



FRAMEWORK DE IDENTIFICAÇÃO DE ZONAS FOCAIS PARA REGIÕES CENTRO-SUL NO CONTEXTO DO PROGRAMA NACIONAL DE BIOCOMBUSTÍVEIS

Alvaro Luz Alves Coutinho

PPGI-EA (ESALQ/CENA), Universidade de São Paulo, Piracicaba, São Paulo, Brasil

alvaro.coutinho@usp.br

Resumo: O artigo explora indicadores socioambientais dos municípios da região canavieira Centro-Sul em um contexto anterior a implementação Programa Nacional de Biocombustíveis (RenovaBio). O objetivo central é identificar Zonas Focais de Políticas Públicas na região, associando a performance das usinas certificadas pelo Renovabio a fatores demográficos, econômicos e ambientais municipais. Por meio de técnicas de *machine learning* não supervisionado, como K-means e LISA, o estudo classifica municípios em três perfis distintos — Eficiente, Padrão e Inadequado — e avalia a correlação entre suas características produtivas, socioeconômicas e emissões de GEE por meio de classificação de Zonas Focais de Política Pública com finalidade de direcionar estudos futuros para averiguar se a implementação da política gera impactos positivos ou negativos nas variáveis utilizadas. O modelo aplicado demonstrou significativo potencial para identificação de padrões geográficos, porém ainda carece de refinamentos para identificação de áreas que divergem do Perfil Padrão.

Palavras-chave: Renovabio, Machine Learning, Políticas Públicas, Biocombustíveis

1. Introdução

A descarbonização das economias até 2030 é um dos maiores desafios globais, impulsionada por tratados como o Protocolo de Kyoto e o Acordo de Paris, que visam mitigar os efeitos das mudanças climáticas e promover uma economia de baixo carbono. No Brasil, o Programa Nacional de Biocombustíveis (RenovaBio), instituído pela Lei nº 13.576/2017, desempenha um papel fundamental na redução das emissões de gases de efeito estufa (GEE), incentivando a produção e o consumo de biocombustíveis de base sustentável. Um dos procedimentos padrões para regulação das emissões de GEE adotadas a partir do Artigo 17 do Protocolo de Kyoto (1997) foram as soluções orientadas pelo mercado, ferramentas de medição chamadas “*cap-and-trade*”. Essa metodologia cria um tipo de *commodity* atrelado aos níveis e estabelecimento de metas de Intensidade de Carbono em um dado setor da economia (BENTO; FILHO, 2023). De forma geral, os mercados de carbono traduzem uma unidade de crédito a uma tonelada de CO₂ equivalente que foi removido da atmosfera ou que gerou redução de emissão (gerados por projetos da atividade econômica) (UNFCCC, 2024). O mercado de Créditos de Descarbonização (CBIOS), produto principal do Renovabio, é central nesse processo, promovendo descarbonização no setor de combustíveis por meio da financeirização desse ativo ambiental (LÁZARO et al., 2023).

A implementação dessa política pública está alinhada com os compromissos climáticos assumidos pelo Brasil no Acordo de Paris, onde o país se comprometeu a reduzir as emissões de GEE em 48% até 2025 e 53% até 2030, tomando como base o ano de 2005



(MAGALHÃES; DOMINGUES, 2016). O setor sucroenergético, responsável pela produção de biocombustíveis em larga escala, desempenha um papel estratégico nesse contexto (NEVES, P. MENDONÇA, M. 2020). Este estudo pretende identificar as Zonas Focais de Política Pública na região Centro-Sul brasileira, ao qual a sustentabilidade e eficiência das usinas de biocombustíveis certificadas pelo Renovabio são relacionadas com fatores externos de escala municipal. Utilizando indicadores como a Nota de Eficiência Energético-Ambiental (NEEA), que mede a redução de emissões de GEE por litro de biocombustível produzido, em associação com indicadores socioambientais de fontes de dados governamentais.

O Renovabio, ao buscar promover a sustentabilidade e a eficiência na produção de biocombustíveis, possui objetivos de longo prazo que podem levar tempo para se materializar em resultados mensuráveis (AGUIAR et al. 2023). Dessa forma, a análise dos dados relativos ao perfil da produção agrícola dos municípios, além de indicadores demográficos, econômicos e ambientais anteriores a adoção da política pública permitirá compreender, em estudos futuros, se as diretrizes dessa política pública estão refletindo em melhorias na eficiência energética, na redução das emissões de gases de efeito estufa (GEE), em indicadores socioeconômicos e na gestão ambiental dos municípios dessa região. Essa avaliação é fundamental para identificar áreas onde a política está funcionando bem, para além de suas métricas intrínsecas, e onde podem ser necessárias intervenções complementares.

Por fim, o trabalho não apenas contribui para a prática, mas também para a literatura acadêmica, oferecendo um estudo de caso contemporâneo com uso de técnicas de *machine learning* sobre a implementação de políticas públicas em um cenário em constante evolução. As conclusões e análises geradas poderão servir como referência para outros países ou regiões que buscam implementar políticas semelhantes. Assim, ao utilizar as mesmas variáveis demográficas, climáticas e ambientais em uma análise *ex-ante*, este trabalho estabelece um piso analítico que fundamenta futuras investigações, possibilitando a avaliação do impacto real da Política Nacional de Biocombustíveis nas regiões estudadas. Essa abordagem holística é essencial para garantir que as iniciativas de biocombustíveis não apenas atendam aos objetivos diretos do regulamento, mas também promovam um desenvolvimento sustentável em amplo sentido e que beneficie as comunidades locais.

2. Fundamentação teórica

Metodologia base do Renovabio

O princípio metodológico de mensuração do Renovabio é a Análise de Ciclo de Vida do biocombustível, na forma de *well-to-wheel* (do poço à roda), seguindo a estrutura, requisitos e diretrizes estabelecidas pela série de normas ISO 14.040:2016 e ISO 14.044:2016 (ABNT, 2014a, 2014b). Neste método de cálculo, considera-se as etapas agrícolas e as etapas industriais para cômputo da Nota de Eficiência Energética Ambiental (NEEA) de determinado volume de biocombustível produzido pelo emissor primário, expresso em gramas de CO₂ equivalente por MJ de energia (Intensidade de Carbono) (ANP, 2021).

Segundo Folegatti-Matsuura et al. (2018), a fase agrícola contribui significativamente para as emissões de GEE de biocombustíveis, sendo as principais práticas que impactam as emissões de GEE são o uso de calcário, o uso de insumos nitrogenados, a queima de resíduos



agrícolas e o consumo de combustível fóssil em operações mecanizadas. Na etapa agrícola mede-se o impacto da produção de matéria prima na emissão de GEE, expresso em kg CO₂ eq/t de matéria prima. Para fins de contabilizar a etapa agrícola neste estudo, utilizou-se a segmentação municipal, referente ao setor agropecuário, da emissão de gases de efeito estufa a partir dos dados do SEEG Municípios. (FORNARO et al, 2023)

O valor de intensidade de dióxido de carbono equivalente resultante da etapa "dentro da porteira" é somado ao impacto da produção do biocombustível dentro da unidade produtiva (usina), neste cálculo considera-se o transporte de matéria prima, o nível de tecnologia da produção de biocombustível, o tipo de combustível produzido etc. Essa nova carga de CO₂eq, resultante do processo de transformação da matéria prima para combustível, sendo expresso em gCO₂eq/m³ e transformado em Intensidade de Carbono (gCO₂eq/MJ). Por fim, a diferença entre a Intensidade de Carbono do Biocombustível e de seu combustível substituto é feita para obter-se a Nota de Eficiência Energética-Ambiental (FOLEGATTI-MATSUURA, 2018).

Novaes et al. (2024) argumenta que, mesmo o Renovabio não quantificando as emissões derivadas da mudança de uso da terra no âmbito do desmatamento, o principal critério de elegibilidade para inserção no programa é relacionado ao cumprimento das obrigações instituídas na Lei nº 12.651/2012, a Lei de Proteção da Vegetação Nativa, por meio da apresentação do Cadastro Ambiental Rural. Não obstante os autores citam que a expansão do plantio da cana-de-açúcar sobre áreas de vegetação nativa tornou-se raro nos últimos 20 anos, sendo mais comum a alteração do uso de pastagens degradadas para canaviais, o que, entretanto, não é uma realidade para a sojicultura, sendo necessário propor discussões nesse sentido (NOVAES et al. 2024). Como forma de identificar se a mudança do uso do solo é uma variável que obedece a padrões geográficos, utilizou-se as métricas de emissão de GEE atribuídas a essa atividade segundo dados do SEEG Municípios.

Para os objetivos deste estudo, utilizamos as planilhas abertas, disponibilizadas na plataforma da ANP, principalmente as que apresentam os certificados de produção eficiente aprovados entre os anos de 2019 e 2023. A modelagem foi criada a partir do relacionamento de dados setoriais e governamentais, em conjunto com bases temáticas de indicadores demográficos do IBGE (valores de 2013, 2015 e 2016), os dados referentes a emissão de GEE (para o ano base de 2019), dados referentes a contratação de apólice de seguros para cultura da cana-de-açúcar na modalidade de produtividade e custeio segundo o Sistema de Subvenção para o Seguro Rural (ano base 2019), valores anuais de produção apresentados pelo IPEADData (valores de 2019 a 2023), área de cobertura do uso do solo na categoria de cana-de-açúcar segundo o MapBiomas (valores de 2019 a 2023), área com sistemas de fertirrigação instalada por município apresentado pela ANA como indicador de tecnificação da cultura (2019) e o Índice de Desenvolvimento Humano Municipal (IDHM) retirado do Atlas Brasil (2010).

Identificação de zonas focais por *machine learning*

O uso de métodos de análise espacial e estatística tem se mostrado fundamental na identificação de padrões e na otimização de recursos em diversas áreas de pesquisa. Zeng et al. (2024) emprega uma estrutura abrangente para identificar Clusters Industriais Inovadores (IIC) por meio da análise de dados de patentes. Nesse estudo, foram utilizados o algoritmo *Latent Dirichlet Allocation* (LDA) para modelagem de tópicos a partir de documentos de



patentes e o algoritmo Multivariate Density-Based Spatial Clustering of Applications with Noise (MDBSCAN). Este último integra distâncias tecnológicas e espaciais, permitindo um delineamento mais preciso dos clusters. O principal objetivo deste trabalho é identificar com precisão o IIC, superando limitações de abordagens anteriores que focam predominantemente nas dimensões geográficas, o que amplia a compreensão da colaboração tecnológica e da competitividade econômica regional (ZENG et al., 2024).

Outro estudo analisa as estratégias de Serviços Médicos de Emergência (EMS) utilizando várias técnicas de análise espacial, como análise de sobreposição, análise de buffer e estimativa de densidade de kernel (KDE), além de métodos de agrupamento e técnicas de autocorrelação espacial (Moran's I e Local Moran's I). O objetivo dessa pesquisa foi investigar padrões espaciais em incidentes de emergência, otimizar a alocação de recursos e aprimorar a capacidade de resposta geral no EMS. Aproveitando dados de código aberto, o estudo visou informar os tomadores de decisão e profissionais de saúde, aumentando a eficiência do planejamento de resposta a emergências (KUO; CHANG, 2023) (CHU; CHEN, 2018). Adicionalmente, a análise de dados de *crowdsourcing* sobre deslizamentos de terra e fluxos de detritos também foi abordada em um estudo que utilizou métodos estatísticos, como Moran's I e Getis-Ord G^* (G_i^*), para identificar grupos espaciais de ocorrências de risco sem viés subjetivo (FENG; XU; SHI, 2022). O modelo DBSCAN foi aplicado para detectar padrões de pontos quentes em tamanhos e formas variadas. O objetivo principal foi demonstrar a viabilidade do uso de fotografias históricas na análise da distribuição espacial de fluxos de detritos e deslizamentos de terra, melhorando a identificação de pontos quentes perigosos.

Tabarej et al (2024) apresentou um novo algoritmo de detecção de pontos de acesso baseado em gráficos, utilizando a teoria dos conjuntos aproximados (GBHSDRS). Esse método categoriza dados em regiões inferiores, de limite e negativas, aprimorando a identificação de pontos ativos significativos e minimizando falsos positivos. A eficácia do algoritmo foi demonstrada por sua aplicação a dados socioeconômicos de Uttar Pradesh, com foco em instalações médicas. Por fim, um estudo que analisou falhas de energia na rede elétrica de Guangzhou entre 2016 e 2018 (ZENG et al., 2024) utilizou a visualização de mapas de calor para análise de frequência, o método I de Moran para avaliar a autocorrelação espacial e o algoritmo DBSCAN para identificar áreas propensas a falhas. Os objetivos dessa pesquisa incluíram analisar padrões de distribuição espacial das falhas de energia, identificar áreas de alto risco e compreender padrões temporais, especialmente os horários de pico de falha relacionados às rotinas diárias. A integração desses métodos visa aprimorar o gerenciamento de riscos e melhorar a confiabilidade dos sistemas de fornecimento de energia.

3. Metodologia

Todos os procedimentos metodológicos realizados neste trabalho foram feitos na plataforma *Google Colab* e programados em linguagem Python. Os dados utilizados são provenientes de repositórios abertos e confiáveis, disponibilizados, em sua maioria, por órgãos e instituições setoriais estatais. A metodologia utilizada neste trabalho teve como referência o método proposto por Silva et al. (2011) e foi composta por 12 etapas principais:


Workshop Internacional
SUSTENTARE & WIPIS 2024
 Sustentabilidade, Indicadores e Gestão de Recursos Hídricos
www.sustentarewipis.com.br

18 a 22
de novembro
 Transmissão online • Evento gratuito

Realização:  
 Apoio Institucional:  

1. **Proposição do framework** para identificação de zonas de interesse considerando fatores socioambientais, demográficos, setoriais e referentes à emissão de gases de efeito estufa e vulnerabilidade climática;
2. **Coleta de dados de fontes diversas**, estes dados foram classificados por grupos temáticos, sendo: grupo MCTI (índices de vulnerabilidade climática extraídos do repositório de dados MCTI AdaptaBrasil); grupo IBGE (dados demográficos e socioeconômicos em escala municipal); grupo AGRO (dados relativos à produção, uso do solo e irrigação para cultura da cana-de-açúcar); grupo GEE (dados de emissão de gases de efeito estufa para o setor agrícola e pela mudança do uso do solo); grupo INPE (dados relativos a focos de queimadas, apólices de seguro contratadas no período, entre outros) e; grupo-alvo ANP (fatores relativos ao Renovabio, sendo as principais variável-resposta deste estudo).
3. **Transformação de dados (*feature engineering*)**. Além das 32 variáveis brutas coletadas, foram propostas a transformação ou correlação entre duas ou mais variáveis para criação de índices. No geral, as transformações propostas foram adotadas para ajustar a dimensionalidade do banco de dados e garantir coesão dos valores de referência.
4. **Identificação de variáveis estatisticamente relevantes (*feature selection*)**. As variáveis separadas por grupos foram submetidas em diversas interações aos algoritmos de clustering K-means e DBSCAN para o teste de hiperparâmetros tendo como resposta o Silhouette Score para identificação da combinação de melhor performance (número de variáveis, número de clusters, valor épsilon e número de samples). Identificou-se que os principais grupos de variáveis para criação de clusters coesos foram o grupo IBGE, grupo AGRO e grupo GEE.
5. **Processamento de dados** para os grupos de variáveis selecionados foram aplicados 3 principais processamentos: i) detecção de outliers, optando pela manutenção dos valores; ii) identificação de valores faltantes; iii) agregamento e transformação em grupos de teste e treino. Foram utilizadas as seguintes bibliotecas de Python para tal: scikit-learn, Pandas, Seaborn and Matplotlib.
6. **Geração de classes da variável resposta** a partir do grupo ANP foi selecionado a Nota de Eficiência Energético Ambiental como principal métrica alvo do modelo. Para tal foram criadas 3 classes utilizando método de classificação por desvio padrão.
7. **Padronização dos dados** utilizando da função *MinMax Scaler* para normalizar as variáveis numéricas em um intervalo entre 0 e 1. Esse método é empregado para que todas as variáveis possam contribuir de forma proporcional ao modelo, evitando a sobreposição de variáveis com escalas numéricas maiores.
8. **Redução de dimensionalidade** com emprego do teste PCA (*Principal Components Analysis*) para identificação das principais variáveis dos grupos selecionados que influem para o resultado e assertividade do modelo de clustering escolhido.
9. **Implementação do modelo** de *clustering* K-means para agrupamento de municípios que apresentam condições demográficas, econômicas e socioambientais similares. O modelo foi implementado utilizando da biblioteca sci-kit learn e o número de clusters definidos foram três, por apresentarem melhores resultados para o Silhouette Score, Dunn Index e Davies-Bouldin Index.
10. **Análise das métricas** de *clustering* por meio da aplicação de testes estatísticos dos agrupamentos identificados tendo NEEA como variável resposta. Utilizou-se de



- ANOVA, teste de Tukey HSD, análise das médias e nível de confiança a partir de *bootstrapping* e representações gráficas de diferenciação entre os grupos (*boxplot* e gráficos de densidade).
11. **Identificação de hotspots**, inicialmente foi feito a extrapolação dos valores de NEEA utilizando a biblioteca kNeighborsRegressor (n=6) para criação de um mapa de valores que compreendam todos os municípios da região Centro-Sul e a partir deste feito o mapa de agrupamentos LISA (*Local Indicator of Spatial Association*, bibliotecas ESDA e SPLOT) definindo 4 grupos de relação espacial.
 12. **Sobreposição dos agrupamentos** K-means e LISA para identificação das Zonas Focais de Política Pública, feito a partir da correlação entre a matriz de grupos, definindo assim 3 padrões principais e os municípios que estão categorizados nestes padrões.

4. Resultados

4.1 Análise de correlação entre variáveis

A investigação prévia realizada a partir do banco de dados devidamente padronizado revelou interações importantes para contextualizar esse estudo. As variáveis relacionadas ao conjunto de dados da ANP revelam padrões significativos sobre a eficiência energética e as emissões de Gases de Efeito Estufa (GEE). A relação positiva entre o volume de produção de etanol certificada e a Nota de Eficiência Energética Ambiental indica que há uma consistência positiva nos indicadores do programa. Ademais, observam-se correlações negativas relevantes entre o volume certificado e as emissões de GEE no setor agro e por mudança de uso do solo. Isso sugere que a expansão do volume de produção certificada e de alta eficiência pode estar ligada a menores emissões de GEE.

A mediana da produção de cana por ano destaca-se ao apresentar uma fortíssima relação com a mediana da área total de cana de açúcar plantada no município, além de uma significativa correlação com a área de cana fertirrigada. Além de uma relação positiva com o número de apólices de seguro contratados no período e com as emissões de GEE no setor agro, sugere que um aumento na produtividade está associado a um nível maior de intensificação e tecnificação de plantio, além de maior cobertura contra possíveis desastres ambientais. Além disso, a produtividade média mantém relações positivas com indicadores demográficos, a área total do município e PIB para o ano de 2013, refletindo a interação entre produção e indicadores econômicos e demográficos.

Em relação às emissões de GEE por mudança do uso do solo, observa-se uma forte relação negativa com o volume certificado de etanol produzido, indicando que quanto maior as emissões por mudanças no uso do solo, menos propriedades estão capacitadas para adesão ao programa Renovabio, sendo que, a única exigência da política é a adequação ao Cadastro Ambiental Rural (CAR). Além disso, as emissões de GEE por mudança de uso do solo possuem uma correlação fortíssima positiva com as emissões do setor agro e uma relação negativa com a área total fertirrigada. A correlação negativa com as emissões de GEE por mudança de uso do solo sugere que áreas onde há maior tecnificação tendem a estar associadas com menores emissões decorrentes da alteração do uso da terra. Não obstante, a relação negativa com o Índice de Desenvolvimento Humano sugere que municípios com



melhores índices de desenvolvimento humano tendem a ter menores emissões associadas ao setor agro e mudança do uso do solo.

O PIB municipal apresenta uma relação positiva fortíssima com a população em 2016 e uma relação média com o IDH-Municipal, sugerindo que um PIB maior está relacionado a um crescimento populacional e a melhores índices de desenvolvimento. Por fim, o Índice de Desenvolvimento Humano Municipal apresenta uma relação negativa com os indicadores de GEE, indicando que municípios com melhores índices de desenvolvimento humano tendem a ter menores emissões nos setores agro e mudança do uso do solo. Sua correlação média positiva com a população e com o PIB em 2013 reforça a relação entre desenvolvimento e crescimento econômico.

4.2 Identificação de clusters

Após a exclusão de variáveis que não demonstraram boa correlação e/ou não impactavam o modelo final (*PCA test*), da padronização dos dados a partir do método *MinMaxScaler* e da realização e validação estatística e paramétrica do algoritmo de *clustering*, realizou-se a elaboração de um gráfico *boxplot* contendo todas as variáveis utilizadas, separadas pelos *clusters*. Ressalta-se que não foram realizadas a exclusão de *outliers* pelo interesse de identificar municípios que destoam dos padrões regionais. A análise realizada utilizando a técnica de *clustering* K-means revelou três perfis distintos de municípios, conforme ilustrado na Figura 1.

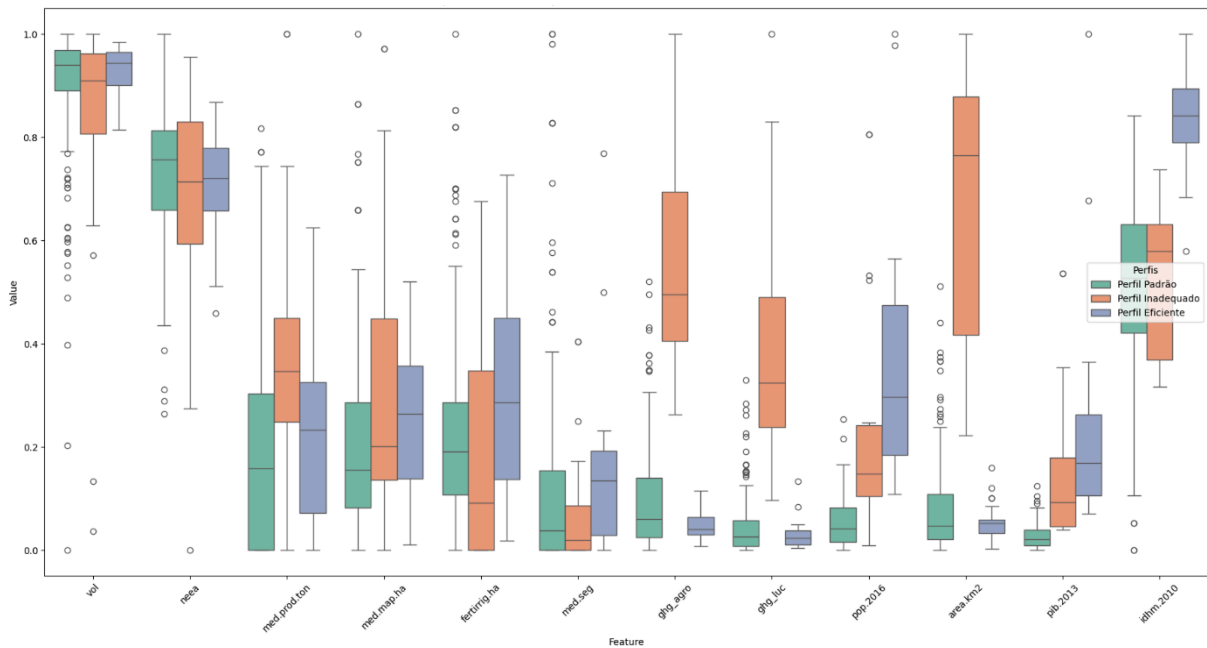


Figura 1: Boxplot das Variáveis por Análise de Cluster (K-means). (Fonte: Elaboração Própria)



- *Perfil Padrão (0)*: as cidades classificadas são caracterizadas por suas menores dimensões em termos de área e população, o que se traduz em uma densidade populacional reduzida. Esses municípios apresentam um Índice de Desenvolvimento Humano (IDH) moderado, e suas emissões de gases de efeito estufa (GEE) são relativamente moderadas. Essa combinação sugere práticas agrícolas que, embora ainda contribuam para a emissão de GEE, não o fazem de maneira excessiva. Além disso, a produtividade da cana-de-açúcar e a área total plantada são baixas, indicando um setor agrícola menos desenvolvido e uma menor dependência socioeconômica deste setor.
- *Perfil Eficiente (1)*: as cidades classificadas sob esse perfil destacam-se por uma população e indicadores econômicos robustos, o que reflete um desenvolvimento urbano dinâmico. Com índices elevados de IDH, essas localidades evidenciam um bom nível de vida e um desenvolvimento social significativo. Notavelmente, as emissões de GEE são mais baixas em comparação aos outros perfis, sugerindo a adoção de práticas agrícolas relativamente mais sustentáveis. Além disso, a alta produtividade e a ampla área plantada de cana-de-açúcar indicam um setor agrícola avançado e tecnificado, contribuindo de forma significativa para a economia local.
- *Perfil Inadequado (2)*: as cidades deste perfil revelam um tamanho médio, com áreas totais e populações que estão em níveis moderados ao se comparar aos outros perfis. O IDH nessas localidades é superior ao do perfil padrão, refletindo melhores condições de vida. Contudo, as emissões de GEE são elevadas, o que aponta para uma forte conexão com a produção agrícola e práticas de uso do solo intensivas com potenciais áreas de desmatamento. Além disso, tanto a produtividade quanto a área cultivada de cana-de-açúcar são mais altas que as observadas no perfil padrão, sugerindo uma economia agrícola mais predominante.

4.3 Métricas de desempenho estatístico do modelo

A Análise de Variância (ANOVA) realizada sobre os clusters obtidos pelo K-means, revela diferenças estatisticamente significativas entre eles em relação à nota de eficiência energética-ambiental (NEEA) como variável de resposta. Os resultados indicam um valor F de 3,998 com um valor p de 0,0197. Esse resultado estatisticamente significativo sugere que pelo menos um dos três clusters é diferente dos outros em termos de NEEA. Esta diferenciação entre clusters sustenta a hipótese de que as regiões de produção de biocombustíveis, segmentadas por indicadores socioeconômicos e ambientais, exibem perfis de eficiência distintos. Dado o valor p abaixo do limite convencional de 0,05, rejeitamos a hipótese nula de ausência de diferença nas médias de NEEA entre os clusters, indicando que esses grupos são significativamente distintos em seu desempenho.

Para explorar mais detalhadamente a natureza dessas diferenças, foi realizado o teste *post-hoc* de Tukey HSD, proporcionando uma visão sobre as relações específicas entre os clusters. As comparações par a par indicam que existe uma diferença significativa nas médias de NEEA entre os perfis (*clusters*) “Padrão” (0) e “Inadequado” (2), assim como entre os perfis “Eficiente” (1) e “Inadequado” (2), com valores p de 0,05 e 0,0143, respectivamente. A diferença de média ajustada para a comparação entre os clusters 0 e 2 é de -0,0915, enquanto a diferença entre os clusters 1 e 2 é de -0,1278. Os limites inferior e superior para

Workshop Internacional
SUSTENTARE & WIPIS 2024
Sustentabilidade, Indicadores e Gestão de Recursos Hídricos
www.sustentarewipis.com.br

18 a 22
de novembro
Transmissão online • Evento gratuito

Realização: PUC CAMPINAS, EESC-USP, Agência das Bacias PCJ
Apoio Institucional: Agência das Bacias PCJ

essas comparações confirmam que a separação entre o cluster 2 e os outros dois clusters é estatisticamente confiável. Em contraste, a comparação entre os clusters 0 e 1 (perfis padrão e eficiente) resultou em um valor não significativo ($p = 0,3857$), sugerindo que esses clusters não diferem substancialmente em termos de suas médias de NEEA.

Esses resultados da ANOVA e do Tukey HSD indicam que o “*Perfil Inadequado*” é estatisticamente distinto dos outros dois clusters, exibindo uma média de NEEA significativamente mais baixa. Esse achado está alinhado com a interpretação de que o cluster 2 pode representar regiões onde a eficiência energética-ambiental é prejudicada por fatores socioeconômicos e ambientais específicos. Em contraste, os clusters 0 e 1 não apresentam diferenças significativas em suas médias de NEEA, sugerindo que podem compartilhar características estruturais semelhantes em relação à eficiência. Essa diferenciação entre os clusters implica uma possível necessidade de abordagens políticas adaptadas para atender às necessidades e ineficiências específicas das regiões do cluster 2.

Para aprofundar a compreensão dos valores médios de NEEA dentro de cada perfil e avaliar a variabilidade nas pontuações de eficiência, foi empregada uma técnica de *bootstrap* para estimar a média e os intervalos de confiança de cada *cluster*. O *bootstrap* é especialmente benéfico aqui, pois permite gerar intervalos de confiança empíricos sem depender de suposições de normalidade, proporcionando estimativas robustas que acomodam potenciais *outliers* ou distribuições assimétricas dos dados.

Os resultados mostram que o “*Perfil Padrão*” possui uma média de NEEA obtida por *bootstrap* de 0,722 com um intervalo de confiança variando de 0,713 a 0,732, indicando um intervalo relativamente estreito e, portanto, um alto grau de consistência nas pontuações de eficiência dentro desse grupo. O “*Perfil Eficiente*” apresenta uma média ligeiramente maior de 0,743, com um intervalo de confiança de 0,720 a 0,764. Esse intervalo sugere, igualmente, um nível moderado de consistência nas pontuações de NEEA dentro do cluster 1, embora com uma variação ligeiramente maior em comparação ao cluster 0. Em contrapartida, o “*Perfil Inadequado*”, identificado anteriormente como distinto nos testes ANOVA e Tukey HSD, mostra a média mais baixa, de 0,640, com um intervalo de confiança mais amplo de 0,556 a 0,716. O intervalo mais largo neste caso sinaliza uma maior variabilidade nas pontuações de NEEA dentro deste cluster, o que pode refletir uma heterogeneidade subjacente nas características socioeconômicas ou ambientais das regiões no cluster 2.

Os intervalos de confiança obtidos por *bootstrap* oferecem um reforço para a separação dos clusters observada na ANOVA. Os intervalos de confiança não sobrepostos entre o “*Perfil Inadequado*” e os outros clusters reforçam a conclusão de que este é fundamentalmente distinto, tanto em termos de média de NEEA mais baixa quanto de maior variabilidade. Essa variabilidade no cluster 2 pode sugerir a necessidade de intervenções específicas para as regiões que apresentam fatores diversificados que contribuem para a baixa eficiência. Além disso, os intervalos de confiança relativamente estreitos nos clusters 0 e 1 indicam um desempenho consistente dentro desses grupos, refletindo práticas de produção ou características ambientais mais homogêneas que contribuem para níveis de eficiência estáveis.

Workshop Internacional
SUSTENTARE & WIPIS 2024
 Sustentabilidade, Indicadores e Gestão de Recursos Hídricos
www.sustentarewipis.com.br

18 a 22 de novembro
 Transmissão online • Evento gratuito

Realização: PUC-CAMPINAS, EESC-USP
 Apoio Institucional: PUCJ, Agência das Bacias PCJ

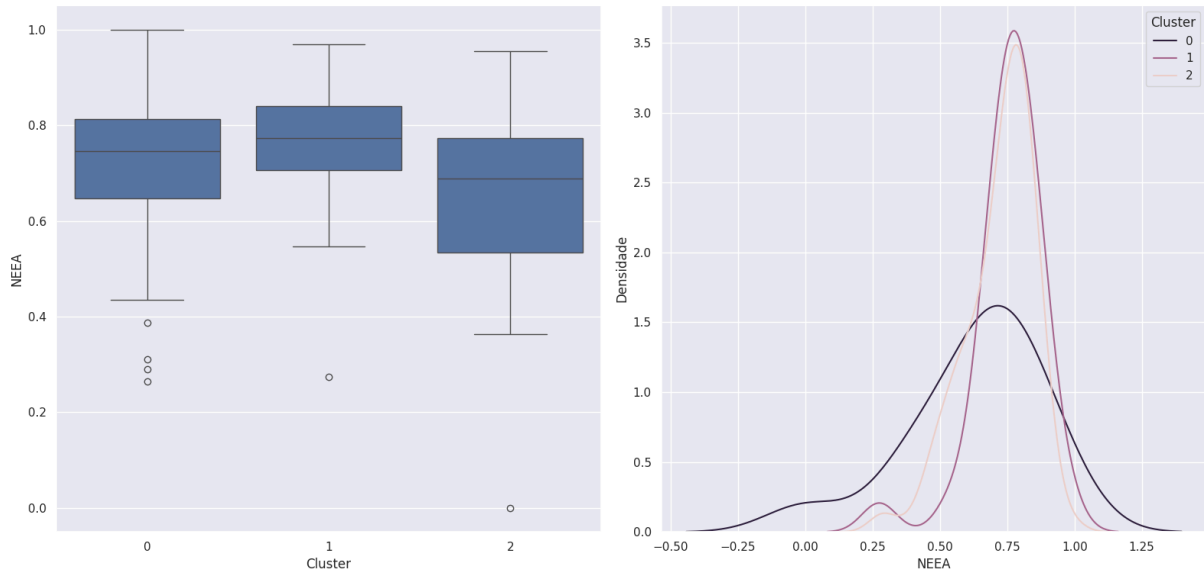


Figura 2: Caracterização - gráfico *Boxplot* e de densidade para os *clusters* e NEEA. (Fonte: Elaboração Própria)

A análise de perfil destaca as características exclusivas de cada cluster com base em diversas variáveis ambientais, econômicas e demográficas, oferecendo uma base para entender as distinções estatísticas observadas. O “*Perfil Padrão*”, por exemplo, emerge como um grupo caracterizado por uma eficiência moderada e um impacto ambiental baixo. As emissões de GEE neste cluster são relativamente baixas, com média do índice em torno de 0,076 para as emissões provenientes de fontes agrícolas. As regiões neste cluster exibem capacidades industriais de produção moderadas, sugerindo que este perfil é composto por regiões com práticas de produção equilibradas e uma pegada ambiental limitada. Esse perfil implica que as regiões no “*Perfil Padrão*” podem representar sistemas de produção de biocombustível menores ou menos tecnificados, onde a eficiência é mantida sem grandes custos ambientais.

O “*Perfil Eficiente*”, que possui a maior média de NEEA, demonstra um equilíbrio distinto entre eficiência e impacto ambiental moderado. As emissões de GEE para este cluster são um pouco mais altas do que as do anterior, com uma média do índice de 0,309, embora permaneçam abaixo dos níveis observados no “*Perfil Inadequado*”. Este agrupamento também apresenta valores moderados para capacidade de produção para etanol instalada, indicando que este grupo inclui regiões com usinas maiores ou mais eficientes. As altas pontuações de eficiência dentro do perfil sugerem que essas regiões podem representar ambientes ideais de produção de etanol, onde as práticas estão bem alinhadas com as metas de desempenho ambiental do RenovaBio. Do ponto de vista das políticas públicas, o “*Perfil Eficiente*” poderia ser posicionado como um ponto de referência para a produção sustentável de biocombustível, onde a produção de energia é maximizada enquanto as emissões são controladas. Formuladores de políticas poderiam usar este cluster para desenvolver diretrizes de melhores práticas ou padrões de eficiência que outras regiões poderiam adotar para melhorar a sustentabilidade da produção.



O “*Perfil Inadequado*”, que foi identificado como distinto tanto estatística quanto operacionalmente, se destaca por sua baixa eficiência e alto impacto ambiental. A média das emissões de GEE provenientes da agricultura neste *cluster* é notavelmente elevada em 0,625, mais do que o dobro das emissões observadas no “*Perfil Eficiente*” e significativamente superior às do “*Perfil Padrão*”. Além disso, este agrupamento mostra altos valores para capacidade industrial instalada, refletindo capacidades de produção substanciais que provavelmente contribuem para as pressões ambientais observadas. Este cluster inclui regiões onde práticas agrícolas ou industriais intensivas parecem comprometer a eficiência enquanto elevam as emissões de GEE, sugerindo que as regiões do “*Perfil Inadequado*” poderiam se beneficiar de políticas direcionadas focadas na redução de emissões, melhoria da eficiência e, possivelmente, atualização tecnológica. A heterogeneidade dentro do *cluster*, como sugerido pelo seu intervalo de confiança mais amplo, implica que as regiões dentro deste cluster podem variar significativamente em termos de sua eficiência e impacto ambiental, reforçando a importância de intervenções personalizadas que abordem ineficiências específicas.

4.4 Análise Integrada de Padrões: K-means e LISA

A metodologia adotada consistiu, primeiramente, na aplicação do algoritmo K-means, que foi utilizado para classificar as regiões em três perfis distintos com base nas características demográficas e na emissão de GEE. O Perfil Eficiente (cluster = 1) representa regiões com alta eficiência (NEEA alta) e bons indicadores socioeconômicos. O Perfil Padrão (cluster = 0) abrange regiões em transição, com características medianas, enquanto o Perfil Inadequado (cluster = 2) identifica regiões com baixa eficiência (NEEA baixa) e perfil socioeconômico desfavorável. Em seguida, o índice LISA foi calculado para avaliar a autocorrelação espacial da NEEA, permitindo identificar padrões locais que podem não ser evidentes em uma análise global. As áreas foram classificadas conforme sua localização em relação aos clusters do K-means, resultando em uma análise mais granular dos padrões de eficiência. Os índices LISA permitiram a identificação de áreas com altos (AA) e baixos (BB) valores de NEEA, bem como áreas com alta eficiência cercadas por regiões de baixa eficiência (AB) e vice-versa (BA), conforme apresenta a Figura 3.

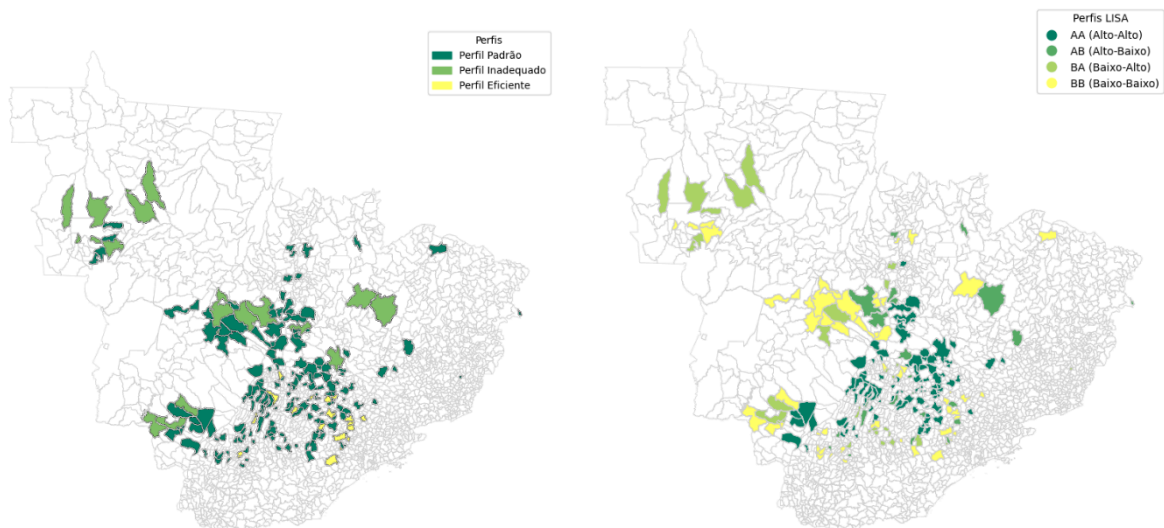




Figura 3: Mapas de Identificação dos clusters K-means e LISA. (Fonte: Elaboração Própria)

A integração dos resultados possibilitou a combinação dos clusters e dos padrões identificados pelo índice LISA, resultando em uma visão abrangente das interações entre a eficiência energética e as variáveis socioeconômicas. As zonas de alta eficiência foram analisadas como áreas que apresentam altos valores de NEEA, sugerindo a adequação das métricas estabelecidas pelo Renovabio e focalizam por ser áreas modelo da política pública. Por outro lado, as zonas críticas de baixa eficiência foram identificadas como áreas que devem ser monitoradas devido sua baixa eficiência e perfil socioeconômico desfavorável. As zonas de transição foram classificadas como regiões medianas que podem, a partir de análises futuras, tornar-se AA ou BB e, a partir dessa metodologia, confirmar a hipótese de interrelações dos indicadores estudados.

A interpretação dos Mapas de Identificação demonstra um perfil regional bem definido onde municípios que se enquadram no Perfil Eficiente encontram-se concentrados na região central e noroeste do estado de São Paulo (Mesorregião de São José do Rio Preto, Ribeirão Preto e Piracicaba) e os valores tendem a reduzir a partir desse núcleo principal. Os resultados obtidos demonstraram um alinhamento entre a NEEA e as variáveis demográficas, evidenciando que as regiões com alto desempenho em NEEA apresentavam antes da adoção da Política Nacional de Biocombustíveis características favoráveis.

Além disso, a análise revelou desvios nas zonas de transição, onde uma maior variação na eficiência indicou a necessidade de monitoramento para identificar o aprimoramento ou deterioração dos indicadores estudados após a vigência do programa. A análise integrada de padrões utilizando DBSCAN e LISA não apenas revelou a relação entre a NEEA e as variáveis demográficas e ambientais, mas também destacou a importância do monitoramento das políticas direcionadas e específicas para cada perfil identificado. Esse entendimento pode informar estratégias para a implementação eficaz do Programa RenovaBio, promovendo um desenvolvimento sustentável e eficiente.

4.5 Identificação das Zonas Focais de Política Pública

Os resultados da análise integrada entre os clusters K-means e LISA permitiram identificar três perfis distintos de regiões com base em suas características de eficiência energética e ambiental, bem como em seus indicadores socioeconômicos e de emissões de gases de efeito estufa (GEE). O Perfil Eficiente (AA & AB com cluster = 1) abrange regiões com alta eficiência energética e ambiental, refletida nos altos valores de Nota de Eficiência Energético Ambiental (NEEA), além de indicadores socioeconômicos favoráveis e baixas emissões de GEE. Essas regiões representam exemplos de municípios com um setor sucroenergético já bem estabelecido, mesmo que de forma anterior à adoção da política pública. Como sugestão para próximos estudos, o foco deve ser na averiguação de manutenção do nível de eficiência.

O Perfil Padrão (AB & BA com cluster = 0) inclui regiões em transição, com perfis demográficos e socioeconômicos medianos e emissões de GEE variáveis. Essas áreas apresentam potencial para melhorar ou piorar e devem ser acompanhadas em estudos



posteriores. As regiões AB, que são áreas eficientes cercadas por vizinhos de menor eficiência, e BA, o oposto, indicam a necessidade de estabilizar essas regiões e evitar a deterioração de sua eficiência. Por fim, o Perfil Inadequado (BA & BB com cluster = 2) engloba as regiões que demonstravam maior vulnerabilidade, com baixa eficiência energética e ambiental (NEEA baixa), perfil socioeconômico desfavorável e altas emissões de GEE. Essas áreas apresentam uma outra zona focal importante para próximas análises a fim de monitorar o avanço ou retrocesso das métricas socioambientais.

As Zonas Focais de Políticas identificadas no estudo sugerem que as áreas classificadas como "Perfil Eficiente" podem ser utilizadas como modelos para a replicação de políticas eficazes em outras regiões. Cidades como Araçatuba (SP), Araras (SP), Cerquilha (SP), Itapetininga (SP), Leme (SP), Piracicaba (SP), Pirassununga (SP), Santa Bárbara d'Oeste (SP) e São João da Boa Vista (SP) demonstram alto desempenho em termos de eficiência energética e ambiental, apresentando boas condições socioeconômicas e baixas emissões de GEE. Por outro lado, as regiões pertencentes ao "Perfil Inadequado", como Uberaba (MG), Dourados (MS), Nova Alvorada do Sul (MS), Ponta Porã (MS), Barra do Bugres (MT), Itumbiara (GO), Jataí (GO), Mineiros (GO) e Quirinópolis (GO), evidenciam a necessidade de intervenções imediatas. Essas áreas apresentam baixa eficiência energética e altos níveis de emissões de GEE, o que as torna prioritárias para ações corretivas, visando tanto a melhoria socioeconômica quanto a redução das emissões e o aumento da sustentabilidade ambiental.

Nas Zonas de Transição, representadas pelo "Perfil Padrão", há um potencial significativo para a implementação de políticas que visem estabilizar ou melhorar a eficiência dessas áreas. Essas regiões, embora ainda não apresentem alto desempenho, possuem características que as tornam candidatas ideais para intervenções que podem promover melhorias contínuas em seus indicadores energéticos e ambientais. O Alinhamento/Discrepância observado entre os clusters AA e o "Perfil Eficiente" demonstra que as áreas de alta eficiência energética e ambiental são sustentadas por condições socioeconômicas favoráveis. No entanto, a discrepância nas zonas AB ou BA em relação aos seus perfis DBSCAN pode evidenciar a influência de fatores socioeconômicos que dificultam ou favorecem a eficiência energética nessas regiões, sugerindo que a análise dessas diferenças pode orientar intervenções políticas mais direcionadas e eficazes.

5. Conclusões

A pesquisa apresenta a caracterização dos padrões regionais antes da implementação do Renovabio, destacando o emprego de diferentes variáveis socioambientais em escala municipal para identificar as principais Zonas de Foco para avaliação do Programa Nacional de Biocombustíveis. O modelo baseado em *clustering* e indicadores espaciais (K-means e LISA) apresenta um potencial importante para análises de caracterização *ex-ante* de políticas públicas, porém, ainda carecem de refinamento paramétrico e teste de novas variáveis para uma maior acurácia estatística.

Como produto analítico, este estudo pode identificar com maior precisão, que as cidades com o perfil "Eficiente" estão concentradas, dentro deste recorte temporal e espacial, principalmente no estado de São Paulo e destacam-se por apresentar alta eficiência energética, baixo nível de emissões de GEE, forte desenvolvimento econômico e melhores índices de qualidade de vida no período imediatamente anterior à implementação do Renovabio, sendo



áreas que devem servir de modelo para identificação das boas práticas associadas ao programa. Por outro lado, o perfil "Inadequado" revela áreas de maior vulnerabilidade, caracterizadas por baixa eficiência energética e altas emissões de GEE. Esses municípios, localizados em estados como Mato Grosso do Sul, Goiás e Minas Gerais, necessitam de acompanhamento dos revisores da política pública para que, caso não demonstrem aprimoramentos, sejam propostas intervenções, que podem incluir a implementação de tecnologias mais avançadas, práticas agrícolas sustentáveis e políticas que incentivem a redução das emissões e a melhoria dos indicadores socioeconômicos.

A pesquisa também sugere que os perfis "Padrão" representam zonas de transição, que possuem potencial para melhora ou deterioração de suas métricas de eficiência energética e ambiental. Essas áreas devem ser monitoradas para identificar se sofreram alteração em seu perfil nos últimos anos. No panorama geral, o estudo destaca a importância de adaptar políticas às condições regionais, já que a eficiência energética não está exclusivamente ligada à adoção de tecnologias, mas também a fatores socioeconômicos e ambientais. A análise integrada oferece uma base sólida para futuras avaliações do impacto do RenovaBio e de outras políticas voltadas à descarbonização da economia, sendo fundamental para guiar intervenções que garantam um desenvolvimento sustentável e inclusivo, beneficiando tanto a economia quanto o meio ambiente.

7. Referências bibliográficas

ABNT. NBR ISO 14040 Gestão ambiental – Avaliação do ciclo de vida – Princípios e estrutura. Rio de Janeiro: versão corrigida, 2014b.

ABNT. NBR ISO 14044 Gestão ambiental — Avaliação do ciclo de vida — Requisitos e orientações. Rio de Janeiro: versão corrigida, 2014a.

AGUIAR, Ana Beatriz; BAPTISTA, Izabella; GUITARRARI, Luiza; CARDOSO, João Victor Marques. Biocombustíveis e Descarbonização: oportunidades político-regulatórias para a transição energética brasileira. Rio de Janeiro: FGV Energia, 2023.

ANP. Despacho ANP nº 322, de 31 de Março de 2023. Brasília: Diário Oficial da União, 2023.

ANP. RenovaBio: assuntos. 2024. Disponível em: <https://www.gov.br/anp/pt-br/assuntos/renovabio>. Acesso em: 09 mar. 2024.

BENTO, José A. N.; FILHO, José E. R. V. O Crédito de Descarbonização (CBIO) e sua relação com o mercado de Carbono. In: FILHO, J. E. R. V.; GASQUES, J. G. (Org.) Agropecuária Brasileira: evolução, resiliência e oportunidades. Rio de Janeiro: Instituto de Pesquisa em Economia Aplicada, 2023. p. 265-288. DOI: <http://dx.doi.org/10.38116/9786556350530cap12>.

BODANSKY, Daniel. Paris Agreement. United Nations Audiovisual Library of International Law. United Nations, 2021.

CHU, H. J.; CHEN, Y. C. Crowdsourcing photograph locations for debris flow hot spot mapping. *Natural Hazards*, v. 90, p. 1259–1276, 2018. DOI: 10.1007/s11069-017-3098-6. Disponível em: <https://doi.org/10.1007/s11069-017-3098-6>. Acesso em: 18 out. 2024.



FENG, G.; XU, Y.; SHI, Y. Exploring spatio-temporal patterns of power faults in Guangzhou. In: 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 2022. p. 490-495. DOI: 10.1109/ICCASIT55263.2022.9987192.

FOLEGATTI-MATSUURA, Marília I. S. et al. RenovaCalc: a calculadora do programa RenovaBio. Brasília: VI Congresso Brasileiro sobre Gestão do Ciclo de Vida, 2018.

FORNARO, Daniel H.; SANTOS, Jeronimo A.; MARJOTTA-MAISTRO, Marta C. Funcionamento da renovacalc: uma visão aplicada à produção do etanol de primeira geração. São Paulo: Revista GeSec, 2023. v. 14, n.4, p. 5437-5459 <http://doi.org/10.7769/gesec.v14i4.1990>

KUO, C. Y.; CHANG, Y. T. Introduction of spatial analysis approaches for Emergency Medical Services data: using the Kaggle dataset as an example.. 4, p. 21-29 + ap1-5, 2023. DOI: 10.30216/JPER.202310_(4).0004. Disponível em: [https://doi.org/10.30216/JPER.202310_\(4\).0004](https://doi.org/10.30216/JPER.202310_(4).0004). Acesso em: 18 out. 2024

MAGALHÃES, Aline Souza; DOMINGUES, Edson Paulo. Aumento da eficiência energética no Brasil: uma opção para uma economia de baixo carbono?. FACE-UFMG, 2016.

NEVES, Pedro D. M.; MENDONÇA, Marcelo R. RenovaBio e o agrohidronegócio canavieiro em Goiás. GeoTextos, vol. 16, n. 1, jul. 2020. p. 85-108

NOVAES, Renan M.L.; AMARAL, Diego R.; MORANDI, Marcelo A. B.; FOLEGATTI-MATSUURA, Marília I. S.; BRANDÃO, Miguel. Comment on “The importance of GHG emissions from land use change for biofuels in Brazil: An assessment for current and 2030 scenarios”. Resources, Conservation & Recycling, 2024.

TABAREJ, M. S.; MINZ, S.; SHAIKH, A. A.; SHUAIB, M.; JERIBI, F.; ALAM, S. Graph-based hotspot detection of socio-economic data using rough-set. Mathematics, v. 12, p. 2031, 2024. DOI: 10.3390/math12132031. Disponível em: <https://doi.org/10.3390/math12132031>. Acesso em: 18 out. 2024.

SILVA, R. F. et al. A data-driven framework for identifying productivity zones and the impact of agricultural droughts in sugarcane using SPI and unsupervised learning. In: IEEE INTERNATIONAL WORKSHOP ON METROLOGY FOR AGRICULTURE AND FORESTRY (MetroAgriFor), 2021, Trento-Bolzano, Italy. Anais [...]. Trento-Bolzano: IEEE, 2021. p. 226-231. DOI: 10.1109/MetroAgriFor52389.2021.9628570.

ZENG, S.; WANG, T.; LIN, W.; CHEN, Z.; XIAO, R. A patent mining approach to accurately identifying innovative industrial clusters based on the multivariate DBSCAN algorithm. Systems, v. 12, p. 321, 2024. DOI: 10.3390/systems12090321. Disponível em: <https://doi.org/10.3390/systems12090321>. Acesso em: 18 out. 2024.