



**ANÁLISE DA ECOEFICIÊNCIA DA AGRICULTURA DOS
MUNICÍPIOS BRASILEIROS: BASEADO NO MODELO
ESTOCÁSTICO-DEA**

GABRIELA MAYUMI SAIKI

**MONOGRAFIA DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO**

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO**

**ANÁLISE DA ECOEFICIÊNCIA DA AGRICULTURA DOS
MUNICÍPIOS BRASILEIROS: BASEADO NO MODELO
ESTOCÁSTICO-DEA**

GABRIELA MAYUMI SAIKI

Orientador: PROF. ANDRÉ LUIZ MARQUES SERRANO, EPR/UNB

MONOGRAFIA DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**PUBLICAÇÃO EPR.PG - XXX/AAAA
BRASÍLIA-DF, 03 DE 07 DE 2023.**

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO**

**ANÁLISE DA ECOEFICIÊNCIA DA AGRICULTURA DOS
MUNICÍPIOS BRASILEIROS: BASEADO NO MODELO
ESTOCÁSTICO-DEA**

GABRIELA MAYUMI SAIKI

MONOGRAFIA DE GRADUAÇÃO SUBMETIDO AO DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE ENGENHEIRO DE PRODUÇÃO.

APROVADA POR:

Prof. André Luiz Marques Serrano, EPR/UnB
Orientador

Prof. Dr. Carlos Rosano Peña, ADM/UnB
Co-orientador

Prof. Dr. Clóvis Neumann, EPR/UnB
Examinador interno

BRASÍLIA, 03 DE 07 DE 2023.

FICHA CATALOGRÁFICA

GABRIELA MAYUMI SAIKI

Análise da Ecoeficiência da Agricultura dos Municípios Brasileiros: Baseado no Modelo Estocástico-DEA

2023xv, 147p., 201x297 mm

(EPR/FT/UnB, Engenheiro de Produção, Engenharia de Produção, 2023)

Monografia de Graduação - Universidade de Brasília

Faculdade de Tecnologia - Departamento de Engenharia de Produção

REFERÊNCIA BIBLIOGRÁFICA

GABRIELA MAYUMI SAIKI (2023) Análise da Ecoeficiência da Agricultura dos Municípios Brasileiros: Baseado no Modelo Estocástico-DEA. Monografia de Graduação em Engenharia de Produção, Publicação xxx/AAAA, Departamento de Engenharia de Produção, Universidade de Brasília, Brasília, DF, 147p.

CESSÃO DE DIREITOS

AUTOR: Gabriela Mayumi Saiki

TÍTULO: Análise da Ecoeficiência da Agricultura dos Municípios Brasileiros: Baseado no Modelo Estocástico-DEA.

GRAU: Engenheiro de Produção ANO: 2023

É concedida à Universidade de Brasília permissão para reproduzir cópias deste monografia de graduação e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor se reserva a outros direitos de publicação e nenhuma parte deste monografia de graduação pode ser reproduzida sem a autorização por escrito do autor.

Gabriela Mayumi Saiki

170010902@aluno.unb.br

Agradecimentos

Agradeço a Deus por todo amor e cuidado para comigo durante todo esse período de formação e desenvolvimento do meu projeto de graduação. Foi meu porto seguro em momentos de angústia e meu melhor amigo em momentos de alegria. Então sei que não poderia fazer nada sem seu imenso amor diário, que me motiva a levantar e tentar mais uma vez todos os dias. No lugar de um ponto final, Ele me deu uma vírgula para recomeçar. E esse projeto de graduação é o resultado de um recomeço.

Agradeço aos meus pais que sempre me ajudaram nos meus estudos como podiam e sempre me deram o melhor que estava ao alcance deles, mesmo em meio as dificuldades sempre estiveram lá da maneira que era possível.

Agradeço aos meus dois irmãos do coração (sem laços de sangue), que também são meus melhores amigos, primos e conselheiros. Fazem a minha vida ser mais leve, alegre e dinâmica.

Agradeço ao meu professor e orientador que me deu essa oportunidade de escrever sobre algo tão relevante, mesmo em meio a tantos compromissos me ensinou coisas novas e enriquecedoras para meu desenvolvimento como aluna e profissional.

Este trabalho analisa a ecoeficiência agropecuária dos municípios brasileiros no período de 2017 por meio da Análise Envoltória de Dados, que é um modelo de análise de eficiência não paramétrico. A DEA tem ganhado força desde os primeiros modelos propostos, o CCR e BCC. Possui uma série de variações publicadas ao longo desses 45 anos, modelos de supereficiência, modelo aditivo, *benchmarking*, *free disposal hull (FDH)*, *Malquist*, DEA estocástico, dentre outros. A DEA tem sido utilizada para diferentes áreas e aplicações como na engenharia, escolas, hospitais, indústrias, agropecuária, universidades, dentre outras unidades produtivas. Se tornando essencial para análise da ecoeficiência da agropecuária, no entanto devido algumas limitações dos modelos (não garantia da qualidade dos dados) torna-se necessário a eliminação de *outliers* dos dados para não impactar de maneira tendenciosa no resultado, dessa forma constituir resultados mais robustos dos escores de ecoeficiência dos municípios brasileiros. Uma solução tratada neste pesquisa é utilizar o *Bootstrap* para eliminar esses *outliers* e construir intervalos de confiança do modelo.

Palavras-chave: Análise Envoltória de Dados. Agropecuária. Ecoeficiência.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	CONTEXTUALIZAÇÃO	1
1.2	JUSTIFICATIVA	1
1.3	OBJETIVOS.....	3
1.3.1	OBJETIVO GERAL.....	3
1.3.2	OBJETIVOS ESPECÍFICOS	3
1.4	METODOLOGIA.....	3
1.5	ESTRUTURA DO TRABALHO	3
2	REVISÃO TEÓRICA	5
2.1	ANÁLISE BIBLIOMÉTRICA ECO-EFICIÊNCIA E AGRICULTURA.....	5
2.1.1	METODOLOGIA DA REVISÃO.....	5
2.1.2	CARACTERÍSTICA DOS ARTIGOS	6
2.1.3	IMPACTO DA CITAÇÃO	7
2.1.4	ANÁLISE DO PAÍS	9
2.1.5	ANÁLISE DOS TEMAS	10
2.1.6	DISCUSSÃO DA REVISÃO	10
2.2	ANÁLISE BIBLIOMÉTRICA <i>Bootsrap</i>	12
2.2.1	METODOLOGIA BIBLIOMÉTRICA <i>Bootstrap</i>	12
2.2.2	RESULTADO GERAL DOS ARTIGOS <i>Boostrap</i>	13
2.2.3	ANÁLISE DAS CITAÇÕES	14
2.2.4	ANÁLISE DO ASSUNTO.....	16
2.2.5	ANÁLISE DOS PAÍSES E AFILIAÇÕES	20
2.2.6	DISCUSSÃO DA REVISÃO BOOTSTRAP	22
3	PROCEDIMENTOS METODOLÓGICOS	24
3.1	MODELO DA ANÁLISE ENVOLTÓRIA DE DADOS - DEA.....	24
3.2	EFICIÊNCIA E PRODUTIVIDADE	25
3.3	MODELO CONSTANTE DE ESCALA - CCR/RCE.....	27
3.4	MODELO VARIÁVEL DE ESCALA - BCC/RVE	30
3.5	ECOEFIÊNCIA E A EFICIÊNCIA.....	31
3.6	ÍNDICE DE DIVERSIDADE DE SHANNON-WEAVE	33

3.7	MODELO ESTOCÁSTICO E DETERMINÍSTICOS - REAMOSTRAGEM E <i>Bootstrap</i>	33
3.8	AVALIAR TIPO DE RETORNO DE ESCALA	35
3.9	CONSTRUÇÃO DE INTERVALOS DE CONFIANÇA COM <i>Bootstrap</i>	36
4	APRESENTAÇÃO DOS RESULTADOS	38
4.1	RETIRADA DOS <i>Outliers</i> COM <i>Bootstrap</i>	39
4.2	DEFINIÇÃO DO TIPO DE RETORNO DE ESCALA.....	41
4.3	ECONOMIA DOS MUNICÍPIOS BRASILEIROS COM INTERVALO DE CONFIANÇA.....	41
5	CONCLUSÃO	48
	REFERÊNCIAS BIBLIOGRÁFICAS.....	50

LISTA DE FIGURAS

2.1	Artigos publicados em 1992 a 2023 ma Categoria de Engenharia Multidisciplinar	6
2.2	Artigos publicados em 1985 a 2023 ma Categoria de Economia	7
2.3	Artigos publicados em periódicos de 1992 a 2023	7
2.4	Artigos que citam artigos locais	8
2.5	Artigos que citam artigos do período de 1775 a 2023	8
2.6	Diagrama de três campos	9
2.7	Países de autores do WoS.....	9
2.8	Nuvem de palavras mais frequentes do WoS	11
2.9	Tabela resumo dos dados da WoS Bootstrap.....	13
2.10	Gráfico de Três Campos Bootstrap.....	14
2.11	Gráfico de Produção Científica Anual Bootstrap	14
2.12	Gráfico de média de citações anuais bootstrap	15
2.13	Gráfico de documentos globalmente citados bootstrap.....	15
2.14	Gráfico de autores mais relevantes bootstrap	15
2.15	Gráfico de Lei de Loka.....	16
2.16	Gráfico cluters por acloppamento	17
2.17	Rede de cluters por acloppamento	17
2.18	Evolução temática	18
2.19	Evolução temática de 2018.....	18
2.20	Evolução temática de 2019.....	19
2.21	Evolução temática de 2020.....	19
2.22	Evolução temática de 2021.....	19
2.23	Evolução temática de 2022.....	20
2.24	Tendência de tópicos	20
2.25	Gráfico de produção dos países ao longo do tempo	21
2.26	Gráfico dos países do autor correspondente.....	22
2.27	Gráfico dos países mais citados	22
2.28	Gráfico das afiliações mais relevantes	22
3.1	Fronteira de possibilidades de produção modelo CCR/RCE	26
3.2	Fronteira de possibilidades de produção modelo BCC/RVE.....	30
4.1	Curvas de densidade das ecoeficiências por Região	40

4.2	Bloxplots com e sem Outliers	40
4.3	Gráfico de Densidade das Ecoeficiências das Regiões	43
4.4	Georreferenciamento dos Escores de Ecoeficiência Região Norte	45
4.5	Georreferenciamento dos Escores de Ecoeficiência Região Nordeste	45
4.6	Georreferenciamento dos Escores de Ecoeficiência Região Centro-Oeste	46
4.7	Georreferenciamento dos Escores de Ecoeficiência Região Sudeste.....	46
4.8	Georreferenciamento dos Escores de Ecoeficiência Região Sul	47

LISTA DE TABELAS

2.1	Análise Bibliométrica dos periódicos WoS	7
2.2	Distribuição dos países dos periódicos WoS	10
2.3	As 30 palavras-chave mais frequentes da categoria EC e EM do WoS	11
2.4	Impacto dos autores locais	16
3.1	Modelo Primal e Dual	31
3.2	Modelo de PPL (Primal-Dual) para Calcular a Ecoeficiência	32
4.1	Estatísticas descritivas (Censo Agropecuário 2017-IBGE) e SEEG	39
4.2	Resultado da retirada de Outliers	39
4.3	Ranking do Índice de Ecoeficiência dos Municípios por Região	42
4.4	Média, mediana e amplitude das ecoeficiências	44
4.5	Ecoeficiência dos Dez Melhores Municípios	44

LISTA DE CÓDIGOS FONTE

Capítulo 1

Introdução

1.1 Contextualização

A agropecuária sempre foi uma das grandes forças econômicas brasileiras responsável pela geração de empregos, de acordo com o Censo Agronegócio de 2017 do Instituto Brasileiro de Geografia e Estatística - IBGE, mais de 15 milhões de pessoas estão trabalhando com atividades agropecuárias, assim como existe cerca de 5 milhões de estabelecimentos com foco na atividade agropecuária. Sendo contribuinte do efeito multiplicador financeiro no País, responsável por parcela considerável do PIB brasileiro.

Desde o século XVI a atividade da produção da cana de açúcar para exportação e a produção de gado para subsidiar a atividade agrícola no período, mostra que desde os princípios da colonização a agropecuária se fez presente na formação econômica brasileira, não é atoa que recentemente mais de 638 milhões de toneladas de cana de açúcar foram produzidas em 2017, seguida pela produção de mais de 100 milhões de toneladas de soja e 88 milhões de grãos de milho. Totalizando cerca de 351 milhões de hectares (3,5 milhões km²) dedicados à agropecuária, o que representa cerca de 41% do território nacional. [IBGE 2017]

1.2 Justificativa

Portanto é imprescindível pensar nos impactos que essa atividade econômica tem causado no nosso país, tanto positivamente (financeiro, geração de empregos e balança comercial), quanto negativamente (exploração de terras, desmatamento e produção de resíduos). Encontrar um equilíbrio entre o desenvolvimento econômico e a preservação do meio ambiente é uma das prerrogativas da ecoeficiência, que procura entregar um produto para atender as necessidades humanas pensando em reduzir os impactos ambientais, otimizando de maneira a minimizar os recursos para produzir mais produtos, dessa forma a ecoeficiência pode ser considerada uma ferramenta prática de grande importância para o desenvolvimento sustentável. [Matsumoto and CHEN 2021]

Dado a imensidão das atividades da agricultura no país, somos geralmente levados a ter um olhar com maior atenção aos grandes produtores quando falamos de desenvolvimento sustentável, ou ecoeficiência. No entanto, entender a realidade dos pequenos produtores é de suma importância, pois não é nítido se suas atividades são sustentáveis, e dado o grande número de pequenos produtores de um país ocasiona impactos relevantes no desempenho do País. [Stepien et al. 2021]

O que torna a Análise Envoltório de Dados (DEA) uma ferramenta útil para esse tipo de análise. Existem dois modelos clássicos, os modelos CCR proposto por [Charnes et al. 1978] e o BCC proposto por [Banker et al. 1984] baseados em retornos constante e variáveis de escala, respectivamente. O modelo possui vantagens e desvantagens, no entanto, uma série de variações tem surgido nos últimos anos para ter resultados cada vez mais precisos, melhorando assim o DEA [LIU et al. 2019].

Na literatura é comum ver variações do modelo DEA aplicado a análise da ecoeficiência. Na literatura, o SBM-Indesejável é utilizado para avaliar a ecoeficiência para analisar o sistema de mineração da China [LIU et al. 2019]. A abordagem meta-fronteira entre saídas indesejáveis, supereficiência e folgas (Meta-USSBM) foi utilizada para calcular a ecoeficiência de províncias da China de 2001 a 2014, com variáveis nas áreas econômica, energético e ambiental [Huang et al. 2019]. A mesma abordagem foi utilizada para avaliar os municípios da China de 2003 a 2013 [Yu et al. 2019]. Avaliar a ecoeficiência propriedades Polônia com uma abordagem de segundo estágio [Stepien et al. 2021]. Medidas de eficiência baseadas em folga e produtos indesejáveis (Super-SBM) foi usada para avaliar 21 cidades da província de Guangdong, China no período de 2005-2014. [Zhou et al. 2018]. Modelo DEA de rede aprimorado com saída indesejável (NDEA) para validar a ecoeficiência de 30 províncias na China no período de 2003 a 2016 [Yu et al. 2020]. Utilizando a supereficiência (SE-DEA) para analisar o impacto da urbanização e ecoeficiência em 281 prefeituras na China no período de 2006 a 2013 [Bai et al. 2019]. A abordagem de dois estágios com DEA para determinar indicadores de ecoeficiência industrial de 30 províncias chinesas de 2005-2015 com uma análise de regressão Tobit de variáveis aleatórias [Matsumoto and CHEN 2021]. Elaboração de um indicador de ecoeficiência por meio da DEA para avaliar a evolução da ecoeficiência de 282 regiões europeias de 2006-2014 [Bianchi et al. 2020]. Avaliação da ecoeficiência de 40 parques industriais por meio da DEA no ano de 2021 [Fan et al. 2017]. Avaliar a ecoeficiência da industrial da metrópole de Ulsan, Coreia 2000-2015 [Shah et al. 2020].

Portanto, existem vários estudos que avaliam a ecoeficiência, mas uma questão que normalmente não é tratada nos modelos são os *outliers*, os quais quando existem no modelo podem influenciar tendenciosamente a consistência dos resultados se não forem identificados e retirados dos dados em análise.

1.3 Objetivos

1.3.1 Objetivo Geral

Analisar quantitativamente quais os municípios e regiões brasileiras são mais ecoeficientes.

1.3.2 Objetivos Específicos

- Estabelecer um modelo consolidado para analisar a ecoeficiência dos municípios brasileiros;
- Identificar as influências dos *outliers* no resultado das análises de ecoeficiências com uso do *bootstrap*;
- Estabelecer intervalos de confiança pelo *bootstrap*;
- Fazer um comparativo dos municípios e regiões brasileiras de maior e menor ecoeficiência na agricultura.

1.4 Metodologia

A metodologia utilizada será a Análise Envoltória de Dados (DEA) desenvolvidos por Charnes (1978) e Banker (1984) baseados na Econometria e na Pesquisa Operacional para calcular as ecoeficiências dos municípios brasileiros, com uso do *bootstrap* para retirada de *outliers* dos dados e para construção de intervalos de confiança. Dessa forma a presente pesquisa visa promover uma avaliação quantitativa para obter parâmetros do cenário do agronegócio baseado no último censo realizado pelo [IBGE 2017], dessa forma comparar a ecoeficiência entre os municípios e estados brasileiros.

1.5 Estrutura do trabalho

O presente trabalho está organizado em 5 etapas, que foram estruturados da seguinte forma:

- I. A primeira etapa contém a introdução do trabalho, contextualização do cenário do agronegócio e a importância da análise DEA para a ecoeficiência, objetivos da pesquisa e metodologia utilizada;
- II. A segunda etapa contém a revisão teórico para embasamento da pesquisa, com uma análise bibliométrica de artigos envolvendo a ecoeficiência, *bootstrap* e agricultura;

- III. A terceira etapa contém os procedimentos metodológicos adotados, uma revisão da literatura contendo seções sobre a análise envoltória de dados, modelos com retornos constantes de escala, modelos com retornos variáveis de escala, eficiência e ecoeficiência, modelos estocásticos e determinísticos, avaliação do tipo de retorno de escala e modelo do *bootstrap* para construção de intervalos de confiança;
- IV. A quarta etapa contém os resultados da pesquisa, retirada dos *outliers*, teste do tipo de retorno de escala, uso do *bootstrap*, construção de intervalos de confiança através do *bootstrap* e análise da ecoeficiência;
- V. A quinta etapa conclui os principais pontos da pesquisa, bem como traz as conclusões mais relevantes.

Capítulo 2

Revisão Teórica

2.1 Análise bibliométrica Eco-eficiência e Agricultura

Diversos estudos foram publicados na área da DEA ao longo dos anos desde seu primeiro modelo proposto. Com diversas aplicações em temas, áreas, países e modelos diferentes para se analisar a eco-eficiência. Por meio da análise bibliométrica, ou revisão bibliométrica é possível avaliar como o modelo DEA evoluiu frente ao desenvolvimento da agropecuária, compreender quais são os principais países que publicam sobre o assunto e quais as principais tendências de estudos e temas.

2.1.1 Metodologia da Revisão

Os artigos se encontram no portal de periódicos da CAPES na plataforma do *Web of Science*, pois apesar de não conter o maior número de publicações de determinado assunto como na *Scopus* ou até mesmo na *Google Scholar*, possui uma base de periódicos de alta qualidade quando se trata das ciências exatas (ciências naturais), tornando essencial para a engenharia, além de registros mais antigos e conseqüentemente mais abrangentes [ELLEGAARD and WALLIN 2015].

A estrutura dos artigos que tratam do tema da presente pesquisa na Plataforma *Web of Science* contém as palavras "*Data*"AND "*envelopment*"AND "*analysis*"OR "*DEA*"AND "*eco-efficiency*" AND "*agriculture*". Campo de pesquisa de *Topic* (*considera o título, resumos e keywords plus*) e categoria de Engenharia multidisciplinar e Economia (EC). Período de 50 anos. Para engenharia multidisciplinar (EM), no entanto os artigos apresentaram resultados a partir do ano de 1992 até 2023, O que reflete em pouco mais de 20 anos de pesquisa. Ao total a pesquisa contém 445 artigos. A segunda categoria (EC) contém artigos a partir do ano de 1985 a 2023. Resultando em 2.873 artigos. Formato do banco de dados utilizado em *Bibtex* contendo todos os registros e referências citadas.

A ciência por trás das análises bibliométricas tem evoluído, isso se deve a avanços em

técnicas de análise, bem como a disposição de ferramentas computacionais que tem um potencial de geração de informação de maneira rápida com alta confiabilidade, se tornando ainda mais importante o uso da análise bibliométrica, principalmente quando se trata de uma grande quantidade de dados que precisam ser analisados, o que a torna uma análise quantitativa. Dessa forma o *RStudio* possui um pacote chamado *Bibliometrix* para análise dos artigos, geração dos gráficos e auxílio na compreensão dos resultados.

A análise contém uma divisão de cinco etapas. A primeira é a caracterização dos artigos que retornaram com a pesquisa por meio das palavras-chave utilizadas e os filtros da pesquisa, com detalhamento da quantidade de autores, fontes, referências citadas, taxa de publicação e quantidade de publicação. A segunda etapa tem a análise do impacto da citação locais, ou globais, assim como as principais referências e como tem evoluído. A terceira etapa contém a análise do País, a contribuição dos países para o aumento de publicações e quais países tem se destacado no campo de pesquisa. A quarta etapa contém a análise dos temas, com as palavras-chaves e assuntos de maior relevância. A quinta e última etapa é a discussão da revisão, contém os principais resultados da análise bibliométrica.

2.1.2 Característica dos Artigos

A primeira categoria (EM) possui 61 fontes, 11.063 referências citadas, cerca de 1.000 autores e uma taxa de crescimento anual de publicações de 8,35%. Descrito na Figura 2.1.

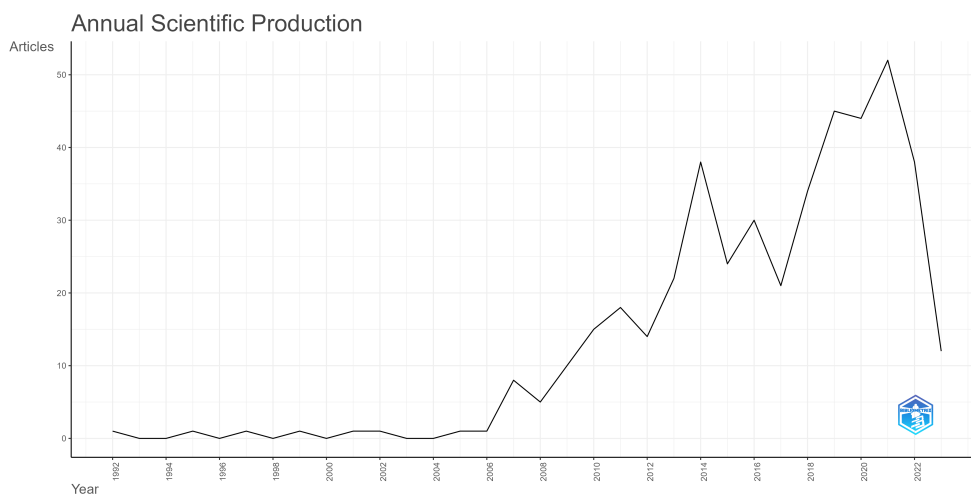


Figura 2.1: Artigos publicados em 1992 a 2023 na Categoria de Engenharia Multidisciplinar

A segunda categoria (EC) possui 301 fontes, 56.935 referências citadas, 4.117 autores e uma taxa de crescimento de 6,51%. Descrito na Figura 2.2)

Os dados da análise bibliométrica estão organizados na Tabela 2.1 para uma melhor visualização dos resultados.

O primeiro nível da análise bibliométrica da categoria de engenharia multidisciplinar apresenta participação de 13,41% dos artigos publicados, no entanto a maior parte são da

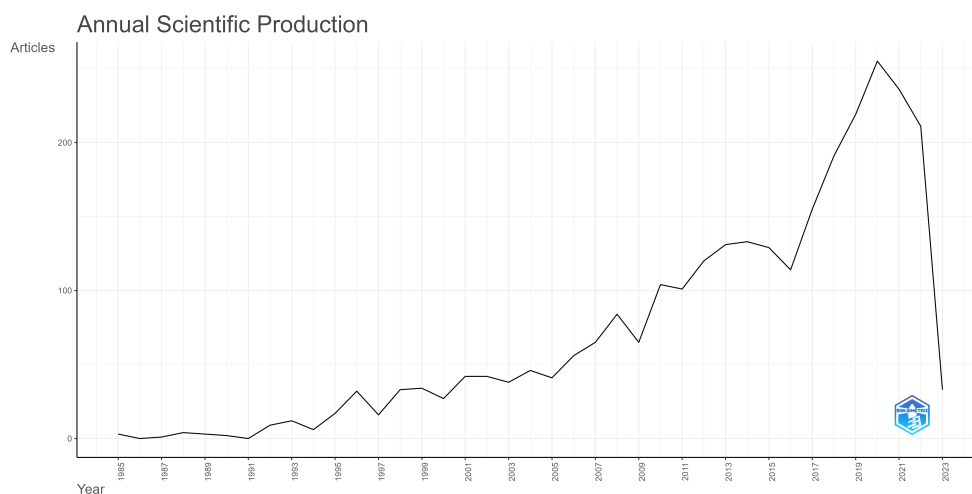


Figura 2.2: Artigos publicados em 1985 a 2023 ma Categoria de Economia

Nível	Categoria	Artigos	Referências	Autores	Taxa de crescimento	Período
Nível 1	Engenharia Multidisciplinar	445	11.063	1.000	8,35%	1992-2023
Nível 2	Economia	2.873	56.935	4.117	6,51%	1985-2023

Tabela 2.1: Análise Bibliométrica dos periódicos WoS

categoria econômica (86,59%), que corresponde ao segundo nível. Percebe-se que independente da categoria ambas possuem com o passar dos anos cada vez mais artigos publicados na área, o que é um indicativo do crescimento do modelo DEA e suas variações na agricultura de 7,39%. Apesar de apresentar um crescimento de publicações ao ano, há poucos artigos publicados por periódicos. (Figura 2.3)

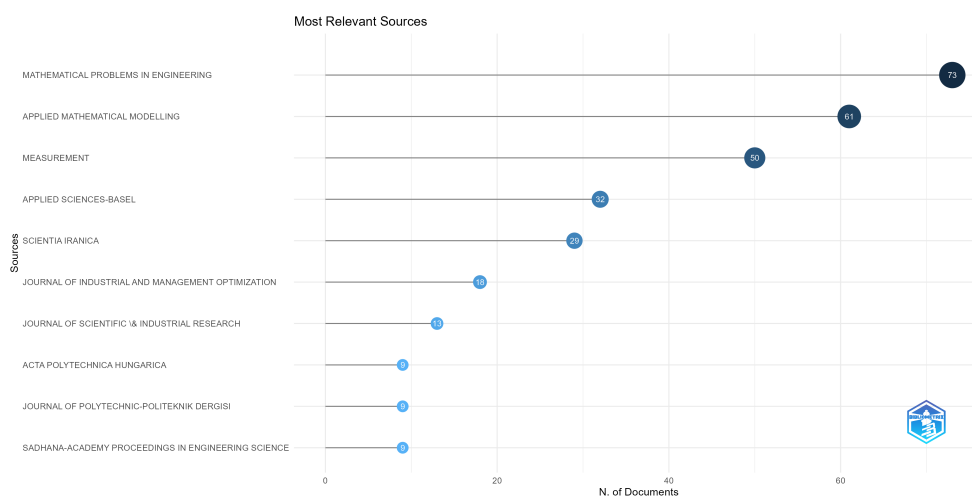


Figura 2.3: Artigos publicados em periódicos de 1992 a 2023

2.1.3 Impacto da Citação

Considerando as referências citadas localmente. A maior parte dos documentos referenciados são do período de 1993 a 2008, exceto apenas de um artigo de 1985, demonstrando

que a maior partes dos artigos utilizados como referência local possuem mais do que dez anos desda publicação. (Figura 2.4)

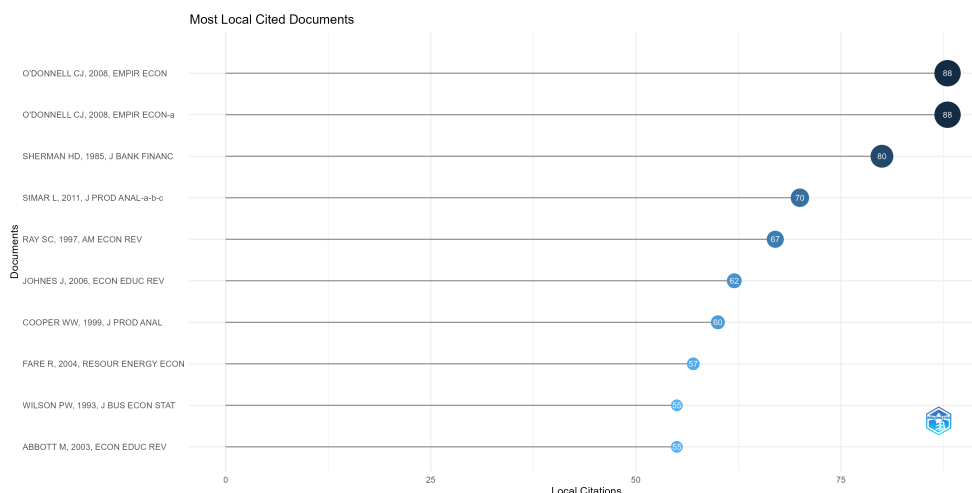


Figura 2.4: Artigos que citam artigos locais

Considerando as referências citadas local e globalmente. A maioria dos documentos referenciados são do período de 2002 a 2018, com pico em 2010, totalizando cerca de 6.297 citações de referências nesse período.

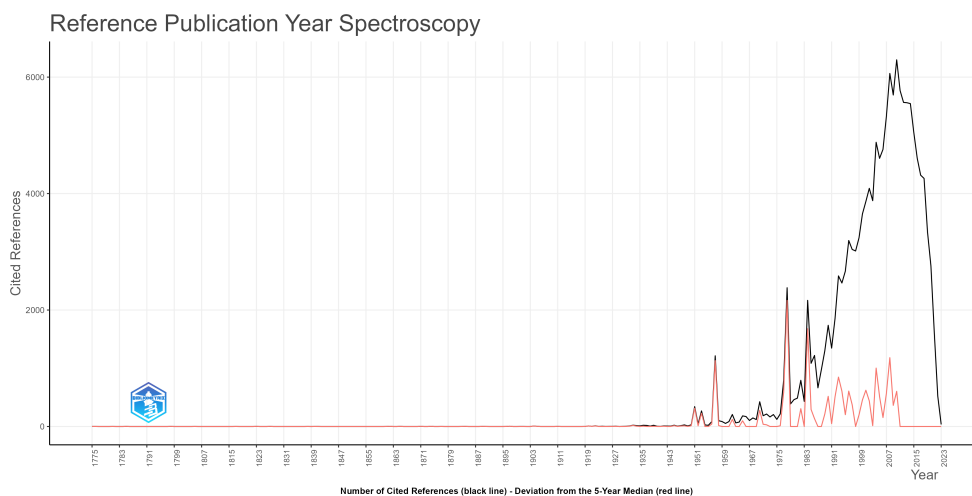


Figura 2.5: Artigos que citam artigos do período de 1775 a 2023

O diagrama de três campos apresenta a relação entre o impacto das principais referências globais, contribuições e palavras-chave utilizado nos artigos plotado na Figura 2.6. As principais referências são [Charnes et al. 1978], [Banker et al. 1984] e [Farrell 1957], dado que são os três principais precursores do modelo DEA. A maioria dos países que tiveram contribuições das principais referências e publicaram sobre o assunto são China, EUA, Espanha e Irã.

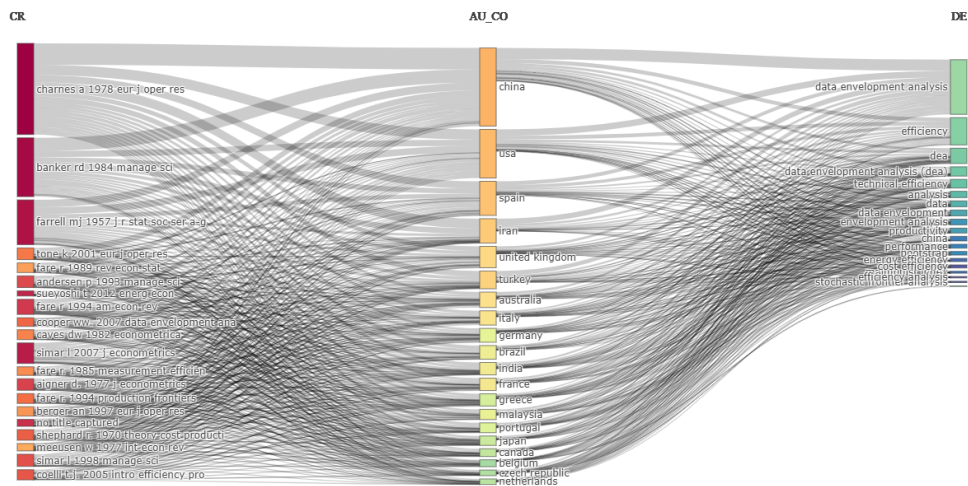


Figura 2.6: Diagrama de três campos

2.1.4 Análise do País

A contribuição dos principais países dos autores que mais publicaram artigos estão a China, EUA, Espanha, Irã, Turquia, Reino Unido, Austrália, Índia, Itália e Grécia. (Figura 2.7)

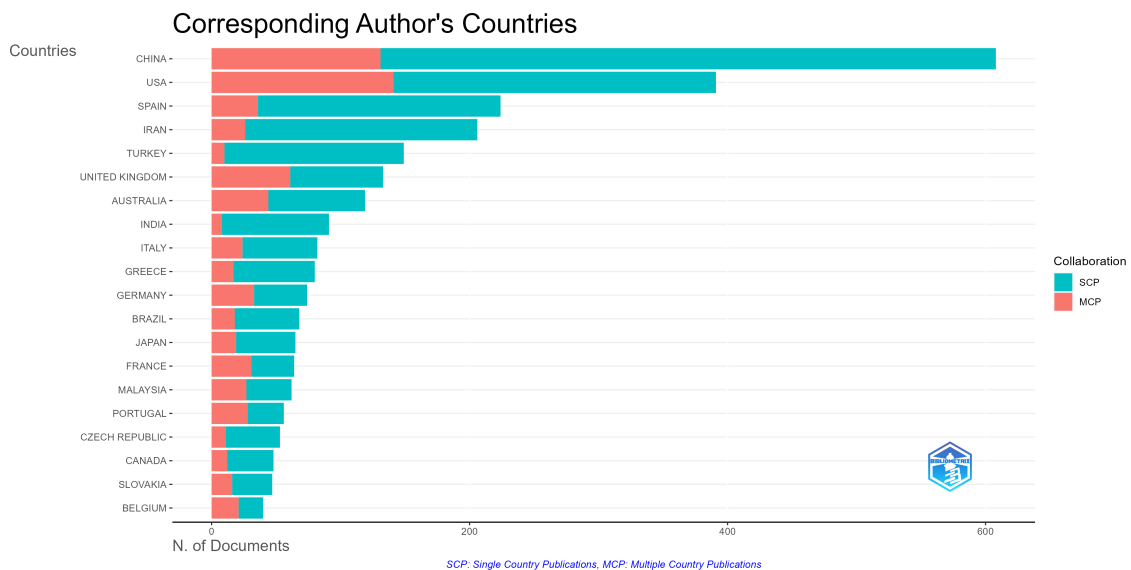


Figura 2.7: Países de autores do WoS

Na tabela 2.2 está a distribuição de frequência de produção científica, número de citações e média de citações dos artigos pelos países com maior produção científica. Verifica-se que nem sempre o País que contém o maior número de publicações terá o mesmo desempenho em relação às citações, esse é o caso da China e EUA, respectivamente com 2.124 e 1.084 produções científicas, mas com 15.620 e 19.306 números de citações de 25,70 e 49,40 em média de citações. Reino Unido e Austrália possuem uma frequência de produção respec-

tivamente de 415 e 330 produções, um valor inferior a Espanha, Irã e Turquia, com uma frequência de 679, 658 e 415 respectivamente, no entanto, possuem número de citações de 7.437 e 6.410, e média de citações de 55,90 e 53,90, enquanto a Espanha, Irã e Turquia possuem um número de citações de 5.992, 3.011 e 1.276, e média de citações de 26,80, 14,60 e 8,60. O que demonstra que os Países que mais publicam, nem sempre serão os mais citados, ou referenciados em trabalhos científicos. (Tabela 2.2)

País	Frequência	Número de citações	Média de citações
China	2124	15620	25,70
EUA	1084	19306	49,40
Espanha	679	5992	26,80
Irã	658	3011	14,60
Turquia	415	1276	8,60
Reino Unido	415	7437	55,90
Austrália	330	6410	53,90
Itália	262	1940	23,70
França	248	2304	36,00
Índia	247	1484	16,30

Tabela 2.2: Distribuição dos países dos periódicos WoS

Analisando-se os últimos anos é demonstrado o quanto que a China e o EUA tem sido de grande importância e relevância na proposição de novos estudos, bem como desenvolvimento de variações do modelo DEA aplicado a agricultura. O que torna a literatura desses dois países de grande importância para serem utilizados como referências para futuros estudos na área.

2.1.5 Análise dos Temas

Em ambas as categorias com assuntos envolvendo a DEA está presente, termos como eficiência, eficiência técnica, produtividade, *bootstrap*, análise de fronteira estocástica, ecoeficiência, saídas indesejáveis e agropecuária, temas abordados na presente pesquisa, verificasse dessa forma que ambas as categorias podem ser utilizadas para contribuir com a estrutura de periódicos base da pesquisa. (Tabela 2.3)

Uma ferramenta muito útil para dimensionar a ocorrência de temas tratados nos artigos é a nuvem de palavras. Figura 2.8

2.1.6 Discussão da Revisão

Portanto, temos a necessidade de novos estudos voltados para a realidade brasileira para se entender a aplicação da análise envoltório de dados no País (Brasil está fora dos principais países que publicam, ou são citados dentro do tema tratado, assim como não possui autores

Palavra-Chave (Assunto)	Ocorrência
data envelopment analysis	1166
efficiency	644
dea	370
data envelopment analysis (dea)	208
technical efficiency	203
analysis	140
data	133
data envelopment	125
productivity	125
envelopment analysis	118
bootstrap	85
performance	65
china	64
malmquist index	64
energy efficiency	62
stochastic frontier analysis	57
cost efficiency	45
efficiency analysis	44
performance evaluation	41
bank efficiency	40
eco-efficiency	39
undesirable outputs	39
benchmarking	38
total factor productivity	37
malmquist productivity index	34
agriculture	33
environmental efficiency	33
innovation	33
network dea	33
directional distance function	32

Tabela 2.3: As 30 palavras-chave mais frequentes da categoria EC e EM do WoS



Figura 2.8: Nuvem de palavras mais frequentes do WoS

relevantes), bem como o uso da ecoeficiência para analisar os municípios brasileiros, o qual a DEA demonstrou ser uma metodologia já consolidado em diversos países para análise de diversas áreas, incluindo a agricultura e apresentou crescimento de aplicações na área. Portanto, sendo de primordial importância para aplicação da pesquisa e análise da ecoeficiência dos municípios brasileiros.

Apesar de nem sempre o maior número de publicações acompanhar o maior número de citações, ficou evidente que quanto maior o número de publicações, há um número maior de citações, com uma menor proporção, em termos de ranqueamento e campanha com outros países. No caso específico da China (país com maior número de publicações) apareceu como uma das palavras-chave com maior ocorrência (64 ocorrências), isso demonstra que

muitos de seus artigos estão focados em análises internas, o que contribuiu para a menor proporção de citação dos artigos. [ELLEGAARD and WALLIN 2015] aponta que países estabelecidos economicamente tendem a ter o campo científico mais estruturado e, portanto, confirma o mesmo comportamento da análise feita, de que países emergentes tendem a ter uma proporção menor de citações.

2.2 Análise Bibliométrica *Bootstrap*

O *Bootstrap* pode ser utilizado para auxiliar na retirada de *outliers* de algum modelo, bem como a construção de intervalos de confiança e teste de significância, por meio de técnicas de re-amostragem. Sua aplicação com a DEA é recente, mas de grande importância considerando que as análises de eficiência não paramétricas serem muito sensíveis a qualquer tipo de ruído. Com a não retirada dos *outliers* do modelo, existe uma grande probabilidade de ter dados enviesados e resultados de pouca importância prática. Embora a análise de eficiência pelo modelo DEA clássico seja conhecido por dispensar o uso estatístico, muitos têm chegado a conclusão da importância do tratamento estatístico dos dados para aplicação do modelo DEA.

Esta seção tem o objetivo de analisar a relação entre a análise envoltória de dados e o *Bootstrap* aplicados a agropecuária, compreender como os estudos tem evoluído e principais temas tratados.

2.2.1 Metodologia Bibliométrica *Bootstrap*

Foi utilizado o portal periódicos CAPES para acessar a plataforma *Web of Science* para encontrar a base de artigos. Foram utilizados o filtro *Topic* (título, resumos e *keywords plus*) e as palavras "*Bootstrap*" AND "*data envelopment analysis*" OR "*DEA*" AND "*agriculture*". Para restringir um recente corte no tempo, foi utilizado o período de 5 anos (2018 a 2023), foi filtrado somente artigos, depois somente publicações com o idioma em inglês. Retornando 772 artigos. Todos baixados no formato Bibtex contendo todos os registros e referências citadas, para evitar qualquer perda de informação e obter resultados mais consistentes, já que a ferramenta a ser utilizada para analisar os dados tem boa aplicação quando o arquivo está em formato Bibtex. Dessa forma é possível obter um número considerável de artigos e estabelecer uma linha de tempo dos temas e aplicações do *Bootstrap* e a DEA.

Devido a grande quantidade de artigos foi utilizado o RStudio, através do pacote *Bibliometrix* para análise dos artigos, elaboração de gráficos e análise de *clusters*. O *Bibliometrix* é uma ferramenta aberta e abrangente, integrada ao *software R*, que permite uma análise de mapeamento científico complexo de maneira automatizada, isso porque cada vez mais a análise bibliométrica tem se tornado algo mais complexo pelo aumento de produções acadêmicas, portanto é crucial sintetizar pesquisas anteriores para se obter um embasamento

teórico consistente e avançar na linha de pesquisa proposta [Aria and Cuccurullo 2017].

2.2.2 Resultado Geral dos Artigos *Bootstrap*

Posteriormente aos dados terem sido baixados, foram carregados no Bibliometrix os dados em formato Bibtex. Na Figura 2.9, temos a tabela com o resumo dos dados disponíveis dos artigos baixados do WoS conforme o retorno da pesquisa e aplicação de filtros feita. A maioria dos dados está disponíveis em sua completude, enquanto resumo, DOI, palavras-chave, ano de publicação e *Keywords Plus* estão com alguns dados faltantes, no entanto, são dados que representam menos de 7% do total e assim não invalidam o resultado. Como o objetivo não é analisar categorias, a falta delas também não prejudicam a análise.

Metadata	Description	Missing Counts	Missing %	Status
C1	Affiliation	0	0.00	Excellent
AU	Author	0	0.00	Excellent
CR	Cited References	0	0.00	Excellent
RP	Corresponding Author	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
SO	Journal	0	0.00	Excellent
LA	Language	0	0.00	Excellent
NR	Number of Cited References	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
AB	Abstract	1	0.13	Good
DI	DOI	30	3.89	Good
DE	Keywords	30	3.89	Good
PY	Publication Year	43	5.57	Good
ID	Keywords Plus	51	6.61	Good
WC	Science Categories	772	100.00	Completely missing

Figura 2.9: Tabela resumo dos dados da WoS Bootstrap

Em resumo, os 772 artigos retorna 355 fontes, 1981 autores, uma média de 3,28 co-autores por documento, sendo que 31,35% é de co-autoria internacional e 29.925 referências. Foi feito então um Gráfico de três campos (Figura 2.10) para analisar as tendências dos principais autores, referências e palavras-chave. As principais referências são os percussores clássicos do modelo [Charnes et al. 1978] e [Banker et al. 1984], verifica-se que todos estão alinhados aos temas agropecuária, *bootstrap*, *bootstrapping*, sustentabilidade, produtividade e ecoeficiência. Um ponto relevante das palavras-chave no gráfico é a presença da China, o que demonstra a importância e a série de estudos publicados envolvendo o País.

Verificou-se uma queda na taxa de publicações durante o período de 2018 a 2023 em 15,16%, no entanto, foi considerado o ano de 2023, que ainda não foi finalizado, se olharmos de 2018 a 2022, percebe-se uma tendência de crescimento em publicações, onde o ano de 2021 é o pico de publicações. (Figura 2.11)

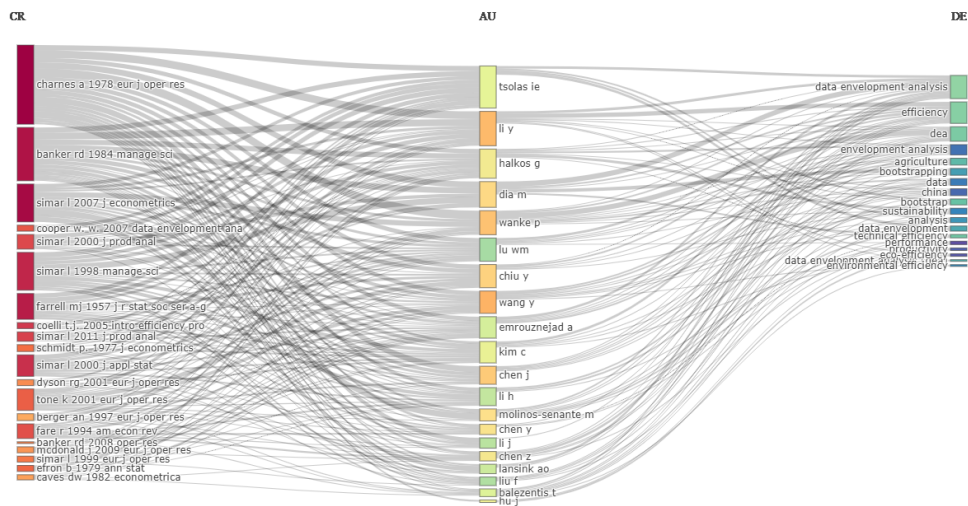


Figura 2.10: Gráfico de Três Campos Boostrap

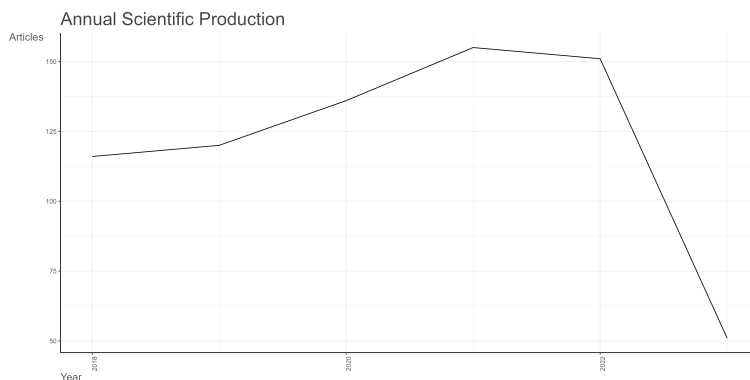


Figura 2.11: Gráfico de Produção Científica Anual Boostrap

2.2.3 Análise das Citações

Nessa seção foi analisado o comportamento das citações locais e globais dos artigos, para identificar as principais referências a serem utilizadas na pesquisa. Em primeiro momento verificou-se a média de citações de artigos a cada ano (Figura 2.12). Percebe-se a diminuição na média das citações, o que pode ser explicado pelo fato de produções mais recentes tendem a ser citadas menos pelo fato de serem publicadas em um menor espaço de tempo, que artigos com datas de publicações mais antigas. Logo, não é um comportamento atípico essa queda.

Analisaram-se as citações globais dos documentos (Figura 2.13), essa etapa identifica quais as referências são importantes para servir como base de novas pesquisas na área, geralmente conhecidos como referências clássicas. As principais referências globais são LI H, 2019, ECOL ECON, CECCHINI L, 2018, J CLEAN PROD e YANG L, 2018, J CLEAN PROD, respectivamente com 118, 117 e 110 citações globais.

Para dar continuidade ao levantamento das principais referências, dado que olhar somente para a quantidade de citações globais e locais podem levar a conclusões equivocadas dos

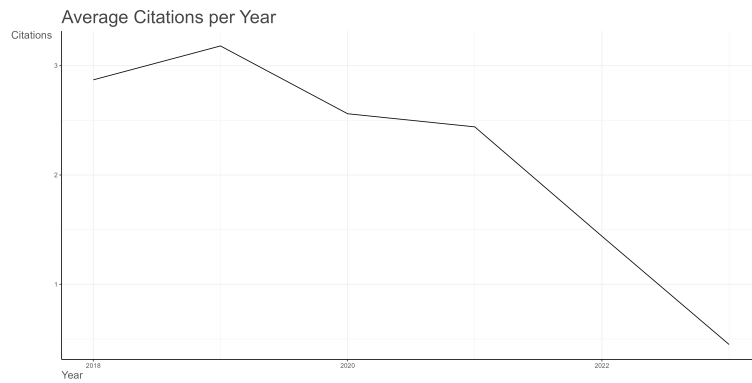


Figura 2.12: Gráfico de média de citações anuais bootstrap

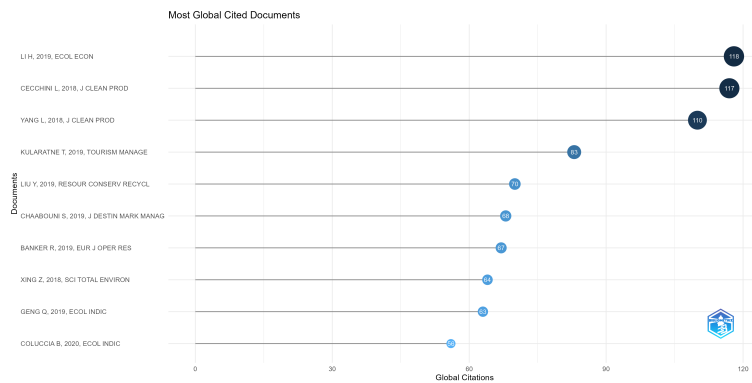


Figura 2.13: Gráfico de documentos globalmente citados bootstrap

principais autores sobre o tema, por tratar de apenas de quantidade de publicações, mas não levar em consideração a qualidade com que essas publicações possuem, foi analisado a relevância dos principais autores (Figura 2.14) e conjuntamente o impacto dos autores locais na Tabela 2.4.

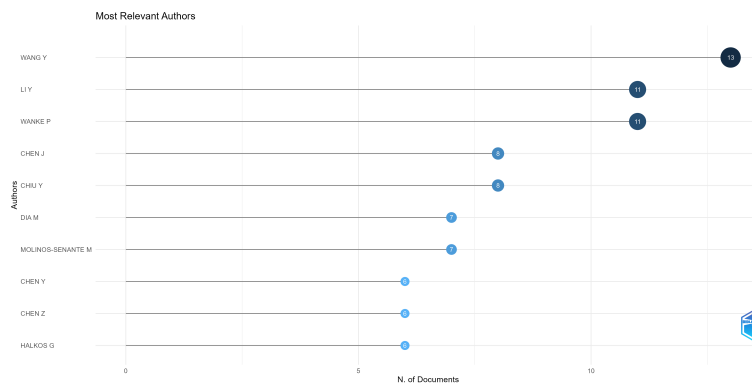


Figura 2.14: Gráfico de autores mais relevantes bootstrap

Se olharmos para relevância dos autores sobre o tema, temos que Wang Y., Li Y. e Wanke P. são os mais relevantes, respectivamente, no entanto, observando o Índice H o autor Wanke P. possui um maior impacto, embora Wang Y. e Li Y. estejam em um segundo grupo formado por quatro autores em segundo lugar na classificação, é Chen J., Wanke P., e Wang Y. que possuem mais citações, isso demonstra que os autores têm aprofundamento na área de

pesquisa, algo que Li Y. não indica ter, devido sua alta relevância, mas um baixo número de citações, o que pode apontar que o autor dedica os seus esforços para pesquisa científica em áreas diversificadas.

Autor	Índice H	Total de citações
WANKE P	6	130
CHEN J	5	167
CHEN Z	5	65
LI Y	5	63
WANG Y	5	118
BALEZENTIS T	4	78
CHEN X	4	50
CHIU Y	4	55
GOMEZ-VEGA M	4	74
HALKOS G	4	71

Tabela 2.4: Impacto dos autores locais

Dessa forma temos a representação gráfica da Lei de Loka (Figura 2.15), que contém a distribuição da quantidade de autores e as suas contribuições em publicação. Assim, podemos obter um parâmetro de produtividade desses autores, onde uma abundância de autores é responsável por um número inferior de publicações (localizados à esquerda do gráfico), enquanto de uma pequena parcela de autores é responsável por contribuir com uma quantidade maior de publicações (localizados à direita do gráfico), ou seja, autores mais improdutivos e produtivos, respectivamente. Obteve-se cerca de 83,1% dos autores são menos produtivos, contribuindo com apenas uma publicação, enquanto menos de 1% (7 autores) tiveram produções maiores que 6 publicações, portanto a maioria dos autores não são pesquisadores da área, o que ocasiona em pouco aprofundamento da maior parte das pesquisas.

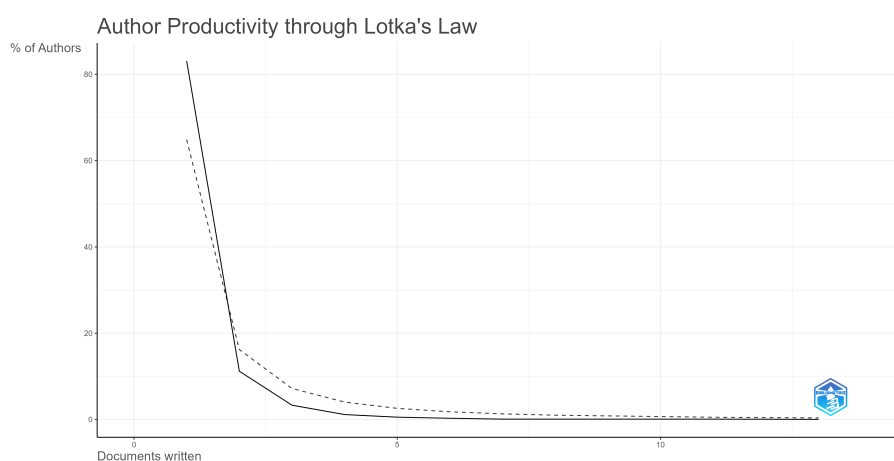


Figura 2.15: Gráfico de Lei de Loka

2.2.4 Análise do Assunto

Essa etapa tem o objetivo de entender quais os principais assuntos tratados nos artigos, bem como as evoluções temáticas e principais pontos abordados.

Foi feito um gráfico para entender os *clusters* das palavras-chave dos autores abordada nos artigos, o que pode ser visualizado na Figura 2.16. Há a presença de dois *clusters*, o primeiro representado pela cor azul formado pelas palavras-chave *data envelopment analysis*, *technical efficiency* e *bootstrap*, o segundo *clusters* representado pela cor verde com as palavras-chave *data envelopment analysis*, *efficiency* e *bootstrap* e o terceiro *clusters* representado pela cor vermelha formado pelas palavras-chave *efficiency*, *dea* e *bootstrap*, portanto vemos pela semelhança das palavras-chave entre os clusters e pela formação de apenas três *clusters*, que os assuntos não são abrangentes, mas extremamente focados em uma linha de pesquisa (DEA e bootstrap).

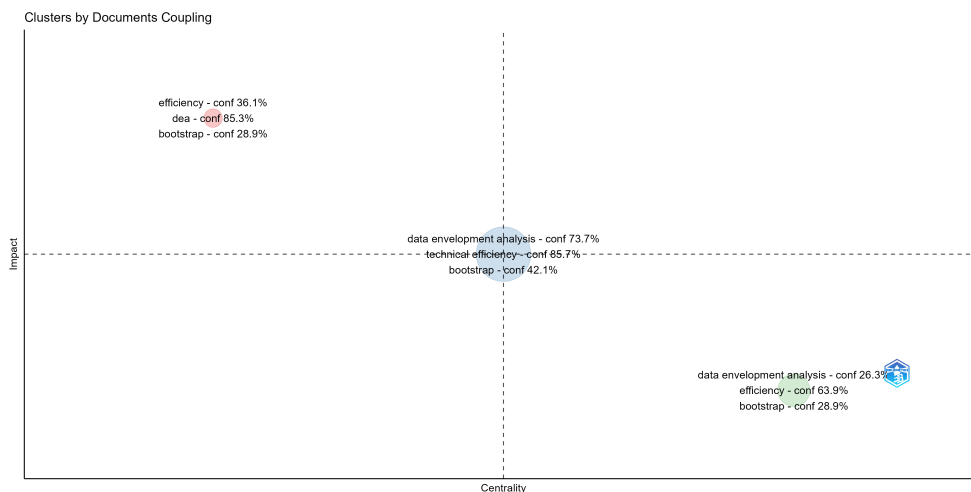


Figura 2.16: Gráfico cluters por acloperamento

Para permitir uma melhor visualização da proximidade dos temas tratados nos artigos foi elaborado a rede de *clusters* dos artigos, que apresenta um comportamento muito semelhante e próximo uns dos outros, principalmente o primeiro e o segundo *clusters* onde essa proximidade é identificada. (Figura 2.17)

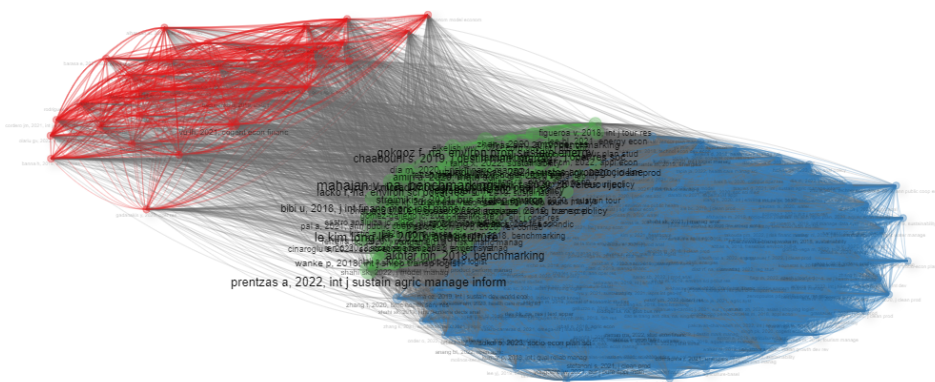


Figura 2.17: Rede de cluters por acloperamento

Avaliando a evolução temática a cada ano percebe-se a presença de uma série de termos envolvendo emissão e emissão de gases do efeito estufa, o que demonstra que muitas pesquisas tem o intuito de avaliar a eficiência ambiental, envolvendo o impacto das atividades

no meio ambiente e que dessa forma aponta uma tendência de estudos envolvendo a DEA e o *bootstrap* com aplicação na eficiência ambiental. (Figura 2.18)

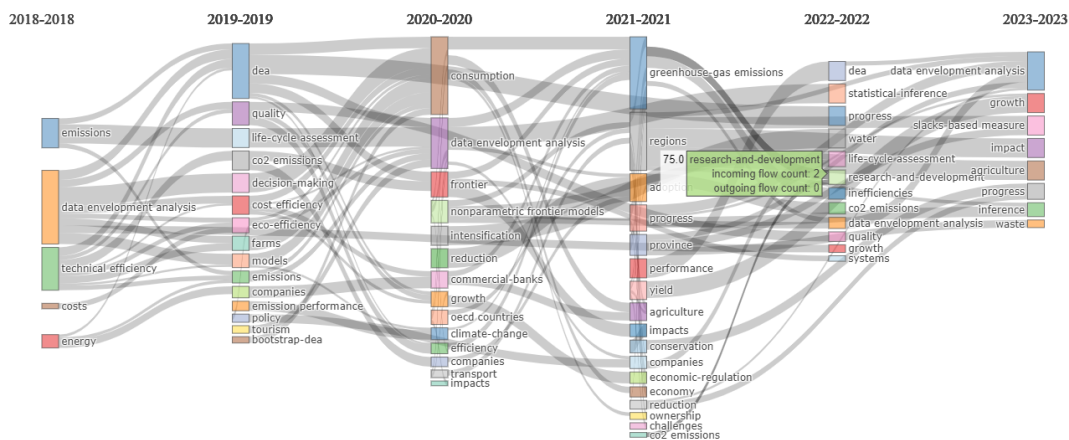


Figura 2.18: Evolução temática

Depois da visão temática global, foi tirado as fatias no tempo e analisado os *clusters* individualmente, procurou olhar temas tratados tirando as palavras-chave envolvendo DEA e eficiência para que pudéssemos ter uma visão de onde a DEA está sendo aplicada e qual medidas de eficiência estão sendo usadas, pois elas estão em todas as fatias e representam os maiores *clusters*. A primeira evolução ocorre no ano de 2018. Os principais temas abordados são energias, custos e desempenho financeiro. (Figura 2.19)

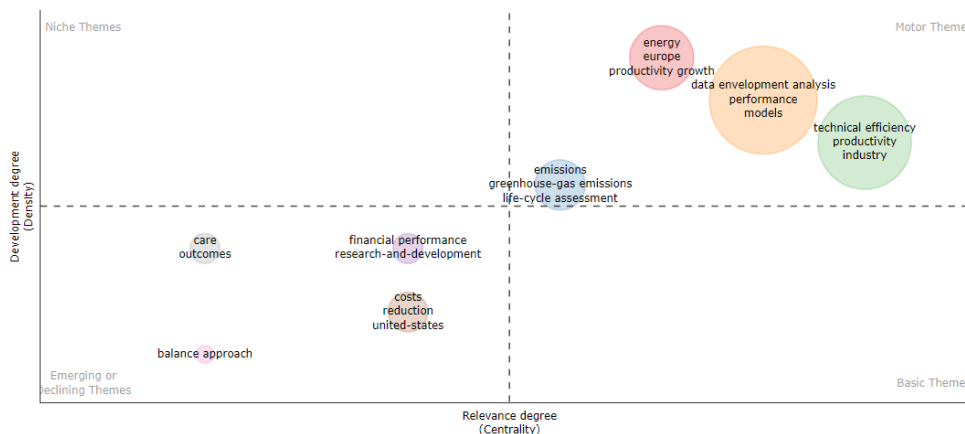


Figura 2.19: Evolução temática de 2018

A segunda fatia de tempo ocorre em 2019 e teve um crescimento de estudos envolvendo emissões, qualidade e eco-eficiência, enquanto eficiência de custos também manteve como tema de destaque no corte de tempo. (Figura 2.20)

A terceira fatia de tempo ocorre em 2020, com temas envolvendo fronteira, consumo, impacto e alterações climáticas como temas de destaques, mantendo assuntos envolvendo os impactos no meio ambiente em evidência. (Figura 2.21)

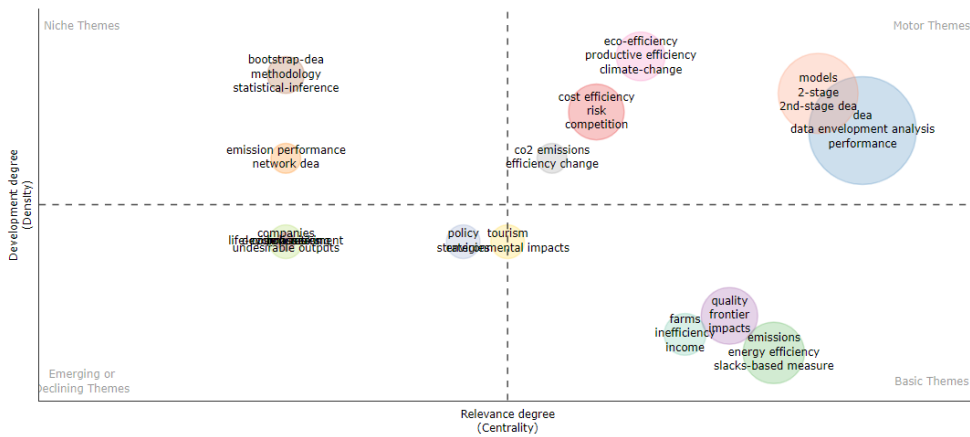


Figura 2.20: Evolução temática de 2019

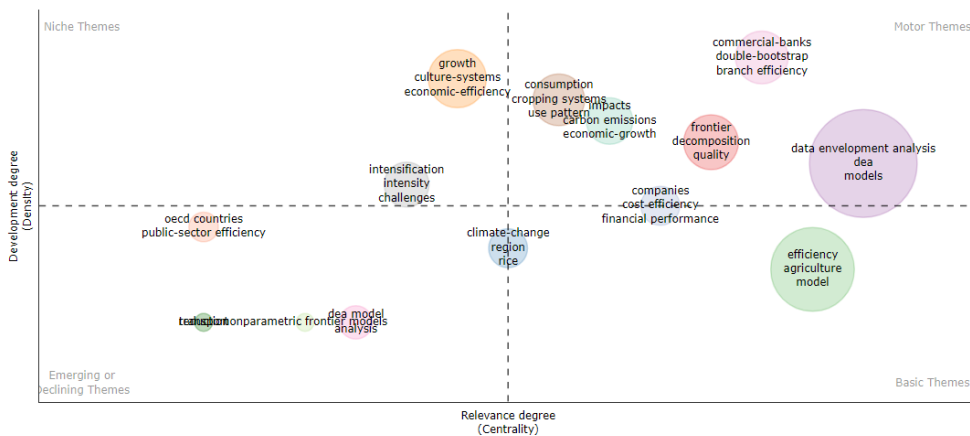


Figura 2.21: Evolução temática de 2020

A quarta fatia de tempo ocorre em 2021, apareceu pela primeira vez em destaque a agricultura e manteve em evidência assuntos envolvendo os impactos ao meio ambiente, com a presença de *clusters* sobre a emissão de gases de efeito estufa. (Figura 2.22)

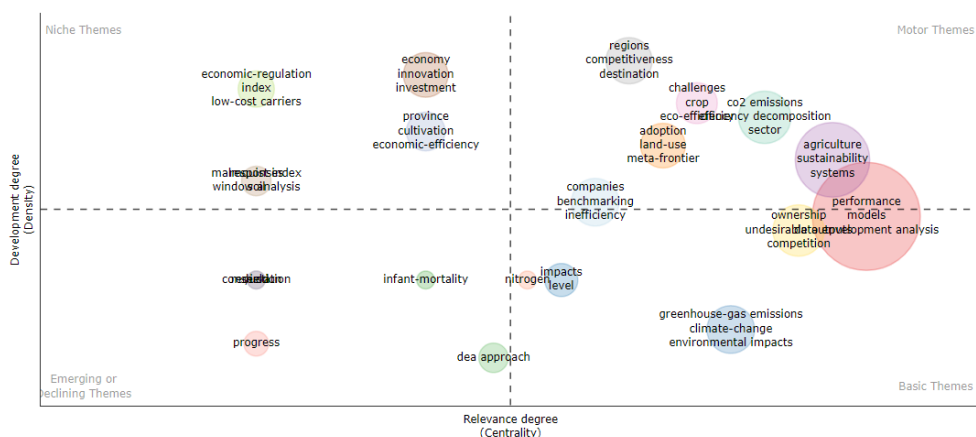


Figura 2.22: Evolução temática de 2021

A última fatia ocorre em 2022 fala sobre crescimento, qualidade e inferências estatísticas, o que pode se considerar um corte que apresentou temas diferentes dos cortes anteriores. (Figura 2.23)

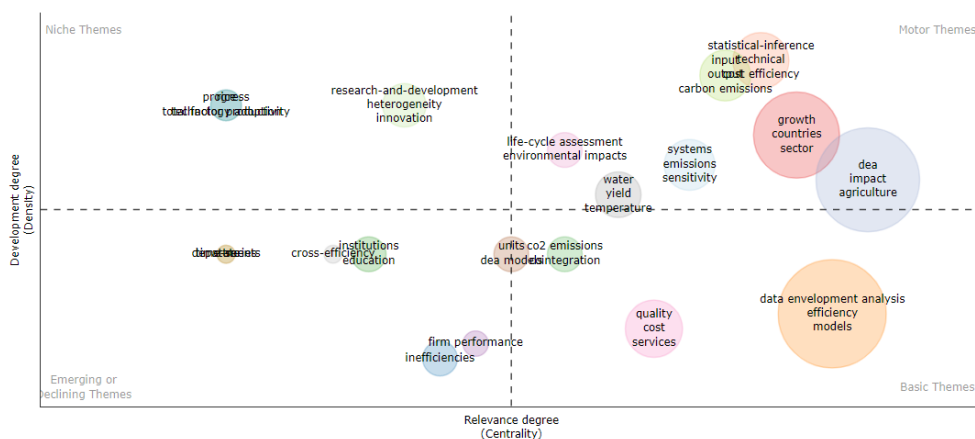


Figura 2.23: Evolução temática de 2022

Outra forma de avaliar a tendência é por meio do gráfico de tendência de tópicos. Ele traz os principais temas tratados em uma linha no tempo para compreender a evolução de cada assunto. Assuntos envolvendo desenvolvimento econômico, agropecuária, desempenho, impacto e geração de resíduos (desperdício) serão temas relevantes para pesquisas tratadas no futuro. (Figura 2.24)

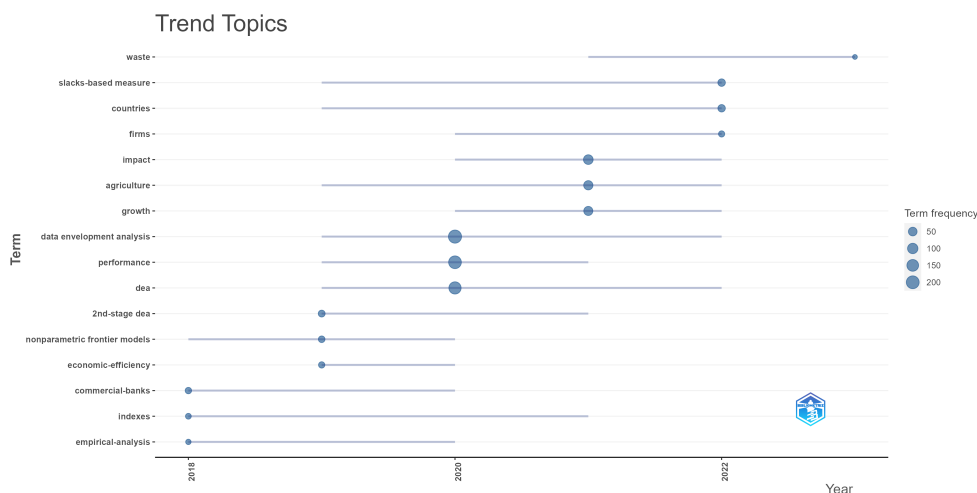


Figura 2.24: Tendência de tópicos

2.2.5 Análise dos Países e Afiliações

Até o momento foi analisado dados gerais dos artigos, citações e assunto dos temas. Com esses resultados já podemos retirar algumas informações dos países, na Tabela 2.4 a maioria dos autores são chineses (maior índice H e citações totais) e na Figura 2.10 a China é o único país que aparece no Gráfico de três campos como tendência entre os principais autores e referências das palavras-chave utilizadas, o que demonstra a importância do país no campo de pesquisa analisado e que tem conotado de maneira significativo no avanço da ciência.

Para obter uma análise focada nos países, essa etapa visa agregar na compreensão, que

até o momento já se teve, das principais influências entre os países e as afiliações. Dessa forma foi traçado o Gráfico de produção dos países ao longo do tempo (Figura 2.25) para verificar como tem sido a evolução das publicações ao longo desses 5 anos, e resultou que a China, assim como já foi compreendido que é um país relevante para se olhar, foi o país que teve o maior crescimento de produção no último anos, tanto que no último ano a China possui um valor de produção cerca de três vezes maior que o segundo colocado (Espanha) e cerca de cinco vezes maior que o terceiro colocado (EUA), o que demonstra não somente a participação de grande parte das publicações, assim como uma importância do país no campo de pesquisa.

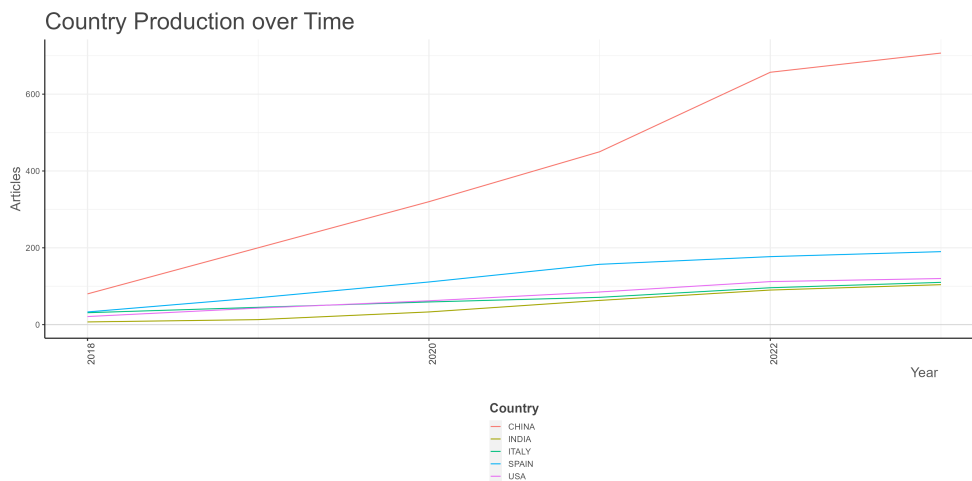


Figura 2.25: Gráfico de produção dos países ao longo do tempo

Se avaliarmos os países do autor correspondente (nacionalidade) também retorna como a China em primeiro lugar, seguida muito atrás pela Espanha, apenas com uma alteração da Índia assumindo o terceiro lugar enquanto os EUA ocupam a quarta colocação (Figura 2.26). A China apresenta tanto um alto valor de MCP apontando uma posição mais periférica em relação às publicações, o que se confirma pelo alto valor SCP, onde mostra a tendência dos autores para a autocitação, o que corresponde a um aprofundamento e maior avanço dos estudos chineses da DEA usando o *bootstrap*.

No Gráfico dos países mais citados (Figura 2.27) a China também ocupa a primeira colocação, o que reafirma a medida de SCP da Figura 2.26 do comportamento de autocitação e citação de outros países. A China possui cerca de 2218 citações enquanto o segundo colocado a Espanha possui 533 citações, seguido pela Itália com 385 citações, Reino Unido com 357 citações e EUA com 304 citações. A diferença da China é grande o suficiente para se somarmos as citações desses quatro países melhor colocados depois da China possuem um total de 1579 citações, o que ainda é menos comparado ao primeiro colocado.

Quando foi analisado as afiliações, apenas a China apresentou mais de uma universidade entre as afiliações mais relevantes, o que dá a ela um desempenho melhor do outros países nesse quesito, em que cada um apresentou apenas uma universidade nacional no *ranking*, no entanto, é de se surpreender que a China tenha apenas duas universidades no *ranking*, dado

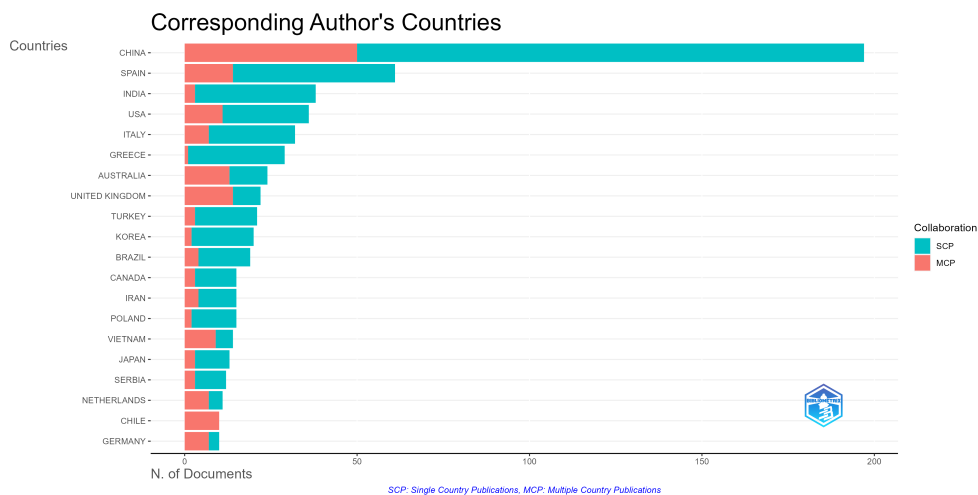


Figura 2.26: Gráfico dos países do autor correspondente

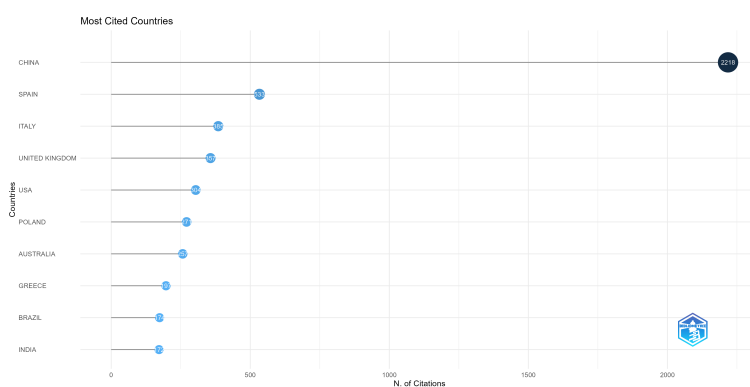


Figura 2.27: Gráfico dos países mais citados

que ela demonstrou supremacia em todas as análises até então feitas. (Figura 2.28)

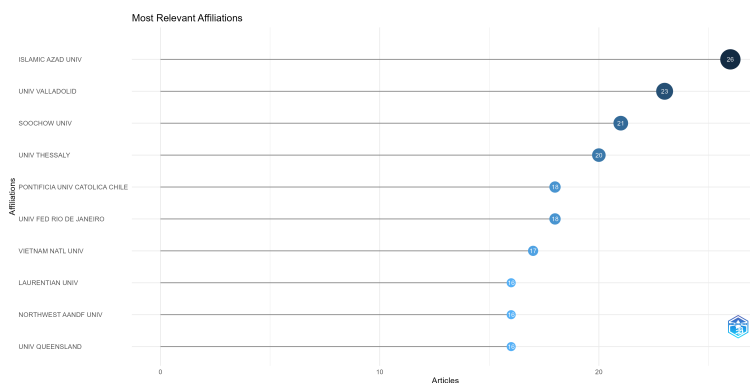


Figura 2.28: Gráfico das afiliações mais relevantes

2.2.6 Discussão da Revisão Bootstrap

A maioria das citações são de autoria chineses, bem como nas análises de países e afiliações a China também apresentou ser a maior referência para a área em estudo, responsável por grande parte das citações, referências, publicações e relevância. O que torna o país como

uma base indispensável para servir como referência para novos estudos sobre a DEA com uso do *bootstrap* aplicados a agricultura.

Há um comportamento de crescimento de assuntos voltados para o desenvolvimento sustentável, diminuição de desperdícios, redução da emissão de gases de efeito estufa e a ecoeficiência. O que pode ter uma relação não somente com o cenário mundial de preocupação ambiental devido a uma série de catástrofes que aconteceram nos últimos dez anos, com maior intensidade e frequência, mas também existe uma influência da maior parte dos estudos serem chineses, pois estes são um dos grandes responsáveis pelo efeito estufa no planeta, e eles tem procurado reduzir o impacto de suas atividades, promovendo novos estudos e soluções para o enfrentamento de sanções e pressões devido o desgaste que o meio ambiente tem sofrido, portanto, verificou-se que a China não tem somente procurado soluções para resolver a questão ambiental, como tem usado a DEA e o *bootstrap* para analisar a eficiência de suas atividades internas do país levando em consideração o impacto causado no meio ambiente.

A presente pesquisa se limitou as pesquisas no espaço-tempo de 2018 a 2023, seria interessante em outras análises bibliométricas ter um olhar em um espaço-tempo mais abrangente para verificar se a China sempre foi dominante e em que momento em específico ela se tornou referência, pois até o momento podemos apenas supor que seja pelo acelerado crescimento industrial que o país teve na última década, afinal para um desenvolvimento econômico existe uma relação com a degradação ambiental de um país.

Portanto, grande parte dos artigos sobre o a DEA e o *bootstrap* apresentam pouca profundidade ainda nos estudos (exceto estudos da China), portanto se torna essencial seu aprofundamento, estudo e novas aplicações, principalmente no Brasil, que não apareceu em destaque em nenhum momento das análises, sendo que o Brasil é um dos países que mais exportam produtos da economia agropecuária, portanto se faz necessário análises aprofundadas no tema, com o uso de metodologias já consolidadas (DEA) para contribuir com um maior conhecimento das principais atividades internas da Nação brasileira para tornar a principal força econômica do País mais produtiva e principalmente dado as novas demandas mundias torná-la ecoeficiente.

Capítulo 3

Procedimentos Metodológicos

3.1 Modelo da Análise Envoltória de Dados - DEA

Baseado em [Farrell 1957] que [Charnes et al. 1978] vai propor o primeiro modelo da Análise Envoltória de Dados (*Data Envelopment Analysis*, DEA), ou Teoria da Fronteira (*Frontier Analysis*), pois tinha a necessidade de propor um novo modelo de análise da produtividade, pois em suma grande parte dos métodos que mensuravam medidas meticulosas falham na sua pouca aplicação a depender do tipo de relação entre as variáveis (múltiplas entradas e saídas) muitas vezes complexas para se compreender.[Bogetoft and Otto 2011] Na literatura há uma conformidade que as análises de micro processo não seja adequada, isso porque as entradas e saídas são normalmente muito bem definidas.[Cook and Zhu 2013]

A DEA utiliza uma abordagem não paramétrica e não restritiva para sistemas de produção modulares e fornece uma avaliação eficiente e equitativa do desempenho, baseada na microeconomia, permite comparar unidades de produção com diferentes estruturas, recursos e resultados, dessa forma identifica as melhores práticas de desempenho presentes nas organizações. É baseada em modelos matemáticos de programação linear e não exige inferências estatísticas, tendências centrais, testes de coeficientes ou regressões. [Cooper et al. 2011]

A DEA é uma técnica de análise de desempenho orientada a dados para avaliar a eficiência de unidades de produção (chamadas de *Decision-Making Unit*, DMU), com a presença de múltiplas entradas convertidas em múltiplas saídas, como empresas, unidade militares, bancos, escolas, universidades, unidades produtivas, hospitais, estados, países, departamentos, regiões, entre outros baseados na teoria microeconômica. Permite comparar o desempenho de cada unidade com as outras unidades em um conjunto de entradas (recursos utilizados, *inputs*) e saídas (resultados produzidos, *outputs*) para determinar a eficiência relativa de cada unidade. [Cooper et al. 2011]

O objetivo é identificar as unidades mais eficientes, usando seus insumos para produzir os mesmos produtos, ao mesmo tempo, em que maximiza sua eficiência. É útil para avaliar a eficiência técnica comparativa, identificando pontos fortes e fracos, e fornecer informações para

melhorias de desempenho de cada DMU por meio do *Benchmarking*. [Cook and Zhu 2013]

É uma metodologia voltada para a fronteira, diferente dos métodos de regressão que utiliza da tendência dos dados, o que permite que a metodologia descubra novos relacionamentos entre os dados de entrada e a saída, que em outras metodologias não fica explícito. De uma maneira direta, sem exigir nenhum tipo de informação, ou suposição formuladas, não necessita adicionar pesos de importância (prioridade) para qualquer variável para propor resultados consistentes. [Cooper et al. 2011]

No entanto, é importante comparar unidades produtivas que exerçam as mesmas funções, produzindo os mesmos produtos ou serviços a partir dos mesmos insumos. Isso garante que a comparação seja justa e significativa.

3.2 Eficiência e produtividade

A eficiência de acordo como uma definição estendida da Lei de Pareto proposto por [Koopmans 1951] é alcançada quando a DMU possui quantidades de insumos e produtos que não podem ser melhorados sem piorar simultaneamente o outro, assim como quando baseada no desempenho de outras DMUs não há melhorias que podem ser feitas sem afetar de maneira negativa os insumos e produtos.

Dessa forma [Farrell 1957] aproximou o conceito de produtividade ao conceito geral de eficiência, que busca encontrar a melhor distribuição de recursos entre as atividades, utilizando os recursos de maneira ótima (sem folgas), reduzindo ao máximo as entradas na unidade produtiva, alinhado com um resultado ótimo (sem escassez), aumentando ao máximo as saídas das DMUs. De maneira simplificada, considerando um processo de produção com apenas um produto. Essa relação pode ser dada por uma razão, onde Q é o produto, X o insumo e P a produtividade.

$$P = \frac{Q}{X} \quad (3.1)$$

E é a partir dos estudos de [Farrell 1957], que posteriormente irão surgir os conceitos de eficiência técnica, alocativa e total (global) como conhecemos hoje, mas já podemos explicá-las baseadas nas ideias iniciais do autor:

- A eficiência técnica leva em consideração do que foi produzido a partir dos insumos disponíveis, com o que poderia ter sido produzido com os mesmos recursos;
- A eficiência alocativa fora chamada de “eficiência de preço”, diferente da eficiência técnica leva em consideração os aspectos custos e preços;
- A eficiência global seria o produto da eficiência técnica e a eficiência alocativa.

Para tornar uma DMU eficiente existem dois tipos de orientação, que podem ser adotados respectivamente, orientação, insumo, ou produto no modelo CCR, temos a Figura 1, que

representa um conjunto de DMUs dentro da fronteira de possibilidades de produção. Sendo y = produto e x = insumo. Onde a eficiência técnica orientada ao insumo pode ser representada pelo segmento de reta $\frac{(yB''B'')}{(yB''B)}$ (mesmo nível de produtividade e redução do insumo). A eficiência técnica orientada ao produto é dada pelo segmento de reta $\frac{(xB'B')}{(xB'B)}$ (mesmo nível de insumo e aumento da produção).

Essa distância do ponto da DMU até a fronteira é baseada no modelo de distância proposto por [Shephard 1970].

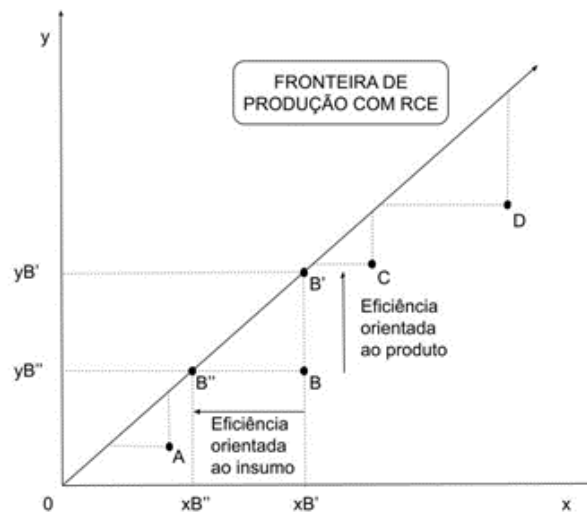


Figura 3.1: Fronteira de possibilidades de produção modelo CCR/RCE

Logo, para calcular as eficiências a partir da Equação 3.1, temos que [FERREIRA and GOMES 2020] vai descrever as etapas da seguinte forma, adicionando pesos às variáveis, sendo W_j ($j = 1, 2, 3, \dots, m$) o peso relacionado ao produto Q_j e V_i ($i = 1, 2, 3, \dots, n$) o peso relacionado ao insumo X_i , e considerando uma DMU k .

Então temos:

$$\theta = \frac{W_j Q_{jk}}{V_i X_{ik}} \quad (3.2)$$

Considerando que existem um, ou mais produtos e insumos, a Equação 3.2 pode ser rescrita.

Então temos:

$$\theta = \frac{\sum_{j=1}^m W_j Q_{jk}}{\sum_{i=1}^n V_i X_{ik}} = \frac{\sum_{j=1}^m (W_1 Q_{1k} + W_2 Q_{2k} + \dots + W_3 Q_{3k})}{\sum_{i=1}^n (V_1 X_{1k} + V_2 X_{2k} + \dots + V_3 X_{3k})} \quad (3.3)$$

Transformando em um problema de programação, baseada em modelo matemático de programação fracionária e considerando r DMUs em análise, temo:

$$\max \theta = \frac{\sum_{j=1}^m W_j Q_{jk}}{\sum_{i=1}^n V_i X_{ik}} = \frac{\sum_{j=1}^m (W_1 Q_{1k} + W_2 Q_{2k} + \dots + W_m Q_{mk})}{\sum_{i=1}^n (V_1 X_{1k} + V_2 X_{2k} + \dots + V_n X_{nk})}$$

sujeito a:

$$\frac{\sum_{j=1}^m W_j Q_{jr}}{\sum_{i=1}^n V_i X_{ir}} = \frac{\sum_{j=1}^m (W_1 Q_{1r} + W_2 Q_{2r} + \dots + W_m Q_{mr})}{\sum_{i=1}^n (V_1 X_{1r} + V_2 X_{2r} + \dots + V_n X_{nr})} \leq 1, \quad \forall r \quad (3.4)$$

$$W_j, V_i \geq 0, \forall j, i$$

Existe um problema com a programação fracionário no caso de calcular a eficiência ótima, pois ela pode considerar um conjunto de soluções ótimas infinitas, pelo fato da equação está explícita em proporções, então é necessário transformar a Equação 3.4 em um problema de programação linear

Para linearizar o modelo podemos igualar a soma do insumo a 1. Logo temos:

$$\max \theta = \sum_{j=1}^m W_j Q_{jk} = \sum_{j=1}^m (W_1 Q_{1k} + W_2 Q_{2k} + \dots + W_m Q_{mk})$$

sujeito a:

$$\sum_{i=1}^n V_i X_{ir} = \sum_{i=1}^n (V_1 X_{1r} + V_2 X_{2r} + \dots + V_n X_{nr}) = 1, \quad \forall r \quad (3.5)$$

$$\sum_{j=1}^m W_j Q_{jr} - \sum_{i=1}^n V_i X_{ir} = \sum_{j=1}^m (W_1 Q_{1r} + W_2 Q_{2r} + \dots + W_m Q_{mr})$$

$$- \sum_{i=1}^n (V_1 X_{1r} + V_2 X_{2r} + \dots + V_n X_{nr}) \leq 0, \quad \forall r$$

$$W_j, V_i \geq 0, \quad \forall j, i$$

3.3 Modelo Constante de Escala - CCR/RCE

O modelo CCR, que trata sobre retornos constantes de escala, foi baseado por [Farrell 1957] e proposto pela primeira vez por Charnes, Cooper, and Rhodes (1978) é uma versão do DEA que avalia o desempenho de uma unidade de produção comparando sua taxa de utilização dos recursos (insumos) com a taxa máxima permitida (frontalmente restrita) pelo modelo radial. Juntando informações e conhecimentos tanto da economia quanto da

engenharia.

A partir da Equação 3.5 chegamos no modelo de multiplicadores CCR, orientado ao insumo:

$$\max \theta = \sum_{j=1}^m W_j Q_{jk} \quad (3.6)$$

sujeito a:

$$\sum_{i=1}^n V_i X_{ik} = 1, \forall k, \quad \sum_{j=1}^m W_j Q_{jr} - \sum_{i=1}^n V_i X_{ir} \leq 0, \forall r, W_j, V_i \geq 0, \forall j, i$$

Onde $X = [X_1, X_2, X_3, \dots, X_r]$ e $Q = [Q_1, Q_2, Q_3, \dots, Q_r]$ representam as matrizes dos inputs e *outputs*, respectivamente de n observações, $V_i = (V_1, V_2, V_3, \dots, V_n)$ e $W_j = (W_1, W_2, W_3, \dots, W_m)$ representam os pesos relacionadas aos inputs e *outputs*, respectivamente de n observações e o θ índice de eficiência orientado ao insumo.

É possível definir a eficiência orientada ao produto pelo modelo dos multiplicadores, que almeja verificar qual a quantidade máxima de produto dado um constante valor de insumo. Embora a decisão de qual orientação a ser tomada não altere algo significativo no valor da eficiência, é importante decidir a orientação por decisões prévias com intuito de reduzir o insumo, ou aumentar a produção, isso depende do caso tratado, podendo existir restrições em um dos casos, obrigando utilizar o outro tipo de orientação, assim como pode ser uma ação tomada pelo grupo em análise. Apesar de ter um funcionamento oposto ao modelo de orientação ao insumo Equação 3.6, existem semelhanças entre os modelos, podendo ser expresso pela equação da seguinte forma:

$$\min \varphi = \sum_{i=1}^n V_i X_{ik} \quad (3.7)$$

sujeito a:

$$\sum_{j=1}^m W_j Q_{jk} = 1, \forall k, \quad \sum_{j=1}^m W_j Q_{jr} - \sum_{i=1}^n V_i X_{ir} \leq 0, \forall r, W_j, V_i \geq 0, \forall j, i$$

Onde φ é o índice de eficiência orientado ao produto.

O modelo demonstrado pela Equação 3.7 possui o modelo dual, conhecido como envoltória. Diferente do modelo do multiplicador onde as variáveis de decisão são os pesos atribuídos aos insumos e produtos (W_j e V_i), no modelo dual temos θ e os pesos λ_r como variável de decisão.

Proposto por [Charnes et al. 1978], descrito-se o modelo dual da eficiência orientada ao insumo.

$$\min \theta \quad (3.8)$$

sujeito a:

$$\sum_{r=1}^p \lambda_r X_{ir} \leq \theta x_{io}, \forall i, \sum_{r=1}^p \lambda_r Q_{mr} \geq qm_o, \forall m, \lambda_r \geq 0, \forall r$$

Onde $X = [X_1, X_2, X_3, \dots, X_r]$ e $Q = [Q_1, Q_2, Q_3, \dots, Q_r]$ representam as matrizes dos inputs e *outputs*, respectivamente de n observações, $x_i = (x_1, x_2, x_3, \dots, x_r)$ e $q_j = (q_1, q_2, q_3, \dots, q_r)$ representam os vetores relacionados às entradas e saídas, respectivamente da DMU _{r} , trânsito, e θ é o índice de eficiência orientado ao insumo.

No modelo envoltório orientado ao produto do CCR também não difere muito do modelo orientado ao insumo Equação 3.8, onde temos que:

$$\max \varphi \quad (3.9)$$

sujeito a:

$$\sum_{i=1}^r \lambda_r X_{ir} \leq x_{io}, \forall i, \sum_{r=1}^p \lambda_r Q_{mr} \geq \varphi qm_o, \forall m, \lambda_r \geq 0, \forall r, i$$

Nesse caso a eficiência técnica vai ser representada pela letra φ onde tanto no modelo envoltório quanto no modelo de multiplicador temos que $\theta = 1/\varphi$.

A eficiência orientada ao insumo e ao produto possuem valores entre 0 e 1, onde temos três classificações possíveis, $\theta < 1$, ou $\varphi < 1$ a eficiência da DMU referente é ineficiente, portanto, aponta que a DMU está fora de uma estrutura razoável, e deve ser tomadas ações para melhoria, $\theta < 1$, ou $\varphi < 1$ e folga maior que zero em algum insumo ou produto, a eficiência da DMU referente é fracamente eficiente, portanto, deve haver melhorias em relação aos insumos, ou produtos para serem feitas na unidade, $\theta = 1$, ou $\varphi = 1$ e folga igual a zero, a eficiência da DMU referente é eficiente, portanto, está sobre o ponto ótimo de utilização dos insumos, ou produtos das DMUs. [Fan et al. 2015]

Um ponto importante para aplicação desse modelo é fato de ser intrínseco entender a natureza do problema porque grande parte das indústrias (setor privado) não são avaliadas pelo modelo CCR, pelo fato dos retornos constantes de escala não serem a natureza do comportamento relacionado ao insumo e produto dessas indústrias. [Yang et al. 2015]

Fazendo assim a necessidade de um modelo que não considera retornos constantes de escala.

3.4 Modelo Variável de Escala - BCC/RVE

Após apresentamos na seção anterior o primeiro modelo clássico DEA-CCR proposto [Charnes et al. 1978], surgiu o modelo BCC, ou RVE proposto por [Banker et al. 1984], consolidando um segundo modelo clássico da DEA-BCC. A principal diferença entre os modelos está na natureza do comportamento das entradas e saídas. O modelo considera rendimento crescente (com taxa de crescimento do produto (y) maior que a taxa de crescimento do insumo (x)); e decrescentes de escala (com taxa de crescimento do produto (y) menor que a taxa de crescimento do insumo (x)).

Para compressão das principais diferenças entre o modelo BCC e o CCR, podemos visualizar na Figura 3.2, a fronteira de possibilidades de produção baseada no modelo BCC, ou RVE.

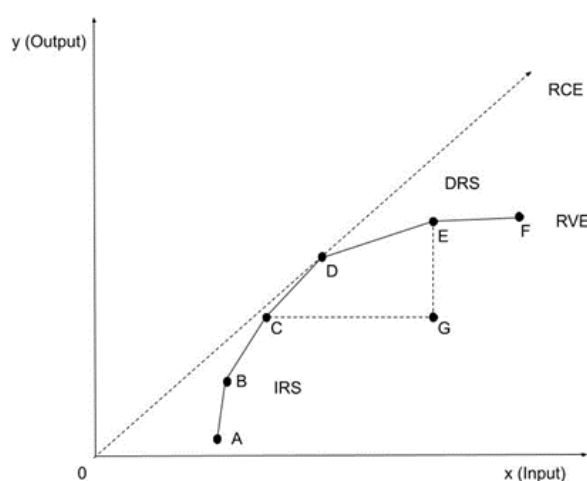


Figura 3.2: Fronteira de possibilidades de produção modelo BCC/RVE

Na figura acima, ABCD (exceto o ponto D) apresentam um comportamento de rendimento de escala crescente, onde os insumos têm uma razão de entrada superior aos produtos. DEF (exceto o ponto D) apresentam um comportamento de rendimento de escala decrescente, onde os insumos possuem uma razão de entrada inferior aos produtos.

Em geral, os valores de eficiência técnica no modelo CCR são inferiores aos valores do modelo BCC. Na Figura 3.2 as DMUs A, B, C, D, E e F são consideradas eficientes pelo modelo BCC, no entanto, apenas a DMU D é considerada eficiente pelo modelo CCR. A DMU G é ineficiente em ambos os modelos, mas para ajustes de insumos, seria necessária uma redução maior para possuir uma eficiência técnica pelo modelo CCR do que pelo modelo BCC. Logo, temos que toda DMU considerada eficiente no modelo CCR, será considerada eficiente no modelo BCC, mas nem toda DMU considerada eficiente no modelo CCR, será considerada eficiente no modelo CCR.

O segmento de reta CD possui um coeficiente linear negativo e sendo estendido irá cortar o eixo das abscissas do quarto quadrante ($W_0 < 0; V_0 > 0$). O segmento OD intercepta a origem ($W_0 = 0; V_0 = 0$) e o segmento de reta DE possui coeficiente linear positivo e sendo

estendido irá cortar o eixo das abscissas do segundo quadrante ($W_0 > 0; V_0 < 0$). Essas variáveis W_0 e V_0 permitem avaliar os retornos variáveis de escala no modelo multiplicador enquanto no modelo envoltório pela restrição de convexidade temos $\lambda_r=1$.

Portanto, pela Equações (3.6, 3.7, 3.8 e 3.9), respectivamente, podemos escrever o modelo BCC dos multiplicadores (dual) e envoltório (primal), orientado ao *input* e *output* da seguinte forma:

Modelo BCC	
Orientação Input (Dual)	
$\max \theta \sum_{j=1}^m W_j Q_{jk} + W_0, \sum_{i=1}^n V_i X_{ik} = 1, \forall k, \sum_{j=1}^m W_j Q_{jr} - \sum_{i=1}^n V_i X_{ir} + W_0 \leq 0, \forall r, W_i, V_i \geq 0, \forall j, i$	
Orientação Output (Dual)	
$\min \varphi \sum_{i=1}^n V_i X_i + V_0, \sum_{j=1}^m W_j Q_{jk} = 1, \forall r, \sum_j j = 1^m W_j Q_{jr} - \sum_{i=1}^n V_i X_{ir} + V_0 \leq 0, \forall r, W_j, V_i \geq 0, \forall j, i$	
Orientação Input (Primal)	
$\min \theta \sum_{r=1}^p \lambda_r X_{ir} \leq \theta_{xi0}, \forall i, \sum_{r=1}^p \lambda_r Q_{mr} \geq q_{m0}, \forall m, \sum_r r = 1^p \lambda_r = 1, \forall r$	
Orientação Output (Primal)	
$\max \varphi \sum_{r=1}^p \lambda_r X_{ir} \leq x_{i0}, \forall i, \sum_{r=1}^p \lambda_r Q_{mr} \geq \varphi q_{m0}, \forall m, \sum_r r = 1^p \lambda_r = 1, \forall r$	

Tabela 3.1: Modelo Primal e Dual

3.5 Ecoeficiência e a Eficiência

O modelo DEA é fundamental para avaliar a ecoeficiência (baseada na Teoria Microeconômica de Firma), que considera entradas e saídas de diferentes naturezas, como econômicas, ambientais e sociais. Esse modelo permite avaliar a eficiência combinando variáveis com unidades de medida distintas, tornando possível uma análise quantitativa dessas variáveis. Parte da ideia de gerar mais saídas com uma menor quantidade de entrada tornando a DMU mais eficiente. [LIU et al. 2019]

Os indicadores que mensuram a ecoeficiência levam em consideração uma medida econômica desejável, alinhada com a entrada de alguma medida ambiental no processo, isso porque o crescimento econômico sempre está alinhado com o aumento da degradação ambiental, sendo necessário um conjunto de tecnologias e legislações aplicáveis ao processo para permitir um crescimento inteligente e sustentável com os menores impactos ambientais possíveis [Bianchi et al. 2020]

Voltando a ideia de [Koopmans 1951] onde a eficiência é alcançada quando a DMU possui quantidades de insumos e produtos que não podem ser melhorados sem piorar simultaneamente

amente o outro, assim como quando baseada no desempenho de outras DMUs não há melhorias que podem ser feitas sem afetar de maneira negativa os insumos e produtos. Portanto a ecoeficiência só poderá ser alcançada quando a utilização dos meus recursos for minimizada sem aumentar os impactos ambientais gerados pelas atividades agroindustriais (emissão de gases poluentes, desmatamento de matas e florestas, ou monocultura).

Logo, um conjunto de entradas X pode produzir um conjunto de saídas Q no entanto a maximização dessas saídas nem sempre serão um objetivo restrito da ecoeficiência, isso porque a variável Q leva em consideração externalidades positivas e saídas indesejáveis, este último geralmente relacionadas aos impactos ambientais causadas pelas atividades industriais.

X pode produzir $(Q_{\text{desejável}}, Q_{\text{indesejável}})$

Para solucionar esse problema de maximização de saídas indesejáveis, podemos tratar o $Y_{\text{indesejável}}$ como o inverso do seu valor original.

X pode produzir $(Q_{\text{desejável}}, Q_{\text{indesejável}})$

Dessa forma, quando o modelo linear maximizar as saídas irá minimizar as saídas indesejáveis. Portanto, podemos calcular a ecoeficiência como um problema de programação linear (PPL) através dos modelos de eficiência demonstrado nas Seções 3.3 e 3.4, como descrito na Tabela 3.2.

Índice de Ecoeficiência
Orientação Input (Dual)
$\max \theta \sum_{j=1}^m W_j Q_{jk} + W_0, \sum_{i=1}^n V_i X_{ik} = 1, \forall k, \sum_{j=1}^m W_j Q_{jr} - \sum_{i=1}^n V_i X_{ir} + W_0 \leq 0, \forall r, W_i, V_i \geq 0, \forall j, i$
Orientação Output (Dual)
$\min \varphi \sum_{i=1}^n V_i X_i + V_0, \sum_{j=1}^m W_j Q_{jk} = 1, \forall r, \sum_{j=1}^m W_j Q_{jr} - \sum_{i=1}^n V_i X_{ir} + V_0 \leq 0, \forall r, W_j, V_i \geq 0, \forall j, i$
Orientação Input (Primal)
$\min \theta \sum_{r=1}^p \lambda_r X_{ir} \leq \theta x_{i0}, \forall i, \sum_{r=1}^p \lambda_r Q_{mr} \geq q_{m0}, \forall m, \sum_{r=1}^p \lambda_r = 1, \forall r$
Orientação Output (Primal)
$\max \varphi \sum_{r=1}^p \lambda_r X_{ir} \leq x_{i0}, \forall i, \sum_{r=1}^p \lambda_r Q_{mr} \geq \varphi q_{m0}, \forall m, \sum_{r=1}^p \lambda_r = 1, \forall r$

Tabela 3.2: Modelo de PPL (Primal-Dual) para Calcular a Ecoeficiência

Onde $\hat{\theta}_{CCR}$ e $\hat{\varphi}_{CCR}$ representam a ecoeficiência orientada ao *Input* e orientada ao *Outputs*, respectivamente, do modelo CCR. O $\hat{\theta}_{BCC}$ e $\hat{\varphi}_{BCC}$ representam a ecoeficiência orientada ao *Input* e orientada ao *Outputs*, respectivamente, do modelo BCC. Temos uma interpretação semelhante do modelo CCR de que $\hat{\theta}_{CCR} = \frac{1}{\hat{\varphi}_{CCR}}$, no entanto, devido à orientação

do modelo BCC, temos que $\hat{\theta}_{BCC} \neq \frac{1}{\hat{\varphi}_{BCC}}$. Dessa forma, a ecoeficiência do modelo BCC sempre será superior ao modelo CCR, causada em um número maior de DMUs ecoeficientes no modelo BCC em comparação com o modelo CCR.

3.6 Índice de diversidade de Shannon-Weave

Uma das variáveis muito importantes para se calcular a ecoeficiência é utilizar o índice de diversidade de Shannon-Weave (IDSW), ele permite capturar os impactos ambientais gerados pela monocultura. Segundo [Beltrán-Esteve et al. 2012] podemos calcular o índice de diversidade da seguinte forma:

$$IDSW = e^{-\sum_{c=1}^C P_{cr} * \ln(P_{cr})} \quad (3.10)$$

Onde P_{cr} representa a proporção P total dos municípios r, que estão dedicados a cultura c, dessa forma o índice terá um valor de 0 a 1, ou seja, quando o valor do índice for igual a 1 aponta uma atividade de monocultura, que é uma atividade que desgasta o solo, portanto foi feito o inverso desse valor por se tratar de uma saída indesejável $1/IDSW$ para calcular a ecoeficiência.

$$y_4 = \frac{1}{e^{\sum_{c=1}^C P_{cr} * \ln(P_{cr})}} \quad (3.11)$$

3.7 Modelo Estocástico e Determinísticos - Reamostragem e *Bootstrap*

O modelo DEA clássico é baseado em dados não paramétricos, sem inferências estatísticas e sem a necessidade de explicar um relacionamento funcional entre as variáveis do processo. Portanto, se tornou comum ao decorrer dos anos, desde a concepção do modelo (CCR e BCC), até pela falta de ferramentas de inferências estatísticas aplicadas a modelos não paramétricos, utilizar o modelo como meio de estimar a eficiência para tornar a análise de eficiência mais robusta, no entanto, atualmente dispomos de uma série de ferramentas baseadas em *bootstrap* para o cálculo de eficiência-DEA com inferências estatísticas, teste de hipóteses e intervalos de confiança. Esse modelo é conhecido como modelo DEA estocástico. O conceito do *bootstrap* foi introduzido por [Efron 1979], depois [Wilson 1993] e [Wilson 1995] introduziu os métodos descritivos para identificar influentes de dados para o cálculo não paramétrico. Esses métodos permitem o uso de inferências estatísticas sem comprometer a natureza não paramétrica do problema, no entanto, exige um trabalho manual, o que torna muitas vezes inviável pela quantidade e diversidade de dados tratados, como no caso dessa pesquisa (5563 observações, 5 *inputs* e 4 *outputs*), a utilização desses métodos

traria uma desvantagem em seu uso. [Stosic 2005]

Já a aplicação do *bootstrap* com o modelo DEA só foi apresentado por [Simar and Wilson 1998]. Onde o *bootstrap* simula uma amostra com aplicação do estimador original, tornando os resultados da simulação replicando a amostra original, por meio de um processo de geração de dados (PGD), em um processo de reamostragem, repetidos várias vezes. O que é algo complexo quando se trata de uma estimação de uma fronteira não paramétrica, ou seja, de um conjunto de variáveis que podem apresentar erros e inferências, levando em consideração que as entradas e saídas são variáveis aleatórias constituída de um processo de geração de dados pouco conhecido.

Portanto [Stosic 2003] propuseram um novo modelo que combina o *bootstrap* (modelo estocástico) e a reamostragem de *Jackknife* (modelo determinístico) para detecção desses *outliers*, onde para cada DMU calcula-se os escores de eficiência de todas as outras DMUs, retira-se a DMU observada, detectam-se os valores discrepantes (*outliers*) e pode tanto remover automaticamente, como manualmente (sendo um número viável) da base de dados.

O autor descreve alguns passos para aplicação da metodologia, organizados aqui em 7 etapas.

1. Primeiro, um algoritmo é implementado baseado na técnica de re-amostragem de *Jackknife*. Eleja de maneira randômica em torno de 10% do conjunto r com $(r = 1, \dots, p)$ para formar um subconjunto que chamaremos de t com $(t = 1, \dots, p)$;
2. Calculando as eficiências das DMUs pelo DEA de $\hat{\theta}_t$ com $t = 1, \dots, p$;
3. Depois deve-se recalculer cada uma das eficiências removendo cada uma das DMUs $\hat{\theta}_{t,b}$ com $t = 1, \dots, p$ e $t \neq b$, onde cada $b = 1, \dots, p$ representa a DMU que foi removida. Calculando assim o desvio padrão ϕ_b de $\hat{\theta}_{t,b}$ em relação a $\hat{\theta}_t$;

$$\sigma_b = \sqrt{\frac{\sum_{t=1, t \neq b}^p (\hat{\theta}_{t,b}^* - \hat{\theta}_t)}{p - 1}} \quad (3.12)$$

4. Repetir (I, II e III) S vezes, acumulando as informações de alavancagem em θ_{bs}

$$n_b = \frac{St}{p} \quad (3.13)$$

5. Calcular a alavancagem (valor que mede o impacto da remoção da DMU do conjunto, dado pelo desvio padrão) média para cada DMU. A ideia é que os *outliers* apresentem um comportamento com uma maior alavancagem do que a média das outras DMUs, portanto será selecionado com uma menor probabilidade;

$$\bar{\sigma}_b = \frac{\sum_{s=1}^{n_b} \sigma_{bs}}{n_b} \quad (3.14)$$

6. Calcular a alavancagem global, por meio do somatório da alavancagem média de cada DMU;

$$\bar{\sigma} = \frac{\sum_{b=1}^p \bar{\sigma}_b}{p} \quad (3.15)$$

7. Depois da re-amostragem de *Jackknife* é aplicado o *bootstrap* para inserir intervalos de confiança e informações, por meio da alavancagem reduzir a probabilidade de se escolher o *outliers* para amostra de re-amostragem estocástica, a probabilidade é calculada baseada na função de *Heaviside*. Dessa forma, a DMU que possuir um valor de alavancagem considerável a alavancagem global, será descartada.

$$P(\bar{\sigma}_b) = \begin{cases} 1, & \text{se } \bar{\sigma}_b < \bar{\sigma}_b \log p \\ 0, & \text{se } \bar{\sigma}_b \geq \bar{\sigma}_b \log p \end{cases} \quad (3.16)$$

3.8 Avaliar Tipo de Retorno de Escala

É estritamente importante entender a natureza do comportamento do retorno de escala para aplicar corretamente o modelo de análise (CCR, OU BCC), podendo haver erros na análise por distorção do resultado a depender do tipo de retorno, causando uma perda da eficiência estatística. Portanto é necessário testar o comportamento das observações para identificar o comportamento da Fronteira tecnológica (F). [Simar and Wilson 2002]

$$H_0 \rightarrow F \text{ é RCE}$$

$$H_1 \rightarrow F \text{ é RVE}$$

Para definir se a hipótese é nula ou não, precisamos calcular a eficiência de cada DMU usando os modelos CCR e BCC. Podemos então definir a eficiência de escala ($\hat{\theta}_E$) da seguinte forma:

$$\hat{\theta}_E = \frac{\hat{\theta}_{CCR}}{\hat{\theta}_{BCC}}, \quad \hat{\theta}_E \leq 1 \quad (3.17)$$

Onde $\hat{\theta}_E$ é aproximadamente igual a 1, isso indica que a DMU L tem um comportamento de Retornos Constantes de Escala (RCE), e a análise de eficiência pode ser feita usando o modelo CCR ($\hat{\theta}_{CCR}$). Se $\hat{\theta}_E$ é significativamente menor que 1, isso indica que a DMU L tem um comportamento de Retornos Variáveis de Escala (RVE), e a análise de eficiência deve ser feita usando o modelo BCC ($\hat{\theta}_{BCC}$). Então temos:

$$H_0 \rightarrow \hat{\theta}_E = 1, \text{ onde } F \text{ é CCR.}$$

$$H_1 \rightarrow \hat{\theta}_E < 1, \text{ onde } F \text{ é BCC.}$$

Os autores [Simar and Wilson 2002] vão propor que a eficiência de escala pode ser calculada da seguinte forma:

$$\hat{\theta}_E = \frac{\sum_{r=1}^p \hat{\theta}_{r,CCR}}{\sum_{r=1}^p \hat{\theta}_{r,BCC}} \quad (3.18)$$

Para denifir se $\hat{\theta}_E$ é significativamente menor que 1, calculamos o valor crítico C_α , onde se $\hat{\theta}_E > C_\alpha$ aceita H_0 . Os autores vão propor usar o método *bootstrap*, com ajuda da função R *boot.sw98*, utilizando o pacote *FEAR*, dessa forma podemos aceitar H_0 , desde que a H_a seja aceita, calculando o valor estimado de $\hat{\theta}_E$ e o valor crítico C_α .

$$H_a \rightarrow \hat{\theta}_E > C_\alpha \text{ é CCR.}$$

$$H_b \rightarrow \hat{\theta}_E < C_\alpha \text{ é BCC.}$$

Logo, mesmo que $\hat{\theta}_E$ seja menor que 1, ao aceitar H_a podemos considerar o modelo como CCR. Pois $\hat{\theta}_E$ não é significativamente menor que 1.

3.9 Construção de Intervalos de Confiança com *Bootstrap*

O modelo DEA clássico é baseado em dados não paramétricos, sem inferências estatísticas e sem a necessidade de explicar um relacionamento funcional entre as variáveis do processo. Portanto se tornou comum ao decorrer dos anos desde a concepção do modelo (CCR e BCC), até pela falta de ferramentas de inferências estatísticas aplicadas a modelos não paramétricos, utilizar o modelo como meio de estimar a eficiência, no entanto atualmente dispomos de uma série de ferramentas baseadas em *bootstrap* para o cálculo de eficiência-DEA com inferências estatísticas, teste de hipóteses e intervalos de confiança. Esse modelo é conhecido como modelo DEA estocástico.

O conceito do *bootstrap* foi introduzido por [Efron 1979], mas sua aplicação com o modelo DEA só foi apresentado por [Simar and Wilson 1998], onde o *bootstrap* simula uma amostra com aplicação do estimador original, tornando os resultados da simulação replicando a amostra original, por meio de um processo de geração de dados (PGD), em um processo de reamostragem, repetidos várias vezes. O que é algo complexo quando se trata de uma estimação de uma fronteira não paramétrica, ou seja, de um conjunto de variáveis que podem apresentar erros e inferências, levando em consideração que as entradas e saídas são variáveis aleatórias constituída de um processo de geração de dados pouco conhecido. O modelo se baseia em 5 passos para aplicação:

1. Para cada observação $(x_i, q_i), i = 1, \dots, n$, calcula o DEA correspondente $\hat{\theta}_{CCR}$ ou $\hat{\theta}_{BCC}$ (representaremos por $\hat{\theta}_i$) usando a programação linear.

2. Por meio de um processo de reamostragem, onde há uma retirada de um conjunto de dados da amostra original de maneira randômica. Usando o método de *bootstrap*, será gerada uma amostra aleatória de tamanho p de $\hat{\theta}_r, r = 1, \dots, p$, fornecendo $\theta_1^*, \dots, \theta_p^*$.
3. Depois a partir dessa amostra aleatória temos $\{X^* = [x_1^*, \dots, x_p^*], Q^* = [q_1^*, \dots, q_p^*]; r = 1, \dots, p\}$, onde $x_r^* = \left(\frac{\hat{\theta}_r}{\theta_r^*}\right) x_r, r = 1, \dots, p$.
4. Então é possível calcular o *bootstrap* estimado $\hat{\theta}_{(i,b)}^*$ de i para $i = 1, \dots, n$, resolvendo o problema linear com restrições de DEA usando:

$$\hat{\theta}_{(i,b)}^* = \min \left\{ \theta \mid \sum_{r=1}^p \lambda_r X_{i,b}^* \leq \theta x_{io}, \sum_{r=1}^p \lambda_r Q_{mr} \geq q_{mo} \forall m, \sum_{r=1}^p \lambda_r = 1, \lambda_r \geq 0, \forall r \right\}$$

5. Depois repita os passos (2) a (4) B vezes para obter o resultado, para cada observação $i = 1, \dots, n$, em um conjunto de estimativas:

$$\{\hat{\theta}_{i,b}^*, b = 1, \dots, B\}$$

Capítulo 4

Apresentação dos Resultados

A pesquisa tem como prerrogativa a análise dos Municípios brasileiros através da Análise Envoltória de Dados (DEA). Totalizando cerca de 5563 Municípios analisados de todos os Estados Brasileiros, mais o Distrito Federal. Subdivididos entre a Região Norte (Acre - AC, Amapá - AP, Amazonas - AM, Pará - PA, Rondônia - RO, Roraima - RR e Tocantins - TO), Região Nordeste (Maranhão - MA, Piauí - PI, Ceará - CE, Rio Grande do Norte - RN, Paraíba - PB, Pernambuco - PE, Alagoas - AL, Sergipe - SE e Bahia - BA), Região Centro-Oeste (Distrito Federal - DF, Goiás - GO, Mato Grosso - MT e Mato Grosso do Sul - MS), Região Sudeste (São Paulo - SP, Rio de Janeiro - RJ, Minas Gerais - MG e Espírito Santo - ES) e Região Sul (Paraná - PR, Santa Catarina - SC e Rio Grande do Sul - RS).

Os dados foram coletados do Instituto Brasileiro de Geografia e Estatística (IBGE), no Censo Agropecuário de 2017 (último publicado até o presente) para coleta das variáveis de entrada, x_1 — Área dos estabelecimentos (hectares), x_2 — Gasto anual com combustíveis e lubrificantes (Mil reais), x_3 — Gasto anual com insumos para produção vegetal e animal (Mil reais), x_4 — Mão de obra ocupada nos estabelecimentos (assalariada e familiar) e x_5 — Outras despesas (mil reais). As variáveis de saídas desejáveis y_1 — Receita bruta anual (Mil reais) e y_2 — Área de matas e florestas naturais e plantadas nos estabelecimentos (hectares) e variáveis para o cálculo do y_4 — Índice de diversidade de *Shannon-Weaver*, uma variável de saída indesejável. O Sistema de Estimativa de Emissões de Gases de Efeito Estufa (SEEG) para coletar a saída indesejável y_3 — Emissões anuais de gases de efeito estufa sector agropecuário (toneladas GWP). Cada variável foi coletada para todos os 5563 municípios brasileiros. O resumo dos dados está na Tabela 4.1.

Os resultados são apresentados mostrando o processo de retirada dos *outliers* com o *bootstrap*, depois vai ser analisado os intervalos de confiança também com uso do *bootstrap* para a construção desses intervalos, definiu-se o tipo de retorno de escala da tecnologia (F) para determinar qual modelo (CCR ou BCC) seria adequado para analisar e gerou as ecoeficiências dos municípios.

	Média	Mediana	Desv.Pad	Máximo	Mínimo
x1	63.147,55	24.480,80	138.808,71	4.810.916,30	0,001000
x2	4.264,12	1.476,00	17.276,88	1.107.997,00	0,042275
x3	51.104,04	18.329,00	109.730,62	1.740.972,06	34,035800
x4	2.715,28	1.768,00	2.944,54	48.246,00	0,001000
x5	3.509,74	955,00	9.881,26	251.959,00	0,001000
y1	88.650,06	41.829,00	172.537,12	3.258.836,00	0,001000
y2	18.222,27	4.773,00	49.424,62	1.125.574,00	8,391978
y3	101.297,16	39.154,58	206.618,66	4.485.536,54	40,518500
y4	0,080088	0,074229	0,041666	1,000000	0,017115

Tabela 4.1: Estatísticas descritivas (Censo Agropecuário 2017-IBGE) e SEEG

4.1 Retirada dos *Outliers* com *Bootstrap*

Devido os dados poderem apresentar erros, dados faltantes disponível no Censo de 2017 (medida de proteção de dados para evitar expor informações de determinadas propriedades importantes a terceiros), ou baixa homoscedasticidade, podem causar distorções nos resultados da análise DEA, ou até mesmo a heterogeneidade das DMUs devido a diferenças tecnológicas, portanto é necessário verificar os *outliers* do modelo e retirar para o resultado não ficar enviesado.

Para aplicação da metodologia do DEA foi utilizado o pacote *jackstrap* no RStudio para análise dos *outliers* e aplicação do *bootstrap*. Disponível na biblioteca R em: [https://cran](https://cran.r-project.org/web/packages/jackstrap/index.html) (SOUSA E MONTE ;2020). Com subconjuntos de 10% do total de DMUs para formar um subconjunto t e número de repetições igual a 1000, calculando as médias de alavancagem, alavancagem global e retiradas dos *outliers*. Segundo as etapas descritas na Seção 3.7.

Como havia uma quantidade grande de dados, optou-se por particionar as análises dos dados por regiões (Norte, Nordeste, Centro-Oeste, Sudeste e sul) para simplificação da análise, devido ao uso considerável de processamento para retirada dos *outliers* caso foi feito de maneira unificada. Foram utilizados a função *jackstrap* para análise, com critérios *Heaviside* e *Kolmogorov-Smirnov* (K-S). Na região sudeste não foram encontrados outliers em nenhum dos critérios. Como o critério de *Heaviside* apresentou dados mais restritos, optou-se pelo método. Para as cinco regiões foram encontradas 143 DMUs consideradas *outliers*. Os resultados estão na Tabela 4.2.

Região	Municípios	Heaviside	K-S	Municípios sem outliers
Norte	450	17	13	433
Nordeste	1794	61	21	1733
Centro-Oeste	467	23	10	444
Sudeste	1662	-	-	1662
Sul	1191	42	21	1149

Tabela 4.2: Resultado da retirada de Outliers

Para compreender a influência dos *outliers* nas medidas foram plotados gráficos de densidade das frequências de cada região que foram encontradas *outliers*. (Figura 4.1)

Portanto, verificou-se que os *outliers* causavam deturpação dos dados, todos enviesados

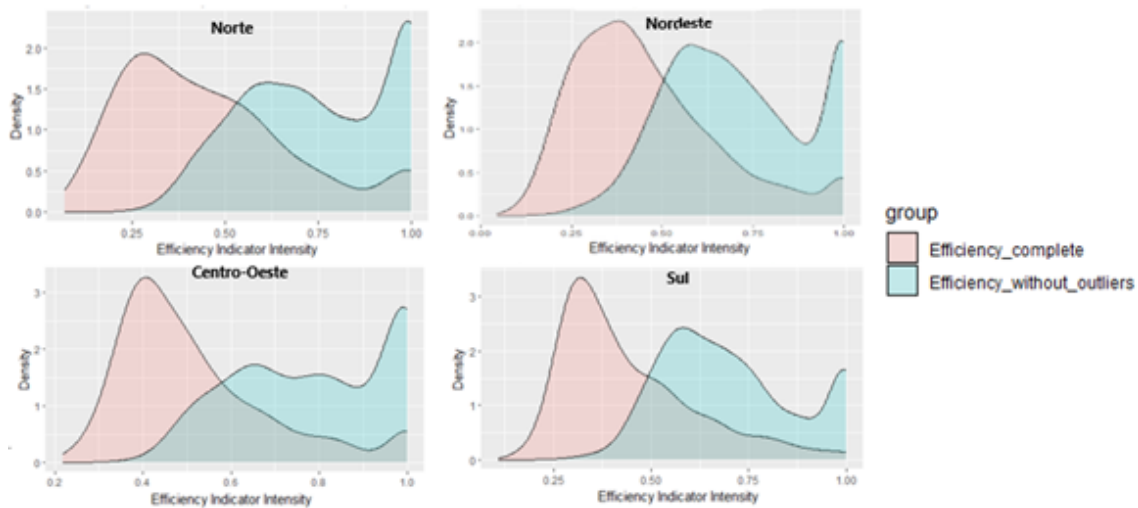


Figura 4.1: Curvas de densidade das ecoeficiências por Região

para valores inferiores de ecoeficiência. Pelo teste de *Wilcoxon bilcaudal* comparando as eficiências com e sem *outliers*, verificou-se que a média e mediana das duas distribuições divergiam ($p - valor < 2.2e^{-16}$, nas regiões que foram identificadas *outliers*). Essas divergências ficam mais explícitas nos gráficos de *bloxplots* das eficiências de cada região. (Figura 4.2)

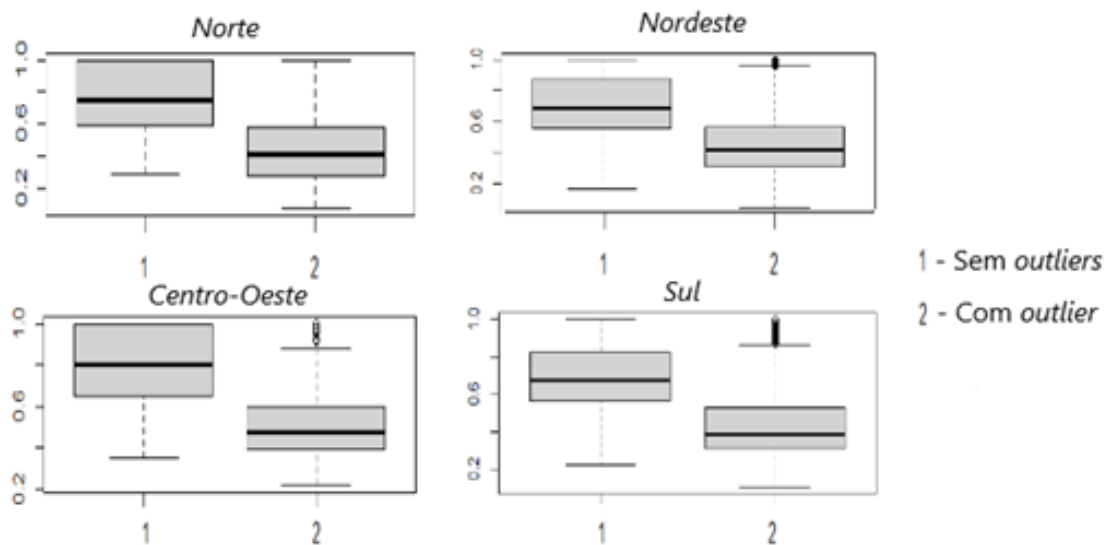


Figura 4.2: Bloxplots com e sem Outliers

Como pode ser observado na Figura 4.1 e na Figura 4.2 as médias de ambos os gráficos são deslocadas para um escore de eficiência mais baixo, puxando as ecoeficiências para um desempenho inferior.

Dessa forma é entendível que os *outliers* influenciavam na consistência dos resultados. Sendo necessária sua retirada para poder analisar a ecoeficiência dos municípios de maneira assertiva, resultando em eficiências mais robustas. As 143 DMUs removidas representam cerca de 2,57% do total, consolidando assim para análise da eficiência 5.420 Municípios.

4.2 Definição do Tipo de Retorno de Escala

A partir da nova amostra resultante dos dados sem *outliers* podemos analisar o tipo de retorno de escala do dado, dessa forma definir qual o melhor modelo (CCR ou BCC), pois usar qualquer retorno de escala (constante, ou variável) pode gerar resultados inadequados, bem como podem causar deturpações nas análises pela avaliação ser de natureza diferente do real comportamento da amostra. Como descrito na seção 3.8 a primeira hipótese H_0 é que a Fronteira tecnologia (F) seja com retornos constantes de escala desde que a eficiência de escala $\hat{\theta}_E$ seja igual a 1 (CCR), a H_1 considera que a Fronteira tecnologia (F) tenha retornos variáveis de escala, desde que a eficiência de escala $\hat{\theta}_E$ seja menor que 1 (BCC). Dado a hipótese alternativa H_a podemos aceitar H_0 caso o valor estimado seja maior que o valor crítico ($\hat{\theta}_E > C_\alpha$).

Foi utilizado a orientação ao *output* devido estarmos lidando com saídas indesejáveis, não é interessante manter as emissões da maneira como estão para minimizar os recursos, mas é mais indicado manter os insumos e maximizar as saídas, que no caso das saídas indesejáveis serão minimizadas pelo modelo. Nessa etapa os dados não foram particionados, portanto, os 5.420 municípios foram analisados os escores de ecoeficiência conjuntamente, com 1000 reamostragens obtidas pelo *bootstrap* e $\alpha = 0.05$.

Calculando o valor estimado obtemos $\hat{\theta}_E = 0.9210$ e o valor crítico $C_\alpha = 0.8385$, sendo assim a H_0 é aceita e a F é considerada com retornos constantes de escala (RCE). Dessa forma se entende que independente do tamanho da propriedade elas podem ser ecoeficientes. O que é um bom indicativo para a produção agrícola, pois cerca de 89% dos agricultores possuem uma propriedade com menos de cem hectares, sendo responsável por 80% da renda rural [E SOUZA and GOMES 2023].

A princípio pode ser algo um pouco difícil de se compreender partindo do ponto em que a maioria dos conjuntos analisados muitas vezes possuem um comportamento variável de escala, principalmente o setor privado quando olhamos para a indústria [Yang et al. 2015]. No entanto, pelos resultados obtidos, a agricultura brasileira apresenta um comportamento diferente e constante, o que confirma uma questão relacionada ao fato da desigualdade na agricultura não ocorrer pelo tamanho das propriedades, mas sim da disponibilidade tecnológica das mesmas [E SOUZA and GOMES 2018].

4.3 Ecoeficiência dos Municípios Brasileiros com Intervalo de Confiança

Aceitando que a tecnologia é CRS foi estimado os índices de ecoeficiências orientado ao *output*. Devido à abundância de dados, a análise foi particionada, em primeiro momento foram colocados os cinco melhores e os cinco piores de cada região com os índices de eco-

eficiência orientado ao *output* na Tabela 4.3, contém os escores de ecoeficiência corrigidos com intervalo de confiança de $\alpha = 95\%$.

Municípios Ecoeficientes	Região	Sem outliers com correção	Intervalo de confiança de 95%		Sem outliers sem correção	Municípios Ineficientes	Região	Sem outliers com correção	Intervalo de confiança de 95%		Sem outliers sem correção
			Mínimo	Máximo					Mínimo	Máximo	
Afuá (PA)	NORTE	0,9340	0,9953	0,9160	1,0000	Aurora do Pará (PA)	NORTE	0,6182	0,6510	0,6056	0,9230
Presidente Figueiredo (AM)	NORTE	0,9336	0,9905	0,9130	0,9999	Ariquemes (RO)	NORTE	0,6160	0,6855	0,6001	0,9643
Manaus (AM)	NORTE	0,9295	0,9898	0,8981	0,9999	Conceição do Araguaia (PA)	NORTE	0,5986	0,6481	0,5841	0,8317
Porto Walter (AC)	NORTE	0,9195	0,9498	0,8938	0,8620	Rio Sono (TO)	NORTE	0,5962	0,6442	0,5796	0,8541
Pimenta Bueno (RO)	NORTE	0,9150	0,9698	0,8895	0,8346	Amajari (RR)	NORTE	0,5169	0,5786	0,4997	1,0000
Araripina (PE)	NORDESTE	0,9818	0,9914	0,9739	0,9660	Mucugê (BA)	NORDESTE	0,7354	0,7568	0,7259	0,8927
Santo Estêvão (BA)	NORDESTE	0,9814	0,9931	0,9715	0,9569	Aldeias Altas (MA)	NORDESTE	0,7214	0,7244	0,7179	0,8656
Macatubas (BA)	NORDESTE	0,9806	0,9974	0,9728	0,9849	Açailândia (MA)	NORDESTE	0,7200	0,7573	0,7000	0,8801
Itapipoca (CE)	NORDESTE	0,9792	0,9958	0,9697	0,9933	Itinga do Maranhão (MA)	NORDESTE	0,7164	0,7410	0,6944	0,9072
Iguatu (CE)	NORDESTE	0,9773	0,9893	0,9683	0,9703	Muquém do São Francisco (BA)	NORDESTE	0,6867	0,7156	0,6704	0,8766
Nova Lacerda (MT)	CENTRO-OESTE	0,9618	0,9947	0,9446	0,9894	Antônio João (MS)	CENTRO-OESTE	0,7037	0,7345	0,6882	0,9119
Ribeirãozinho (MT)	CENTRO-OESTE	0,9533	0,9947	0,9214	0,8947	Planaltina (GO)	CENTRO-OESTE	0,6836	0,7204	0,6660	0,9602
Inocência (MS)	CENTRO-OESTE	0,9528	0,9952	0,9307	0,9087	Amambai (MS)	CENTRO-OESTE	0,6758	0,6953	0,6581	0,8595
Turvânia (GO)	CENTRO-OESTE	0,9520	0,9953	0,9304	0,8886	Nova Andradina (MS)	CENTRO-OESTE	0,6474	0,6676	0,6339	0,8136
Edealina (GO)	CENTRO-OESTE	0,9519	0,9888	0,9188	0,8639	Nova Alvorada do Sul (MS)	CENTRO-OESTE	0,6447	0,6709	0,6244	0,7986
Nova Campina (SP)	SUDESTE	0,9004	0,9261	0,8780	0,9943	Jaíba (MG)	SUDESTE	0,4163	0,4764	0,4191	0,9010
Itatinga (SP)	SUDESTE	0,8888	0,9539	0,8621	0,9985	Morro Agudo (SP)	SUDESTE	0,4139	0,4874	0,4218	0,9198
Cananéia (SP)	SUDESTE	0,8656	0,9513	0,8467	0,9999	Paraguacu Paulista (SP)	SUDESTE	0,4124	0,4679	0,4043	0,8796
Josépolis (MG)	SUDESTE	0,8502	0,9521	0,8321	1,0000	Ecoporanga (ES)	SUDESTE	0,4105	0,4996	0,4345	0,9789
Carbonita (MG)	SUDESTE	0,8432	0,9732	0,8373	0,9999	Ataléia (MG)	SUDESTE	0,3781	0,4490	0,3981	0,9701
Caxias do Sul (RS)	SUL	0,9400	0,9912	0,9094	0,9999	Eldorado do Sul (RS)	SUL	0,5720	0,6321	0,5583	0,9700
Giruá (RS)	SUL	0,9221	0,9663	0,8998	0,9778	Santo Inácio (PR)	SUL	0,5667	0,6041	0,5516	0,8438
Barão do Triunfo (RS)	SUL	0,9174	0,9949	0,9119	0,9999	Iguaraçu (PR)	SUL	0,5383	0,6088	0,5325	0,9484
Mafra (SC)	SUL	0,9165	0,9755	0,8947	0,9877	Colorado (PR)	SUL	0,5130	0,5448	0,5016	0,7876
Congonhinhas (PR)	SUL	0,9163	0,9935	0,8976	0,9999	Colorado (RS)	SUL	0,5130	0,5448	0,5016	0,7876

Tabela 4.3: Ranking do Índice de Ecoeficiência dos Municípios por Região

Verificou-se para as DMUs ineficientes que os valores não corrigidos beneficiavam os escores de ecoeficiência, tornando os escores da DMUs ineficientes mais altos comparados ao escore de corrigidos das mesmas DMUs. Assim como os escores de ecoeficiência sem correção das DMUs eficientes eram subestimados, ou seja, apresentavam valores abaixo do escore com valores corrigidos. Podemos visualizar essa diferença nos escores corrigidos e não corrigidos na Figura 4.3.

As regiões apresentaram médias, medianas e amplitudes interquartil dos dados sem correção maiores que os dados com correção. Dessa forma, obtemos dados mais robustos. Na Tabela 4.4 podemos ver a estatística descritiva das eficiências calculadas com *outliers*, sem *outliers* e com correção sem *outliers*.

Dessa forma vemos que a retirada dos *outliers*, a parametrização dos dados e a correção dos dados, gera escores de eficiências mais robustos e sendo assim resultados mais próximos de representar a realidade de cada municípios e região brasileira. Apenas a região Sudeste não apresentou diferença entre estar com, ou sem *outliers*, pois nos municípios dessa região não foram identificados *outliers* pelo *jackstrap*.

Depois foi feita a análise para definir qual Região possui os municípios mais ecoeficientes, no entanto, seria injusto comparar a região que possui mais municípios ecoeficien-

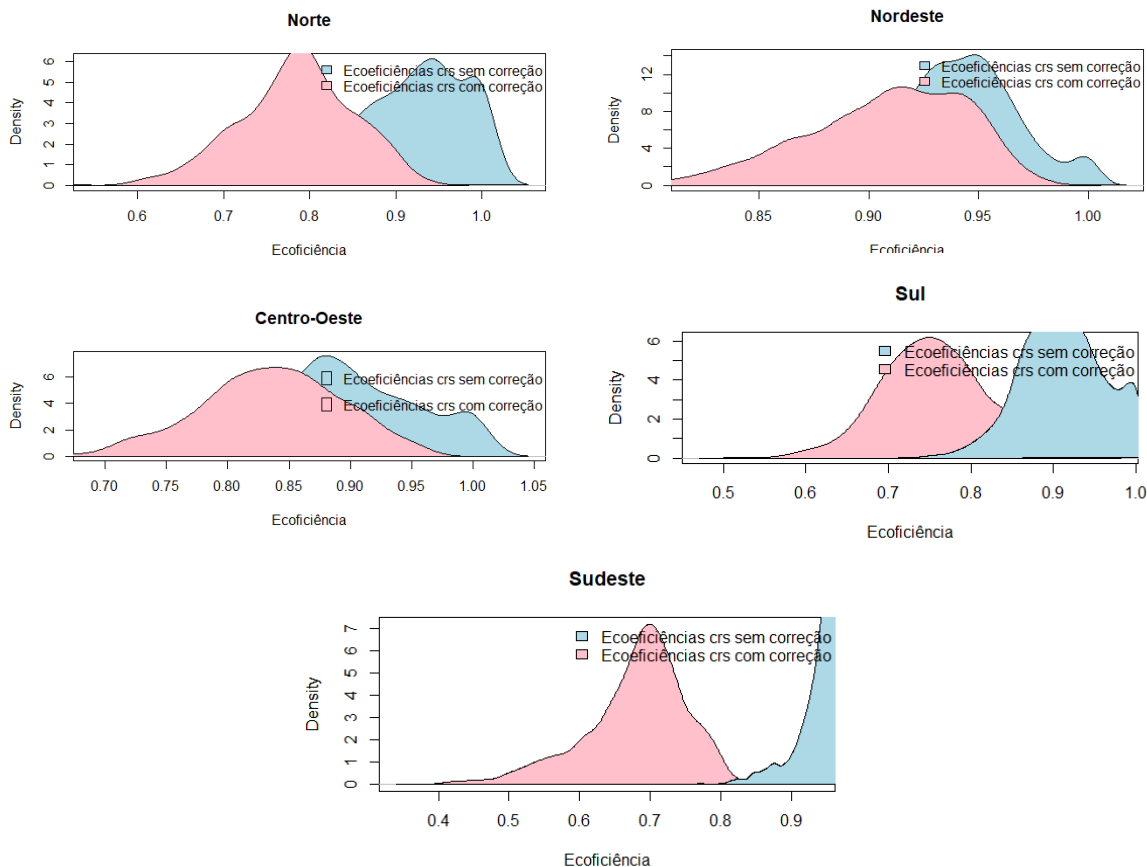


Figura 4.3: Gráfico de Densidade das Ecoeficiências das Regiões

tes dado que tem regiões que possuem mais do que o dobro de municípios do que outras regiões, esse é o caso da região Nordeste que possui 1.733 municípios, valor aproximadamente 4 vezes maior que as regiões Norte e Centro-Oeste, então foi feito um comparativo da ecoeficiência orientado ao *output* entre as cinco melhores DMUs de cada Região.

No comparativo entre os dez melhores municípios, verificou-se que a Região Sudeste, Sul e norte, foi as que saíram melhores, respectivamente e a Região Nordeste e Centro-Oeste (não apareceu entre os dez melhores municípios na Tabela 4.5) foram as que obtiveram pior desempenho.

Em relação aos municípios mais bem colocados, os valores com correção apresentou valores acima dos escores sem correção, assim como algumas DMUs apresentaram valores acima com escore corrigido, portanto pelo intervalo de confiança podemos diferenciar a ecoeficiência entre essas unidades. Dessa forma entende-se a importância de definir um intervalo de confiança para os escores.

A média geral dos municípios é 0,7891 (valor sem *outliers* e com viés de confiança), o que é diferente das médias regionais do índice de ecoeficiência das regiões Norte (0,7839), Nordeste (0,9067), Centro oeste (0,8348), sudeste (0,6802) e sul (0,7526), um ponto interessante das médias regionais é que a região nordeste apresenta em média valores de escores maiores que todas as regiões, o que demonstra que a região apesar de ter poucos municípios

Estatística descritiva	Região				
	Norte	Nordeste	Centro-Oeste	Sudeste	Sul
Média com outlier	0.7343	0.8909	0.6408	0.9589	0.8428
Média sem outlier	0.9169	0.9376	0.9011	0.9589	0.9108
Média com correção	0.7839	0.9067	0.8348	0.6802	0.7526
Mediana com outlier	0.7571	0.8912	0.6259	0.9661	0.8393
Mediana sem outlier	0.9273	0.9391	0.8947	0.9661	0.9116
Mediana com correção	0.7881	0.9125	0.8359	0.6915	0.7548
Amplitude com outlier	0.2801	0.0650	0.2317	0.0309	0.1112
Amplitude sem outlier	0.0895	0.0380	0.0785	0.0309	0.0713
Amplitude com correção	0.0838	0.0532	0.0761	0.0829	0.0864

Tabela 4.4: Média, mediana e amplitude das ecoeficiências

Municípios Ecoeficientes	Região	Sem outliers com correção	Intervalo de confiança de 95%		Sem outliers sem correção
			Máximo	Mínimo	
			Araripina (PE)	NORDESTE	
Santo Estêvão (BA)	NORDESTE	0,9814	0,9931	0,9715	0,9569
Macaúbas (BA)	NORDESTE	0,9806	0,9974	0,9728	0,9849
Itapipoca (CE)	NORDESTE	0,9792	0,9958	0,9697	0,9933
Iguatu (CE)	NORDESTE	0,9773	0,9893	0,9683	0,9703
Nova Lacerda (MT)	CENTRO-OESTE	0,9618	0,9947	0,9446	0,9894
Ribeirãozinho (MT)	CENTRO-OESTE	0,9533	0,9947	0,9214	0,8947
Inocência (MS)	CENTRO-OESTE	0,9528	0,9952	0,9307	0,9087
Turvânia (GO)	CENTRO-OESTE	0,9520	0,9953	0,9304	0,8886
Edealina (GO)	CENTRO-OESTE	0,9519	0,9888	0,9188	0,8639

Tabela 4.5: Ecoeficiência dos Dez Melhores Municípios

ecoeficientes, a maioria está mais perto da fronteira produtiva comparada as outras regiões, isso releva uma maior homogeneidade entre as áreas de atividade agrícola da região. Diferente da região Sudeste e sul que possuem as menores médias por região de 0,6802 e 0,7526, respectivamente, abaixo da média geral, o que demonstra que os municípios que não fazem parte da fronteira produtiva dessas regiões estão mais distantes da fronteira ecoeficiente, apontando uma maior heterogeneidade entre as áreas de atividade agrícola da região. Esse é o caso do município de Ataléia (MG) com um escore de ecoeficiência de 0,3782. Portanto, a região Sudeste e sul apresentam bons desempenhos para algumas DMUs, mas também representam os menores desempenhos municipais.

Isso pode ser visualizados nos mapas georreferenciados, gerados por meio do pacote *geobr* do *RStudio* (Disponível na biblioteca R em: <https://cran.r-project.org/package=geobr>) [PEREIRA et al. 2022], ele puxa dados do IBGE para fazer o georreferenciamento, dessa forma combina-se os escores calculados para a ecoeficiência e as informações de georreferenciamento disponível IBGE de 2017 e plota-se os gráficos de escala disponíveis (Figuras 4.4, 4.5, 4.6, 4.5 e 4.8). Os locais em branco são os municípios faltantes que não apresentam os escores de ecoeficiência por serem considerados *outliers*.

Na região Norte as DMUs mais ineficientes se localizam na parte litorânea da região e próxima à região Nordeste, conforme a localização vai indo para a parte interna da região em análise, os escores de ecoeficiência tende a aumentar. No entanto, a escala do mapa na totalidade também apresenta homogeneidade. (Figura 4.4)

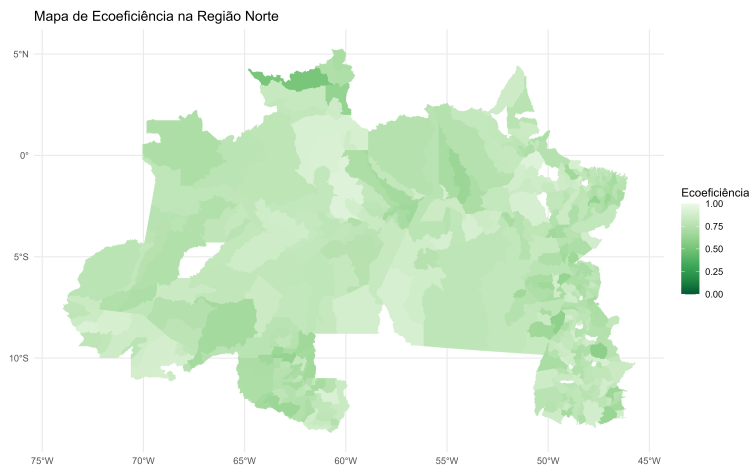


Figura 4.4: Georreferenciamento dos Escores de Ecoeficiência Região Norte

Na região Nordeste, assim como na região Norte, os escores das DMUs mais ineficientes ficam mais próximas do litoral, ou próxima a divisas de outros estados, no entanto não apresenta um contraste significativo de escala, apresentando escores mais homogêneos. (Figura 4.5)

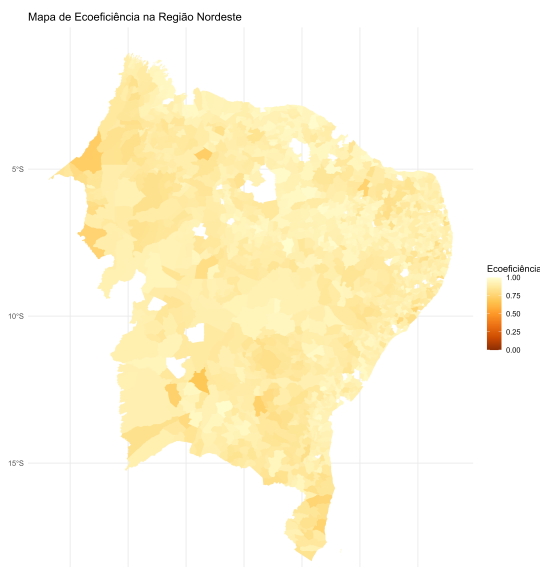


Figura 4.5: Georreferenciamento dos Escores de Ecoeficiência Região Nordeste

Na região Centro-Oeste percebe-se uma heterogeneidade maior entre os escores comparados a região Norte e nordeste, com um contraste maior entre as escalas de cada município, ou seja, tanto as DMUs eficientes quanto as menos eficientes estão distribuídas de maneira randômica pelo espaçamento geográfico. Embora a escala dos municípios dessa Região, na maioria, apresente escores mais próximos de 1. (Figura 4.6)

Na região Sudeste é visualizado um maior contraste, o que aponta uma alta heterogeneidade nos dados entre as escalas dos escores de ecoeficiência e uma tendência das DMUs mais ineficientes da região estarem localizados em áreas de fronteira a região Centro-Oeste e Nordeste, já as DMUs eficientes tendem a ficar na parte interna da região. (Figura 4.7)

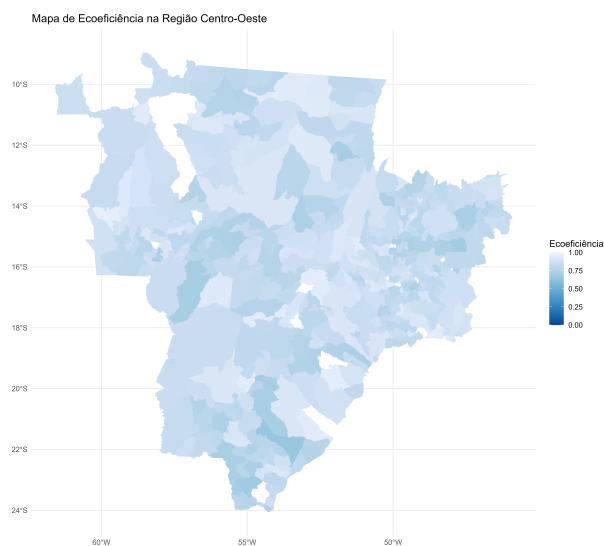


Figura 4.6: Georreferenciamento dos Escores de Ecoeficiência Região Centro-Oeste

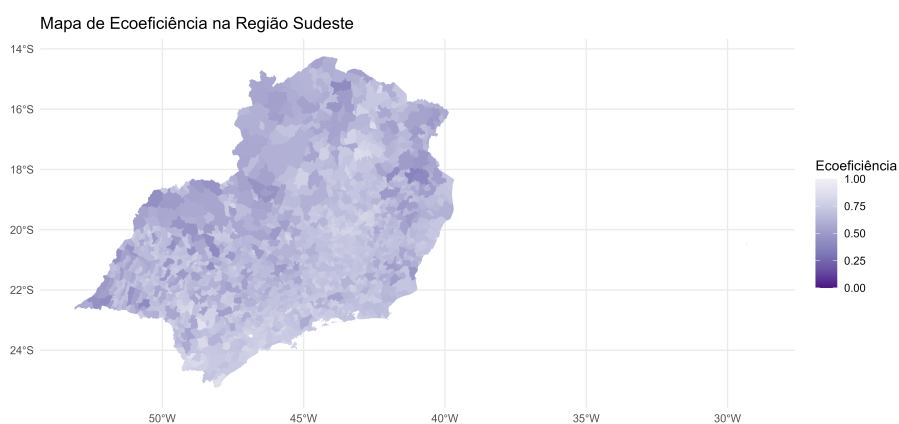


Figura 4.7: Georreferenciamento dos Escores de Ecoeficiência Região Sudeste

Na região Sul, assim como na região Sudeste, os escores de ecoeficiência apresentam uma heterogeneidade maior entre os municípios, percebe-se que os escores tendem a apresentar escalas mais elevadas na parte mais ao sul do gráfico. (Figura 4.8)

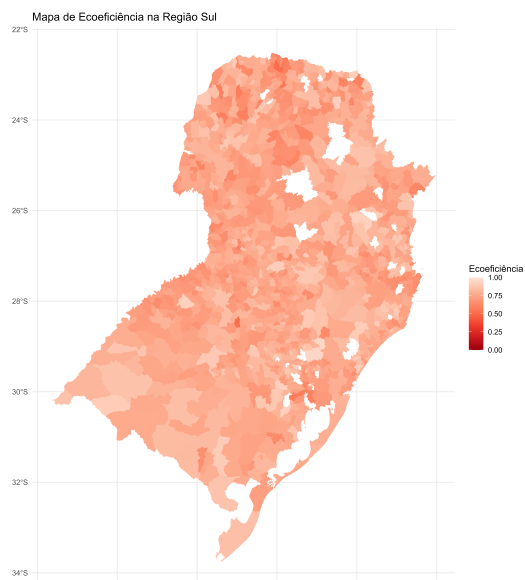


Figura 4.8: Georreferenciamento dos Escores de Ecoeficiência Região Sul

Capítulo 5

Conclusão

Neste trabalho foi utilizado o *bootstrap* para verificar *outliers* do modelo, bem como para estimar um viés de confiança das eficiências calculadas para 5420 municípios brasileiros. Utilizou-se o pacote *Jackstrap* para identificação e eliminação dos *outliers* que influenciava na qualidade dos resultados. Em tese o modelo DEA é sensível aos *outliers* e isso pode ser comprovado pela presente pesquisa, demonstrando que os *outliers* causavam uma mudança na média dos escores de eficiência para valores inferiores comparado aos escores de eficiência com a presença dos *outliers*. Depois foi feito um teste de retorno de escala para identificar o comportamento da tecnologia (F) do modelo, como resultado foi obtemos o modelo CCR para depois ser estimado os escores de eficiência por meio da função *boot.sw98*.

Como resultado, obtivemos diferenças entre as curvas de densidade dos conjuntos de eficiência dos dados com a presença de *outliers* e dos dados sem *outliers* para todas as regiões, exceto para a região Sudeste, onde não foi encontrado nenhum *outlier*. Dessa forma ocasionou na eliminação de 143 DMUs do conjunto de dados originais, que deslocavam a média das curvas de densidade. O que demonstra que a presença desses *outliers* enviesavam os resultados e portanto é necessário muita cautela para interpretar resultados quando estes estão presentes no modelo, o que ressalta a importância da utilização do *bootstrap* na análise.

A definição do tipo de retorno de escala como modelo CCR, é um resultado relevante para entendermos a agricultura brasileira, onde há uma grande disponibilização de terras, que são distribuídas de maneira heterogênea entre a maior parte dos produtores rurais, pois demonstra que o pequeno, médio e grande produtor podem se tornar igualmente ecoeficientes. Isso ficou mais claro com o cálculo dos escores de ecoeficiência dos municípios, a estimação do valor crítico em comparação com a ecoeficiência de escala. Analisou-se que tanto produtores de grande e pequeno porte foram considerados ecoeficientes, o que afirma a capacidade do pequeno produtor de se tornar ecoeficiente, no entanto é importante ressaltar que a maior parte dos municípios considerados ecoeficientes na análise eram de grande e médio porte, isso se deve ao fato de terem um maior acesso a tecnologia.

A análise permite o ranqueamento dos dez municípios mais ecoeficientes, sendo Arari-

pina (PE), Santo Estêvão (BA), Macaúbas (BA), Itapipoca (CE), Iguatu (CE), Nova Lacerda (MT), Ribeirãozinho (MT), Inocência (MS), Turvânia (GO) e Edealina (GO), respectivamente, com destaques para a região Nordeste e Centro-Oeste, portanto por meio das estatística descritiva e dos mapas de georreferenciamento entende-se que as regiões que apresentam escores de ecoeficientes mais altos no ranqueamento, também apresentam uma maior homogeneidade nos escores em comparação com outros municípios da mesma região, o que aponta os municípios da região Nordeste e Centro-Oeste (melhores colocadas no ranqueamento) estão em um nível mais próximo para a ecoeficiência do que em outras regiões.

Entre os municípios, olhando para os escores calculados com e sem correção, analisou-se a necessidade de estabelecer um intervalo de confiança para a correta classificação dessas DMUs, demonstrando a necessidade de se ter um cuidado ao se calcular os escores de ecoeficiência de maneira determinística, sem o cuidado no tratamento dos dados, como a retiradas dos *outliers*, parametrização dos dados e o uso do *bootstrap* para estabelecer um intervalo de confiança.

Referências Bibliográficas

- [Aria and Cuccurullo 2017] Aria, M. and Cuccurullo, C. (2017). bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of informetrics*, 11(4):959–975.
- [Bai et al. 2019] Bai, Y. Deng, X., Gibson, J., Zhao, Z., and Xu, H. (2019). Exploring the relationship between urbanization and urban eco-efficiency: Evidence from prefecture-level cities in china. *Journal of cleaner production*, 195:1487–1496.
- [Banker et al. 1984] Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9):1078–1092.
- [Beltrán-Esteve et al. 2012] Beltrán-Esteve, M., Gómez-Limón, J. A., and Picazo-Tadeo, A. J. (2012). Assessing the impact of agri-environmental schemes on the eco-efficiency of rain-fed agriculture. *Spanish Journal of Agricultural Research*, 10(4):911–925.
- [Bianchi et al. 2020] Bianchi, M., Del Valle, I., and Tapia, C. (2020). Measuring eco-efficiency in european regions: Evidence from a territorial perspective. *Journal of Cleaner Production*, 276:123246.
- [Bogetoft and Otto 2011] Bogetoft, P. and Otto, L. (2011). Benchmarking with dea, sfa, and r. *Springer Science & Business Media*.
- [Charnes et al. 1978] Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2((6)):429–444.
- [Cook and Zhu 2013] Cook, W. D. and Zhu, J. (2013). Data envelopment analysis: Balanced benchmarking. *CreateSpace Independent Publishing Platform*.
- [Cooper et al. 2011] Cooper, W. W., Seiford, L. M., and Zhu, J. (2011). Handbook on data envelopment analysis. *International Series in Operations Research & Management Science*.
- [E SOUZA and GOMES 2018] E SOUZA, G. S. and GOMES, E. G. (2018). A stochastic production frontier analysis of the brazilian agriculture in the presence of an endogenous covariate. *In: Operations Research and Enterprise Systems: 7th International Conference, ICORES 2018, Funchal, Madeira, Portugal*, pages 3–14.

- [E SOUZA and GOMES 2023] E SOUZA, G. S. and GOMES, E. G. (2023). Assessing the influence of external factors on agricultural production in brazil. *Socio-Economic Planning Sciences*, 85:101440.
- [Efron 1979] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *annals of statistics*. *Socio-economic planning sciences*, pages 7:1–26.
- [ELLEGAARD and WALLIN 2015] ELLEGAARD, O. and WALLIN, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact. *Scientometrics*, 105:1809–1831.
- [Fan et al. 2015] Fan, J. L., Zhang, X., Zhang, J., and Peng, S. (2015). Efficiency evaluation of co2 utilization technologies in china. *Journal of CO2 Utilization*, pages 33–41.
- [Fan et al. 2017] Fan, Y., Bai, B., Qiao, Q., Kang, P., Zhang, Y., and Guo, J. (2017). Study on eco-efficiency of industrial parks in china based on data envelopment analysis. *Journal of environmental management*, pages 107–115.
- [Farrell 1957] Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical*, pages 253–281.
- [FERREIRA and GOMES 2020] FERREIRA, C. M. and GOMES, A. P. (2020). Introdução à análise envoltória de dados: Teoria modelos e aplicações (2° ed.). *UFV*.
- [Huang et al. 2019] Huang, J. X., J., Yu, Y., and Zhang, N. (2019). Composite eco-efficiency indicators for china based on data envelopment analysis. *Ecological indicators*, pages 674–697.
- [IBGE 2017] IBGE (2017). Instituto brasileiro de geografia e estatística. *Censo Agropecuário 2017: Resultados Definitivos*. Acesso em 2023, disponível em IBGE: <https://sidra.ibge.gov.br/pesquisa/censo-agropecuario/censo-agropecuario-2017/resultados-definitivos>.
- [Koopmans 1951] Koopmans, T. C. (1951). An analysis of production as an efficient combination of activities. *Activity analysis of production and allocation*.
- [LIU et al. 2019] LIU, X., GUO, P., and GUO, S. (2019). Assessing the eco-efficiency of a circular economy system in china’s coal mining areas: Emergy and data envelopment analysis. *Journal of Cleaner Production*, pages 1101–1109.
- [Matsumoto and CHEN 2021] Matsumoto, K. and CHEN, Y. (2021). Industrial eco-efficiency and its determinants in china: a two-stage approach. *Ecological Indicators*, page 108072.
- [Shah et al. 2020] Shah, I. H., Dong, L., and Park, H. (2020). Tracking urban sustainability transition: An eco-efficiency analysis on eco-industrial development in ulsan, korea. *Journal of cleaner production*, page 121286.

- [Shephard 1970] Shephard, R. W. (1970). The theory of cost and production function. *Princeton University Press*, page 308.
- [Simar and Wilson 1998] Simar, L. and Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management science*, pages 49–61.
- [Simar and Wilson 2002] Simar, L. and Wilson, P. W. (2002). Non-parametric tests of returns to scale. *European Journal*, pages 115–132.
- [Stepien et al. 2021] Stepien, S., Czyżewski, B., Sapa, A., Borychowski, M., Poczta, W., and Poczta-Wajda, A. (2021). Eco-efficiency of small-scale farming in poland and its institutional drivers. *Journal of cleaner production*, page 123721.
- [Stosic 2003] Stosic, B. D. (2003). Jackstrapping dea scores for robust efficiency measurement. *Anais do XXV Encontro Brasileiro de Econometria SBE*, pages 1525–1540.
- [Stosic 2005] Stosic, B. D. (2005). Technical efficiency of the brazilian municipalities. *Journal of Productivity Analysis*, pages 157–181.
- [Wilson 1993] Wilson, P. (1993). Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis*, pages 57–45.
- [Wilson 1995] Wilson, P. (1995). Detecting influential observations in deterministic non-parametric frontiers models. *Journal of Business and Economic Statistics*, pages 319–323.
- [Yang et al. 2015] Yang, L., Ouyang, H., Fang, K., Ye, L., and Zhang, J. (2015). Evaluation of regional environmental efficiencies in china based on super-efficiency-dea. *Ecological Indicators*, pages 13–19.
- [Yu et al. 2020] Yu, S., Liu, J., and Li, L. (2020). Evaluating provincial eco-efficiency in china: an improved network data envelopment analysis model with undesirable output. *Environmental science and pollution research*, pages 6886–6903.
- [Yu et al. 2019] Yu, Y., Peng, C., and Li, Y. (2019). Do neighboring prefectures matter in promoting eco-efficiency? empirical evidence from china. *Technological Forecasting and Social Change*, pages 456–465.
- [Zhou et al. 2018] Zhou, C., Shi, C., Wang, S., and Zhang, G. (2018). Estimation of eco-efficiency and its influencing factors in guangdong province based on super-sbm and panel regression models. *Ecological Indicators*, pages 67–80.

Apêndice