Modelling the moisture status of habitats by using NDVI on the example of the Cerrado and Atlantic Forest biomes borderland (Brazil)

Adam Młynarczyk [1-2],
Monika Konatowska [3],
Wojciech Kowalewski [4],
Sławomir Królewicz [1],
Kelly Cristina Tonello [2],
Rogerio Hartung Toppa [2],
Marcos Roberto Martines [2],
Jan Piekarczyk [1],
Paweł Rutkowski [3]

The Brazilian Cerrado and the Atlantic Forest are important global biodiversity hotspots that are highly diverse in geological structure, soil, and climatic conditions, all of which directly affect vegetation diversity. The conservation of these biomes depends on recognizing variations in their humidity levels. Given the correlation between water access and plant health, as illustrated by NDVI, we assessed the feasibility of building an NDVI-based model to detect variations in habitat moisture. Using various statistical algorithms, the correctness of the NDVI-based habitat moisture assessment model was confirmed. In addition, it was determined that UMAP was the most favourable of the algorithms employed. Our method provides a practical, efficient tool for assessing habitat moisture that can benefit fields such as ecology, conservation biology, and land management.

Keywords: Ipanema National Forest, NDVI, UMAP Algorithm, Habitat Moisture Index

Introduction

All ecosystems on Earth depend on water. Several methods are used for detecting differences in the moisture status of terrestrial ecosystems and monitoring changes in their hydrological conditions, such as remote sensing (Lin et al. 2022), ground-based measurements (Weaver et al. 2019), hydrological models (Bai et al. 2019), and carbon and water flux measurements (Chen et al. 2020). All these methods can identify moisture differences in terrestrial biomes and monitor hydrological changes over time. However, there is a persistent demand for cost-effective, easy-to-apply techniques, particularly for valuable natural areas. Among the most important biomes on the globe are the Brazilian Cerrado, considered a biodiversity hotspot, and the Atlantic Forest. The Cerrado is situated in central Brazil, and is the second-largest of Brazil's major biomes, following Amazonia. It comprises woodlands, savannas, grasslands, as well as gallery and dry forests (Klink & Machado 2005). Kisselle et al. (2002) distinguish two types of Cerrado: the *cerrado sensu stricto* (20-50% canopy cover) and the *cerrado campo sujo* (open, scrubland, <10% canopy cover). This biome hosts some 4,800 species of plants and vertebrates found nowhere else (Strassburg et al. 2016). Despite its enormous importance for species conservation, as little as 19.8% of the Cerrado remains undisturbed, the majority having been transformed into pasture and other forms of agricultural land use (Strassburg et al. 2017). The idea of monitoring the Brazilian Cerrado using spectral indices dates back many years. In 2004, Ferreira and Huete described the seasonal dynamics of this biome using NDVI and the Soil Adjusted Vegetation Index (SAVI), analysed in five seasonal periods: (1) summer rains (from January to March), (2) dry winter (from April to August), (3) end-of-the-drought period (September), (4) beginning of the rainy season (October), and (5) spring rains (November and December). These authors identified three major spatial domains of the Cerrado: (i) the "true" Cerrado formations and pasture sites, (ii) the forested areas (seasonal broadleaf and dense Cerrado woodland), and (iii) agricultural crops. They concluded that "the use of multiple indices should improve vegetation studies".

The Atlantic Forest (AF) is the second-largest tropical forest in South America and one of the most biodiverse biomes worldwide. Brazil owns most of the AF (93% of the total area), with the remainder belonging to Paraguay (5.3%) and Argentina (1.7%) (Marques et al. 2021). The AF is classified as one of three major biodiversity hotspots, together with the Cape Floristic Region and Polynesia-Micronesia region, and particularly vulnerable to climate change (Béllard et al. 2014). The altitudes of the AF vary from sea level to 2891 m, with the relief ranging from depressions to mountains. The climate is extremely diverse, including tropical rainforest (Af); tropical monsoon (Am); tropical savannah (Aw); arid hot steppe (BSh); and temperate, without dry season; with hot (Cfa) and warm (Cfb) summers and temperate dry winters (Cw). Soils vary widely, from dystrophic to eutrophic (Marques et al. 2021).

A common feature of the AF is the existence of two annual cycles. According to Da Silva et al. (2021), vegetation in Caparaó National Park (Brazil) and its buffer zone is highly dependent on precipitation, with a
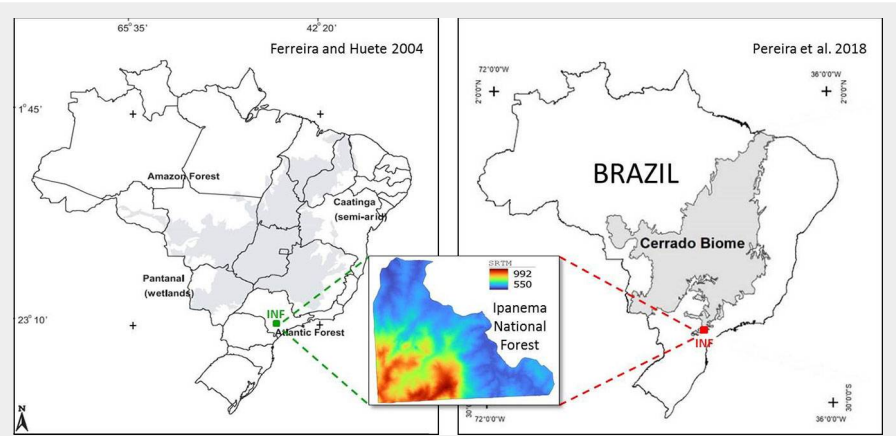
Fig. 1 - Location of the Ipanema National Forest (INF). Location of the INF on the background of the Cerrado biome (highlighted in grey), according to Ferreira & Huete (2004) (left) and Pereira et al. (2018) (right); in the middle relief map of the INF.

strong correlation between precipitation and vegetation indices. Tests of the relationships between forest and vegetation indices conducted in the Atlantic Rainforest on the Atlantic slope of Serra do Mar (Freitas et al. 2005) concluded that MVI5 (moisture vegetation index using Landsat's band 5) and MVI7 (moisture vegetation index using Landsat's band 7) showed the best performances in dense humid forests. In contrast, NDVI was a good indicator for deciduous and dry forests. However, the usefulness of NDVI for ecological research has been demonstrated by multiple authors worldwide (Pravalie et al. 2022, Zhang et al. 2022). Studies on the vegetation at the border of the Cerrado and the Atlantic Forest should help protect these biomes and clarify the relationships between vegetation and habitat conditions.

In this study, we examined the use of NDVI to assess moisture conditions in ecosystems, using the Ipanema National Forest (INF) as a case study. This forest is located at the boundary between the Brazilian Cerrado and the Atlantic Forest biomes. It was assumed that (i) the water accessible to plants is affected by the sum and distribution of atmospheric precipitation and the differentiated water retention

capacity depending on the variability of the terrain properties (land relief, geology, soils), (ii) the humidity level of the habitat affects the diversity of the vegetation cover, and (iii) the NDVI values of the vegetation depends on the amount of available water. Based on these assumptions, the hypothesis was formulated that habitat moisture levels, influenced by vegetation, could lead to differences in NDVI. These variations in NDVI values could then be used to assess differences in habitat moisture. The aim of the study was to develop a mathematical model based on NDVI, vegetation types, and various habitat conditions to monitor changes in habitat moisture. NDVI has not yet been used in this context. This information can help understand how ecosystems are responding to climate change and other environmental stressors.

Material and methods

Study area

The Ipanema National Forest (INF) is a poorly studied protected area (Londe et al. 2022). It encompasses approximately 5,000 ha, of which 75% is covered by semideciduous (90% of all forests) and rainforests in several successional stages (Cavarzere et

al. 2017). Due to extensive exploitation of tree species over the last two centuries, the vegetation consists of secondary forests (Saint-Hilaire 1976). The topography of the INF is highly diversified, and the soil is clay and rocky (Rossi 2017). From September to February, the climate is wet, while it is dry from March to August (Cavarzere et al. 2017). According to Ferreira & Huete (2004), the territory lies within the Atlantic Forest biome; according to Pereira et al. (2018), at the southern range of the Cerrado biome (Fig. 1).

Satellite data

Image data from the Sentinel-2 (A, B) satellites of the European Space Agency (ESA) were used for the analysis. The satellites are equipped with an MSI sensor with characteristics listed on the ESA website (https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi). They are installed on two platforms, namely S2A and S2B. Image data were imported into Google Earth Engine™ (GEE) and subjected to the automatic cloud cover assessment. Data from the Level L2A product were used for the analysis. It means that each image pixel for the spectral band contains a calibrated reflectance of the Earth's surface and was created as a result of geometric correction that accounts for atmospheric radiation changes. In GEE, images with a cloud ratio of less than 10% with a defined INF area were selected. For this purpose, the SCL (Scene Classification Layer) was used. On this layer, areas with the following values were identified: shadows -3, clouds with low probability -7, medium -8, high -9, cirrus clouds -10. Areas with medium and high probability were excluded when their number of pixels was greater than 50. In total, 61 images from 2018 to 2021 were selected. For all images, NDVI was utilized for calculations. As shown in Tab. 1, the differences between Sentinel 2A and 2B are not significant and were therefore considered insignificant for the purposes of these studies.

NDVI was calculated in GEE. The multilayer file was downloaded and analyzed using QGIS ver. 3.22.8. The quality of RGB images was visually assessed, and images with visible defects were discarded. In addition, a plug-in from the website http://terrabrasilis.dpi.inpe.br/ was installed to enable the addition of another layer of the Cerrado forest ranges and Atlantic Forest biomes. Habitat layers were taken from Willmersdorf (2017). The habitat and biome range layers were multiplied by and subtracted from each other, and the resulting boundary layer was used for calculations and for visualizing NDVI values. The calculations were performed using the zonal statistics plug-in, using the boundary layer and the NDVI map. The plug-in is a standard QGIS plug-in that converts a raster layer to a vector layer and calculates variable statistics for a given area, such as sum, mean, deviation, and median.

Tab. 1 - Sentinel 2 channels used in the NDVI calculations (https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument).

Sentinel 2 Bands	B4	B8
Spatial resolution (m)	10	10
Sentinel 2A central wavelength (nm)	664.5	835.1
Sentinel 2B central wavelength (nm)	665.0	833.0
Sentinel 2A bandwidth (nm)	38.0	145.0
Sentinel 2B bandwidth (nm)	39.0	133.0
Other Characteristics	pigment chlorophyll absorptions in red band (Yadav & Borana 2019)	responsive to canopy structural variations, canopy type, and architecture (Yadav & Borana 2019)

Determination of the habitat moisture index

To determine the habitat moisture index, the following data were used: (i) data on precipitation for the years 2018-2021 taken from the Fezenda Ipanema Airport meteorological station (METAR), located 300 m from the eastern border of the research area; this data provided the total precipitation per hour, with 24 measurements per day. (ii) A relief map generated based on the global elevation model SRTM 30 m (Farr et al. 2007). (iii) A map of geological formations obtained from Willmersdorf (2017). (iv) A soil map obtained from Willmersdorf (2017). (v) A map of vegetation types generated based on Willmersdorf (2017).

The classification of habitats into wet, humid, mesic, and dry categories is based on the total atmospheric precipitation, which was compiled as average values for 14-day periods. This classification was related to changes in NDVI values, as illustrated in Fig. S2 (Supplementary material). According to the yearly trend, it was assumed that the data for the first quarter of the year (January-March) correspond to the wet period, and those for the third quarter (July-September) correspond to the dry period. The second and fourth quarters (April-June and October-December, respectively) were considered transitional periods, and classified as "humid" and "mesic", respectively.

Similarly, a subdivision of the year into 4 quarters was adopted for analyzing NDVI data: (i) January to March; (ii) April to June; (iii) July to September; (iv) October to December. Each quarter was assigned colors corresponding to four classes of decreasing humidity: (i) wet - blue and light blue; (ii) humid - green; (iii) mesic - yellow; (iv) dry - red. To calculate the humidity index of habitats, it was assumed that: (i) class "g1" (wet) are areas where the contour of the blue and light blue colours are present in all four quarters of the year; (ii) class "g2" (humid) are areas where the contour of the blue and light blue colours are present in three quarters; (iii) class "g3" (mesic) are areas where the contour of the blue and light blue colours are present in only two quarters; (iv) class "g4" (dry) are areas where the blue and light blue colours are present in only one quarter of the year. By employing applied mathematical algorithms, a model was developed to determine the contours of the areas for g1-g4 classes based on the images. Finally, the habitat moisture index was determined by linking the contours of the habitat moisture levels determined by the model with vegetation, precipitation, soils, terrain, and geological formations.

The model validation involved analyzing results in relation to soil and terrain features that are closely associated with moisture, *e.g.*, the relief of a river valley (Fig. 2g) which collects precipitation water, and fluvisols (Fig. 2c), which are soils particularly dependent on the river water level.

Statistical analysis

The primary goal of the statistical analysis was to assess whether the 38 parameters, categorized into four groups (altitude, vegetation, soils, and geology), could validate the classification of INF habitats into four moisture groups (g1-g4). These groups were based on habitat moisture as derived from NDVI and other related parameters. The aim was to confirm the accuracy of the NDVI-based habitat humidity assessment model, rather than prove that NDVI indicates humidity.

Three tools were used for the above purpose: the study of the covariance matrix, two dimensionality reduction algorithms (PCA and UMAP), and a g1-g4 classifier built using the adopted parameters. The UMAP algorithm (https://arxiv.org/abs/1802.03426) is far more promising than PCA. UMAP focuses on a low-dimensional representation that best reflects the topological structure of the original (high-dimensional) data. It is an alternative to the t-SNE method, a commonly used nonlinear dimension reduction algorithm. UMAP produces similar or better representations, as it retains more global data features, and its output is more stable. Moreover, UMAP is more efficient than t-SNE in both dimensionality and data size.

To construct a classification model of classes g1-g4 based on 38 parameters, we tested five classification algorithms: (i) linear discriminant analysis (LDA); (ii) quantitative descriptive analysis (QDA); (iii) logistic regression; (iv) support vector machine (SVM); (v) Random forest.

All calculations were made using Python and the scikit-learn library (https://scikit-learn.org/stable/). The dataset used consisted of 510 cases, classified into classes g1-g4 according to 38 parameters: 11 classes of vegetation cover, 10 classes of altitude, nine classes of soil, and eight classes of geological formations. The INF area was divided into 1×1 km squares, creating distinct sets in which classes g1 through g4 were established. Each square contained a variety of habitat conditions and vegetation types, allowing us to treat each square as a separate sample (Fig. S1 in Supplementary material). For each tested model, learning curves were generated (Géron 2019) and the accuracy and mean square error of the model were recorded.

The methods were tested on training and validation datasets as a function of the amount of training data. The "LearningCurveDisplay.from_estimator" function, which visualises the method's error as negative values, was used to generate the curves, where higher values indicate higher quality (https://scikit-learn.org/stable/modules/model_evaluation.html#scoring-parameter). The cross-validation level was set to 30. Graphs were created for seven tested algorithms. As for the SVM algorithm, three kernels were tested: linear functions, polynomial functions (*e.g.*, a fifth-degree polynomial), and radial basis functions.

Results

The analysis of NDVI-based images showed that INF vegetation is clearly rainfall-dependent, as reflected by its partition into dry and wet seasons (Fig. S2 in Supplementary material). It also highlighted the relationships between vegetation, soil types, geological formations, and relief.

Fig. 2 shows the diversity of vegetation cover (2a, 2b), soil types (2c, 2d), geological formations (2e, 2f), and relief (2g, 2h) compared to NDVI diversity.

Statistical analysis

Covariance matrix analysis (Fig. 3) revealed that: (i) altitude parameters are weakly or very weakly correlated with parameters belonging to other groups; the maximum correlation values and the percentage of zeros for individual groups were, respectively, 0.09 and 0.7 with the parameters of the vegetation group, 0.18 and 0.73 with the group of geological parameters, and 0.19 and 0.74 with the group of soil parameters; (ii) the group of vegetation parameters was fairly correlated with the group of geological parameters and the group of soil parameters; the maximum correlation values and the percentage of zeros for individual groups were respectively: 0.49 and 0.33 with the group of geological parameters, 0.41 and 0.13 with the group of soil parameters; (iii) the group of geological parameters was fairly correlated with the group of soil parameters; the maximum correlation value was 0.72, while the percentage of zeros was 0.1. This suggests considering class parameterisation limited to the altitude block and one of the other three blocks.

The PCA algorithm results indicated that trying to reduce the parameter space dimension with a linear model is likely to fail. Fig. S3 (Supplementary material) shows the percentage distribution of the degree of explanation for all principal components. The quantitative impact of the original parameter values for the first ten principal components is visualised in Fig. S4. It is clearly visible that the three blocks (vegetation, geology, and soil) equally strongly affect their construction, which suggests their mutual redundancy.

Fig. S5 illustrates the 510 examined cases projected onto the first two UMAP components. The poor separation of groups g1 and g4 does not concern the natural habitat conditions in the field, but the presentation of the UMAP analysis results in a multidimensional space (38 dimensions). The overlap between groups g1 and g4 may reflect their narrow range of habitat conditions (although completely different), and thus fewer occurrences, which makes the groups more mathematically similar to each other. However, it is worth noting that on the horizontal axis, within the range 4 to 5.5, there is an apparent clustering of points representing group g1.

The separation of groups g1 to g4 is not perfect, but it is adequate to validate the

a: vegetation cover
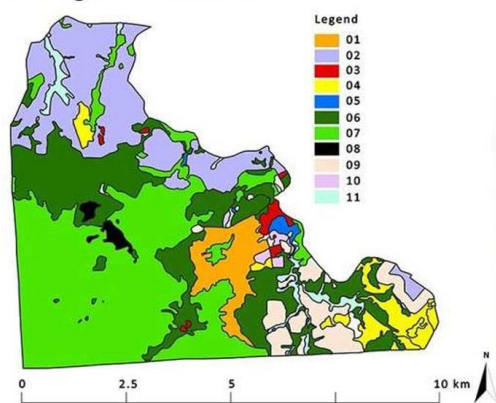
b: NDVI - vegetation cover
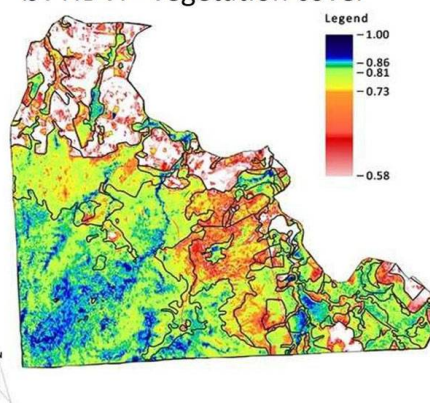
c: soils types
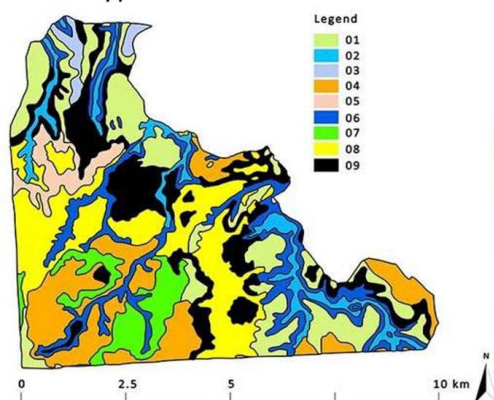
d: NDVI – soil types
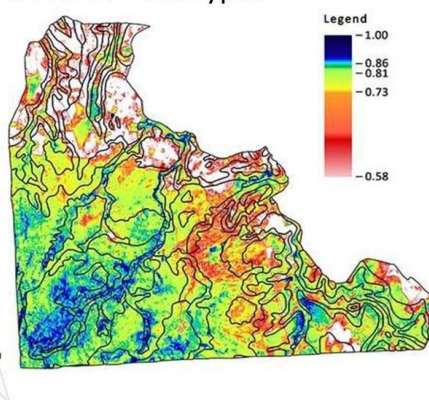
e: geology
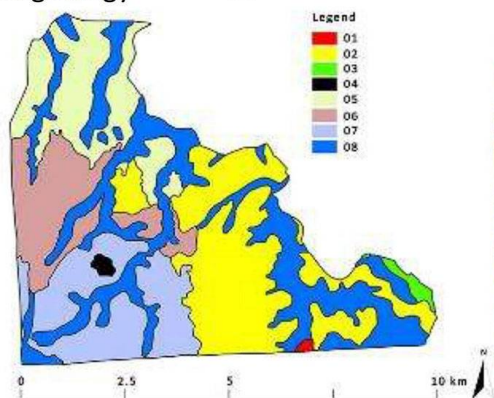
f: NDVI - geology
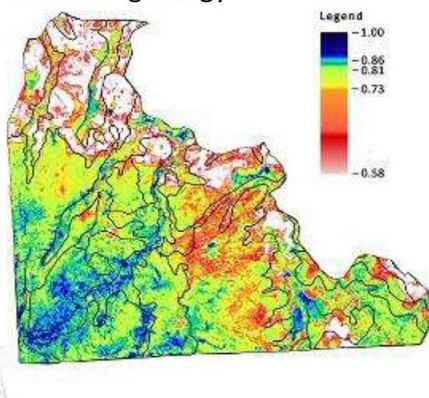
g: relief
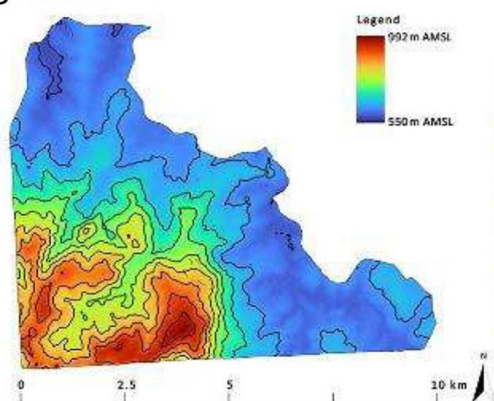
h: NDVI - relief
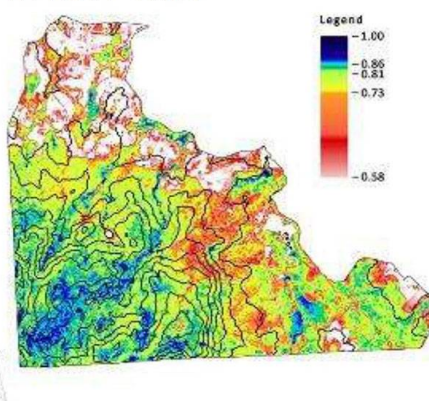
Fig. 2 - Comparison of differentiation of INF vegetation cover (2a), soil types (2c), geology (2e) and relief (2g) on the background of average (2018-2021) NDVI values in relation to the contours of vegetation (2b), soils (2d), geology (2f) and relief (2h). Legend to Figure 2a: 1 - Rocky Outcrop (Grasses/Shrubs) [Portuguese: *Afloramento Rochoso (Gramineas/Arbustivas)*]; 2 - Agriculture (Pasture/Cultivation) [*Agropecuaria (Pastagens/Cultivos)*]; 3 - Urban Area (Buildings/Towns) [*Área Urbana (Edificações/Vilas)*]; 4 - Cerrado (*Campo Sujo*); 5 -Water Bodies (Reservoirs) [*Corpo d'Água (Reservatórios)*]; 6 - Forest in Regeneration (*Floresta em Regeneração*); 7 - Seasonal Semideciduous Forest (*Floresta Estacional Semidecidual*); 8 - Mining (Cavages/Buildings) [*Mineração (Cavas/Edificações)*]; 9 - Reforestation (Exotic) [*Reflorestamento (Exóticas)*]; 10 - Reforestation (Native) [*Resflorestamento (Nativas)*]; 11 - Floodplain [*Várzea (Alagados/Brejos)*]. Legend to Figure 2c: 1 - Acrisols (In Portuguese *Argissolo Vermelho Amarelo Distrófico*); 2 - Gleysols (*Gleissolo*); 3 - Ferralsols (*Latossolo Amarelo*); 4 - Ferralsols (*Latossolo Vermelho Distrófico*); 5 - *Neossolo Litólico Afloramento* (no equivalent in IUSS/WRB 2015); 6 - Fluvisols (*Neossolo Flúvico*); 7 - *Neossolo Latólico Listico* (no equivalent in IUSS/WRB 2015); 8 - *Neossolo Litólico Distrófico* (IUSS/WRB 2015); 9 - *Neossolo Litólico Húmido* (IUSS/WRB 2015). Legend to Figure 2e: 1 - Amphibolite (Portuguese: *Anfibolito*); 2 - Conglomeratic Sandstone (*Arenito Conclomeratico*); 3 - Claystones (*Argilito*); 4 - Basic eruptive: basalt (*Eruptivas basicas basalto*); 5 - Stratified rock, verve (*Folhelho Varvito*); 6 - Porphyroid-granite (*Granito porfiroide*); 7 - Alkaline Intrusives: Shonkinite-porphyry (*Intrusivos Alcalinas Shonkinito-porfito*); 8 - Quaternary (*Quaternario*).

Fig. 3 - Covariance matrix. Green boxes indicate high parameter dependence; orange boxes indicate fields excluded from the correlation analysis; the individual blocks of the tested parameters were separated using sand, blue, and grey colours.
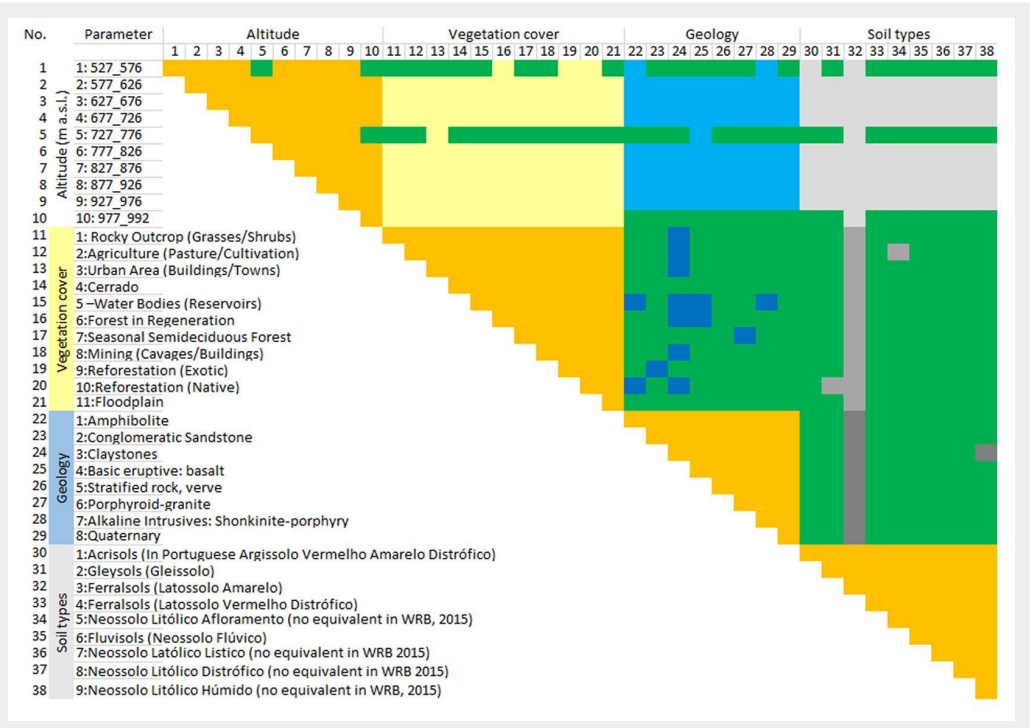
selection criteria used for distinguishing these moisture groups. Given the different ontologies of the 38 parameters, the universal Euclidean metric was employed to measure the distance between them. Additionally, a value of seven was assumed for the number of neighbors when estimating the similarity measure. This choice suggests a reasonable level of "local" calculations within the 38-dimensional space.

The accuracy and the error rate of the learning curves are presented in Fig. S6 (Supplementary material), where the averaged (smoothed) curves for the training and validation sets are reported, along with their variance (indicated by the lighter bands surrounding the curves). Based on the analysis of these graphs, we can draw the following key considerations. (i) The 510 samples used are adequate to achieve stable results, as indicated by the flattening curves. (ii) Two methods, LDA and SVM-poly, demonstrated abnormally low accuracy. The significant increase in SVM-poly's accuracy around 470 training cases, after prior stabilization, is likely a numerical anomaly. (iii) In all cases, except for SVM-poly, the variance of the training set decreases (improving or worsening) as its size increases. This suggests that the information quality of the training set is relatively good. (iv) A high variance in the validation set coupled with a low variance in the training set indicates that the model is overfitting, meaning it was too rigidly tailored to the training data. This is evident in all cases except SVM-RBF and QDA. In QDA, a training set larger than 200 cases increased the variance, with additional data adding only noise, and results stabilizing before reaching 200. Possible strategies to reduce overtraining include testing a wider range of standard regularisation parame-

ters and modifying the parameter set by removing some and/or adding new ones, which is more difficult both conceptually and in terms of cost. The analysis of covariance described above supports the reduction of highly correlated groups.

A relatively small variance was observed in the validation set across specific ranges of training set size for the random forest algorithm. A general upward trend in this variance (similar to QDA) suggests that information noise increases as the size of the training set grows.

The above results prompted us to test three algorithms using a reduced set of parameters, consisting of the two blocks height and vegetation. In the case of the QDA algorithm, this set proved numerically unstable. The learning curves for the other

two cases (SVM-RBF and random forest) are shown in Fig. S7 and Fig. S8 (Supplementary material). In both cases, minimal loss in accuracy was observed, along with a reduction in the validation set's variance, especially with random forests. This demonstrates that the SVM-RBF and random forests algorithms achieve fairly good classification of the defined moisture classes with the adopted set of parameters, thereby confirming the adequacy of the class g1-g4 definitions.

The significance of the altitudinal parameters group was further confirmed when a classifier was developed using only parameters from the vegetation, geological, and soil groups. This is illustrated by the learning curve for the random forest algorithm (Fig. S8), which reveals a marked decrease
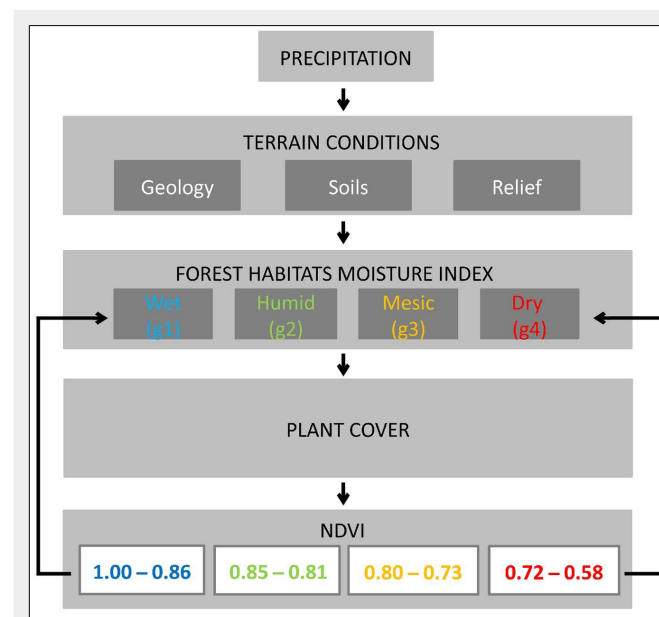
Fig. 4 - Relationships among individual parameters, NDVI values, and habitat moisture classes for INF.

in accuracy and a considerable increase in variance for both the training and validation sets.

Summary of results

NDVI value ranges were assigned to the different humidity levels in the study area based on the aforementioned analyses. When the average annual NDVI value was in the range 1.00-0.86, the habitat was considered wet (g1); when it was 0.85-0.81, the habitat was considered moist (g2); habitats with average annual NDVI within the range 0.80-0.73 were considered mesic (g3); while when average annual NDVI was below 0.73, the habitat was considered dry (g4).

The diagram of the relationships among individual parameters, NDVI values, and habitat moisture classes for the INF is shown in Fig. 4.

Discussion

Studies on NDVI have revealed complex relationships among this index, meteorological factors, soil moisture, and vegetation cover type (Chen et al. 2014, Piedallu et al. 2019, Giuseppi et al. 2021, Zhang et al. 2022). Our results from the Ipanema National Forest confirmed that a complex set of factors, including climate, relief, altitude, geology, and soil type, affects vegetation cover and its moisture status. According to Mazón et al. (2020), the effects of elevation on various forest characteristics remain poorly understood. Regardless of habitat conditions, forests are primarily found in the higher mountainous regions of the study area, likely due to the inaccessibility of these areas, which makes them unsuitable for agricultural development. At the same time, seasonal semideciduous forests (Fig. 2) had the highest average NDVI values from 2018 to 2021. This is due to the river valley's relief (Fig. 2g), where fluvisols (Fig. 2c) developed by collecting water from precipitation, thus supporting the link between terrain and soil. Gleysols, associated with high humidity (Sposito et al. 2008), are also found in this area. However, the average NDVI values for 2018-2021 in these soils are markedly lower (Fig. 2c). This may be due to different vegetation cover, which is mainly composed of plantations of alien species on gleysols. The most planted forest species in South America is eucalyptus (Freitas et al. 2020), which may change the natural water regime (Reichert et al. 2021). Eucalypt monocultures are also present in the INF (Londe et al. 2022).

Among the different types of plant cover, *cerrado campo sujo* is noted in the INF (Fig. 2a), which would confirm the range of this biome given by Pereira et al. 2018 (Fig. 1). At the same time, these are lower, more easily accessible mountain locations, therefore their transformation may be affected by human activities. The share of agricultural and urbanised areas in the lower part of the study area seems to confirm it (Fig. 2a). At the same time, it is a low-moisture

area (Fig. 2b) that corresponds to the characteristics of the Cerrado described by Ferreira & Huete (2004). These authors noted that the Cerrado region extends much farther north of the INF. Given the 14-year gap between the studies conducted by Ferreira & Huete (2004) and Pereira et al. (2018), this suggests that the increasing human pressure in the INF surrounding has forced forests into more remote areas.

The Cerrado biome, characterized by its canopy cover, animal grazing, land conversion into pasture and agriculture, and moisture conditions, is similar to the type of vegetation noted in the central part of Poland (European xero-thermophile oak woods – European Commission 2013), a region where the lowest precipitation and NDVI values were shown (Fig. S9 in Supplementary material). This is also the most deforested lowland part of the country, where human activity has led to the disappearance of natural hornbeam and oak forests. At the same time, it is unclear whether xero-thermophile oak woods thrive in the central part of the country because of the drier climate or if deforestation has caused the climate to become drier. To some extent, a similar phenomenon may occur at the border between the Cerrado and the Atlantic Forest, which is strongly influenced by human activity.

Our results confirm the relationship between NDVI and the moisture content of forest areas. Fig. S9 (Supplementary material) shows that the adopted method for assessing habitat moisture status is universal and can be applied worldwide.

Comparing the Brazilian Cerrado to the European xerothermophile oak woods involves examining how human activities impact forest management. This includes practices such as livestock grazing and the conversion of forested areas into agricultural land and pastures. Although a more comprehensive analysis of this comparison could be the focus of a separate study, this example illustrates that geographically distant forest ecosystems can be compared quickly and affordably. However, to effectively interpret the data, it is necessary to use a combination of satellite techniques and terrestrial data. Key factors to consider include vegetation cover diversity, soil diversity, and terrain variations. This approach was exemplified by the case study of the Ipanema National Forest.

Conclusions

The Normalized Difference Vegetation Index (NDVI) is widely used for estimating vegetation greenness and moisture content. However, our results proved that NDVI can also be used to build reliable models for assessing habitat moisture. Among the algorithms tested, UMAP was found to be the most effective. By using average NDVI values from 2018 to 2021, we classified the study area into four moisture classes: g1 (wet habitat – average NDVI = 1.00-0.86), g2 (moist – 0.85-0.81), g3 (mesic

– 0.80-0.73), and g4 (dry – average NDVI value < 0.73).

The method used in this study is easy to apply, cost-effective, and provides high-resolution (10×10 m) data applicable to large geographical areas. The results showed that the method can be applied across diverse environmental conditions, including different vegetation cover types, soil types, and geological sediments. It can be used to assess habitat moisture diversity, compare moisture levels across habitats, and monitor them over time. For example, our study identified the boundary between two critical biomes, the Atlantic Forest and Cerrado, in the study area and suggested increased anthrogenic pressure in the surroundings of the INF, forcing forests into less accessible areas.

Overall, this method provides a practical, efficient tool for assessing habitat moisture that can benefit research fields such as ecology, conservation biology, and land management.

Conflicts of Interest

The authors declare no conflict of interest.

References

Bai Y, Ochuodho TO, Jian Y (2019). Impact of land use and climate change on water-related ecosystem services in Kentucky, USA. Ecological Indicators 102: 51-64. - doi: 10.1016/j.ecolind.2019.01.079

Béllard C, Leclerc C, Leroy B, Bakkenes M, Veloz S, Thuiller W, Courchamp F (2014). Vulnerability of biodiversity hotspots to global change. Global Ecology and Biogeography 23: 1376-1386. - doi: 10.1111/geb.12228

Cavarzere V, Silveira LF, Tonetti VR, Develey P, Ubaid FK, Regalado LB, Figueiredo LFA (2017). Museum collections indicate bird defaunation in a biodiversity hotspot. Biota Neotropica 17 (4): 1-19. - doi: 10.1590/1676-0611-BN-2017-0404

Chen N, Wang A, An J, Zhang Y, Ji R, Jia Q, Zhao Z, Guan D (2020). Modeling canopy carbon and water fluxes using a multilayered model over a temperate meadow in Inner Mongolia. International Journal of Plant Production 14: 141-154. - doi: 10.1007/s42106-019-00074-4

Chen T, De Jeu RAM, LiunYY, Van Der Werf GR, Dolman AJ (2014). Using satellite based soil

moisture to quantify the water driven variability in NDVI: a case study over mainland Australia. Remote Sensing of Environment 140: 330-338. - doi: 10.1016/j.rse.2013.08.022

Da Silva RG, Santos AR, Pelúzio JBE, Fiedler NC, Juvanhol RS, Souza KB, Branco ERF (2021). Vegetation trends in a protected area of the Brazilian Atlantic forest. Ecological Engineering 162 (1): 106180. - doi: 10.1016/j.ecoleng.2021.106180

European Commission (2013). Interpretation manual of European Union Habitats, vers. EUR28. DG Environment, European Commission, Brussels, Belgium, pp. 144.

Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Alsdorf DE (2007). The shuttle radar topography mission. Reviews of Geophysics 45 (2): 33. - doi: 10.1029/2005RG000183

Ferreira LG, Huete AR (2004). Assessing the seasonal dynamics of the Brazilian Cerrado vegetation through the use of spectral vegetation indices. International Journal of Remote Sensing 25 (10): 1837-1860. - doi: 10.1080/0143116031000101530

Freitas CH, Elli EF, Sentelhas PC, Carneiro RL, Alvares CA (2020). Adaptation, calibration and evaluation of a simple agrometeorological model for wood eucalyptus productivity estimation. European Journal of Forest Research 139 (5): 759-776. - doi: 10.1007/s10342-020-01283-7

Freitas SR, Mello MCS, Cruz CBM (2005). Relationships between forest structure and vegetation indices in Atlantic Rainforest. Forest Ecology and Management 218 (1-3): 353-362. - doi: 10.1016/j.foreco.2005.08.036

Géron A (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc., Sebastopol, Ukraine, pp. 848.

Giuseppi A, Germanà R, Fiorini F, Delli Priscoli F, Pietrabissa A (2021). UAV patrolling for wildfire monitoring by a dynamic Voronoi tessellation on satellite data. Drones 5 (130): 1-19. - doi: 10.3390/drones5040130

IUSS/WRB (2015). World reference base for soil resources 2014, update 2015. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps, World Soil Resources Reports no. 106, FAO, Rome, Italy, pp. 192. [online] URL: http://www.fao.org/3/i3794en/I3794en.pdf

Kisselle KW, Zepp RG, Burke RA, Pinto AS, Bustamante MMC, Opsahl S, Varella RF, Viana LT (2002). Seasonal soil fluxes of carbon monoxide in burned and unburned Brazilian savannas. Journal of Geophysical Research: Atmospheres 107 (D20): 8051. - doi: 10.1029/2001JD000638

Klink CA, Machado RB (2005). Conservation of the Brazilian Cerrado. Conservation Biology 19 (3): 707-713. - doi: 10.1111/j.1523-1739.2005.00702.x

Lin M, Hou L, Qi Z, Wan L (2022). Impacts of climate change and human activities on vegetation NDVI in China's Mu Us sandy land during 2000-2019. Ecological Indicators 142: 109164. - doi: 10.1016/j.ecolind.2022.109164

Londe V, Ferreira FS, Martins FR (2022). Structure of a semideciduous seasonal forest in the national forest of Ipanema, Brazil: contributing to the floristic knowledge of a poorly studied protected area. Research in Ecology 4 (3): 10-19. - doi: 10.30564/re.v4i3.4916

Marques MCM, Trindade W, Bohn A, Grelle CEV (2021). The Atlantic Forest: an introduction to the megadiverse forest of South America. In: "The Atlantic Forest" (Marques MCM, Grelle CEV eds). Springer, Cham, Switzerland, pp. 3-23. - doi: 10.1007/978-3-030-55322-7_1

Mazón MM, Klanderud K, Finegan B, Veintimilla D, Bermeo D, Murrieta E, Delgado D, Sheila D (2020). How forest structure varies with elevation in old growth and secondary forest in Costa Rica. Forest Ecology and Management 469: 118191. - doi: 10.1016/j.foreco.2020.118191

Mróz W (2010). Monitoring siedlisk przyrodniczych. Przewodnik metodyczny [Monitoring of natural habitats. Methodological guide]. Part I. GIO Ed., Warszawa, Poland, pp. 321. [in Polish]. [online] URL: http://siedliska.gios.gov.pl/images/pliki_pdf/publikacje/Monitoring-siedlisk-przyrodniczych.-Przewodnik-metodyczny.-Cz-I.pdf

Pereira O, Ferreira L, Pinto F, Baumgarten L (2018). Assessing pasture degradation in the Brazilian Cerrado based on the analysis of MODIS NDVI time-series. Remote Sensing 10 (11): 1761. - doi: 10.3390/rs10111761

Piedallu C, Chéret V, Denux JP, Perez V, Azcona JS, Seynave I, Gégout JC (2019). Soil and climate differently impact NDVI patterns according to the season and the stand type. Science of The Total Environment 651: 2874-2885. - doi: 10.1016/j.scitotenv.2018.10.052

Pravalie R, Sîrodoev I, Nita IA, Patriche C, Dumitrascu M, Rosca B, Tiscovschi A, Bandoc G, Savulescu I, Manoiu V, Birsan MV (2022). NDVI-based ecological dynamics of forest vegetation and its relationship to climate change in Romania during 1987-2018. Ecological Indicators 136: 108629. - doi: 10.1016/j.ecolind.2022.108629

Reichert JM, Prevedello J, Gubiani PI, Vogelmann ES, Reinert DJ, Consensa COB, Soares JCW, Srinivasan R (2021). Eucalyptus tree stockings effect on water balance and use efficiency in subtropical sandy soil. Forest Ecology and Management 497: 119473. - doi: 10.1016/j.foreco.2021.119473

Rossi M (2017). Mapa pedológico do Estado de São Paulo: revisado e ampliado [Soil map of the State of São Paulo: revised and expanded]. São Paulo, Instituto Florestal, vol. 1, pp. 118. [in Portuguese]

Saint-Hilaire A (1976). Viagem à Província de São Paulo [Voyage to the province of São Paulo]. EDUSP, São Paulo, Brazil, pp. 312. [in Portuguese]

Sposito G, Chesworth W, Evans L, Chesworth W, Spaargaren O (2008). Gleysols. In: "Encyclopedia of Soil Science" (Chesworth W ed). Encyclopedia of Earth Sciences Series, Springer, Dordrecht, Netherlands, pp. 299-300. - doi: 10.1007/978-1-4020-3995-9_252

Strassburg BBN, Latawiec A, Balmford A (2016). Brazil: urgent action on Cerrado extinctions. Nature 540 (7632): 199. - doi: 10.1038/540199a

Strassburg BBN, Brooks T, Feltran-Barbieri R, Iri-barrem A, Crouzeilles R, Loyola R, Latawiec AE, Oliveira-Filho FJB, Scaramuzza CAM, Scarano FR, Soares-Filho B, Balmford A (2017). Moment of truth for the Cerrado hotspot. Nature Ecology and Evolution 1 (4): 99. - doi: 10.1038/s41559-017-0099.

Yadav SK, Borana SL (2019). Modis derived NDVI based time series analysis of vegetation in the Jodhpur area. In: Proceedings of the "ISPRS-GEOGLAM-ISRS Joint Int. Workshop on "Earth Observations for Agricultural Monitoring". New Delhi (India) 18-20 Feb 2019. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-3/W6, pp. 535-539. - doi: 10.5194/isprs-archives-XLII-3-W6-535-2019

Weaver D, Strong K, Walker KA, Sioris C, Schneider M, McElroy CT, Vömel H, Sommer M, Weigel K, Rozanov A, Burrows JP, Read WG, Fishbein E, Stiller G (2019). Comparison of ground-based and satellite measurements of water vapour vertical profiles over Ellesmere Island, Nunavut. Atmospheric Measurement Techniques 12: 4039-4063. - doi: 10.5194/amt-12-4039-2019

Willmersdorf O (2017). Plano de manejo da Floresta Nacional de Ipanema [Management plan for the Ipanema National Forest]. Instituto Chico Mendes – ICMBio, Revisão, Iperó, vol. 1, pp. 306. [in Portuguese] [online] URL: http://www.icmbio.gov.br/flonaipanema/quem-somos.html?id=19

Zhang X, Zhang B, Yao Y, Wang J, Yu F, Liu J, Li J (2022). Dynamics and climatic drivers of evergreen vegetation in the Qinling-Daba Mountains of China. Ecological Indicators 136: 108625. - doi: 10.1016/j.ecolind.2022.108625

Supplementary Material

Fig. S1 - Division of the study area into 1×1 km squares.

Fig. S2 - NDVI imagery of INF on the background of a graph of average 14-day precipitation in 2021.

Fig. S3 - Percentage distribution of the degree of explanation for all principal components.

Fig. S4 - PCA results for the first 10 principal components.

Fig. S5 - The result of data analysis using the UMAP method.

Fig. S6 - Learning curve graphs.

Fig. S7 - Learning curve plots for reduced versions of SVM-RBF and random forest.

Fig. S8 - Learning curve for the random forest algorithm excluding the height parameter group.

Fig. S9 - Range of xero-thermophile oak woods in Poland, according to Mróz (2010).

Link: Mlynarczyk_4413@suppl001.pdf