots show that VHI in the current month is the strongest predictor of drought stress in the following month. This indicates that VHI does not change significantly from month to month, capturing the slow progression of plant health. This result supports the inclusion of VHI in combined drought indices, since it measures a current impact on the vegetation status and a continuous influence on drought intensity the following month. After VHI, either precipitation and/or SPEI appear among the first three features in order of importance, indicating that the water balance components are crucial to assess the likelihood of that meteorological drought conditions will propagate to agricultural drought the following month. While precipitation accounts for the input of water, SPEI considers the balance of precipitation and evapotranspiration, and soil moisture conditions as an indirect assessment, relative to the climatological average during that time of year. This results in SPEI being more able to identify longer and more severe droughts than SPI, which has been shown in multiple studies, especially for arid and semi-arid regions (Tirivarombo et al., 2018, Mwinjuma et al., 2025, Lotfirad et al., 2022).

Correlation analysis of climate variables and probability of drought shows how the climate of different biomes can have varying strengths of relationship with probability of drought. RZSM shows a high correlation with average probability of drought across the Caatinga biome. This is likely because Caatinga soils are generally shallow (0–50 cm) which limits water infiltration and storage capacity, hence fluctuations in soil moisture are more likely to have a significant effect on vegetation health (Lopes Ribeiro et al., 2021). Given previous research which shows decreasing soil moisture trends across the biome (Lopes Ribeiro et al., 2021), this may also affect the vegetation health to a greater extent. Precipitation is strongly correlated with vegetation health both across the Caatinga and Cerrado. Again, shallow soils likely mean that precipitation takes a more rapid effect on vegetation health in the Caatinga biome.

Spatio-temporal dynamics of climate are indicated by the EOF analysis. This reveals that modes of variation in VHI, RZSM and precipitation are often correlated on a coarse scale across Brazil. Most notably, the northeast exhibits specific patterns of variation in VHI, RZSM and precipitation which are distinct to the rest of Brazil. This can be used to better understand how drought may propagate in different regions from meteorological drought to agricultural drought stress. EOF 3 shows common patterns of variation between rainfall, RZSM and VHI in the northeast, therefore indicating that the chance of propagation in this region will often be more likely. This finding coincides with higher correlations between SPEI, RZSM and VHI in the northeast (Gallear et al., 2025) and therefore indicates that forecasting systems should treat northeast Brazil as a separate system when training models to forecast VHI or the integrated drought index (IDI) (which uses VHI, SPI, and RZSM).

We have shown that the impact of meteorological drought varies across biomes. Agricultural land is unevenly distributed across each biome with large areas of soybean farmland found in the Cerrado and Pampa regions. Therefore, forecasting would likely be most important for soybean growth in the Cerrado region. However, most of the Caatinga's rural population depends on agriculture (José Maria Cardoso da Silva, 2019). For this reason, it is important to include regions which may not be so intensely farmed as agricultural drought forecasts are still relevant and useful to smaller scale farming operations. The analysis of model performance shows that monthly performance is consistent (despite anomaly of October in the Caatinga). Therefore, model performance in most biomes can be relied upon for critical months relevant to the cropping season calendar (with the example of soybean given here in this study in Figure 1). If future work is to improve the relevance of forecasts for agriculture, improved skill may be obtained by focusing on specific growing regions where sufficient data allows for robust monthly model performance.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

6. Conclusions

Model evaluation in this work demonstrates strong performance across all Brazilian biomes, with accuracy, precision, and recall scores generally exceeding 0.8. However, the Caatinga biome requires special attention; while median performance is high, it exhibits significant spatial variability, reflecting its inherent drought susceptibility and greater prediction uncertainty, a concern underscored by recent studies showing increasing aridity trends (Tomasella et al., 2025). The most powerful predictor for next-month drought stress is the current month's Vegetation Health Index (VHI), owing to its slow-changing nature (inertia). High inertia in VHI allows for high skill forecasting stress for the subsequent month in other biomes such as Cerrado, Amazonia and Pampa. Strong forecasts are especially important for the Cerrado where most Soybean farming occurs, an important export for the Brazilian economy. Water balance components like precipitation, root zone soil moisture (RZSM), and SPEI are also identified as crucial drivers, with a greater number of droughts forecasted at more severe thresholds in Brazil's northeast Caatinga biome.

The distinct behaviour observed in the Caatinga is rooted in its unique regional climate patterns and physical geography. The biome's characteristically shallow soils limit water storage capacity, making vegetation health highly and rapidly responsive to changes in RZSM and precipitation. Large-scale climate analysis confirms that Northeast Brazil functions as a separate system regarding key climate variables. Therefore, the findings strongly support that drought forecasting models should treat Northeast Brazil as a distinct system, training models specifically for this region to improve the prediction accuracy of vegetation health and integrated drought indices.

Acknowledgements

This work was funded by the Met Office Climate Science for Service Partnership (CSSP) Brazil project under the International Science Partnerships Fund (ISPF).

The authors also acknowledge support from the Growing Health (BB/X010953/1) and AgZero+ (NE/W005050/1) Institute Strategic Programmes both funded by the BBSRC.

Data availability

All data are available from publicly available and free-to-access data repositories. ERA 5 data were obtained from https://doi.org/10.24381/cds.f17050d7 (Hersbach et al., 2023), NASA GRACE data were obtained from https://doi.org/10.5067/UH653SEZR9VQ (Beaudoing et al., 2021; Li et al., 2019), CHIRPS data were obtained from https://data.chc.ucsb.edu/products/CHIRPS-2.0/ (Funk et al., 2014, 2015) and VHI data from https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/vh_ftp.php (NOAA, 2025; Kogan, 1997), SPEI data was calculated using a combination of ERA5 data and CHIRPS data.

7. References

ALLEN, R. G., PEREIRA, L. S., RAES, D. & SMITH, M. 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome,* 300**,** D05109.

ALVARES, C. A., STAPE, J. L., SENTELHAS, P. C., GONÇALVES, J. D. M. & SPAROVEK, G. 2013. Köppen's climate classification map for Brazil. *Meteorologische zeitschrift,* 22**,** 711–728.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

BARONA, E., RAMANKUTTY, N., HYMAN, G. & COOMES, O. T. 2010. The role of pasture and soybean in deforestation of the Brazilian Amazon. *Environmental Research Letters,* 5**,** 024002.
BOPAR. 2025. *Producers must pay attention to planting and harvesting deadlines, warns Mapa* [Online]. Available: https://www.bopar.com.br/noticias-destaque/df-produtores-devem-ficar-atentos-aos-prazos-de-plantio-e-colheita-das-culturas-alerta-mapa [Accessed].
BREIMAN, L. 2001. Random forests. *Machine learning,* 45**,** 5–32.
CATTELAN, A. J. & DALL'AGNOL, A. 2018. The rapid soybean growth in Brazil. *Ocl,* 25**,** D102.
CHALLINOR, A., SLINGO, J., WHEELER, T., CRAUFURD, P. & GRIMES, D. 2003. Toward a combined seasonal weather and crop productivity forecasting system: determination of the working spatial scale. *Journal of Applied Meteorology,* 42**,** 175–192.
CHAWLA, N. V. 2010. Data mining for imbalanced datasets: An overview. *Data mining and knowledge discovery handbook***,** 875–886.
CHEN, H., COVERT, I. C., LUNDBERG, S. M. & LEE, S.-I. 2023. Algorithms to estimate Shapley value feature attributions. *Nature Machine Intelligence,* 5**,** 590–601.
CIRILO, J. A. 2008. Public water resources policy for the semi-arid region. *estudos avançados,* 22**,** 61–82.
CUNHA, A. P. M., ZERI, M., DEUSDARÁ LEAL, K., COSTA, L., CUARTAS, L. A., MARENGO, J. A., TOMASELLA, J., VIEIRA, R. M., BARBOSA, A. A. & CUNNINGHAM, C. 2019. Extreme drought events over Brazil from 2011 to 2019. *Atmosphere,* 10**,** 642.
DALEZIOS, N., BLANTA, A., SPYROPOULOS, N. & TARQUIS, A. 2014. Risk identification of agricultural drought for sustainable agroecosystems. *Natural Hazards and Earth System Sciences,* 14**,** 2435–2448.
DE OLIVEIRA-JÚNIOR, J. F., DA SILVA JUNIOR, C. A., TEODORO, P. E., ROSSI, F. S., BLANCO, C. J. C., LIMA, M., DE GOIS, G., CORREIA FILHO, W. L. F., DE BARROS SANTIAGO, D. & DOS SANTOS VANDERLEY, M. H. G. 2021. Confronting CHIRPS dataset and in situ stations in the detection of wet and drought conditions in the Brazilian Midwest. *International Journal of Climatology,* 41**,** 4478–4493.
DELERCE, S., DORADO, H., GRILLON, A., REBOLLEDO, M. C., PRAGER, S. D., PATIÑO, V. H., GARCÉS VARÓN, G. & JIMÉNEZ, D. 2016. Assessing Weather-Yield Relationships in Rice at Local Scale Using Data Mining Approaches. *PLOS ONE,* 11**,** e0161620.
FERREIRA BARBOSA, M. L., HADDAD, I., DA SILVA NASCIMENTO, A. L., MÁXIMO DA SILVA, G., MOURA DA VEIGA, R., HOFFMANN, T. B., ROSANE DE SOUZA, A., DALAGNOL, R., SUSIN STREHER, A. & SOUZA PEREIRA, F. R. 2022. Compound impact of land use and extreme climate on the 2020 fire record of the Brazilian Pantanal. *Global Ecology and Biogeography,* 31**,** 1960–1975.
FUNK, C., PETERSON, P., LANDSFELD, M., PEDREROS, D., VERDIN, J., SHUKLA, S., HUSAK, G., ROWLAND, J., HARRISON, L. & HOELL, A. 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data,* 2**,** 1–21.
GALLEAR, J. W. 2023. *Using machine learning and process-based crop modelling for regional scale prediction.* University of Leeds.
GALLEAR, J. W., VALADARES GALDOS, M., ZERI, M. & HARTLEY, A. 2025. Evaluation of machine learning approaches for large-scale agricultural drought forecasts to improve monitoring and preparedness in Brazil. *Natural Hazards and Earth System Sciences,* 25**,** 1521–1541.
GIDEY, E., DIKINYA, O., SEBEGO, R., SEGOSEBE, E. & ZENEBE, A. 2018. Analysis of the long-term agricultural drought onset, cessation, duration, frequency, severity and spatial extent using Vegetation Health Index (VHI) in Raya and its environs, Northern Ethiopia. *Environmental Systems Research,* 7**,** 13.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

HERSBACH, H. 2023. ERA5 monthly averaged data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), Electronic resource: 10.24381/cds. f17050d7 (2023). Accessed 2023–03–01.
IPCC 2021. Climate Change 2021: The physical Science Basis. Contribution of working roup I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.
JOSÉ MARIA CARDOSO DA SILVA, I. R. L., MARCELO TABARELLI (ed.) 2019. *Caatinga: The largest tropical dry forest region in South America*
KLOOS, S., YUAN, Y., CASTELLI, M. & MENZEL, A. 2021. Agricultural drought detection with MODIS based vegetation health indices in southeast Germany. *Remote Sensing,* 13**,** 3907.
KOGAN, F., ADAMENKO, T. & GUO, W. 2013. Global and regional drought dynamics in the climate warming era. *Remote Sensing Letters,* 4**,** 364–372.
KOGAN, F. N. 1997. Global drought watch from space. *Bulletin of the American meteorological society,* 78**,** 621–636.
LEES, T., TSENG, G., ATZBERGER, C., REECE, S. & DADSON, S. 2022. Deep learning for vegetation health forecasting: a case study in Kenya. *Remote Sensing,* 14**,** 698.
LI, B., RODELL, M., KUMAR, S., BEAUDOING, H. K., GETIRANA, A., ZAITCHIK, B. F., DE GONCALVES, L. G., COSSETIN, C., BHANJA, S. & MUKHERJEE, A. 2019. Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research,* 55**,** 7564–7586.
LOPES RIBEIRO, F., GUEVARA, M., VÁZQUEZ-LULE, A., CUNHA, A. P., ZERI, M. & VARGAS, R. 2021. The impact of drought on soil moisture trends across Brazilian biomes. *Natural Hazards and Earth System Sciences,* 21**,** 879–892.
LOTFIRAD, M., ESMAEILI-GISAVANDANI, H. & ADIB, A. 2022. Drought monitoring and prediction using SPI, SPEI, and random forest model in various climates of Iran. *Journal of Water and Climate Change,* 13**,** 383–406.
MARENGO, J., A, CUNHA, A., P, ESPINOZA, J.-C., FU, R., SCHÖNGART, J., JIMENEZ, J., C, COSTA, M., C, RIBEIRO, J., M, WONGCHUIG, S. & ZHAO, S. 2024. The Drought of Amazonia in 2023-2024. *American Journal of Climate Change,* 13**,** 567–597.
MARENGO, J. A., CUNHA, A. P., CUARTAS, L. A., DEUSDARÁ LEAL, K. R., BROEDEL, E., SELUCHI, M. E., MICHELIN, C. M., DE PRAGA BAIÃO, C. F., CHUCHÓN ANGULO, E., ALMEIDA, E. K., KAZMIERCZAK, M. L., MATEUS, N. P. A., SILVA, R. C. & BENDER, F. 2021. Extreme Drought in the Brazilian Pantanal in 2019–2020: Characterization, Causes, and Impacts. *Frontiers in Water,* Volume 3 - 2021.
MARENGO, J. A., GALDOS, M. V., CHALLINOR, A., CUNHA, A. P., MARIN, F. R., VIANNA, M. D. S., ALVALA, R. C., ALVES, L. M., MORAES, O. L. & BENDER, F. 2022. Drought in Northeast Brazil: A review of agricultural and policy adaptation options for food security. *Climate Resilience and Sustainability,* 1**,** e17.
MARENGO, J. A., TORRES, R. R. & ALVES, L. M. 2017. Drought in Northeast Brazil—past, present, and future. *Theoretical and Applied Climatology,* 129**,** 1189–1200.
MARSLAND, S. 2011. *Machine learning: an algorithmic perspective*, Chapman and Hall/CRC.
MEDEIROS, R., ANDRADE, J., RAMOS, D., MOURA, M., PÉREZ-MARIN, A. M., DOS SANTOS, C. A., DA SILVA, B. B. & CUNHA, J. 2022. Remote sensing phenology of the Brazilian caatinga and its environmental drivers. *Remote Sensing,* 14**,** 2637.
MIYAMOTO, B. C. B. 2024. *Estimation of agricultural revenue losses due to droughts in the state of Rio Grande do Sul.* Statistics, Federal University of Rio Grande do Sul.
MOLNAR, C. 2025. *Interpretable machine learning*.
MWINJUMA, M., WANG, R., MTUPILI, M. & TWAHA, M. 2025. Comparisons of SPI and SPEI in capturing drought dynamics: A Global assessment across arid and humid regions. *Atmospheric Research***,** 108475.
OVERBECK, G. E., MÜLLER, S. C., FIDELIS, A., PFADENHAUER, J., PILLAR, V. D., BLANCO, C. C., BOLDRINI, I. I., BOTH, R. & FORNECK, E. D. 2007. Brazil's

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

neglected biome: the South Brazilian Campos. *Perspectives in Plant Ecology, Evolution and Systematics,* 9**,** 101–116.
RODRÍGUEZ-PÉREZ, R. & BAJORATH, J. 2020. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of computer-aided molecular design,* 34**,** 1013–1026.
SANTOS E SILVA, C. M. S., BEZERRA, B. G., MENDES, K. R., MUTTI, P. R., RODRIGUES, D. T., COSTA, G. B., DE OLIVEIRA, P. E. S., REIS, J., MARQUES, T. V. & FERREIRA, R. R. 2024. Rainfall and rain pulse role on energy, water vapor and CO2 exchanges in a tropical semiarid environment. *Agricultural and Forest Meteorology,* 345**,** 109829.
SONG, X.-P., HANSEN, M. C., POTAPOV, P., ADUSEI, B., PICKERING, J., ADAMI, M., LIMA, A., ZALLES, V., STEHMAN, S. V. & DI BELLA, C. M. 2021. Massive soybean expansion in South America since 2000 and implications for conservation. *Nature sustainability,* 4**,** 784–792.
TABARELLI, M., PINTO, L. P., SILVA, J. M., HIROTA, M. & BEDE, L. 2005. Challenges and opportunities for biodiversity conservation in the Brazilian Atlantic Forest. *Conservation Biology,* 19**,** 695–700.
TANG, F. H., NGUYEN, T. H., CONCHEDDA, G., CASSE, L., TUBIELLO, F. N. & MAGGI, F. 2024. CROPGRIDS: a global geo-referenced dataset of 173 crops. *Scientific Data,* 11**,** 413.
TIRIVAROMBO, S., OSUPILE, D. & ELIASSON, P. 2018. Drought monitoring and analysis: standardised precipitation evapotranspiration index (SPEI) and standardised precipitation index (SPI). *Physics and Chemistry of the Earth, Parts a/b/c,* 106**,** 1–10.
TOMASELLA, J., DO AMARAL CUNHA, A. M., ZERI, M. & COSTA, L. C. 2025. Changes in the aridity index across Brazilian biomes. *Science of The Total Environment,* 989**,** 179869.
VICENTE-SERRANO, S. M., BEGUERÍA, S. & LÓPEZ-MORENO, J. I. 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate,* 23**,** 1696–1718.
WILKS, D. S. 2006. *Statistical methods in the atmospheric sciences*, Academic press.
ZERI, M., WILLIAMS, K., CUNHA, A. P. M., CUNHA-ZERI, G., VIANNA, M. S., BLYTH, E. M., MARTHEWS, T. R., HAYMAN, G. D., COSTA, J. M. & MARENGO, J. A. 2022. Importance of including soil moisture in drought monitoring over the Brazilian semiarid region: An evaluation using the JULES model, in situ observations, and remote sensing. *Climate Resilience and Sustainability,* 1**,** e7.

Preprint not peer reviewed

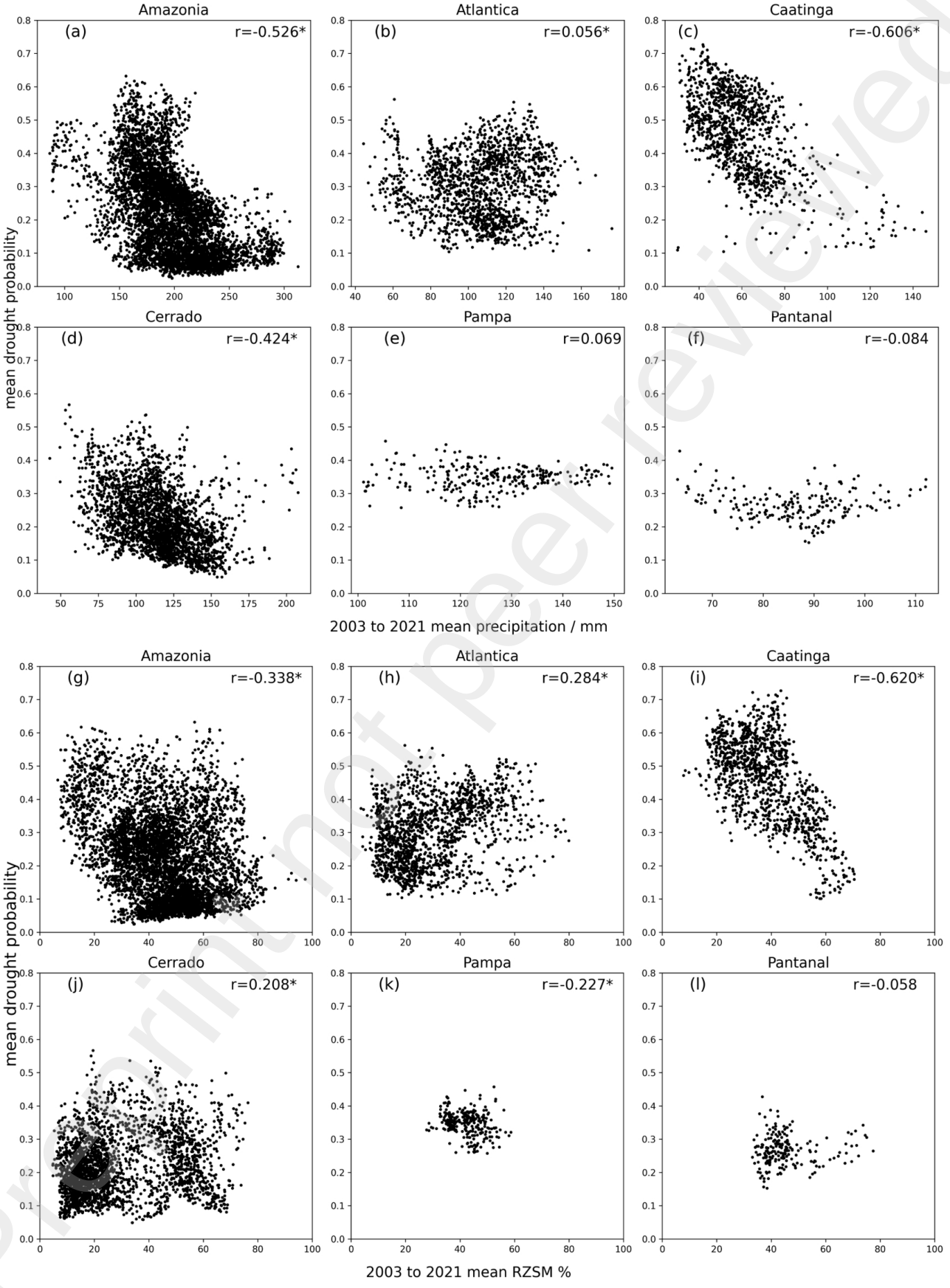
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758

Table 1. *Variables used in this study with data sources and abbreviations used throughout the text.*

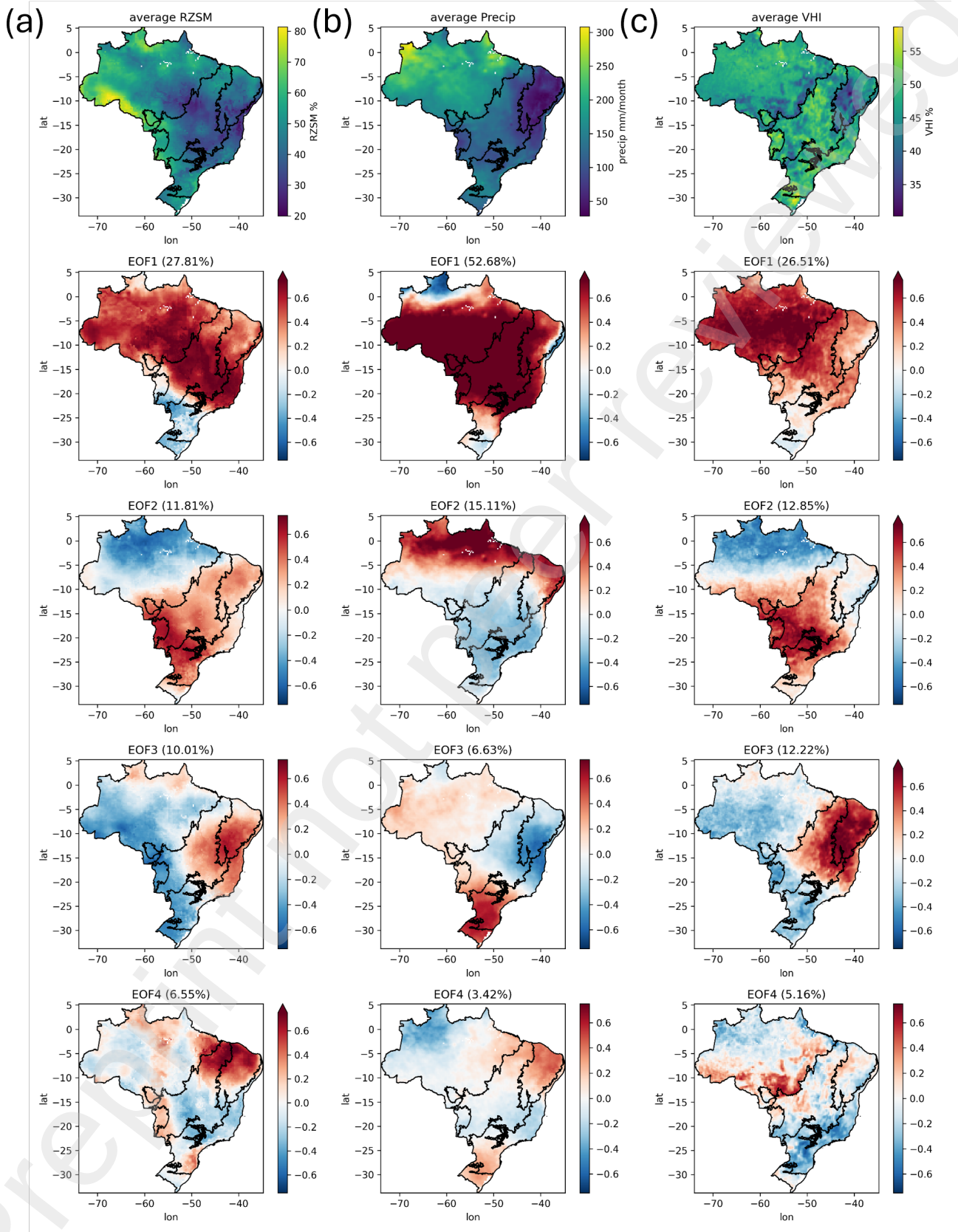
Variable (units)	Abbreviation	Source	Usage
Precipitation (mm/month)	Precip	CHIRPS	Feature
2 metre temperature (K)	T2M	ERA5	Feature
Potential evaporation (kg m^{-2})	pev	ERA5	Feature
Mean surface downward long-wave radiation flux (W m^{-2})	longrad	ERA5	Feature
Root Zone Soil Moisture (%)	RZSM	NASA GRACE	Feature
Vegetation health index (%)	VHI	NOAA STAR	Next month's value used to derive dependent variable
Standardized evapotranspiration -precipitation index (Unitless)	SPEI	Calculated from CHIRPS precipitation data and ERA5 data	Feature

Preprint not peer reviewed

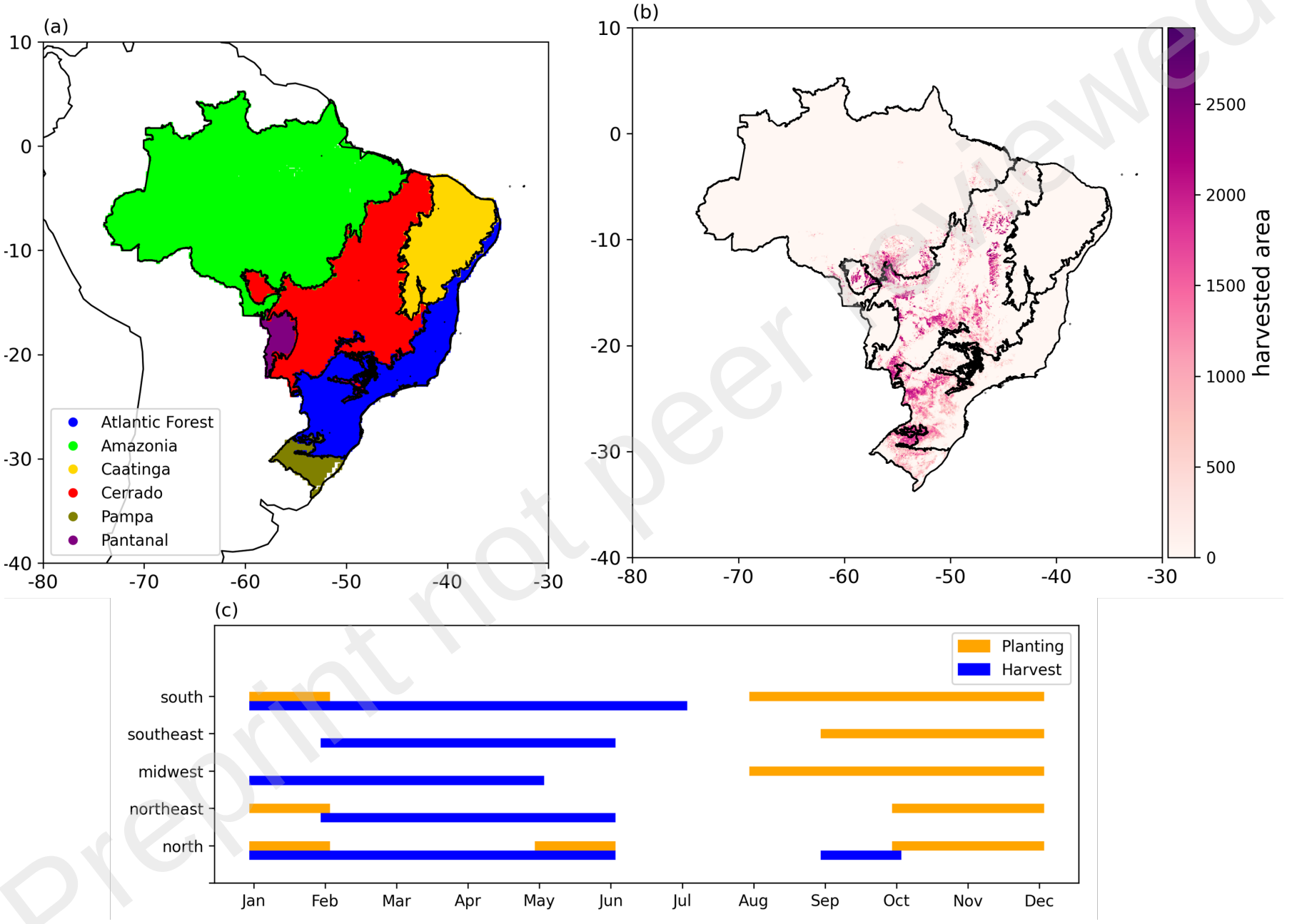
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=5655758



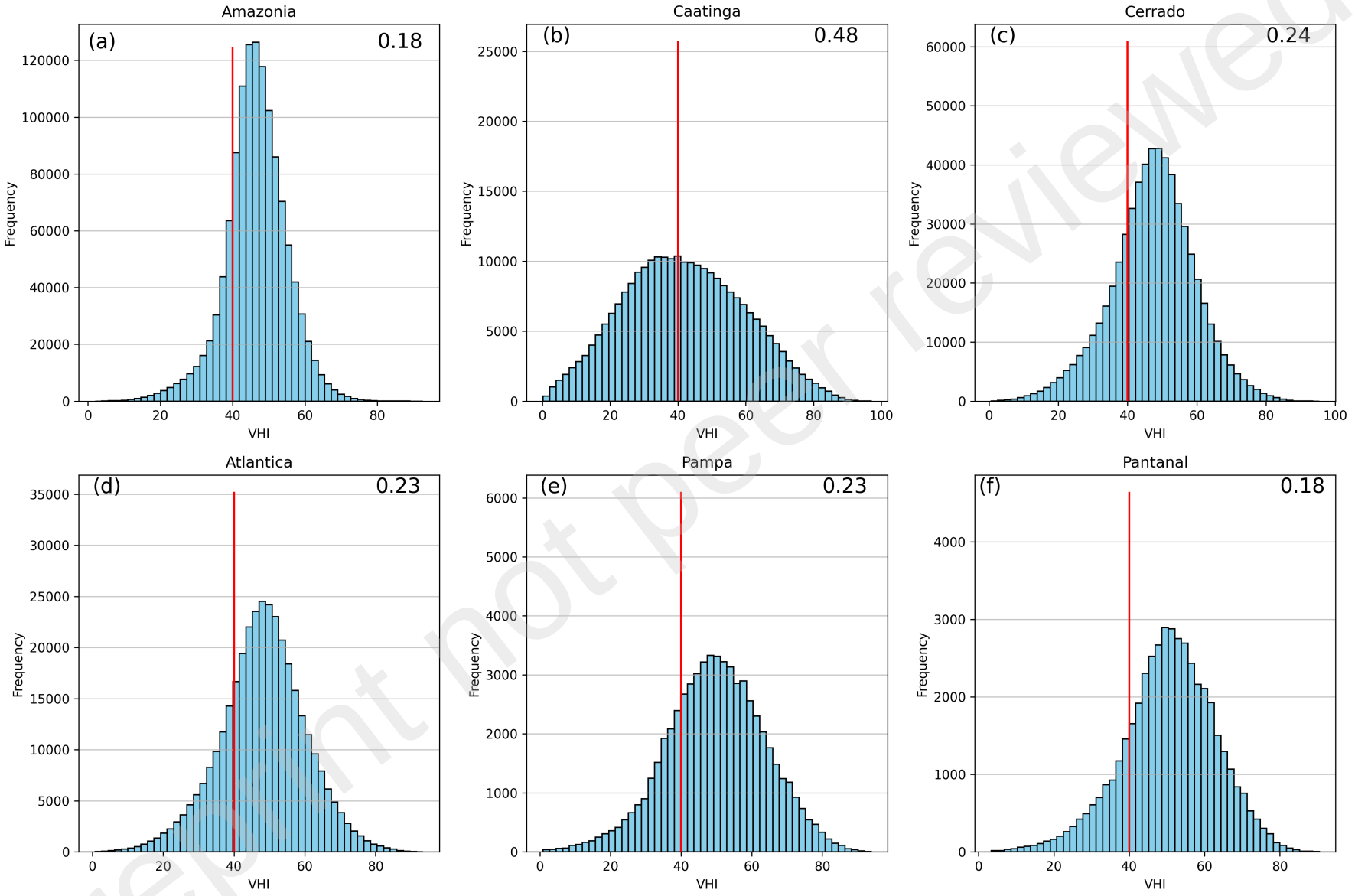
Preprint not peer reviewed



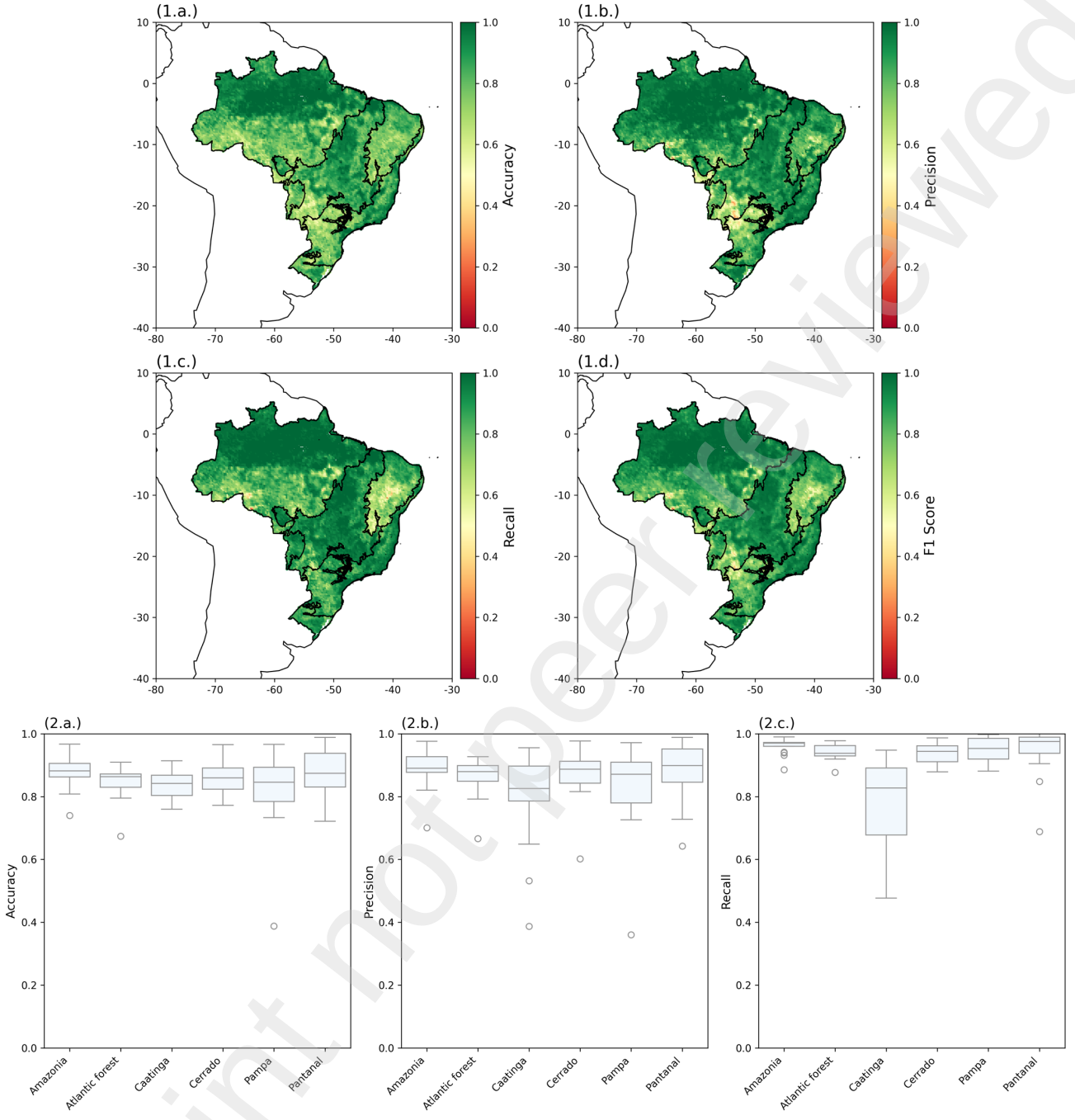
Preprint not peer reviewed



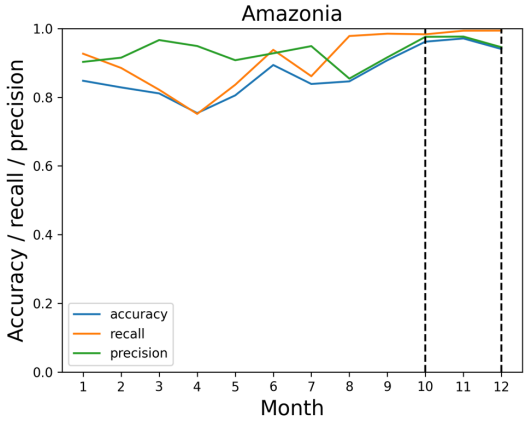
Preprint not peer reviewed

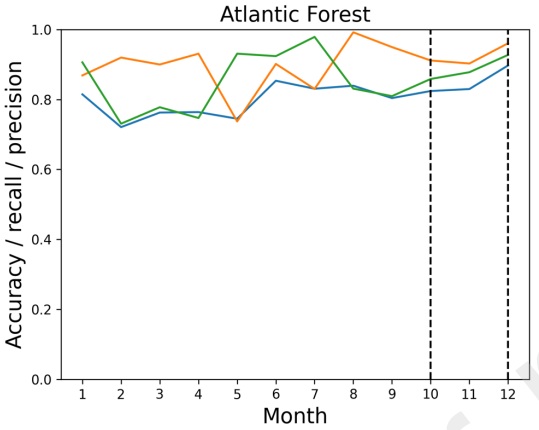


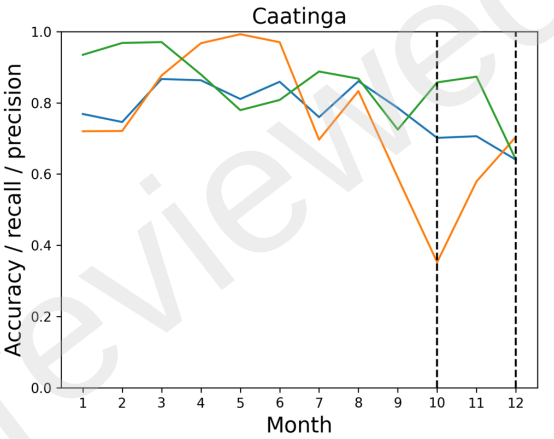
Preprint not peer reviewed

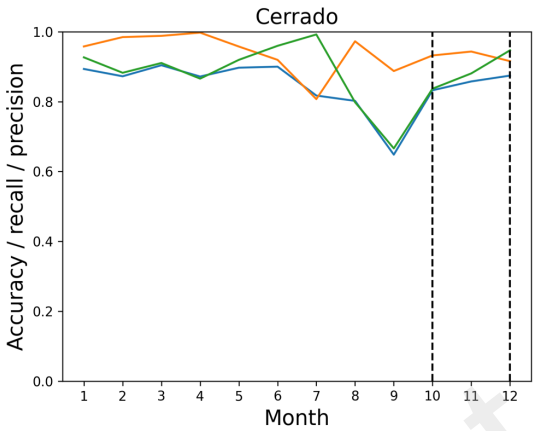


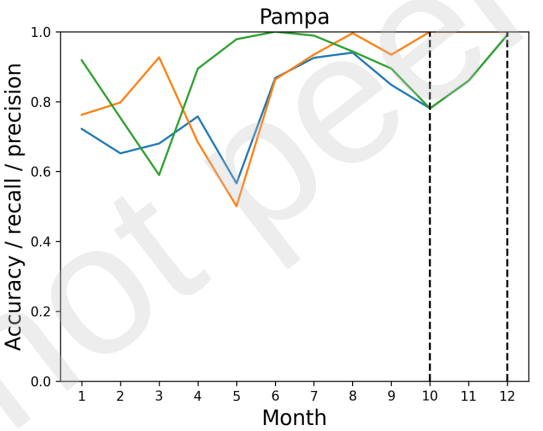
Preprint not peer reviewed

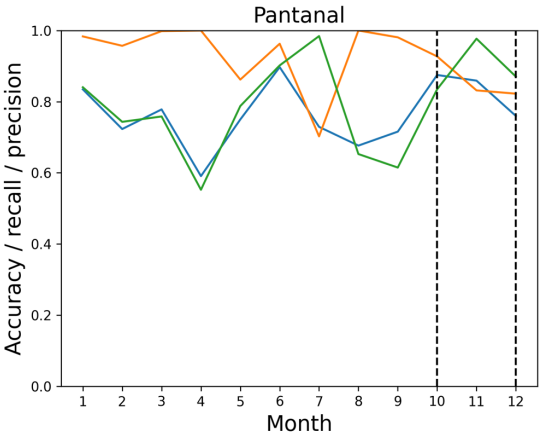




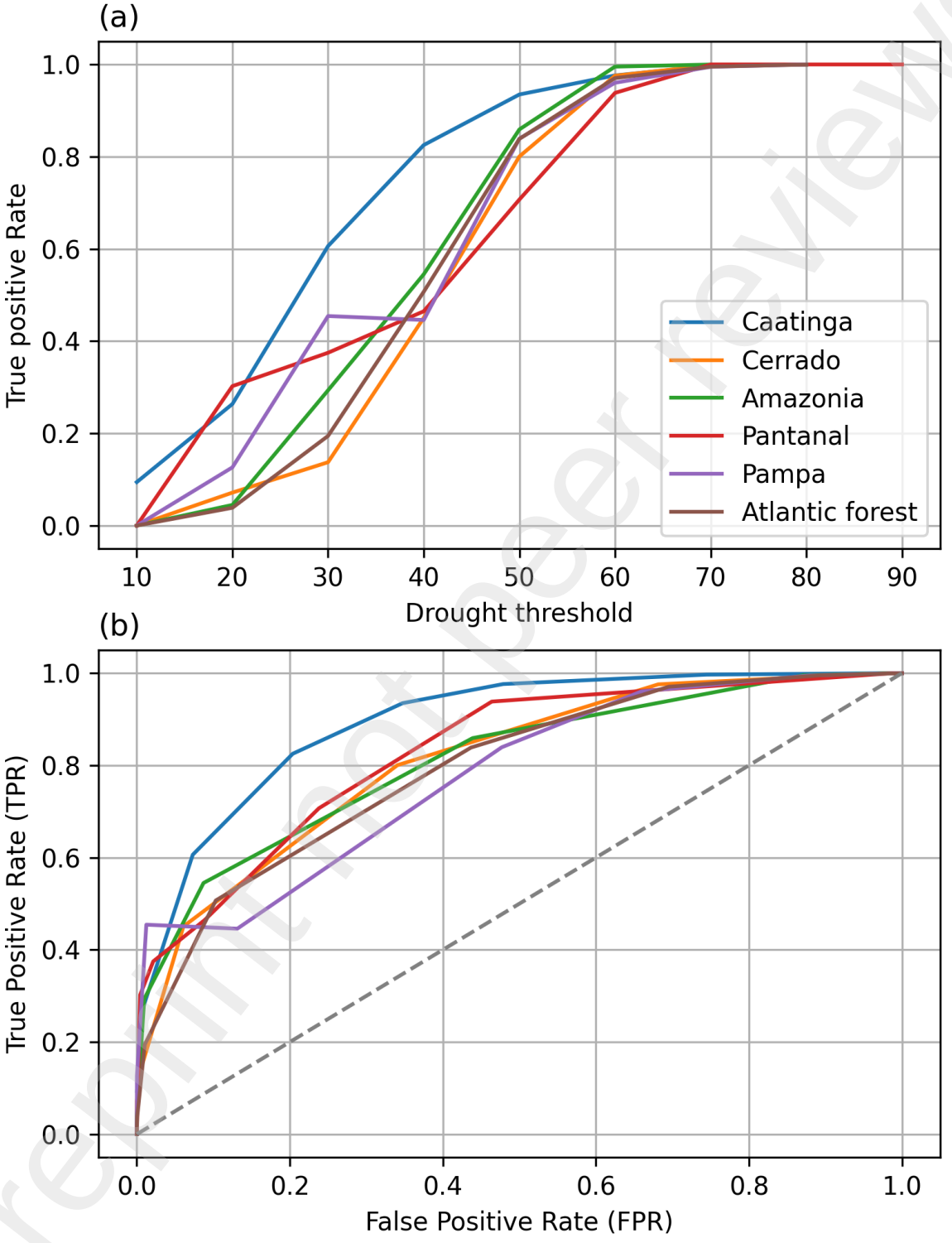




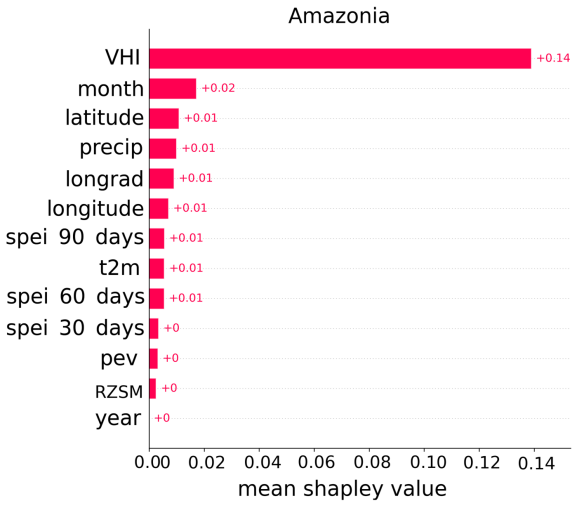


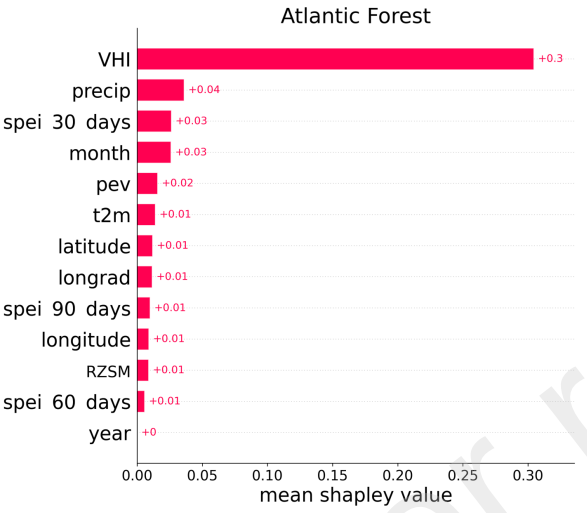


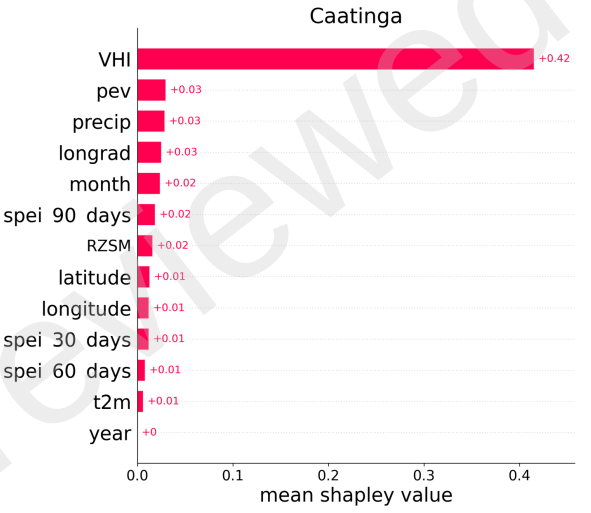
Preprint not peer reviewed

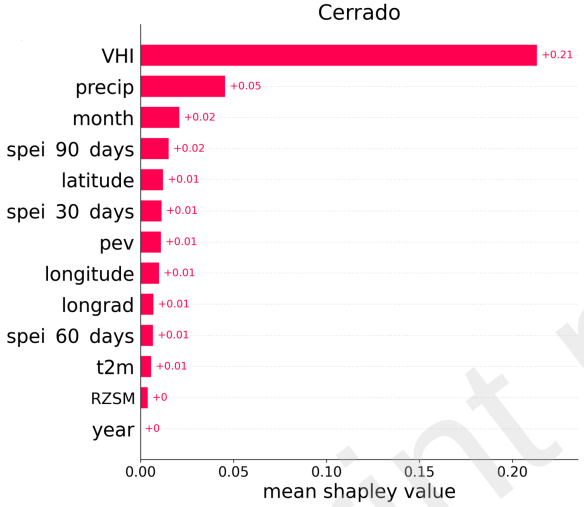


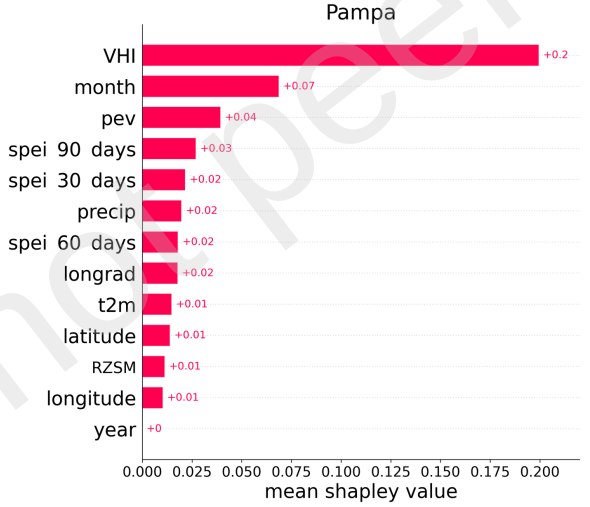
Preprint not peer reviewed

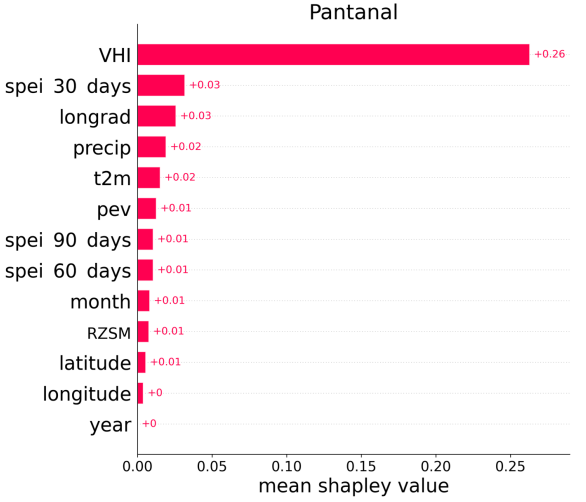












Preprint not peer reviewed