



MAPEAMENTO E CLASSIFICAÇÃO DA DENSIDADE DA COBERTURA VEGETAL DO SEMIÁRIDO BRASILEIRO: REGRESSÃO LINEAR APLICADA A PARTIR DO ÍNDICE MSAVI2 E ORTOFOTOS DO PE3D

Riclaudio Silva Santos

Doutorando do Programa de Pós-Graduação em Geografia - PPGeo, da Universidade Federal de Pernambuco - UFPE, Pernambuco, Brasil

riclaudio.silva@gmail.com

Lucas Costa de Souza Cavalcanti

Professor do Programa de Pós-Graduação em Geografia - PPGeo, da Universidade Federal de Pernambuco - UFPE, Pernambuco, Brasil

lucas.cavalcanti@ufpe.br

RESUMO – Este artigo traz uma aplicação do geoprocessamento aliado a estatística, utilizados em conjunto para o mapeamento e classificação da densidade da cobertura vegetal do semiárido brasileiro. Foram utilizadas informações do índice de vegetação MSAVI2 e ortofotos do Projeto Pernambuco Tridimensional (PE3D), aplicando análise bivariada e regressão linear para extrapolação dos dados espaciais. Por meio dessa metodologia foi possível estender as informações obtidas nas ortofotos da Serra Negra, entre os municípios de Floresta, Inajá e Tacaratu (Pernambuco), e assim classificar a densidade da vegetação do semiárido brasileiro. A análise estatística permitiu ainda validar os dados correlacionados e confirmar a classificação da densidade da cobertura vegetal proposta para a região semiárida em questão.

Palavras-chave: Geoprocessamento; Geoestatística; Estatística; Análise Bivariada.

MAPPING AND CLASSIFICATION OF VEGETATION COVER DENSITY OF THE BRAZILIAN SEMIARID: LINEAR REGRESSION APPLIED FROM THE MSAVI2 INDEX AND PE3D ORTHOPHOTOS

ABSTRACT – This article brings an application of geoprocessing combined with statistics, used together for the mapping and classification of vegetation cover density in the Brazilian semiarid. Information from the MSAVI2 vegetation index and orthophotos from the Three-Dimensional Pernambuco Project (PE3D) were used, applying bivariate analysis and linear regression for extrapolation of spatial data. Through this methodology it was possible to extend the information obtained through the orthophotos of Serra Negra, between the municipalities of Floresta, Inajá and Tacaratu (Pernambuco), and thus classify the density of vegetation in the Brazilian semi-arid region. The statistical analysis also allowed validating the correlated data and confirming the classification of the density of the vegetation cover proposed for the semi-arid region in question.

Keywords: Geoprocessing; Geostatistics; Statistics; Bivariate Analysis.

INTRODUÇÃO

As representações cartográficas desempenham um papel fundamental como ferramenta de relevância para a Geografia na análise do espaço. Os mapas permitem ao pesquisador espacializar um conjunto abrangente de informações, abordando uma diversidade de temas, sejam eles de cunho

natural, referentes a uma área específica (tais como vegetação, geomorfologia, geologia, distribuição pluviométrica, pedologia, entre outros), ou mesmo relacionados às características da sociedade (como densidade demográfica, uso do solo e outras atividades de origem antrópica).

Com o avanço tecnológico e a aplicação de satélites, o Sensoriamento Remoto e o Geoprocessamento, empregados por meio dos Sistemas de Informações Geográficas (SIG), consolidaram-se como importantes aliados dos geógrafos. As imagens adquiridas por meio do sensoriamento remoto proporcionam uma visão panorâmica multitemporal de vastas áreas da superfície terrestre. Essa perspectiva sinóptica do ambiente ou da paisagem viabiliza estudos regionais, abrangendo diversas áreas do conhecimento (FLORENZANO, 2005).

E aliado ao geoprocessamento, os métodos estatísticos contribuem de maneira significativa para a análise e interpretação dos dados espaciais. A geoestatística é utilizada, por exemplo, na construção de modelos preditivos e de regressão que podem ser aplicados em previsões espaciais. Contando ainda com ferramentas para a validação dos dados e confirmação das conclusões obtidas (CÂMARA e MEDEIROS, 1998).

Alicerçado nesses pressupostos, o estudo aqui apresentado tem por objetivo classificar a densidade da vegetação do semiárido brasileiro, por meio da geoestatística, utilizando dados de uma localidade, e extrapolando os mesmos, a partir de regressão linear, para toda a região do semiárido.

A área escolhida como referência para aplicação da metodologia de classificação da cobertura vegetal encontra-se entre os municípios de Floresta, Inajá e Tacaratu, no semiárido pernambucano. A escolha dessa área se deve a disponibilidade de ortofotos, e de diferentes contextos de cobertura vegetal, que vão desde áreas de solo exposto assim como cobertura vegetal densa, com destaque para a Serra Negra.

Para a análise da área de referência foram utilizadas ortofotos do Projeto Pernambuco Tridimensional (PE3D), devido sua alta resolução, permitindo observar a densidade da cobertura vegetal. Foram utilizadas também imagens do satélite Landsat 8, para geração do índice MSAVI2, utilizado na classificação da densidade da cobertura vegetal para todo o semiárido brasileiro.

É importante ressaltar o contexto geomorfológico/geológico do recorte geográfico escolhido, bacia sedimentar com capeamento predominantemente siliciclástico, e solos derivados, não necessariamente representativo de todo o semiárido. Entretanto, a captação da clorofila realizada pelo MSAVI2 pode ser estendida a toda vegetação do semiárido brasileiro.

MATERIAIS E MÉTODOS

A metodologia utilizada tem por objetivo relacionar o percentual da densidade de cobertura da vegetação da área de referência com os valores do Índice MSAVI2 (Índice de Vegetação Ajustado ao Solo Modificado 2), buscando identificar os intervalos do índice citado para cada contexto de cobertura da vegetação (Muito Esparsa, Esparsa, Aberta e Densa). Aplicando a análise bivariada entre esses dois dados (densidade da vegetação observada na área de referência e o valor do MSAVI2), e após isso, por meio da regressão linear, extrapolar os dados para todo o semiárido brasileiro.

Para a classificação da densidade da vegetação foram utilizadas ortofotos adquiridas a partir do PE3D (<http://www.pe3d.pe.gov.br/>) e imagens do satélite Landsat 8, ambas referentes ao ano de 2015. A utilização das ortofotos do PE3D se deve a sua alta resolução que permite a identificação visual da cobertura vegetação, e sua classificação em porcentagem de área vegetada. Assim, optou-se por classificar a vegetação em Muito Esparsa (menor ou igual a 10% de cobertura vegetal), Esparsa (acima de 10% a 30% de cobertura vegetal), Aberta (acima de 30% a 70% de cobertura vegetal), e Densa (acima de 70% de cobertura vegetal), seguindo referência já aplicada na literatura (SCARTH et al., 2019).

A escolha do índice MSAVI2 se deve ao fato de que o mesmo já tem apresentado bons resultados para pesquisas de ambientes semiáridos (CAVALCANTI e SANTOS, 2021), e diferentemente do MSAVI tradicional que faz uso da variável L, estimada com base na quantidade de vegetação existente, o mesmo não necessita desta variável, tornando o resultado do índice mais independente, eliminando a necessidade de encontrar a linha do solo a partir de um gráfico de espaço de feições ou mesmo especificar explicitamente o fator de correção de brilho do solo (QI et al., 1994).

Quanto as técnicas estatísticas utilizadas na pesquisa, destaca-se a análise bivariada e a regressão linear. A análise bivariada teve por objetivo verificar a correlação entre as duas variáveis estudadas (densidade da vegetação observada na área de referência e o valor do MSAVI2). Essa técnica inclui a análise de coeficientes de correlação, gráficos de dispersão, e testes de validação dos dados.

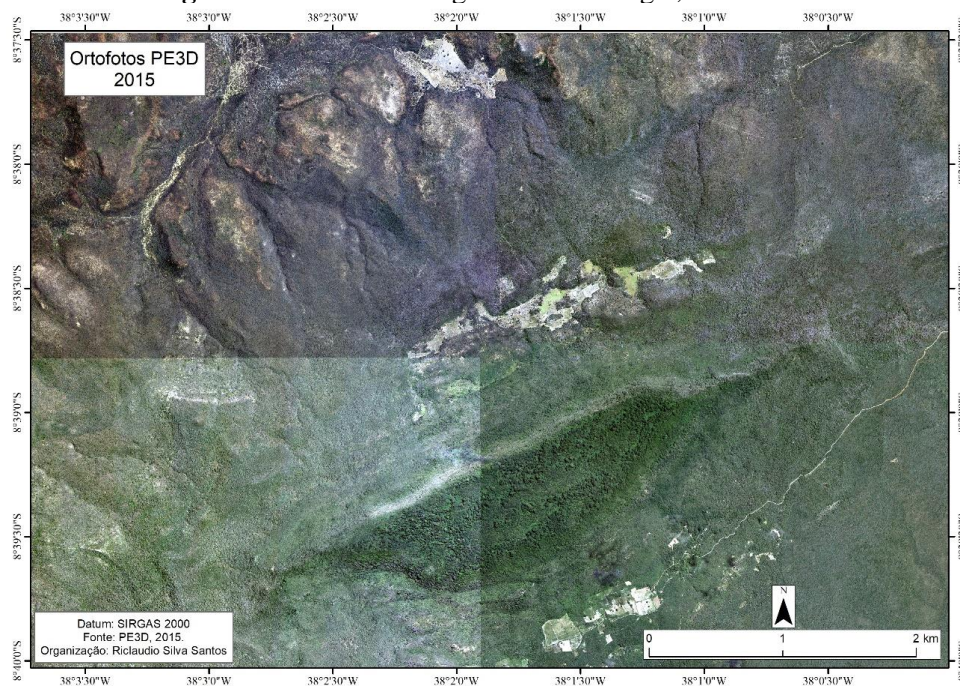
A regressão linear, por sua vez, é uma técnica estatística que se propõe modelar a relação entre uma variável dependente (também chamada de variável resposta, neste caso a porcentagem de cobertura da vegetação) e a variável independente (também chamadas de variáveis preditoras, neste caso o valor do MSAVI2). O objetivo é encontrar uma equação linear que descreva essa relação, permitindo estimar o valor da variável dependente (porcentagem de cobertura da vegetação) a partir do valor da variável independente (valor do MSAVI2).

O software utilizado para a análise dos dados foi o Past (versão 4.08). O mesmo é um software livre para análise de dados científicos, com funções para manipulação de dados, plotagem, estatística univariada e multivariada, análise ecológica, séries temporais, análise espacial, morfometria e estratigrafia. Distribuído gratuitamente pela Universidade de Oslo (<https://www.nhm.uio.no/english/research/resources/past/>).

RESULTADOS E DISCUSSÃO

Primeiramente foram selecionadas as ortofotos do PE3D da área de referência (Figura 1), onde será possível obter o valor (em porcentagem) da densidade da cobertura vegetal e posteriormente o valor do MSAVI2 correspondente.

Figura 1. Ortofotos da Região da Serra Negra, PE3D 2015.



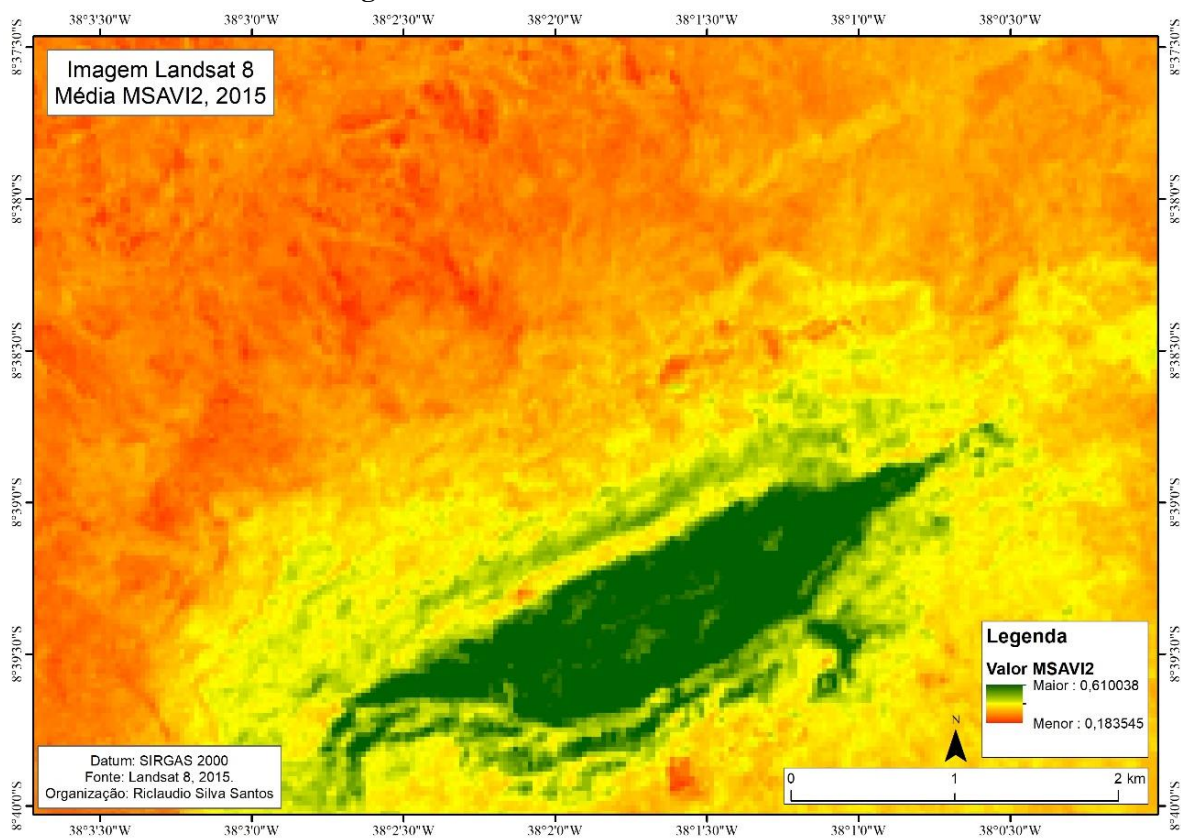
Org. Autores (2023).

Posteriormente foi realizado o cálculo do MSAVI2 a partir de imagem Landsat 8 (Figura 2), referente ao ano 2015, mesmo ano de realização do imageamento do PE3D. O Índice de Vegetação Ajustado ao Solo Modificado 2 (MSAVI2) foi calculado utilizando a fórmula a seguir:

$$MSAVI2 = \frac{(2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - RED)})}{2}$$

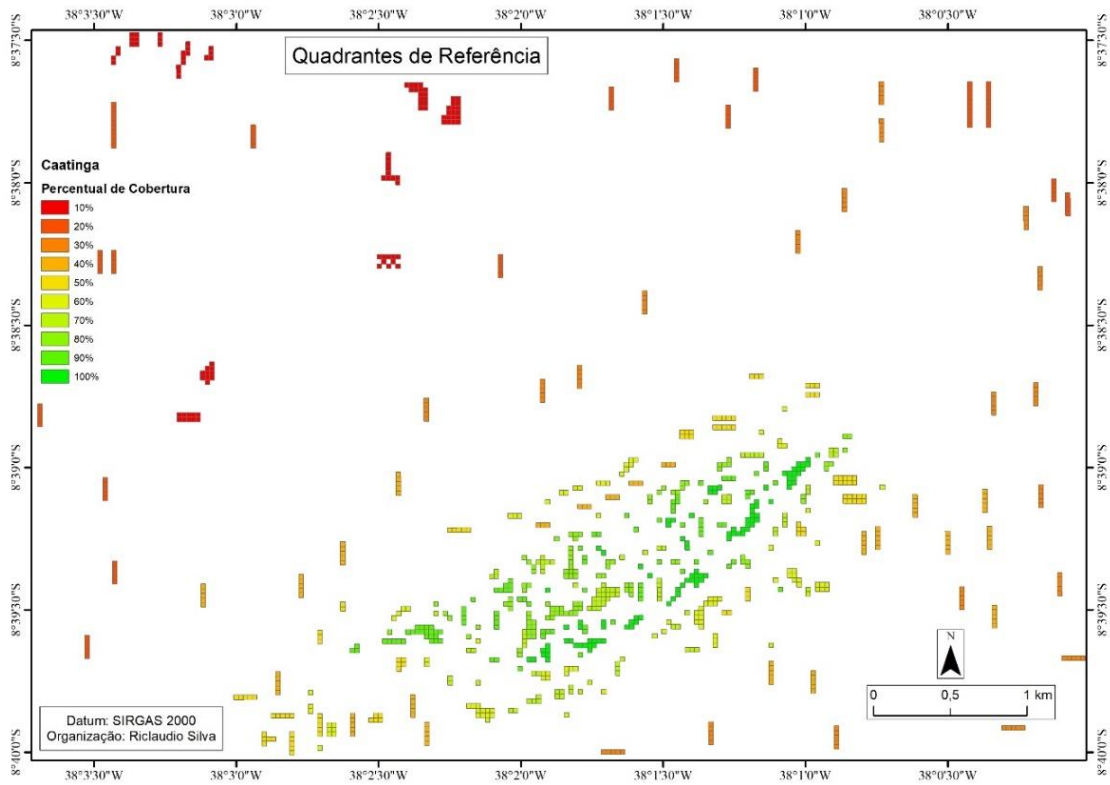
Onde NIR é a banda do infravermelho próximo (banda 5 do Landsat 8) e RED é a banda do vermelho (banda 4 do Landsat 8). Os dados do MSAVI2 foram obtidos na plataforma do Google Earth Engine (<https://code.earthengine.google.com/>).

Figura 2. MSAVI2 da área de referência.



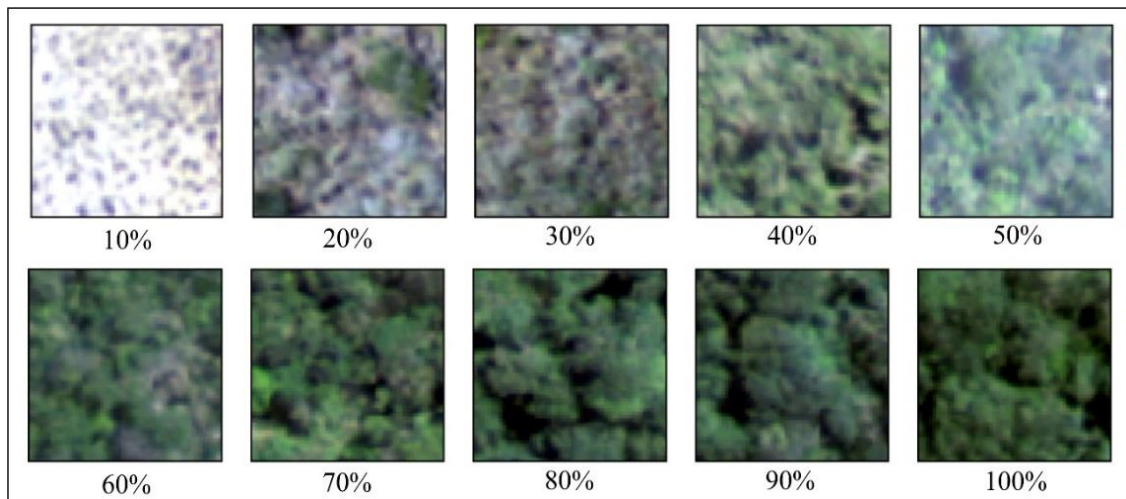
Tendo as informações da densidade da vegetação a partir do PE3D, e dos valores de MSAVI2 a partir da imagem Landsat 8, foi criada uma malha com 1000 (mil) quadrantes para extrair os respectivos valores das duas variáveis que serão utilizadas posteriormente na análise bivariada e na regressão linear (Figura 3). Para adquirir a porcentagem da cobertura da vegetação foram utilizados quadrantes com dimensão de 30m x 30m (correspondente a resolução do pixel das imagens do Satélite Landsat 8). Esses quadrantes foram sobrepostos às ortofotos do PE3D, os mesmos foram então divididos em 10 grupos contendo 100 quadrantes cada (correspondendo a 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% e 100% de cobertura vegetal), como observado na Figura 4.

Figura 3. Quadrantes de Referência.



Org. Autores (2023).

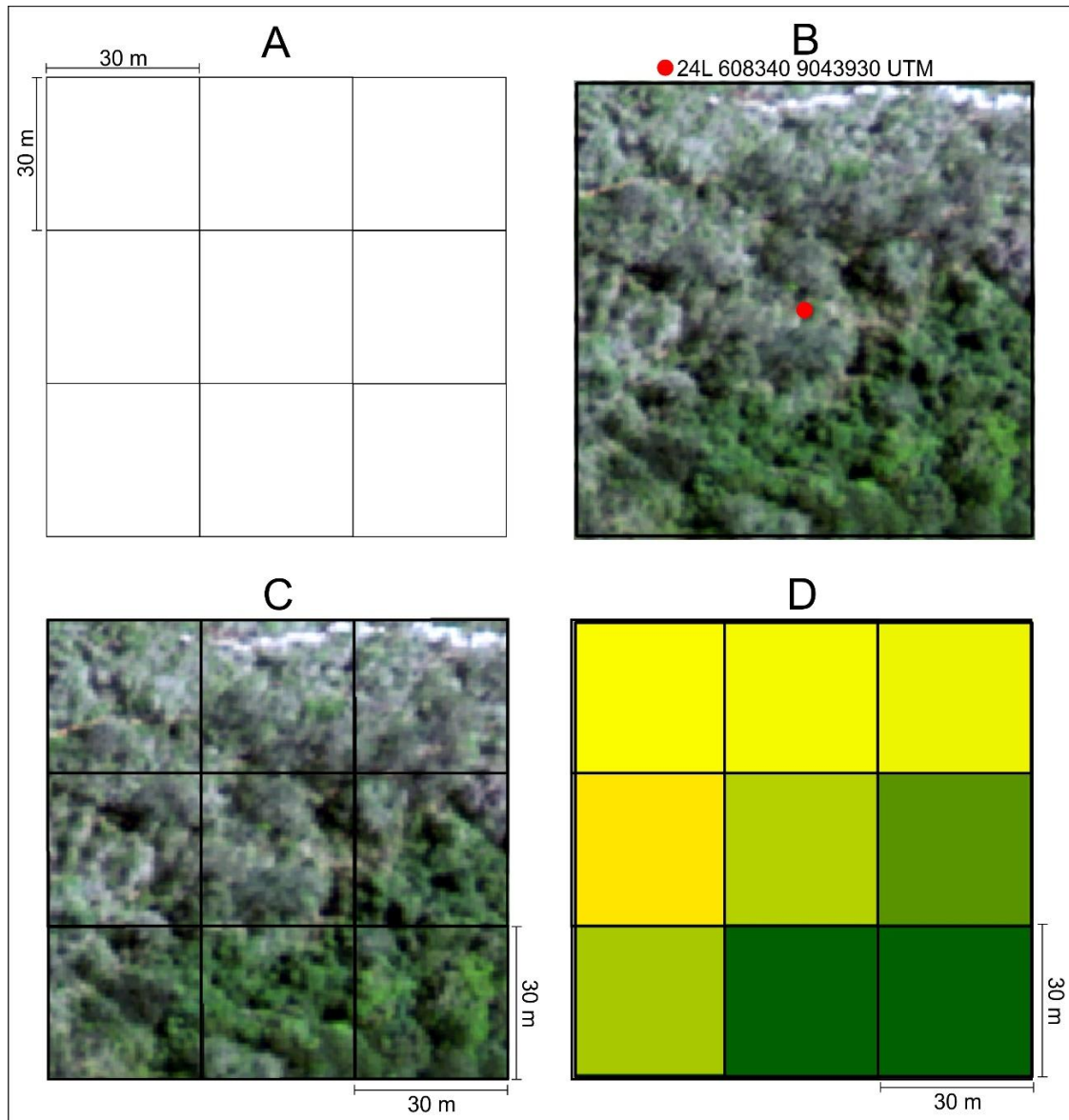
Figura 4. Exemplos das classes dos índices de cobertura vegetal, a partir de ortofotos do PE3D.



Org. Autores (2023).

Para um melhor entendimento, o modelo explicativo da pesquisa pode ser observado na Figura 5, observa-se a sobreposição dos quadrantes de referência sobre a imagem do PE3D e do Landsat 8, onde foram adquiridos os dados de percentual de cobertura e valor do índice MSAVI2, respectivamente.

Figura 5. A- Malha de quadrantes com dimensão de 30 m por 30 m; B- Ortofoto PE3D, 2015, com localização geográfica; C- Ortofoto do PE3D com malha de quadrantes sobreposta; D- Resultado do MSAVI2 (Média anual de 2015, Imagens Landsat 8, com malha de quadrantes sobreposta.



Org. Autores (2023).

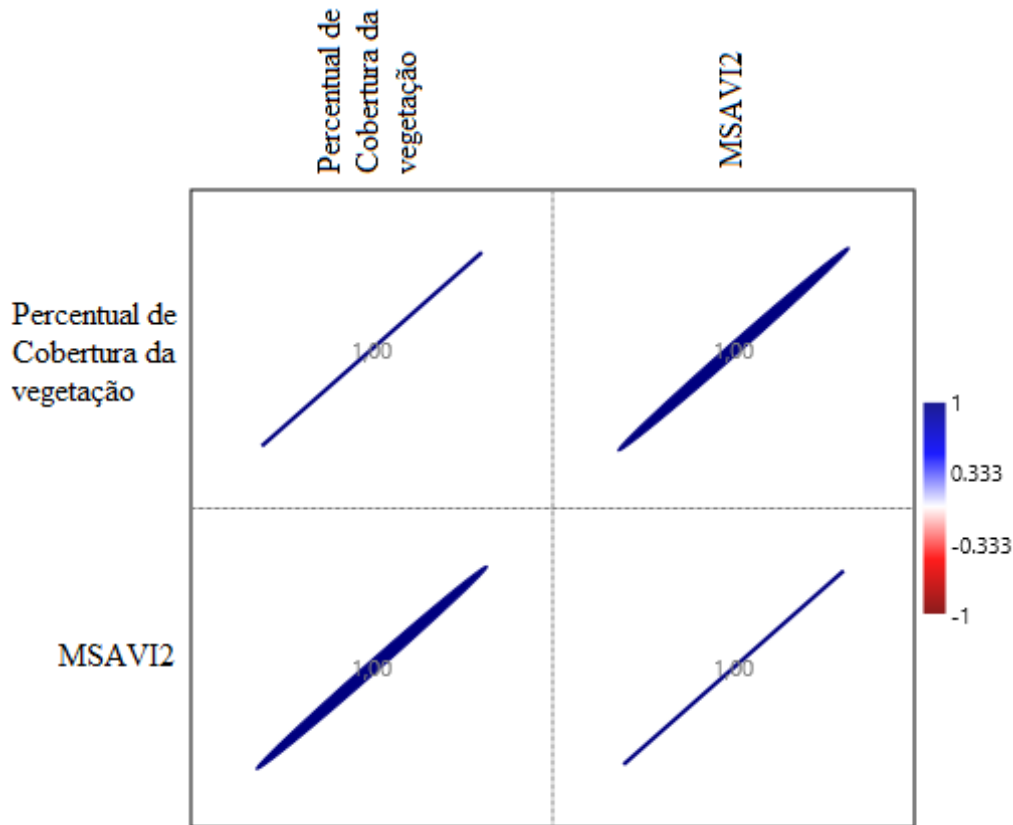
Em posse das informações de porcentagem de cobertura vegetal e do valor do MSAVI2 foi gerado então uma tabela contendo a base amostral para as técnicas estatísticas. Sendo necessário, primeiramente, validar a correlação dos dados que serão utilizados na regressão linear, o que será demonstrado a seguir.

Validação dos dados para construção do modelo de regressão linear

Após a aquisição dos dados de porcentagem de cobertura vegetal e do MSAVI2, se fez necessário realizar as análises estatísticas para confirmar a validação dos dados e construção do modelo de classificação da vegetação a partir dos intervalos do MSAVI2 por meio da regressão linear.

O primeiro passo foi verificar a presença de correlação amostral entre os dados em questão (Correlação de Pearson), para que se pudesse afirmar que existe uma associação entre os dados de percentual de cobertura e os dados do MSAVI2, assim, foi calculado o coeficiente de correlação por meio da ferramenta Univariate/Correlation do software Past (Figura 6).

Figura 6. Resultado da correlação entre os dados de Percentual de Cobertura da vegetação e dos



valores de MSAVI2, por meio do software Past 4.08.

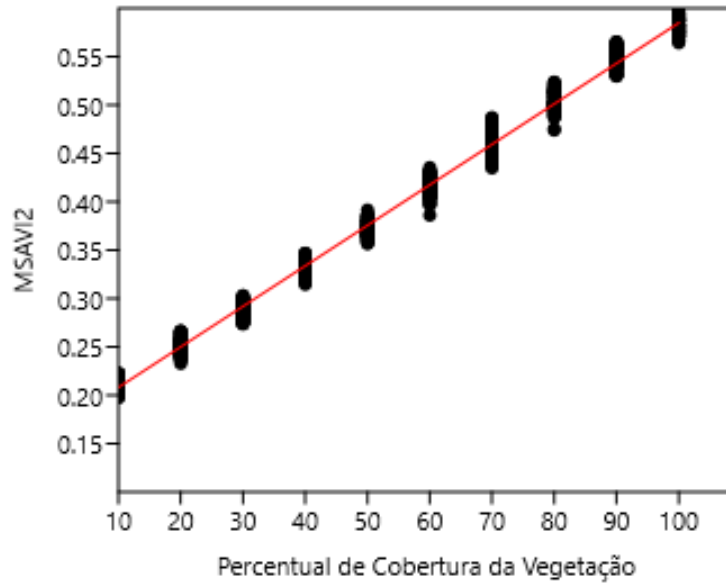
Org. Autores (2023).

Como observado na figura anterior, os dados apresentaram correlação de 1, ou seja, uma correlação perfeita positiva entre as duas variáveis. Entretanto, para se validar a correlação de Pearson para a criação do modelo de regressão linear se faz necessário validar se os dados obtidos podem realmente ser analisados utilizando a correlação de Pearson. E isto se dá pela verificação de quatro pressupostos dos dados (COHEN, 1988; HOFFMANN, 2016).

A primeira suposição demanda que as variáveis devem ser medidas no intervalo ou nível de razão (ou seja, é necessário que elas sejam contínuas), fator que se encaixa nos dados de percentual de cobertura da vegetação e do MSAVI2.

A segunda suposição diz respeito a necessidade de verificar se há uma relação linear entre as duas variáveis, e isto pode ser observado no gráfico de dispersão apresentado a seguir na Figura 7.

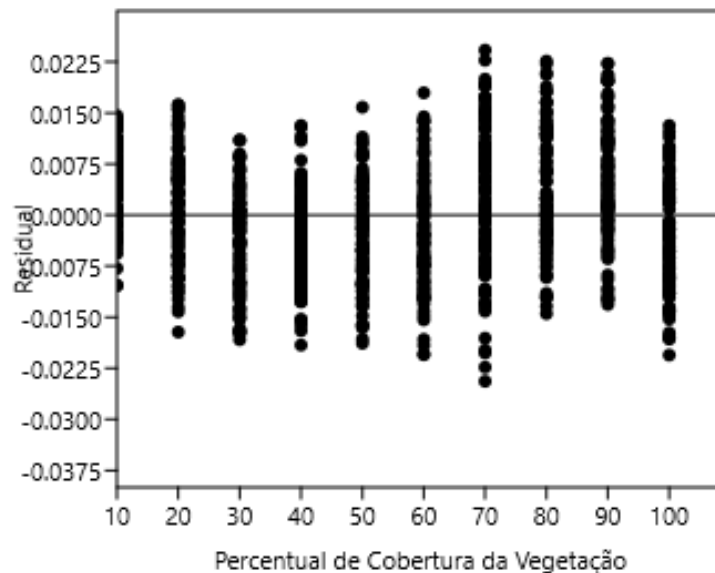
Figura 7. Gráfico de dispersão dos dados, a partir do software Past 4.08.



Org. Autores (2023).

A terceira suposição para validação dos dados diz respeito a ausência de valores discrepantes significativos (Outliers). Os Outliers são pontos de dados únicos dentro dos dados que não seguem o padrão usual, e que por vezes, a depender do tamanho do banco amostral, podem interferir no resultado final por possuírem valores discrepantes. Entretanto, ao observar a Figura 8, observa-se que o gráfico de residuais não apresentou Outliers significativos. E o banco amostral contendo 1000 pontos de dados protege os dados da influência de possíveis outliers.

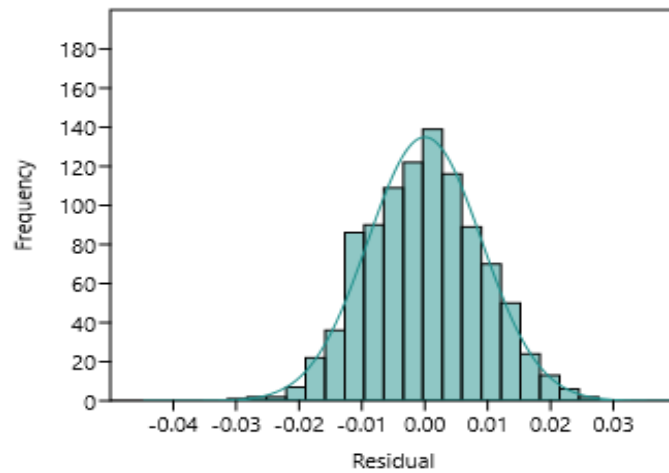
Figura 8. Gráfico de residuais, a partir do software Past 4.08.



Org. Autores (2023).

Por fim, a quarta suposição diz respeito a natureza da distribuição dos dados, é necessário que os dados tenham distribuição aproximadamente normal. Para isto é necessário testar a normalidade de cada variável por meio do teste de normalidade Shapiro-Wilk. Ao calcular o Shapiro-Wilk por meio do Past foram obtidos os valores de 0,9352 para os dados de percentual de cobertura da vegetação, e 0,9411 para os dados de MSAVI2. Para que uma distribuição seja considerada como normal é necessário que o resultado de Shapiro-Wilk seja maior que 0,05. Sendo assim, os resultados de Shapiro-Wilk confirmam que os dados possuem uma distribuição normal e que, portanto, podem ser usados coeficientes como Pearson e modelos de regressão linear. Os dados podem ainda ser observado por meio de um histograma, verificando sua distribuição normal (Figura 9).

Figura 9. Histograma com distribuição normal dos dados utilizados, a partir do software Past 4.08.



Org. Autores (2023).

Após a validação dos dados para a criação do modelo, os mesmos foram processados visando encontrar os intervalos de MSAVI2 correspondente a cobertura da vegetação. Na figura 10 pode-se observar os resultados obtidos por meio do software Past, com o uso da ferramenta Model/Linear/Bivariate.

Figura 10. Resultado da regressão linear com os dados de Percentual de Cobertura da Vegetação e MSAVI2, a partir do software Past 4.08.

| Ordinary Least Squares Regression: Percentual de Cobertura da Vegetação-MSAVI2 | | | |
|--|------------------------|-----------------------|------------|
| Slope α : | 0,0041857 | Std. error α : | 1,0039E-05 |
| t : | 416,96 | p (slope): | 0 |
| Intercept β : | 0,16633 | Std. error β : | 0,0006202 |
| 95% bootstrapped confidence intervals ($N=1999$): | | | |
| Slope α : | (0,0041658, 0,0042054) | | |
| Intercept β : | (0,16527, 0,16741) | | |
| Correlation: | | | |
| r : | 0,99718 | | |
| r^2 : | 0,99437 | | |
| t : | 416,96 | | |
| p (uncorr.): | 0 | | |
| Permutation p : | 0,0001 | | |

Org. Autores (2023).

Onde $r = 0,99718$ sugere uma relação linear forte e positiva. $r^2 = 0,99437$ indica que o modelo com 1000 quadrantes conseguiu um ajuste excelente. E $t = 416,96$ indica que a relação entre o Percentual de Cobertura da Vegetação e MSAVI2 não é obra do acaso, mas uma relação real.

O $p = 2,57$ no teste de Breusch-Pagan indica que não há heterocedasticidade nos dados (valor acima de 0,05). O erro padrão da estimativa é de 0,0090558. Dividindo o erro padrão da estimativa pela média do MSAVI2 (0,39527) temos o percentual de erro do modelo, que é de 0,0229. O Slope = 0,0041857 indica que para cada 1% de percentual de cobertura da vegetação, o MSAVI2 aumenta 0,0041857. O Intercept = 0,16633 sugere que, quando o MSAVI2 for 0,16633 o Percentual de Cobertura da Vegetação será 0%.

Com isso, o limite para uma cobertura muito esparsa (<10% de percentual de cobertura) é igual a $0,16633 + 10 * 0,0041857$. O valor para a cobertura esparsa (10% a 30% de percentual de cobertura) é igual a $0,16633 + 30 * 0,0041857$. E o valor para a cobertura aberta (30% a 70% de percentual de cobertura) é igual a $0,16633 + 70 * 0,0041857$. Assim, temos:

Tabela 1. Relação entre o Percentual de Cobertura Vegetal, o Tipo de Cobertura Vegetal, e o Valor de MSAVI2.

| Percentual de Cobertura Vegetal | Tipo de Cobertura Vegetal | Valor MSAVI2 |
|---------------------------------|---------------------------|---------------------|
| < = 10% | Muito Esparsa | < = 0,208187 |
| 10% a 30% | Esparsa | 0,208187 a 0,291901 |
| 30% a 70% | Aberta | 0,291901 a 0,459329 |
| > 70% | Densa | > 0,459329 |

Fonte: Autores (2023).

A classe de vegetação Muito Esparsa corresponde a um valor de MSAVI2 menor ou igual a 0,208187, a classe de vegetação Esparsa corresponde a um valor de MSAVI2 de 0,208187 a 0,291901, a classe de vegetação Aberta corresponde a um valor de MSAVI2 de 0,291901 a 0,459329, e a classe de vegetação Densa corresponde a um valor de MSAVI2 maior que 0,459329.

Ainda de acordo com o Intercept = 0,16633, que sugere 0% de cobertura vegetal, pode-se identificar ainda os Corpos Hídricos a partir desse valor do MSAVI2.

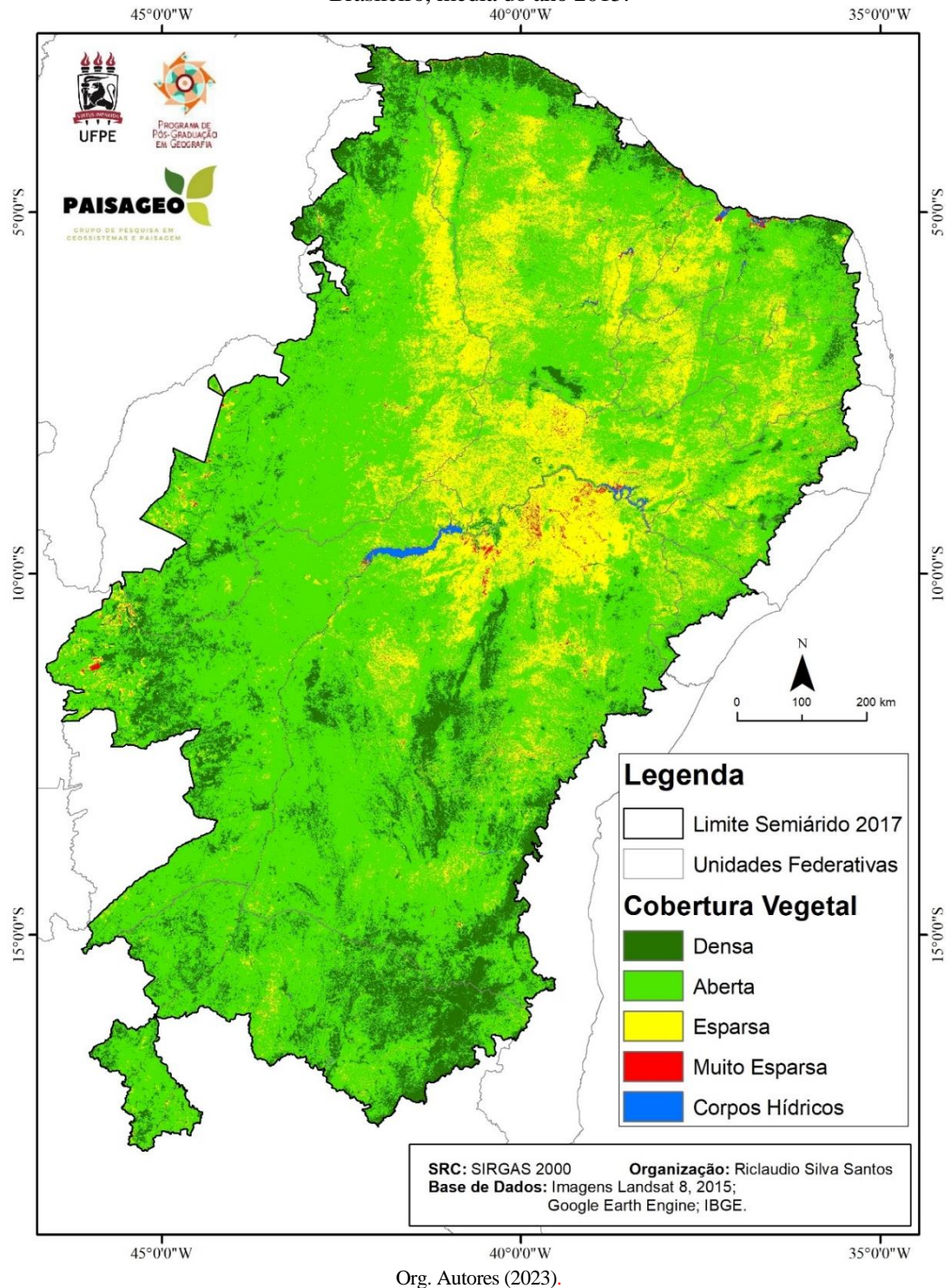
Classificação da densidade da vegetação do Semiárido Brasileiro

Após a aplicação do modelo de regressão linear foi possível classificar o histograma do MSAVI2 e extrapolar as informações sobre a densidade da vegetação para todo o semiárido brasileiro.

Na Figura 11 é possível observar o mapa da vegetação do semiárido brasileiro, com as classes da densidade da vegetação, a partir da média do MSAVI2, referente a média do ano 2015.

A presente classificação mostrou-se concordante com o esperado para a região em questão. As áreas de exceção, com maior densidade de vegetação, como a porção nordeste do Araripe e serras menores foram demarcadas corretamente a partir da presente classificação. A área com concentração de vegetação Muito Esparsa corresponde a região do Núcleo de Desertificação de Cabrobó. E os Corpos Hídricos também foram identificados, com destaque para o Rio São Francisco.

Figura 11. Mapa do Índice de Vegetação Ajustado ao Solo Modificado 2 (MSAVI2), Semiárido Brasileiro, média do ano 2015.



CONSIDERAÇÕES FINAIS

O presente texto buscou demonstrar como a estatística pode ser utilizada para auxiliar no mapeamento e no estudo da paisagem, de forma a agilizar a pesquisa e fornecer dados inéditos que podem ser transformados em informação. A aplicação da análise bivariada e da

regressão linear permitiu gerar informações para todo o semiárido brasileiro a partir de dados de uma área de referência menor, utilizando dados desta última para extrapolar as informações para a região semiárida maior.

Além da validação estatística dos dados para construção da regressão linear, a mesma também se mostrou concordante com o esperado para a região semiárida estudada, a partir da classificação dos valores do MSAVI2, que mapeou com sucesso os diferentes contextos paisagísticos da área de estudo, dentro da sua proposta e limitações.

A utilização de um banco de dados amostral contendo 1000 (mil) quadrantes foi de fundamental importância para gerar um modelo de regressão linear com alta correlação e baixo erro padrão e conseqüente baixo percentual de erro do modelo.

Cabe ressaltar que os valores de MSAVI2 estipulados neste trabalho para classificar a vegetação se referem à vegetação da caatinga, com imageamento do satélite Landsat 8. Para outros tipos de vegetação, assim como para imagens provenientes de outros satélites, ou mesmo outros índices de vegetação, os valores para classificação da vegetação serão diferentes.

A metodologia aqui aplicada, utilizando do geoprocessamento e de análises estatísticas, pode ser reproduzido para outras pesquisas, analisando diferentes variáveis que sejam de utilidade para o pesquisador, desde que sejam respeitados os pressupostos dos dados. A primeira suposição demanda que as variáveis devem ser medidas no intervalo ou nível de razão (ou seja, é necessário que elas sejam contínuas). A segunda suposição diz respeito a necessidade de verificar se há uma relação linear entre as duas variáveis. A terceira suposição para validação dos dados diz respeito a ausência de valores discrepantes significativos (Outliers). E a quarta suposição diz respeito a natureza da distribuição dos dados, é necessário que os dados tenham distribuição aproximadamente normal.

Desta forma, cabe ao pesquisador, a luz dos objetivos da sua pesquisa, utilizar dessas (e outras) técnicas da estatística, para criar novas informações a partir dos dados que lhe são de interesse. Ferramentas valiosas para o estudo das variáveis geográficas, auxiliando nos mapeamentos de correlações, criação de modelos preditivos, análise regional, conservação ambiental, e outros campos da Geografia.

REFERÊNCIAS

- CÂMARA, G.; MEDEIROS, J. S. de. GEOPROCESSAMENTO PARA PROJETOS AMBIENTAIS. 2ª Edição - Revisada e Ampliada. São José Dos Campos, SP. 1998.
- CAVALCANTI, L. C. de S.; SANTOS, R. S. Domínio das Caatingas: considerações a partir de uma cartografia comparada. In: SOUZA, S. O.; LIMA, K. C.; VALEZIO, E. V.; SAMPAIO, S. A. (org.). Perspectivas e desafios do sertão nordestino. – 1. ed. Vol. 1 - Senhor do Bonfim-Ba: Univasf, 2021.
- COHEN, J. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers. 1988.
- FLORENZANO, Teresa Galloti. Geotecnologias na Geografia Aplicada: Difusão e Acesso. Revista do Departamento de Geografia, 17, p.24-29, 2005.
- HOFFMANN, R. Análise de regressão: uma introdução à econometria [recurso eletrônico]. 5. ed. Piracicaba, 2016.
- QI, J.; KERR, Y.; CHEHBOUNI, A. External Factor Consideration in Vegetation Index Development. In Proc. Of Physical Measurements and Signatures in Remote Sensing. ISPRS, 723-730, 1994.
- SCARTH, P.; ARMSTON, J.; LUCAS, R.; BUNTING, P.; A Structural Classification of Australian Vegetation Using ICESat/GLAS, ALOS PALSAR, and Landsat Sensor Data. Remote Sensing. 11, 147. 2019.