

CLASSIFICAÇÃO DE SISTEMAS DE INTEGRAÇÃO LAVOURA-PECUÁRIA (ILP) NO SUL DO MARANHÃO: UMA ABORDAGEM COM MACHINE LEARNING E SENSORIAMENTO REMOTO

Ana Beatriz Freitas da Silva¹, Vinicius Morales da Silva², Hilton Luís Ferraz da Silveira³, Lauro Rodrigues Nogueira Júnior⁴

^{1,2,3,4}Embrapa Territorial, Avenida Soldado Passarinho n° 303 – Jardim Chapadão – Campinas, São Paulo, ^{1,2}Instituto de Geociências, Departamento de Geografia - Unicamp - Rua Carlos Gomes, 250 - Cidade Universitária - Campinas, São Paulo
^{1,2}{anabf.silva, vinicius.morales}@colaborador.embrapa.br, ³lauro.nogueira@embrapa.br e ⁴hilton.ferraz@embrapa.br.

RESUMO

Utilizou-se o algoritmo *Random Forest* (RF) para testar a capacidade do classificador na identificação de sistemas integrados Lavoura-Pecuária (ILP). O estudo foi realizado na Fazenda Nova Balsa e no seu entorno, no Município de Tasso Fragoso (MA). Foi organizado um banco de dados (BD) com 518 pontos, sendo 259 de ILP e 259 áreas sem integração. Para cada ponto, foram coletados valores de reflectância nas bandas do vermelho, azul, infravermelho próximo e médio (MOD13Q1), NDVI e EVI (SATVeg) para os períodos de janeiro, fevereiro e maio-junho. O BD, foi utilizado para treinamento e validação do RF na classificação de ILP. Como resultado obteve-se AUC (1,0), acurácia (0,995), F1-Score (0,995), precisão (0,995), Recall (0,995) e MCC (0,989). Concluiu-se que, apesar da boa acurácia, o grupo amostral limitado gerou viés, mas a banda do infravermelho próximo de janeiro foi a mais assertiva na classificação de ILPs.

Palavras-chave – *random forest*; MODIS; Plano ABC.

ABSTRACT

The *Random Forest* (RF) algorithm was used to test the classifier's ability to identify Crop-Livestock Integration (CLI) systems. The study was conducted on the Nova Balsa farm and its surroundings, in the municipality of Tasso Fragoso (MA). A database (DB) was organized with 518 points, comprising 259 CLI points and 259 non-integration areas. For each point, reflectance values were collected for red, blue, near-infrared, and mid-infrared bands (MOD13Q1), as well as NDVI and EVI indices (SATVeg) for the periods of January, February, and May-June. The DB was used for training and validating the RF classifier for CLI identification. The results yielded AUC (1.0), accuracy (0.995), F1-score (0.995), precision (0.995), recall (0.995), and MCC (0.989). It was concluded that, despite high accuracy, the limited sample group introduced bias; however, the near-infrared band in January was the most effective for CLI classification.

Key-words – *random forest*; MODIS; LCA: Low-Carbon Agriculture Plan.

1. INTRODUÇÃO

A partir dos anos 2010, há a chegada de novas políticas ambientais no Brasil, como o caso do Plano ABC - Agricultura de Baixo Carbono para a agricultura [1]. Programas como esse visam mitigar o impacto das mudanças climáticas e controlar a emissão de gases de efeito estufa pela adoção de práticas e manejos que possam emitir menos ou sequestrar carbono. Nesse sentido, os sistemas de produção integrados, como a integração lavoura-pecuária-floresta (ILPF), surgem como uma alternativa [2].

A ILPF é um sistema de produção que visa agrupar diferentes métodos numa mesma área, sendo eles lavoura, pecuários e florestais, resultando em diferentes combinações, através de rotação, sucessão ou por cultivo consorciado, por exemplo: agropastoril (ILP), silvipastoril (IPF), silviagrícola (ILF) e agrossilvipastoril (ILPF) [2]. A integração de sistemas produtivos busca otimizar o uso da terra, combinando aumento da produtividade com a conservação de recursos naturais, resultando em benefícios ambientais, sociais e econômicos, como bem-estar animal, mitigação de gases de efeito estufa, qualificação profissional e redução de custos [3,4].

Dentro do monitoramento do uso do solo, o mapeamento dos alvos agropecuários envolve uma elevada complexidade, em função dos tipos de culturas, dos calendários e dos diversos manejos utilizados. Nos sistemas integrados esta complexidade é ainda maior visto que há uma diversidade de práticas agropecuárias, o que acaba por dificultar o monitoramento através do sensoriamento remoto [5].

A região do sul do Maranhão se apresenta como uma área de franca expansão dos plantios de soja, milho e algodão que ocorriam nos platôs, durante a safra principal e sem a participação da irrigação. Até recentemente, a prática pastoril estava limitada às pequenas propriedades localizadas nas partes baixas do relevo. Entretanto, a ILP, onde o gado se alimenta da cultura de cobertura no período que antecede o plantio da safra principal, tem se tornado mais comum, mostrando uma mudança no comportamento nas grandes fazendas da região.

Neste sentido, o seguinte estudo focou-se na compreensão e análise do uso da integração lavoura-pecuária (ILP) na região sul do Maranhão, a fim de testar o uso de *machine learning* e *data mining*, e atingir o objetivo de conseguir a separabilidade das classes de ILP e não-ILP.

2. MATERIAL E MÉTODOS

2.1 Área de Estudo

A região de estudo (Figura 1) concentra-se ao sul do Maranhão, na cabeceira do rio Marcelino, integrante da bacia hidrográfica do rio Parnaíba, mas possui enfoque na Fazenda Nova Balsa no município de Tasso Fragoso, com 3331 hectares, e talhões que variam de 28,22 hectares a 157,61 hectares, sendo a maioria superior a 100 hectares. Além da área em questão, talhões de fazendas ao redor foram usados para o estudo comparativo. O bioma característico é Cerrado, e o clima da região pode ser classificado como Aw segundo a definição climática de Köppen [6], do tipo Tropical, com baixa variação de temperatura ao longo do ano, tendência de chuvas no verão e períodos de secas no inverno. É uma região cuja agricultura se baseia em soja, algodão e milho.

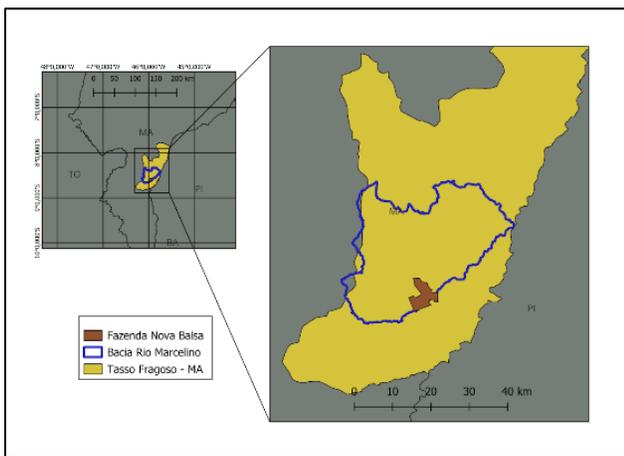


Figura 1. Mapa da região de estudo

2.2 Banco de dados de agricultura

Para organizar o banco de dados a ser utilizado neste trabalho, os pontos foram divididos em dois tipos: com e sem a presença de ILP.

Devido ao campo realizado previamente (de 08/03 a 16/03/2024), sabe-se que a Fazenda Nova Balsa realiza a integração de sistemas Lavoura-Pecuária (ILP) em todos os seus talhões, aonde foram então coletados pontos representativos de áreas de ILP. Em fazendas vizinhas onde não havia integração (agricultura convencional), foram coletados outros pontos. No total, 518 pontos foram coletados, sendo 259 de ILP e 259 de agricultura convencional.

A fim de entender o comportamento espectro-temporal dos ILPs e dos cultivos convencionais, foram selecionados aleatoriamente inicialmente 8 pontos do banco de dados coletados, sendo 4 pontos de ILP na área de estudo e 4 em talhões de agricultura convencional para serem trabalhados no Sistema de Análise Temporal de Vegetação (SATVeg). Este último “é uma ferramenta Web desenvolvida pela Embrapa Agricultura Digital, destinada ao acesso e visualização de perfis temporais dos índices vegetativos NDVI e EVI do sensor MODIS” [7]. Assim, observou-se os valores de NDVI no período de um ano (safra de 2022-2023) para os pontos, e as

datas que apresentavam maior disparidade entre as duas classes foram selecionadas para o estudo.

Em virtude do tamanho dos talhões foi escolhido o produto MOD13Q1 do satélite MODIS, com resolução espacial de 250 metros e resolução temporal de 16 dias. Por meio da plataforma *Google Earth Engine* (GEE) foram extraídas as bandas correspondentes ao vermelho, azul, infravermelho próximo e médio, além dos índices NDVI e EVI.

Os valores espectrais e dos índices de vegetação adquiridos dos 518 pontos foram organizados em uma tabela de treinamento e validação, com suas respectivas classificações acerca do sistema usado - se eram ILP ou não.

2.3 Random Forest

Para testar a separabilidade das duas classes de manejo, foi escolhido o algoritmo *Random Forest* (RF) de aprendizado de máquina, disponível no *software* livre de mineração e análise de dados *Orange*. Este se baseia num fluxograma através de interface visual para programação avançada, sem a necessidade de codificação, portanto, somente é necessário colocar os dados de entrada e selecionar o algoritmo de interesse.

O banco de dados com os pontos foi separado em dois grupos de maneira aleatória. O primeiro conjunto, contendo 333 pontos, foi submetido como treinamento ao classificador definindo-se a floresta com 200 árvores de decisões e máximo de até 5 ramificações [8]. Os restantes dos pontos (185) foram utilizados para validar a classificação e foram apurados utilizando os índices de acurácia e precisão: AUC - Área sob Curva ROC; Acurácia; *F1-Score*; Precisão; *Recall*; e Coeficiente de Correlação de Matthews.

Utilizou-se ainda o índice *Gini Decrease* para elencar quais bandas, índices de vegetação e meses apresentaram melhor desempenho na separação das classes.

3. RESULTADOS E DISCUSSÃO

3.1 Gráfico de valores do SATVeg

Ao colocar os valores dos 8 pontos selecionados em um gráfico de análise (Gráfico 1) no período de 16/10/2022 a 16/10/2023, é possível observar o comportamento espectral durante a safra e perceber as diferenças entre os pontos de ILP e não-ILP. As datas que apresentaram maior discrepância espectral (sinalizadas no Gráfico 1) foram 01/01/2023, 18/02/2023 e 09/05/2023 até 26/06/2023.

Na agricultura tradicional, sendo representada pelos pontos de não-ILP, percebe-se três diferentes ciclos, que são apontados pelas datas escolhidas. Esse comportamento espectral diz respeito à rotação de culturas feita pelos produtores, variando entre o cultivo de milho e soja, a fim de realizar uma manutenção da saúde do solo, evitando a disseminação de doenças e pragas, e manter a fertilidade do solo. A semeadura de soja, por exemplo, cultura essa que representa a maioria na região, tende a ser semeada a partir da segunda quinzena de outubro ou, no mais tardar, no mês de novembro, pois trata-se do período em que se inicia as chuvas regularmente [9].

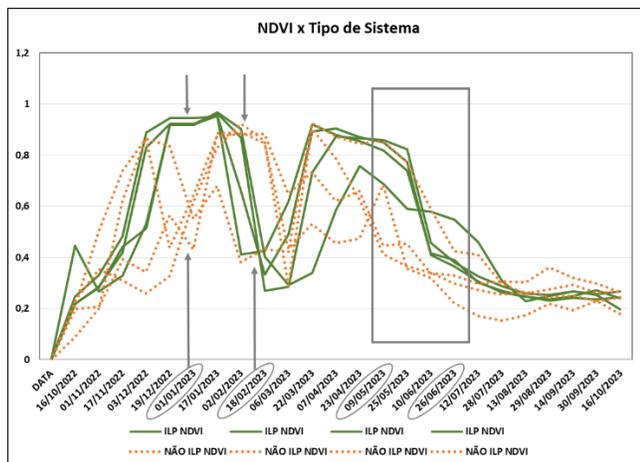


Gráfico 1. NDVI da safra 2022/2023 para pontos de ILPs e não-ILPs (agricultura convencional).

Já para os pontos de ILP, percebe-se dois grandes ciclos, sendo o primeiro iniciado em novembro e com término em fevereiro, e o segundo inicia-se em março e com fim em junho. Entende-se esse efeito como característico dos sistemas integrados Lavoura-Pecuária, uma vez que há o consórcio de culturas com forrageiras, tendo então um primeiro plantio de espécies como milho e sorgo, seguidas de braquiária, para que se obtenha o uso da forrageira para produção de palha para o plantio direto ou para a alimentação animal [3].

3.2 Classificação Random Forest

A aplicação do RF nos dados de validação nos valores de Área sob Curva ROC, Acurácia, F1-Score, Precisão, Recall e Coeficiente de Correlação de Matthews são mostrados na Tabela 1 e na Matriz de Confusão (Tabela 2). Apesar de terem valores que indicam um bom desempenho, é importante notificar que se trata de um grupo amostral reduzido e específico, então as chances de acerto foram enviesadas.

O Gráfico 2 mostra *Gini Decrease* dos 5 atributos mais bem ranqueados no processo de classificação do RF. Pode-se entender através dos resultados que a performance das bandas e índices de vegetação escolhidos possuem valor variado ao longo do período de um ano. Essa discriminação de valores é compreendida pela dinâmica fenológica-espectral, ou seja, pelas mudanças de reflectância e absorvância das fitofisionomias ao longo do ciclo fenológico.

| AUC | CA | F1 | Prec | Recall | MCC |
|-------|-------|-------|-------|--------|-------|
| 1.000 | 0.995 | 0.995 | 0.995 | 0.995 | 0.989 |

Tabela 1. Resultados dos atributos: AUC - Área sob Curva ROC; Acurácia; F1-Score; Precisão; Recall; e Coeficiente de Correlação de Matthews.

| | NÃO-ILP | ILP | Σ |
|----------|---------|-----|----------|
| NÃO-ILP | 83 | 1 | 84 |
| ILP | 0 | 101 | 101 |
| Σ | 83 | 102 | 185 |

Tabela 2. Matriz de confusão para resultados do *Random Forest*.

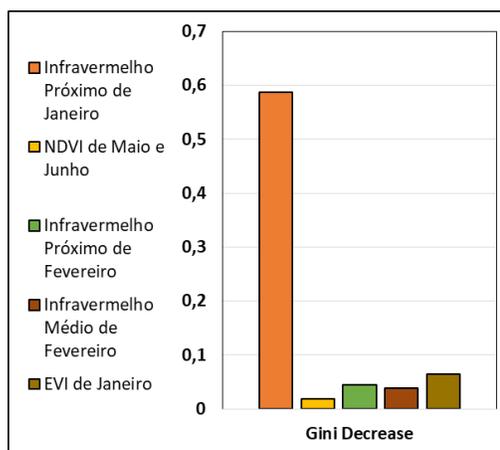


Gráfico 2. *Gini Decrease* dos atributos com mais acerto na diferenciação de sistemas integrados.

A banda do infravermelho próximo (*Near-infrared spectroscopy* - NIR) se mostrou relevante, especialmente nos meses de janeiro e fevereiro. No NIR, as folhas verdes saudáveis apresentam altos valores de reflectância, alta transmitância e baixa absorvância [10] e se deve ao espalhamento da radiação no interior das folhas em função da estrutura celular [11]. Na região de estudo, durante o período de janeiro e o período de fevereiro há discrepância entre os valores, pois no primeiro momento, há o ápice da vegetação para ILPs e colheita para não-ILPs, e o inverso acontece para o segundo mês. É nesse mesmo cenário que aparece o resultado do infravermelho médio, pois a reflectância foliar é inversamente relacionada com a quantidade total de água presente nas folhas [10]. Assim, há reflectância e espalhamento da radiação com base na umidade da planta, o que acaba por representar as folhas no seu período inicial de crescimento onde estão mais saudáveis e hidratadas.

Nos meses de maio e junho, o NDVI destacou-se devido à sua capacidade de representar a variação sazonal e fenológica da vegetação [11], uma vez que ele é eficaz para monitorar a densidade da cobertura vegetal, o vigor da vegetação e as mudanças na fotossíntese, que ocorrem conforme o período de crescimento e a senescência. Em maio e junho, o NDVI auxilia na identificação de áreas onde a vegetação atinge seu pico de desenvolvimento, ou começa a apresentar sinais de senescência das folhas, onde a clorofila, responsável pela pigmentação verde, praticamente desaparece, tornando-se então predominante os pigmentos amarelos (carotenos e xantofilas) [10]. Esse comportamento espectral é perceptível no Gráfico 1, onde as linhas de agricultura tradicional, não-ILP, apresentam queda no decorrer dos meses, devido à senescência da vegetação, enquanto os pontos de produção integrada mantêm-se estável, reiterando a presença de um segundo ciclo.

Por fim, em janeiro também se destacou a métrica de EVI. O Índice de Vegetação Melhorado (*Enhanced Vegetation Index* - EVI) trata-se de um melhoramento, a fim de reduzir a interferência do solo e da atmosfera, o que acaba por melhorar a sensibilidade de sua detecção em regiões com maiores densidades de biomassa e otimiza o sinal de vegetação [11]. Nesse sentido, o EVI foi capaz de capturar pequenas variações de cobertura verde inicial, mas não o suficiente para superar ou apresentar o mesmo desenvolvimento atingido pelo NDVI no período de janeiro.

4. CONCLUSÃO

A pesquisa destacou a relevância do monitoramento da integração lavoura-pecuária (ILP), pois há a necessidade do conhecimento da variabilidade espectral, temporal e espacial do uso e da cobertura da terra, a fim de entender as mudanças nos sistemas produtivos e ambientais [12]. Portanto, o estudo reitera que o uso de metodologias como *machine learning* e *data mining* associadas ao sensoriamento remoto são ferramentas possibilitadoras de monitoramento. Contudo, percebe-se que há um viés nos resultados devido a quantidade de amostras reduzida, mas isso não impossibilitou a avaliação de possíveis valores de atributos para identificação de padrões e diferenciações de uso de sistemas integrados. Assim, tem-se como resultado a potencialidade do uso de dados espectrais do infravermelho próximo no mês de janeiro para identificar ILPs na região de estudo.

5. REFERÊNCIAS

[1] BRASIL. Decreto n.º 7.390, de 9 de dezembro de 2010. *Regulamenta os arts. 6º, 11 e 12 da Lei n.º 12.187, de 29 de dezembro de 2009, que institui a Política Nacional sobre Mudança do Clima - PNMC*. Portal da Legislação, Brasília, DF, 2010. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/decreto/d7390.htm. Acesso em: 4 out. 2024.

[2] SILVA, I. M. da; ARAKAKI, K. K. *Carbono florestal em sistemas de integração lavoura-pecuária-floresta*. Revista de Política Agrícola, Brasília, Ano XXI, n. 4, p. 111-123, out./nov./dez. 2012.

Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/84997/1/Carbono-florestal-em-sistemas-de-integracao-lavoura-pecuaria-floresta.pdf>. Acesso em: 4 out. 2024.

[3] BALBINO, L. C.; CORTEZ, J. W.; MORAES, A. **Integração lavoura-pecuária-floresta na região Sul: análise econômica e de risco**. Brasília: Embrapa, 2011. 64 p. (Documentos / Embrapa Solos, 155). Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/82226/1/0000005512-ILPF-REGIAO-SUL.pdf> >. Acesso em: 01 set. 2024.

[4] REDE ILPF. Portfólio de Projetos da Rede ILPF. Disponível em: <https://redeilpf.org.br/portifolio-de-projetos-da-rede-ilpf> Acesso em: 29 set. 2024.

[5] KUCHLER P. *et al.* **Monitoring complex integrated crop – livestock systems at regional scale in Brazil**. *Remote Sensing*, 14:1648, 2022.

[6] CORRÊA, Wellington; CARVALHO, Marcus Willame Lopes; MENDES, Telmo José. **Atualização da classificação climática e balanço hídrico climatológico no estado do Maranhão**. Revista Brasileira de Climatologia, [S.L.], v. 32, p. 517-543, 25 abr. 2023. ABCLima (Brazilian Association of Climatology). <http://dx.doi.org/10.55761/abclima.v32i19.16727>

[7] EMBRAPA AGRICULTURA DIGITAL. “Sistema de Análise Temporal da Vegetação - SATVeg”. Campinas, 2015. Disponível em: <https://www.satveg.cnptia.embrapa.br/satveg/login.html> > Acesso em: 30 set. 2024.

[8] RODRIGUEZ-GALIANO, V. F.; GHIMIRE, B.; ROGAN, J.; CHICA-OLMO, M.; RIGOL-SANCHEZ, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 67, p. 93-104, 2012. DOI: 10.1016/j.isprsjprs.2011.11.002.

[9] FILHO, E. P. **A cultura da soja no sul do Maranhão**. Londrina: EMBRAPA-CNPSo, 1995. 34 p. (Documentos, 84). Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/445817/1/Documentos84.pdf> >. Acesso em: 09 out. 2024.

[10] FORMAGGIO, Antonio Roberto; SANCHES, Ieda Del'Arco. **Sensoriamento Remoto em Agricultura**. São Paulo: Editora Oficina de Textos, 2017.

[11] PONZONI, Flávio Jorge; SHIMABUKURO, Yosio Edemir. **Sensoriamento Remoto no Estudo da Vegetação**. 2010. ed. São José dos Campos: Parêntese, 2009.

[12] ANDRADE, Ricardo Guimarães et al. Geotecnologias aplicadas em sistemas de produção integrada e apoio a políticas públicas. In: BRASÍLIA - DF. Davi José Bungenstab. Empresa Brasileira de Pesquisa Agropecuária. **ILPF: inovação com integração de lavoura, pecuária e floresta**. Brasília: Embrapa, 2019. p. 263-279.