

ANÁLISE DE COMPONENTES PRINCIPAIS PARA O ESTUDO DAS CONCENTRAÇÕES DE POLUENTES ORIUNDAS DE ESTIMATIVAS DE EMISSÕES DE QUEIMADAS NA MESORREGIÃO SUDESTES DE MT

1. INTRODUÇÃO

O Estado de Mato Grosso é formado por 141 municípios que se dividem em cinco mesorregiões: Norte(55 municípios), Nordeste(25 municípios), Sudeste(22 municípios), Centro-Sul (17 municípios) e Sudeste(22 municípios).Este estudo terá como base a mesorregião Sudeste do estado.

Nos últimos anos vem crescendo os casos de doenças respiratórias causadas por queimadas nos meses de agosto a outubro, que é justamente a época das secas no estado de Mato Grosso fazendo que ocorra muitos focos de incêndio.

Com isso a emissão de poluentes na atmosfera é muito grande, alguns poluentes como CO , $PM_{2.5}$, NO_2 , SO_2 e O_3 , os mais encontrados, além disso também temos fenômenos naturais que proporcionam uma maior incidência de doenças, como precipitação acumulada, velocidade e direção do vento, temperatura do ar, umidade relativa do ar e focos de queimada de vegetação, fazendo com que muitas variáveis sejam avaliadas, em geral, são correlacionadas entre si. Diante esse contexto, pode-se questionar: quais os impactos das emissões de poluentes atmosféricos na saúde humana? Para responder este questionamento utilizaremos a estatística multivariada verificando se ocorre uma correlação entre as variáveis em estudo.

O uso de técnicas univariadas faz com que sejam ignoradas certas relações entre as variáveis. A análise multivariada é uma ferramenta que permite estatisticamente avaliar simultaneamente múltiplas informações de um objeto ou indivíduo sob investigação gerando uma gama de cruzamentos entre suas variáveis dependentes ou independentes (HONGYU et al, 2015).

Por causa disso se optou por usar as técnicas de análises multivariadas que são mais indicadas para uma quantidade enorme de variáveis, qualquer pesquisador que examine apenas relações entre duas variáveis e evite o uso de análise multivariada estará ignorando poderosas ferramentas que podem lhe dar informações potencialmente úteis (HAIR et al, 2006).

Uma dessas técnicas da Estatística Multivariada é a de Análise de Componentes Principais (PCA). Esta técnica consiste em transformar um conjunto de variáveis originais em um novo conjunto de variáveis, em que cada variável é chamada de componente principal (SARTORIO et al, 2021), onde os seus autovetores representam o peso de cada variável para a composição do componente e o autovalor explica a contribuição de cada componente para explicar a variação total dos dados (GOMES et al, 2004). A visualização gráfica da matriz de dados permite extrair conclusões extremamente úteis que não poderiam ser obtidas através da análise univariada (SARTORIO et al, 2021).

2. PROBLEMA DE PESQUISA E OBJETIVO

O estado de Mato Grosso sofre todo ano com o aumento das queimadas ocorrendo com isso um aumento das doenças respiratórias na população. Nos últimos anos o estado vem batendo o recorde em focos de queimadas em toda a região.

Os incêndios florestais que acontecem na zona rural prejudicam muito a área urbana, o vento leva a fumaça e dificulta ainda mais a situação respiratórias das pessoas, crianças e idosos aumentando os atendimentos nos hospitais.

Desta forma, a presente pesquisa tem como objetivo analisar os altos índices de emissões de poluentes oriundos das queimadas dos anos de 2015 a 2018 da mesorregião sudeste de MT, em relação ao processo de altas taxas de doenças respiratórias ocorridas na época das queimadas, utilizando para isso da estatística multivariada.

3. FUNDAMENTAÇÃO TEÓRICA

3.1 Práticas de queimadas

Por milhares de anos, o fogo tem sido um importante integrante dos processos biogeoquímicos, influenciando nas interações terra-atmosfera, estando presente em forma de instrumento de manejo em mais de 50% dos ecossistemas terrestres. Ele pode ter origem de várias formas, mas os grandes incêndios florestais são de forma acidental ou até mesmo criminosa tornando-se às vezes incontroláveis (LOTUFO, 2020).

No Brasil, os incêndios florestais e as queimadas são problemas antigos, relacionados, principalmente, à cultura do uso do fogo como ferramenta no modelo de cultura desenvolvido pela colonização (GONÇALVES, 2005).

Para algumas populações indígenas que ocupam originalmente a região leste do Estado de Mato Grosso, o manejo do fogo faz parte da cultura, utilizado como ferramenta tradicional para limpeza de áreas e para a prática da caça (MISTRY & BIZERRIL, 2011). O estado de Mato Grosso possui muitas reservas indígenas fazendo com que a tradição se prevaleça e que ocorra um maior descontrole das queimadas no estado.

O ar geralmente é afetado pelos grandes incêndios fazendo com que ocorra um grande prejuízo para a saúde das pessoas pela liberação de gases e partículas para a atmosfera. O conselho Nacional do Meio Ambiente (CONAMA), pela Resolução n 491 de 19/11/2018, estabeleceu padrões de qualidade do ar, fazendo com que as concentrações de poluentes não prejudiquem a segurança e o bem estar da população (SISAM).

As partículas finas liberadas pela fumaça de queimadas causam preocupação por que permanecem por mais tempo na atmosfera e, em função de sua capacidade de dispersão, podem ser transportadas a grandes distâncias. Elas podem se depositar nos brônquios terminais e nos alvéolos dos pulmões causando ou agravando problemas respiratórios, e contribuindo para o aumento nas admissões hospitalares e emergenciais, utilização de medicamentos e aumento da mortalidades, especialmente entre idosos e crianças (FERREIRA, 2016).

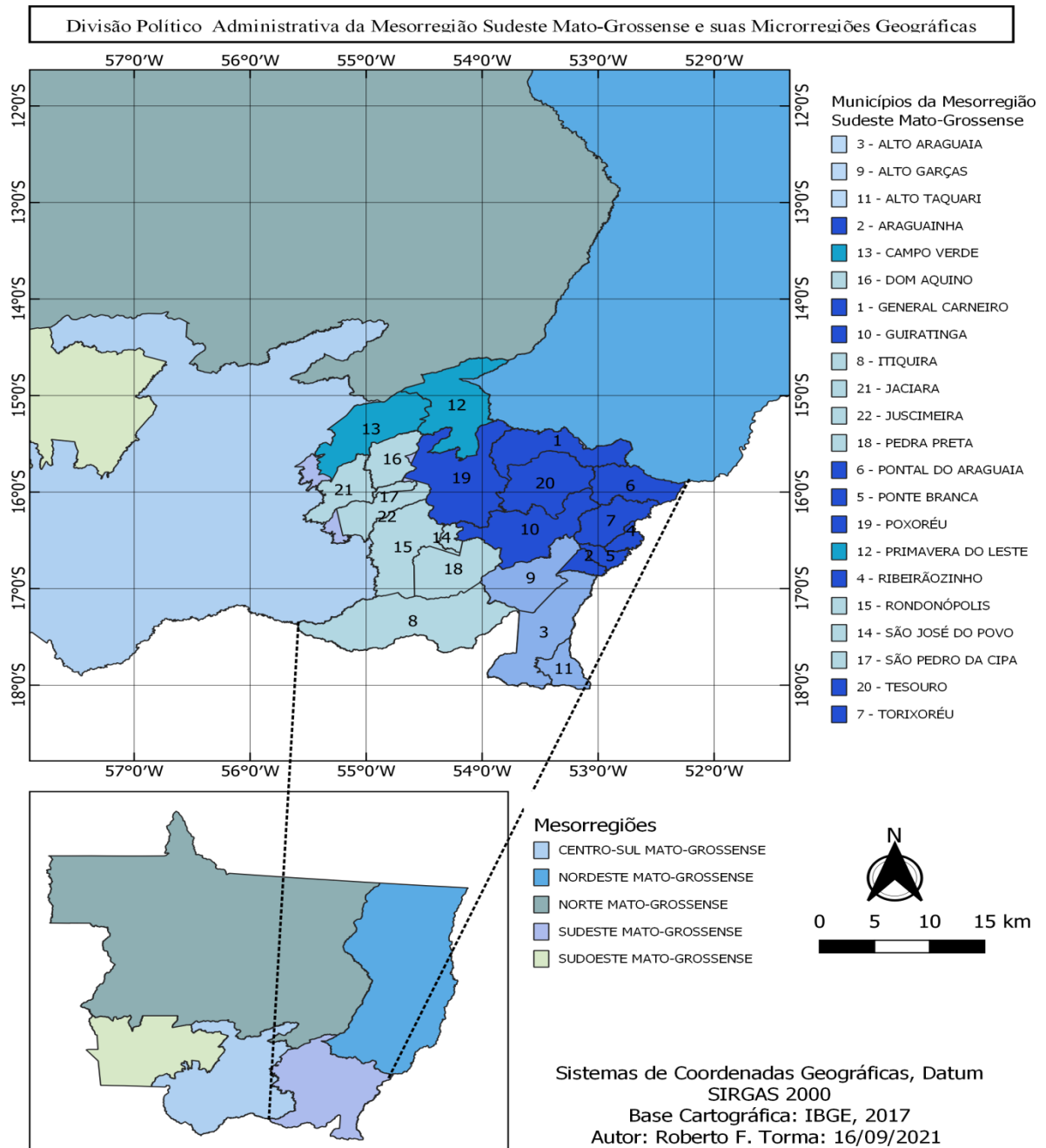
3.2 Mesorregiões

O Estado de Mato Grosso é dividido em cinco mesorregiões que são subdivisões do estado que congrega diversos municípios com similaridades econômicas e sociais, que por sua vez, são subdivididas em microrregiões. As cinco mesorregiões do estado se dividem em Norte(55 municípios), Nordeste(25 municípios), Sudeste(22 municípios), Centro-Sul (17 municípios) e Sudeste(22 municípios) totalizando 141 municípios do estado.

A pesquisa se foca na mesorregião Sudeste do estado que é composta por 4 microrregiões: Primavera do Leste que contém os municípios de Campos Verde, Primavera do Leste, Tesouro composta pelos municípios de Araguaína, General Carneiro, Guiratinga, Pontal do Araguaia, Ponte Branca, Poxoréu, Ribeirãozinho, Tesouro e Torixoréu,

microrregião de Rondonópolis composta pelos municípios de Dom Aquino, Itiquira, Jaciara, Juscimeira, Pedra Preta, Rondonópolis, São José do Povo, São Pedro da Cipa e a microrregião de Alto Araguaia que compreende os municípios de Alta Araguaia, Alto Garças e Alto Taquari.

Figura 1 - Mapa da Mesorregião Sudeste e as microrregiões geográficas



3.3 Análise de Componentes Principais

Para essa pesquisa optou-se pela estatística multivariada utilizando a técnica da análise de componentes principais (PCA) que consiste em transformar um conjunto de variáveis originais em outro conjunto de variáveis denominadas de componentes principais.

A estatística multivariada ou análise multivariada é um conjunto de métodos estatísticos que torna possível a análise simultânea de medidas múltiplas para cada indivíduo, objeto ou fenômeno observado (BETARELLI, 2021).

Um desses métodos é chamado de Análise de Componentes Principais (do inglês Principal Component Analysis). Este método tem a finalidade de reduzir a quantidade de variáveis estudadas através de combinações lineares das variáveis originais (CARVALHO, 2019).

A análise de componentes principais é associada à idéia de redução de massa de dados, com menor perda possível da informação utilizando-se com isso o processo de ortogonalização de vetores. Procura-se redistribuir a variação observada nos eixos originais de forma a se obter um conjunto de eixos ortogonais não correlacionados. Esta técnica pode ser utilizada para geração de índices e agrupamento de indivíduos.

A análise agrupa os indivíduos de acordo com sua variação, isto é, os indivíduos são agrupados segundo suas variâncias, ou seja, segundo seu comportamento dentro da população, representado pela variação do conjunto de características que define o indivíduo (VARELLA, 2008).

Os objetivos da análise de componentes principais são: identificação dos principais fatores que explicam uma certa variável resposta, seleção das variáveis mais pertinentes dentre um número grande de variáveis, afim de reduzir o trabalho de obtenção de dados em estudos posteriores, agrupamento de objetos ou indivíduos por meio da mensuração, de suas similaridades concernentes às variáveis observadas, num sistema de coordenadas independentes (DUARTE, 1998)

4. DISCUSSÃO

Os dados para a pesquisa foram baixados do Sistema de Informações Ambientais (SISAM) que faz parte do INPE. Cada planilha corresponde aos anos de 2015 a 2018 possuindo 11 variáveis: CO (Monóxido de Carbono), O₃ (Ozônio), NO₂ (Dióxido de Nitrogênio), SO₂ (Dióxido de Enxofre), PM₂₅ (Material particulado), direção e velocidade do vento, temperatura do ar, umidade relativa do ar, precipitação acumulada, focos de queimadas da vegetação.

Estes dados correspondem a mesorregião sudeste do estado de Mato Grosso formada por 22 municípios. Estes dados foram filtrados, ou seja, aplicou-se a média aritmética dos meses para que fosse possível fazer a análise de componentes principais. Para este estudo vamos utilizar o software SPSS v. 26.

O teste de esfericidade de Bartlett mede se existe correlação suficientemente forte para que a análise possa ser aplicada. Temos aqui um teste de hipóteses: H₀: a matriz de correlação é uma matriz identidade, não há correlação suficiente entre as variáveis. Análise não é adequada; H₁: a análise é adequada, existe correlação. Como o p-valor (sig) é menor que o nível de significância =5%, rejeitamos H₀, isto significa admitir que a análise é adequada (SILVA ET AL, 2021).

O teste KMO (Kaiser Meyer Olkin) mede se a base de dados se é adequada para o estudo. De modo geral, KMO com valores baixos significam que o tamanho da amostra é

inadequado para o uso desta ferramenta. O valor obtido neste teste deve ser $KMO \geq 0,5$, para que a ferramenta seja considerada possível ao problema(SILVA et al, 2012)

Teste de KMO e Bartlett		
Medida Kaiser-Mayer-Olkin de adequação de amostragem		0,483
Teste de esfericidade de Aprox. Qui-quadrado		184,524
Bartlett	Gl	55
	Sig	0,000

Tabela: 1- Teste de Bartlett e KMO

Vamos definir o número de componentes principais que serão utilizados para estabelecer o percentual da variância total explicada associada ao seu respectivo autovalor.

Variância Total explicada			
Autovalores iniciais			
Componentes	autovalor	% de Variância	% cumulativa
1	4,029	36,625	36,625
2	2,125	19,316	55,941
3	1,326	12,059	68,000
4	1,225	11,137	79,137
5	1,000	9,094	88,231
6	0,624	5,672	93,903
7	0,330	3,000	96,904
8	0,211	1,922	98,826
9	0,096	0,876	99,702
10	0,022	0,200	99,901
11	0,011	0,099	100,000

Tabela: 2 - Método de extração de componente principal

Podemos verificar que foi gerado 11 componentes principais, ou seja o número de componentes principais gerados é igual ao número de variáveis originais. Na coluna dos autovalores representa a informação carregada por cada componente.

Na coluna da variância, temos a variância explicada por cada componente, onde podemos observar que o primeiro componente é aquele que possui a maior variância o segundo a segunda maior variância e assim por diante.

O gráfico a seguir ilustra a variabilidade dos dados, mostrando a distribuição entre os eixos de ordenação, permitindo identificar o número de eixo mais significativos. Podemos verificar que o decaimento ocorre até o componente cinco, onde a partir deste componente a análise representa um incremento relativamente pequeno em termos da representatividade dos autovalores(SILVA et al, 2012).

Nota-se que no começo da análise a quantidade de variáveis demandava um total de 11, com o método houve uma redução para 5 variáveis.

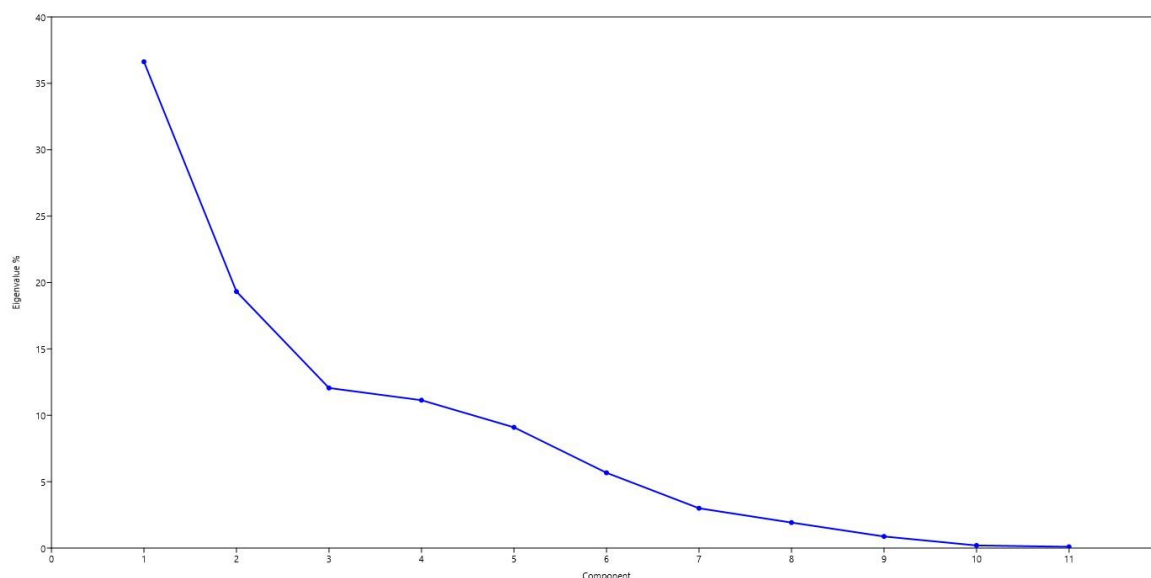


Gráfico: 1- Gráfico de escarpa (Scree Plot)

O gráfico acima ilustra a variância dos componentes principais, ou seja os autovalores lotados em ordem decrescente de magnitude, neste caso costuma-se selecionar os componentes que estão localizados anteriormente ao ponto de inflexão que é aquele ponto em que ocorre uma queda significativa da variância explicada em relação ao componente anterior. Outra convenção é utilizar os componentes principais que possui um autovalor superior a 1, neste caso utilizaremos os cinco primeiros componentes principais, que representa 88,231% da amostra, conforme a tabela acima.

A tabela a seguir representa a cumunalidade que é a porção da variância que uma variável compartilha com todas as outras variáveis consideradas, ou seja, é a variância explicada pelos fatores comuns (ROCHA, 2015). Nota-se que quanto mais próximo de um é mais significativo.

Camunalidades		
Fatores	Inicial	Extração
Co	1,000	0,977
No2	1,000	0,854
O3	1,000	0,946
Pm25	1,000	0,871
So2	1,000	0,850
Precipitação	1,000	0,929
Temperatura	1,000	0,959
Umidade relativa	1,000	0,723
Vento direção	1,000	0,965
Vento velocidade	1,000	0,751
Focos queimada	1,000	0,881

Tabela: 3- Método de extração: análise de componente principal

Para o cálculo da matriz de correlação foi utilizado os parâmetros do coeficiente de correlação e Pearson(ρ). o qual possui os seguintes níveis de correlação (ROCHA, 2015):

- $\rho > 0,70$ - Forte correlação.
- $0,30 < \rho < 0,70$ - Correlação moderada
- $0 < \rho < 0,30$ - Fraca correlação

O coeficiente de correlação de Pearson é uma medida do grau de relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. Quanto mais próximo de 1 existe uma relação linear perfeita e quanto mais próximo de -1 também existe uma relação perfeita só que inversa, ou seja, quanto uma aumenta a outra diminui. O valor zero significa que não ocorre uma relação entre as variáveis.

Podemos verificar também graficamente esta relação entre as variáveis. Quanto mais próximo de 1 o diâmetro do círculo fica maior e de cor azul mais forte. Assim temos também que quando se aproxima do -1 temos a cor vermelha.

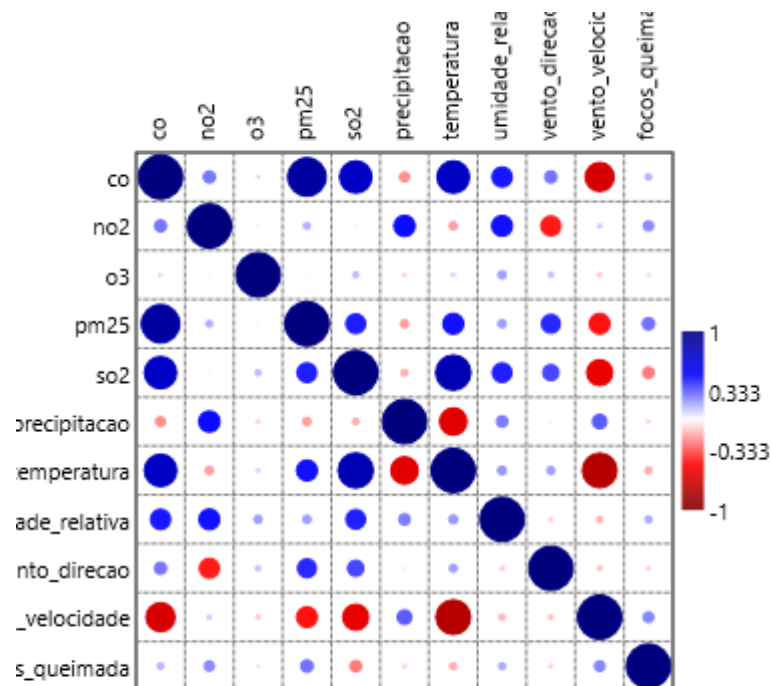


Gráfico: 2 - Representação gráfica da matriz de correlação

Matriz de Correlação											
	Co	No2	O3	Pm25	So2	Precipitação	Temperatura	Umidade relativa	Vento direção	Vento velocidade	Focos queimada
Co	1	0,268	0,068	0,867	0,730	-0,217	0,735	0,455	0,270	-0,655	0,144
No2	0,268	1	0,022	0,154	-0,025	0,478	-0,188	0,467	-0,449	0,092	0,227
O3	0,068	0,022	1	0,018	0,131	-0,079	0,077	0,188	0,115	-0,097	-0,065
Pm25	0,867	0,154	0,018	1	0,442	-0,188	0,469	0,186	0,423	-0,465	0,275
So2	0,730	-0,025	0,131	0,442	1	-0,153	0,796	0,440	0,361	-0,585	-0,256
Precipitação	-0,217	0,478	-0,079	-0,188	-0,153	1	-0,617	0,254	0,033	0,323	-0,080
Temperatura	0,735	-0,188	0,077	0,469	0,796	-0,617	1	0,200	0,183	-0,784	-0,158
Umidade relativa	0,455	0,467	0,188	0,186	0,440	0,254	0,200	1	-0,088	-0,140	0,160
Vento direção	0,270	-0,449	0,115	0,423	0,361	0,033	0,183	-0,088	1	-0,111	-0,085
Vento velocidade	-0,655	0,092	-0,097	-0,465	-0,585	0,323	-0,784	-0,140	-0,111	1	0,235
Focos queimada	0,144	0,227	-0,065	0,275	-0,256	-0,080	-0,158	0,160	-0,085	0,235	1

Tabela: 4 - Tabela de correlação

Observando a matriz de componentes , podemos verificar os componentes principais, conforme critério adotado na tabela 1, valores maiores ou iguais a 1, por isso reduzimos 11 variáveis para 5.

As primeiras componentes principais tem o maior peso que no caso com esses componentes podemos interpretar a variabilidade dos dados, ou seja a componente principal 1 tem variância 92,20% dos dados, a componente 2 tem 31,0% dos dados, etc..

Matriz de componente					
Componentes					
	1	2	3	4	5
Co	0,922	0,310	0,156	0,038	-0,067
Temperatura	0,904	-0,238	-0,041	-0,287	-0,022
So2	0,850	0,027	-0,332	0,082	-0,094
Vento_velocidade	-0,795	0,148	0,146	0,248	0,121
Pm25	0,743	0,216	0,434	0,285	-0,052
No2	-0,032	0,900	-0,060	-0,184	-0,077
Umidade_relativa	0,362	0,703	-0,276	-0,032	0,144
Precipitacao	-0,398	0,573	-0,385	0,467	-0,278
Focos_queimada	-0,056	0,414	0,795	0,013	0,274
Vento_direção	0,369	-0,314	0,040	0,854	0,018
O3	0,158	0,035	-0,345	0,102	0,889

Tabela: 5 - matriz de componentes

5. CONCLUSÃO

A técnica da análise de componentes principais, utilizada neste estudo, mostrou-se efetiva pelo fato de ter reduzido a quantidade de variáveis de 11 para 5 componentes que explicam parte satisfatória dos dados originais. Foi possível sintetizar, num determinado período, quais os principais poluentes que prejudicam substancialmente a saúde das pessoas durante o período de queimadas.

Os dois primeiro componentes explicam mais que 55% da variabilidade total das variáveis originais isto significa que o fenômeno sob estudo pode ser muito simplificado.

O gráfico biplot mostra a posição relativa de cada componente e mostra a posição relativa das amostras. Através da posição relativa das amostras podemos verificar se existe um comportamento conceituado de uma variável em uma amostra. Pelo gráfico podemos verificar que as variáveis Co e Pm25 tem um correlação forte

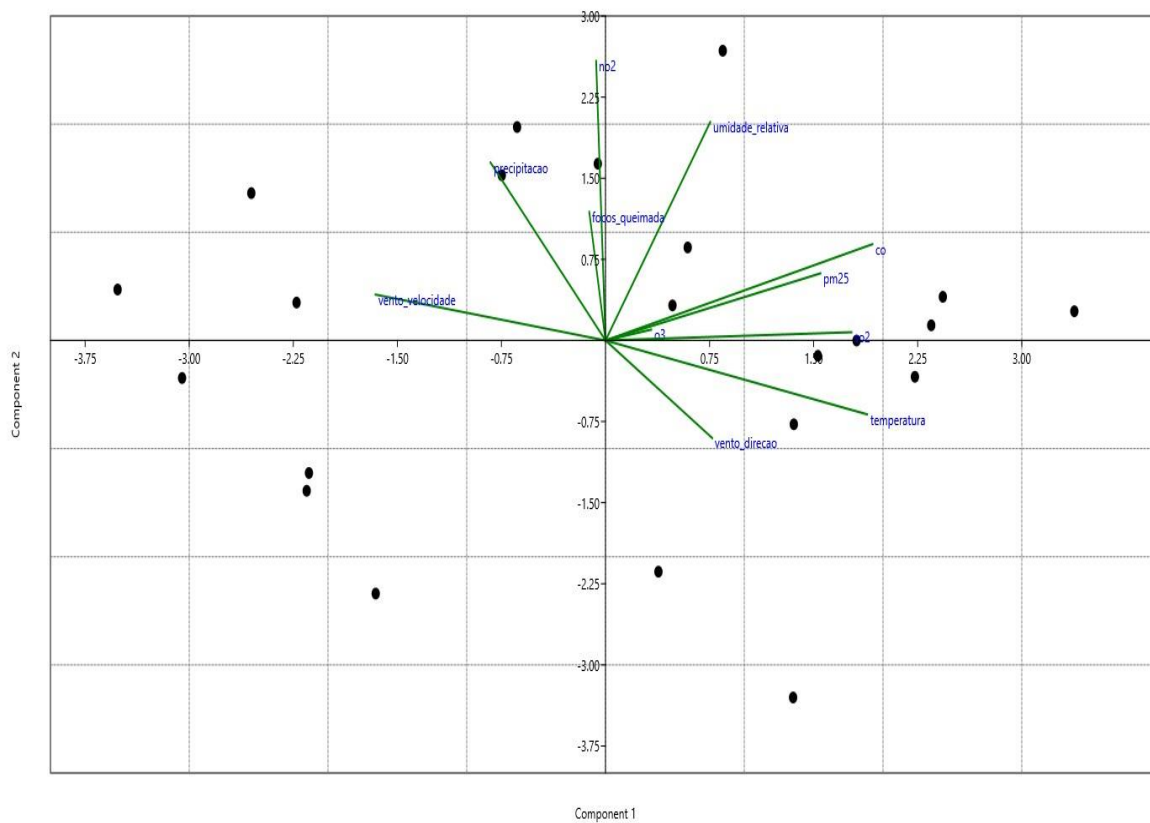


Gráfico: 3 - biplot

6. REFERÊNCIAS BIBLIOGRÁFICAS

BETARELLI JR, Admir Antônio. **Análise Multivariada I**. Disponível em: https://www2.ufjf.br/lates/files/2016/12/Conte%3%bado-1-%e2%80%93N_Introdu%3%a7%c3%a3o-%c3%a0-Estat%3%adstica-Multivariada-e-an%3%a1lise-dos-dados.pdf. Acesso em: 18/09/2021.

CARVALHO, Luiz H M. **Análise multivariada: uma síntese dos principais métodos**. Disponível em: <https://operdata.com.br/blog/analise-multivariada/>. Acesso em: 19/09/2021.

DUARTE, João Batista. **Introdução à análise de componentes principais**.

VARALLA, Carlos A.A. **Análise de componentes principais**.

GOMES, J. B. V. et al . **Análise de componentes principais de atributos físicos, químicos e mineralógicos de solos do bioma cerrado**. Rev. Bras. Ciênc. Solo, Viçosa , v. 28, n. 1, fev. 2004 .

GONÇALVES, JS. **A prática da queimada no saber tradicional e na concepção científica de risco: estudo sobre o uso do fogo por pequenos produtores do Norte de Minas Gerais**. Viçosa: UFV; 2005. 139 p.

HAIR JR., J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Análise Multivariada de Dados**. Tradução de A.S. Sant'anna e A. Cloves Neto, 5. ed. Porto Alegre: Bookman, 2006. 593 p.

HONGYU, Kuang, SANDANIELO, Vera L. M., OLIVERIRA JR., Gilmar J.. **Análise de Componentes Principais: resumo teórico, aplicação e interpretação**. E&S - Engineering and Science, v. 1, n. 5, p. 89-90. ISSN: 2358-5390.

LOTUFO, Juliana B da Silva et al. Índices Espectrais e Temperatura de Superfície em Áreas Queimadas no Parque Estadual do Araguaia em Mato Grosso. **Revista Brasileira de Geografia Física**. <https://periodicos.ufpe.br/revistas/rbgfe/article/viewFile/244539/34932> Acesso em: 13/09/2021.

MISTRY J, BIZERRIL M. **Por que é importante entender as inter-relações entre pessoas, fogo e áreas protegidas? Biodiversidade Brasileira**. 2011; 1(2): 40-49. Número Temático: Ecologia e Manejo de Fogo em Áreas Protegidas.

ROCHA, Luiz F. **Análise estatística multivariada: nova metodologia de cálculo do ICMS verde no estado do Pará**. Disponível em: <https://www.semas.pa.gov.br/wp-content/uploads/2015/11/AN%C3%81LISE-ESTAT%C3%8DSTICA-MULTIVARIADA.pdf>. Acesso em: 19/09/2021.

SARTORIO, Simone D., SCAVAZZA, Ana L, PESSOA, Cleucione de O..**Aspectos fitotécnicos da gérbera de corte por meio da análise de componentes principais**. Disponível em: https://cancri.ead.unesp.br/sigeve/paginas/baixar_trabalho_aprovado.php?id=500. Acesso em: 17/09/2021.

SILVA, A.M.C. et al. 2013. **Material particulado originário de queimadas e doenças respiratórias**.Rev. Saúde Pública, 47:345-352.

SILVA, T.E.V et al. **Análise de componentes principais aplica a avaliação discente: um estudo de caso em ambientes virtuais de aprendizagem.** Disponível em: <https://siaiap32.univali.br/seer/index.php/acotb/article/view/6552>. Acesso em : 18/09/2021.