



<http://dx.doi.org/10.21707/ga.v11.n01a11>

ANÁLISE DAS VARIÁVEIS CLIMÁTICAS DAS ESTAÇÕES AGROMETEOROLÓGICAS DO ESTADO DE SERGIPE ATRAVÉS DE MÉTODOS MULTIVARIADOS (2011 – 2013)

CRISLAINE MONIQUE ALVES DA SILVA^{1*}, ÍKARO DANIEL DE CARVALHO BARRETO², ESDRAS ADRIANO BARBOSA
DOS SANTOS³, PÉRICLES DE FARIAS BORGES⁴, LÁZARO SOUTO DE ARAÚJO⁴, EUCYMARA FRANÇA NUNES
SANTOS³

¹Graduada em Estatística. Universidade Federal de Sergipe - Centro de Ciências e Tecnologia - Departamento de Estatística e Ciências Atuariais- Cidade Universitária Prof. José Aloísio de Campos, Av. Marechal Rondon, s/n Jardim Rosa Elze - CEP: 49100-000 - São Cristóvão/SE, Brasil.

²Doutorando da Universidade Federal Rural de Pernambuco - Departamento de Informática e Estatística. Programa de Pós-Graduação em Biometria e Estatística Aplicada, Rua Dom Manoel de Medeiros, s/n, Dois Irmãos - CEP: 52171-900 - Recife/PE, Brasil.

³Docente da Universidade Federal de Sergipe - Centro de Ciências e Tecnologia - Departamento de Estatística e Ciências Atuariais- Cidade Universitária Prof. José Aloísio de Campos, Av. Marechal Rondon, s/n Jardim Rosa Elze - CEP: 49100-000 - São Cristóvão/SE, Brasil.

⁴Docente da Universidade Federal da Paraíba - Departamento de Ciências Fundamentais e Sociais - Laboratório de Matemática e Estatística-Centro de Ciências Agrárias - Campus II, Rodovia PB - 079, CEP: 58397 - 000, Areia-PB, Brasil.

*Autor para correspondência: crislaine_monique@hotmail.com

Recebido em 02 de fevereiro de 2016. Aceito em 29 de setembro de 2016. Publicado em 31 de março de 2017.

RESUMO - Neste artigo aplicamos a análise multivariada para explicar o clima no estado de Sergipe através das variáveis climáticas utilizando a técnica de componentes principais e análise de agrupamento. Utilizou-se um banco de dados com informações extraídas através do SINDA (Sistema Nacional de Dados Ambientais), coletadas no período de janeiro de 2011 a dezembro de 2013 com 6570 observações, considerando as seguintes variáveis climáticas: direção do vento, direção da velocidade dos ventos máxima, pluviosidade, pressão atmosférica, radiação solar acumulada, temperatura do ar, temperatura do ar máxima, temperatura do ar mínima, umidade relativa do ar e velocidade máxima dos ventos. Os resultados mostraram que da relação de 10 variáveis climáticas analisadas foram escolhidas apenas três: umidade relativa do ar, temperatura do ar máxima e velocidade máxima dos ventos, as quais obtiveram um percentual de explicação de 75,15%. Quanto ao agrupamento das Plataformas de Coletas de Dados, em relação às variáveis climáticas, foi possível obter três agrupamentos entre os municípios: No primeiro Estância; no segundo Itabaiana e Japaratuba e no terceiro Glória, Riachão do Dantas e Poço Redondo. Sendo considerado que para este estudo a melhor técnica de agrupamento hierárquica utilizada foi o método da Ligação Média. Para este resultado o método das médias produz melhores partições que os métodos de ligação simples e completa.

PALAVRAS-CHAVE: COMPONENTES PRINCIPAIS; AGRUPAMENTOS; CLIMA.

ANALYSIS OF CLIMATE VARIABLES BY MULTIVARIATE METHODS (2011 - 2013) IN THE AGROMETEOROLOGY STATIONS IN THE STATE OF SERGIPE, BRAZIL

ABSTRACT - In this study, the multivariate analysis was performed to explain the climate in the state of Sergipe, through climate variables, using the technique of principal components and cluster analysis. It was used a data base with information obtained from the National System of Environmental Data (SINDA), collected from January 2011 to December 2013, with 6,570 observations, considering the following climatic variables: wind direction, direction of the maximum wind speed, rainfall, atmospheric pressure, accumulated solar radiation, air temperature, maximum air temperature, minimum air temperature, air relative humidity and maximum wind speed. Of the 10 analyzed climate variables, only 3 were chosen (air relative humidity, maximum air temperature, and maximum wind speed), which had a percentage of explanation of 75.15%. Regarding the clustering of the data collection platforms, as the climate variables, three clusterings were obtained

among the municipalities: Estância in the first clustering, Itabaiana and Japaratuba in the second one, and Glória, Riachão do Dantas and Poço Redondo in the third one. The average link method was the best hierarchical clustering technique used in this study. For this result the average method produces better partitions than the single and complete link methods.

KEYWORDS: *PRINCIPAL COMPONENTS; CLUSTERING; CLIMATE.*

ANÁLISIS DE LAS VARIABLES CLIMÁTICAS DE LAS ESTACIONES AGROMETEOROLÓGICAS DEL ESTADO DE SERGIPE A TRAVÉS DE MÉTODOS MULTIVARIADOS (2011 - 2013)

RESUMEN - En este estudio se aplicó el análisis multivariado para explicar el clima en el estado de Sergipe, Brasil, a través de las variables climáticas utilizando la técnica de componentes principales y análisis de conglomerados. Se utilizó una base de datos con información extraída del Sistema Nacional de Datos Ambientales (SINDA), de enero 2011 a diciembre 2013, con 6.570 observaciones, teniendo en cuenta las siguientes variables climáticas: la dirección del viento, la dirección máxima de la velocidad de los vientos, la precipitación atmosférica, la presión atmosférica, la radiación solar acumulada, la temperatura del aire, la temperatura máxima del aire, la temperatura mínima del aire, humedad relativa del aire y velocidad máxima de los vientos. Los resultados mostraron que de las 10 variables climáticas analizadas, se eligieron solamente 3 (la humedad relativa del aire, la temperatura máxima del aire y la velocidad máxima del viento), las cuales tuvieron un porcentaje del 75,15% explicación. En cuanto al conglomerado de las plataformas de colectas de datos, con relación a las variables climáticas, se obtuvieron tres conglomerados entre los municipios: Estância en el primero, Itabaiana y Japaratuba en el segundo y Glória, Riachão do Dantas y Poço Redondo en el tercero. El método de vinculación pro medio fue la mejor técnica de conglomerado jerárquico utilizada en este estudio. Para este resultado, el método del promedio produce particiones mejores que los métodos de vinculación simples y completa.

PALABRAS CLAVE: *COMPONENTES PRINCIPALES; CONGLOMERADOS; CLIMA.*

INTRODUÇÃO

A análise climática de locais ou regiões costuma ser apresentada em termos de frequência de ocorrência de um ou de vários elementos climáticos ou pela distribuição espacial em termos do traçado de isolinhas. Esta análise visa proporcionar um melhor aproveitamento e uma visualização mais fácil dos dados, já que os resultados finais obtidos podem ser representados em formas de figuras de fácil entendimento.

Em termos climatológicos, normalmente apresentam-se tabelas de valores dos elementos: temperatura do ar, umidade relativa do ar, direção e velocidade do vento, pressão atmosférica, etc. Um exemplo disto são os dados publicados pela Superintendência do Desenvolvimento do Nordeste – SUDENE (1990), para a área do polígono das secas e pelo Departamento Nacional de Meteorologia - DNMET).

As constantes mudanças de temperatura têm ocasionado transtornos e até causado alguns prejuízos a população mundial. Medida por meteorologistas, a temperatura do ar indica mudanças causadas pela energia solar e pela superfície. São esses processos que ocorrem na atmosfera que determinam a variação climática. Esses dados climáticos podem ser coletados em estações e postos meteorológicos, uma estação de coleta de dados meteorológicos é um local adequado e bem planejado, que contém um conjunto de instrumentos aptos a fornecerem: a temperatura, a direção e velocidade dos ventos, etc (Andrade et al. 2011).

Trabalhos recentes têm dado ênfase e mostrado alguns indicadores de mudanças climáticas em escala regional e global. Constata-se que estão sendo observadas coerentes mudanças da temperatura média e da precipitação pluvial em várias partes do mundo, especialmente nos últimos 40-50 anos devido, em particular,

às melhorias tecnológicas implantadas que acarretaram uma melhora significativa na qualidade das medidas de todos os parâmetros climáticos (Alexandre et al. 2006).

A implantação das Plataformas de Coleta de Dados (PCD), em Sergipe fez do Estado o primeiro do país a oferecer, via internet, a previsão meteorológica para todos os seus 75 municípios (Andrade et al. 2011).

Assim sendo, o emprego de técnicas multivariadas seria o mais indicado, visto que uma de suas atribuições é analisar grandes conjuntos de dados referentes a diversas variáveis. Além disso, através das análises multivariadas é possível simplificar a estrutura de variabilidade dos dados (Mingoti 2005), facilitando a interpretação dos mesmos. Outra vantagem do emprego de técnicas de análise estatística multivariada, segundo Nonato et al. (2007) é a possível otimização da rede de amostragem proposta bem como da frequência de amostragem e do número de parâmetros analisados, sem perda de informação.

O agrupamento de locais em termos de elementos climáticos é importante quando da necessidade de transferência de resultados de pesquisas em agroclimatologia, visto que as recomendações para essa transferência se baseiam na existência de semelhanças climáticas entre os locais.

Neste trabalho objetivou-se agrupar e classificar os locais mais similares, baseado em observações coletadas no período de janeiro de 2011 a dezembro de 2013 de variáveis agrometeorológicas com dados diários das seis Plataformas de Coleta de Dados (PCD's) do Estado de Sergipe.

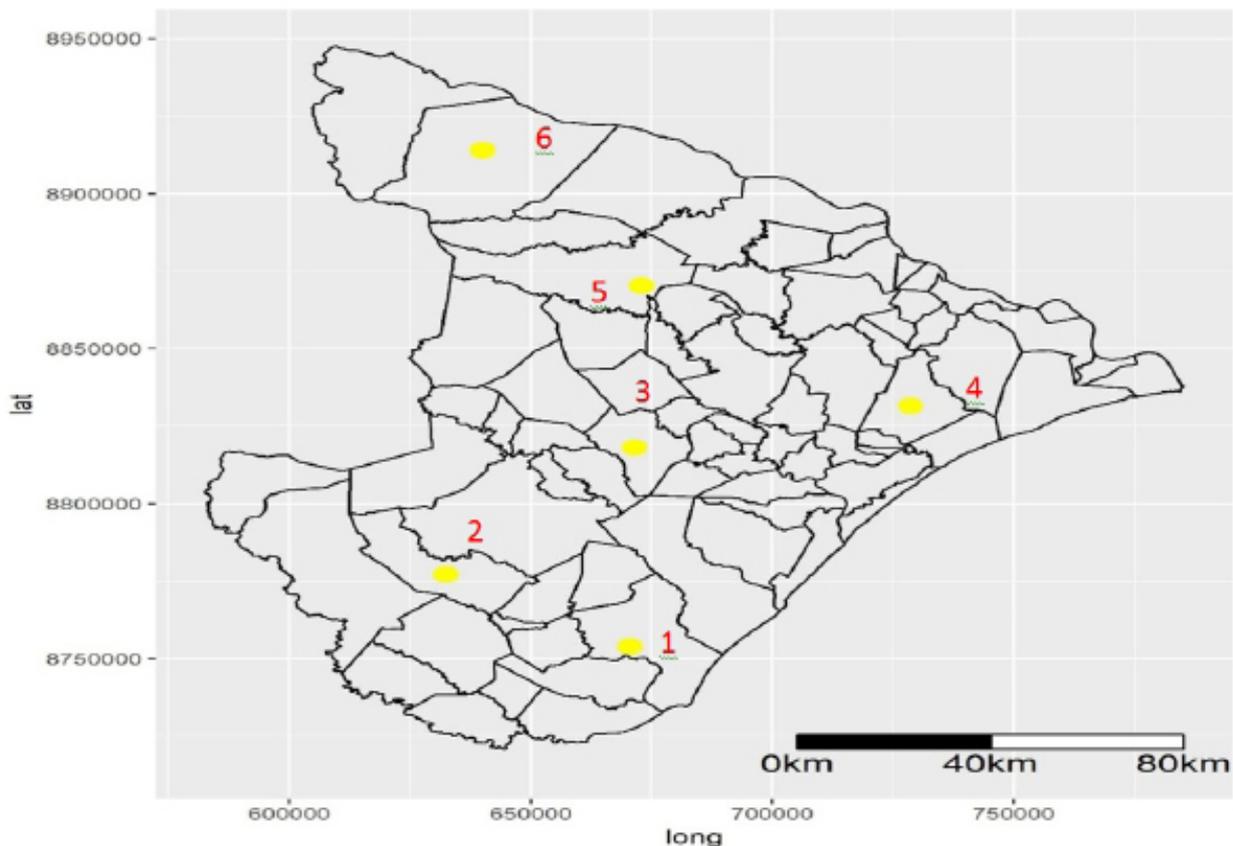
MATERIAL E MÉTODOS

As estações de monitoramento de coleta dos dados agrometeorológicos são então localizadas em pontos estratégicos do estado, sendo distribuídas conforme a Figura 1. As PCD's em estudo captam os dados em um raio de 35 km de sua localização, fazendo assim a cobertura de grande parte do estado (Tabela 1).

Tabela 1 - Cobertura das estações agrometeorológicas dos municípios do Estado de Sergipe.

ESTAÇÕES	COBERTURA
ESTÂNCIA	Estância, Itaporanga d'Ajuda, Indiaroba, Umbaúba, Arauá, Boquim, Pedrinhas e Santa Luzia.
ITABAIANA	Itabaiana, Ribeirópolis, Frei Paulo, Macambira, Moita Bonita, Malhador, Areia Branca, Campo do Brito e São Domingos.
GLÓRIA	Glória, Monte Alegre, Gararu, Graccho Cardoso, Feira Nova, Nossa Senhora Aparecida e Carira.
POÇO REDONDO	Poço Redondo, Canindé do São Francisco, Porto da Folha e porção de Monte Alegre.
RIACHÃO DO DANTAS	Riachão do Dantas, Lagarto, Simão Dias, Poço Verde, Tobias Barreto, Boquim, porção de Pedrinhas, Itabaianinha e Tomar do Geru.
JAPARATUBA	Japaratuba, Carmópolis, General Maynard, Maruim, Capela, Divina Pastora, Siriri, Japoatã, Pacatuba, Pirambu e Muribeca.

Figura 1 – Localização da distribuição das Plataformas de Coletas de Dados das estações agrometeorológicas no estado de Sergipe.



Nota: 1: Estância, 2: Riachão do Dantas, 3: Itabaiana, 4: Japarutuba, 5: Glória, 6: Poço Redondo.

Análise de Componentes Principais

A análise de componentes principais consiste em transformar um conjunto de variáveis originais em outro conjunto de variáveis de mesma dimensão denominadas de componentes principais. Essa análise é associada à ideia de redução de massa de dados, com menor perda possível da informação. Esta técnica pode ser utilizada para geração de índices e agrupamento de indivíduos (Varela 2008).

Geometricamente, as combinações lineares das variáveis originais representam a seleção de um novo sistema de coordenada obtido pela rotação do sistema original com coordenadas X_1, X_2, \dots, X_p . Os novos eixos Y_1, Y_2, \dots, Y_p representam a direção com variabilidade máxima e permite uma interpretação mais simples da estrutura da matriz de covariância (Johnson e Wichern 1998).

De modo geral, os principais objetivos da análise de componentes principais são reduzir o número de variáveis, melhorar a interpretação e analisar quais variáveis ou conjuntos de variáveis explicam a maior parte da variabilidade total, revelando que tipo de relacionamento existe entre elas. Além disso, ela é capaz de revelar informações que a princípio não se poderiam suspeitar. No entanto, segundo Johnson e Wichern (1998), esta análise funciona mais como um meio para o fim do que propriamente um fim, sendo muito útil como método auxiliar em Regressão, Análise Fatorial e Análise de Agrupamentos.

A obtenção das componentes principais depende somente da matriz de covariância " Σ " ou da matriz de correlação " ρ " de X_1, X_2, \dots, X_p , não dependendo da suposição de normalidade (Johnson and Wichern 1998).

Assim, o que ocorre quando a distribuição de probabilidades do vetor aleatório em estudo é normal p -variada, é que as componentes principais, além de serem não correlacionadas e independentes, têm distribuição normal.

A análise de componentes principais é realizada a partir da matriz de correlação quando as unidades e escalas de mensuração são diferentes e no caso de uma variável apresentar variância muito maior do que as das outras. Outro modo equivalente para solucionar este problema é, por exemplo, primeiramente padronizar ou normalizar os dados (média = 0, variância = 1) e então realizar a análise a partir da matriz de covariância.

As componentes principais são as combinações lineares não correlacionadas Y_1, Y_2, \dots, Y_p e são derivadas em ordem decrescente de importância, ou seja, a primeira componente principal (Y_1) será responsável pela maior variância contida em todas as componentes e a última (Y_p), conseqüentemente, será responsável pela menor variância restante. As variâncias de cada componente principal são na verdade os autovalores (λ) da matriz de covariância - ou correlação dependendo do caso - sendo ordenadas do maior número para o menor.

Quando a finalidade da aplicação da técnica é a redução da dimensionalidade do espaço amostral, isto é, a sumarização da informação das “ p -variáveis” originais em “ k ” componentes principais, faz-se necessário estabelecer critérios para a seleção do número “ k ”, que é o número de componentes principais a serem retidas no sistema. Geralmente são considerados os seguintes critérios (Marques 2006) e (Mingoti 2005):

1) *ScreePlot* (Cattell 1966): Por este critério, procura-se no gráfico um “ponto de salto”, que estaria representando um decréscimo de importância em relação à variância total.

2) Análise da representatividade em relação à variância total: Esta porcentagem da variância total é um valor pré-determinado pelo pesquisador, não havendo um limite definido, podendo ser escolhido de acordo com a natureza do fenômeno investigado. Em algumas situações, é possível obter-se uma porcentagem de explicação de variância total acima de 90% ou 95% com apenas 1 ou 2 componentes, enquanto que em outras, é necessário um número muito maior. Além disso, em alguns casos torna-se necessário trabalhar com porcentagens de explicação abaixo de 90%.

3) Critério de Kaiser (1958): A idéia básica do critério é manter no sistema novas dimensões que representem pelo menos a informação de variância de uma variável original.

Análise de Agrupamentos

A análise de agrupamentos, também conhecida como análise de conglomerados, classificação ou *cluster*, tem como objetivo dividir os elementos da amostra, ou população, em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis (características) que neles foram medidas, e os elementos em grupos diferentes sejam heterogêneos em relação a estas mesmas características (Mingoti 2005).

O critério básico que utilizamos para agrupar objetos são suas semelhanças. Em geral, objetos muito semelhantes são ditos similares e aqueles com poucas semelhanças, são denominados dissimilares. A semelhança entre os objetos é quantificada por meio de uma medida de proximidade, que engloba tanto as medidas de similaridades quanto as de dissimilaridade. Nas medidas de similaridades quanto maior for o valor mensurado maior vai ser a semelhança entre os objetos considerados. Nas medidas de dissimilaridade a interpretação é o oposto, ou seja, quanto maior o valor mensurado menos semelhante são os objetos e quanto menor, mais semelhantes eles são (Bussab et al.1990).

Entre as medidas de dissimilaridade a mais comumente utilizada é a distância Euclidiana, os dois elementos amostrais são comparados em cada variável pertencente ao vetor de observações.

Muitos algoritmos foram propostos com o objetivo de formar grupos, devido à inexistência de critérios universalmente aceitos. Segundo (Bussab et al.1990) todos os métodos baseiam-se em duas ideias básicas: coesão interna dos objetos e isolamento externo entre os grupos. Uma dificuldade com a qual nos deparamos para utilizar um critério único que obedeça a esses princípios é a impossibilidade computacional de examinar todos os possíveis agrupamentos, mesmo com os mais rápidos possantes computadores, se o número de objetos for moderado ou grande.

As técnicas de agrupamentos hierárquicas são realizadas por série de junções sucessivas tanto como por séries de divisões sucessivas. Os métodos aglomerativos hierárquicos começam com objetos individuais e há inicialmente tantos grupos quanto objetos.

Os objetos mais similares são agrupados inicialmente, e esses grupos fundem-se de acordo com suas similaridades. Eventualmente, abrindo o critério de similaridade os subgrupos vão se unindo a outros subgrupos até formar um grupo único (Marques 2006).

O modo de se agrupar os objetos semelhantes é realizado por meio de ligações: Ligações Simples ou Vizinho mais próximo. Neste tipo de ligação, unem-se os dois grupos com menor distância ou maior similaridade. Isto é, a distância entre dois grupos C_1 e C_2 é a menor distância entre dois elementos amostrais de cada grupo segundo Mingoti (apud Sneath 1957), Ligações Completas ou Vizinho mais distante. O procedimento adotado no caso de ligações completas é muito parecido com o caso de ligações simples, diferenciando-se apenas que a distância entre dois grupos é determinada pela distância máxima de dois elementos, uma de cada grupo. Método das Médias das Distâncias, Segundo Mingoti (2005), este método trata a distância entre dois conglomerados (ou grupos) como a média das distâncias entre todos os pares de elementos que podem ser formados com os elementos dos dois conglomerados que estão sendo comparados. Método do Centróide é direto e simples. Para fazer o agrupamento, no entanto, em cada passo é necessário voltar-se aos dados originais para o cálculo da matriz de distâncias, o que exige um tempo computacional maior do que nos outros métodos. Ao contrário dos três métodos expostos anteriormente, o método do centróide não pode ser usado em situações nas quais se dispões apenas da matriz de distâncias entre os n elementos amostrais, e Método de Ward (Marques 2003).

Coefficiente de Correlação Cofenética

Segundo França (apud Chiguti 2005) uma forma de avaliar a validade da informação gerada pela função de ligação é compará-la com os dados originais da distância. Se o agrupamento é válido, a ligação dos objetos no agrupamento tem uma forte correlação com as distâncias entre objetos no vetor de distâncias. A função cofenética compara esses dois conjuntos de valores e calcula sua correlação. A melhor solução para um agrupamento tem correlação cofenética igual a 1.

RESULTADOS

Foi obtida uma análise exploratória dos dados onde observou-se o comportamento das variáveis sem distinção de estação de monitoramento. Destacando a média das variáveis direção dos ventos (DirVento), direção

da velocidade dos ventos máxima (VentoMax), umidade relativa do ar (UmidRel), velocidade do vento máximo (VelVentoMax), pressão atmosférica (Pressaoatm), radiação solar (Radsol) e pluviosidade (Pluvio), onde podemos observar que os valores extremos da temperatura do ar (TempAr), mínimo (TempArMin) e máximo (TempArMax), influenciaram significativamente no valor da média visto na Tabela 2.

Tabela 2 - Análise descritiva das variáveis climáticas analisadas do estado de Sergipe: direção dos ventos (DirVento), direção da velocidade dos ventos máxima (VentoMax), umidade relativa do ar (UmidRel), velocidade do vento máximo (VelVentoMax), pressão atmosférica (Pressaoatm), radiação solar (Radsol) e pluviosidade (Pluvio), extremos da temperatura do ar (TempAr), mínimo (TempArMin) e máximo (TempArMax).

	N	Média	Mínimo	Máximo	Variância	Desvio Padrão	Coficiente de variação
DirVento	6570	293,57	0,00	630	49885,27	223,35	76,08
VentoMax	6570	298,41	0,00	630	50528,11	224,78	75,33
Pluvio	6570	69,63	0,20	1379	21222,29	145,68	209,2
PressaoAtm	6570	996,08	845,00	1054,60	387,82	19,69	1,98
RadSol	6570	5,62	0,00	26,69	15,47	3,933	70
TempAr	6570	25,03	18,50	33	4,38	2,09	8,36
TempArMax	6570	33,34	7,50	87,5	157,28	12,54	37,62
TempArMin	6570	24,60	9,20	87,5	219,36	14,81	60,2
UmidRel	6570	78,93	37,40	127,	92,53	9,62	12,19
VelVentoMax	6570	20,64	0,00	51,1	401,33	20,03	97,04

Os coeficientes de variações das variáveis climáticas pressão atmosférica, temperatura do ar e umidade relativa estão abaixo de 25% indicando uma homogeneidade para estas variáveis.

Ainda pela análise da estatística descritiva, verificou-se que há uma grande variabilidade dos dados coletados, fato esse observado também no valor máximo e mínimo de cada observação.

Foi previamente analisadas dez variáveis, onde nenhuma apresentou uma distribuição normal aplicando o teste de Kolmogorov-Smirnov.

Segundo (Mingoti 2005), quando a distribuição de probabilidade do vetor aleatório em estudo é normal p-variada, as componentes principais, além de não correlacionadas, são independentes e têm distribuição normal. Entretanto, a suposição de normalidade não é requisito necessário para que a técnica de análise de componentes principais possa ser utilizada.

Onde a obtenção das componentes principais envolve a decomposição da matriz de covariâncias do vetor aleatório de interesse. Além disso, se uma variável aleatória X puder ser representada pela soma de quaisquer n variáveis aleatórias independentes, então esta soma, para n suficientemente grande, terá distribuição aproximadamente normal.

Obteve-se a estimação do número de componentes principais, através da matriz de correlação das variáveis climáticas analisadas como mostrada na Tabela 3.

Observou-se na matriz de correlação obtida através da análise de componentes principais uma alta correlação positiva entre as variáveis ventomax em relação a dirvento, dirvento em relação a velventmax e temparmax em relação a temparmax. Podemos ver que existe uma correlação negativa entre as variáveis velventmax e pluvio, umidrel e temparmax, sendo correlações baixa entre as variáveis dirvento e pressaoatm, e ventomax e radsol.

Tabela 3 - Coeficientes de correlação simples das variáveis climáticas analisadas do estado de Sergipe. As variáveis podem ser consultadas na legenda da Tabela 2.

	dirvento	ventomax	pluvio	pressaoatm	radsol	tempar	temparmax	temparmin	umidrel	velventmax
dirvento	1,0000									
ventomax	0,9817	1,0000								
pluvio	-0,0791	-0,0771	1,0000							
pressaoatm	0,5552	0,5423	-0,398	1,0000						
radsol	0,6374	0,6437	-0,0139	0,4338	1,0000					
tempar	0,1027	0,0880	0,0213	0,0743	0,2830	1,0000				
temparmax	0,3338	0,3246	-0,0582	0,3544	0,3516	0,3506	1,000			
temparmin	0,3397	0,3315	-0,590	0,3575	0,3185	0,2709	0,504	1,0000		
umidrel	0,0694	0,0672	-0,0439	0,1717	-0,2609	-0,3972	0,1185	0,2164	1,0000	
velventmax	0,9747	0,9653	-0,0932	0,5453	0,6225	0,1262	0,3453	0,3447	0,866	1,0000

Foram calculados os autovetores e autovalores das variáveis climáticas através da matriz de correlação. Os autovalores em ordem decrescente definem a importância das componentes principais. Os autovalores correspondem a variância explicada por cada uma das componentes principais. Para a escolha do número de componentes, adotou-se o critério de Kaiser (Kaiser 1958), no qual o número de autovalores é igual ao número de autovalores maiores que 1.

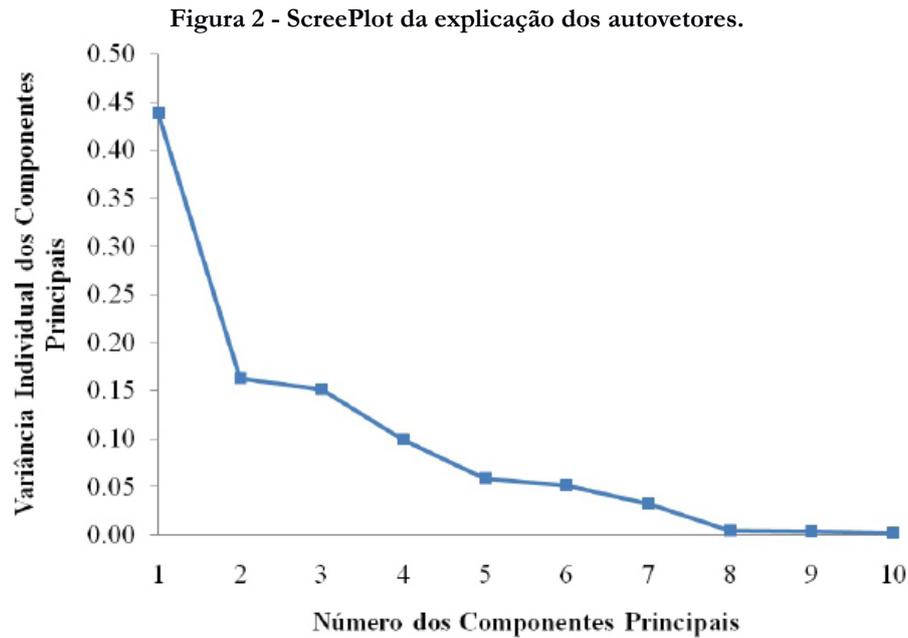
Obtiveram-se três componentes principais e a porcentagem da variância explicada por elas foi igual a 75,15%, conforme mostra destaquena Tabela 4.

Tabela 4 - Autovalores das variáveis climáticas da matriz de correlação e porcentagem de explicação das variâncias totais. Em negrito, componentes principais e a porcentagem da variância explicada por elas.

Componente Principal	Autovalor	Diferença	Proporção Individual	Proporção Acumulada
1	4,37859760	2,75393599	43,79	43,79
2	1,62466161	0,11304761	16,25	60,03
3	1,51161400	0,52426889	15,12	75,15
4	0,98734511	0,40813766	9,87	85,02
5	0,57920745	0,07265362	5,79	90,81
6	0,50655383	0,18798806	5,07	95,88
7	0,31856576	0,27369183	3,19	99,07
8	0,04487393	0,01269882	0,45	99,51
9	0,03217511	0,01576952	0,32	99,84
10	0,01640559		0,16	100

Observou-se que a partir da quarta componente principal a variância explicada é praticamente insignificante, quando comparada com as variâncias das componentes de um a três.

A Figura 2 mostra o screeplot da explicação dos autovetores das variáveis climáticas acumulados pelas componentes.



Observou-se na Tabela 5 que as variáveis climáticas que são passíveis de descarte são aquelas que possuem maior correlação das sete últimas componentes principais, como é o caso das variáveis: pluvio, pressaoatm, tempair, radsol, tempairmin, tempairmax, velventmax e dirvento, destacadas em negrito abaixo.

Tabela 5 - Coeficientes dos Componentes passíveis de descarte das variáveis climáticas.

VARIÁVEIS	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10
dirvento	0,007729	-0,188961	0,122977	-0,143801	-0,081502	0,156177	-0,796895
ventomax	0,009030	-0,204560	0,547871	-0,124808	-0,169841	0,583266	0,096239
pluvio	0,987346	-0,073258	0,015377	-0,036799	0,007053	-0,010526	0,002726
pressaoatm	0,095621	0,910740	0,014173	-0,176284	-0,009924	-0,003544	0,014250
radsol	0,017443	0,066306	-0,490068	0,724429	0,010318	-0,025478	-0,010232
tempair	-0,059215	0,095723	0,710542	0,223806	-0,048840	0,032954	0,000732
tempairmax	0,008531	-0,114607	-0,174224	-0,149194	0,670281	0,222460	-0,025136
tempairmin	0,023661	-0,139488	-0,180667	-0,111268	-0,667268	-0,222921	0,027410
umidrel	0,106607	-0,038923	0,369020	0,553729	0,047376	0,043025	-0,011097
velventmax	-0,007460	-0,199921	0,175983	-0,123175	0,255153	-0,729738	0,250945

Para a escolha das melhores variáveis climáticas, observamos na Tabela 6 que nas três primeiras componentes principais, são aquelas que tiverem maiores escores.

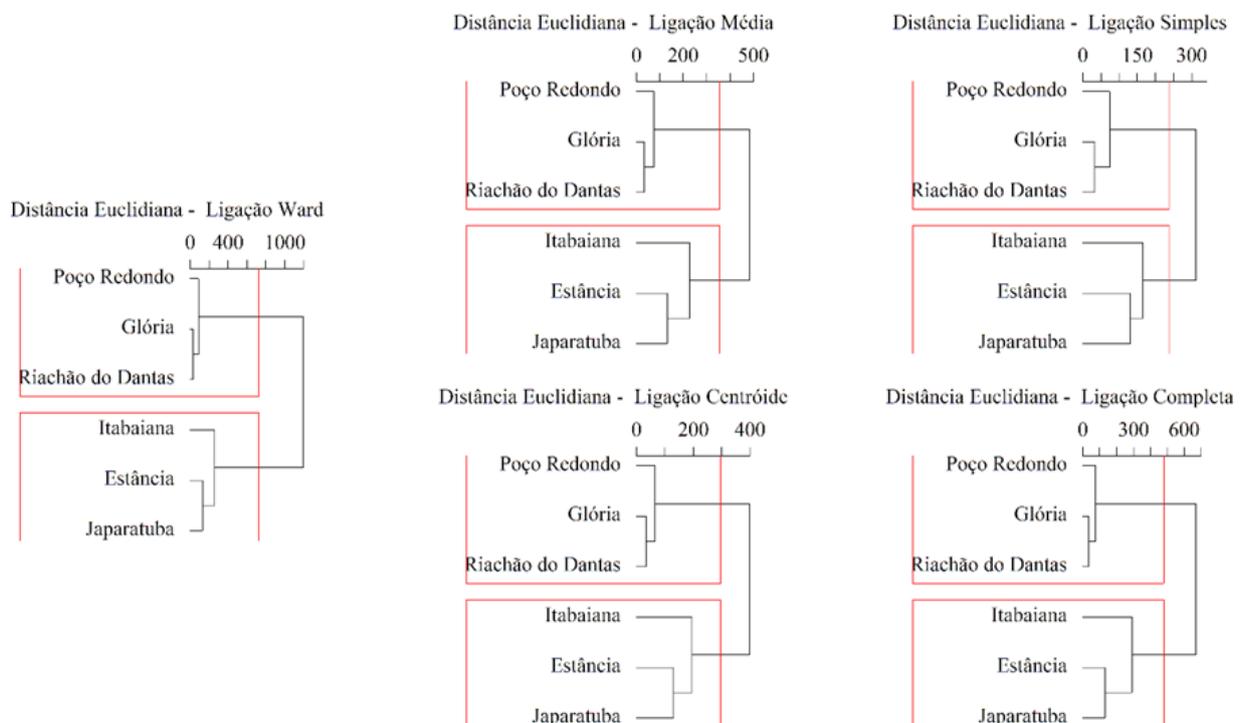
Podemos observar que na primeira componente as variáveis dirvento, ventomax e velventmax obtiveram valores próximos, porém a variável escolhida foi ventomax, pois não foi variável de possível descarte. Na segunda componente observamos tempair, tempairmax e tempairmin que são entre si altamente correlacionadas, mas tempairmax possui maior score em relação as demais. Já na terceira componente a variável escolhida foi umidrel.

Tabela 6 - Coeficientes dos três Componentes Principais das variáveis climáticas.

Variáveis	Prin 1	Prin 2	Prin 3
dirvento	0,437646	-0,262746	-0,042275
ventomax	0,434590	-0,270547	-0,044305
pluvio	-0,047445	0,032830	-0,121221
pressaoatm	0,325486	-0,071220	0,137129
radsol	0,355031	-0,017306	-0,321136
tempar	0,128457	0,458348	-0,449795
temparmax	0,299366	0,547827	0,212246
temparmin	0,296510	0,515504	0,292588
umidrel	0,032943	-0,116438	0,724891
velventmax	0,436158	-0,248638	-0,034415

A partir das variáveis encontradas na aplicação dos componentes principais calculamos a matriz de distância euclidiana, onde foi aplicados os seguintes métodos hierárquicos aglomerativos: método do vizinho mais próximo, método do vizinho mais distante, método das média das distâncias, método do centroide e método de Ward, que estão apresentados nos dendogramas da Figura 3.

Figura 3 - Análise de Agrupamentos formado pelos municípios do estado de Sergipe.



Observando os dendogramas dos métodos hierárquicos apresentados nas Figuras 3A, B e E, foi considerada a presença de dois grupos. Já com base nos dendogramas do método das médias e centroide nas Figuras 3B e

C, percebe-se a formação de três grupos, tendo em vista que na PCD de Estância formou um grupo disperso, uma vez que as cidades

abrangidas possuem baixos índices pluviométricos.

Baseando-se nas correlações cofenéticas dos agrupamentos hierárquicos apresentados na Tabela 7 pode-se avaliar que o método das distâncias proporcionou o melhor agrupamento das estações, apresentando a maior correlação cofenética (0,8828).

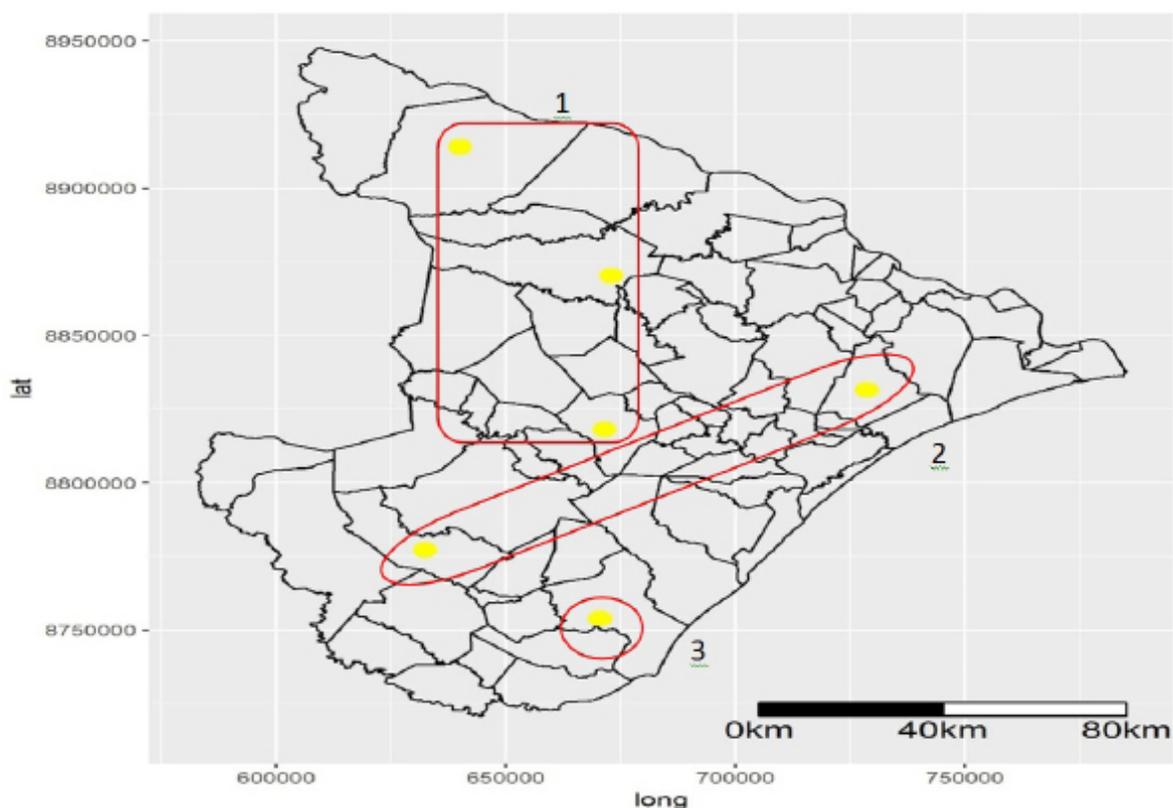
Tabela 7 - Correlação Cofenética das variáveis climáticas através de métodos de Agrupamentos

Método do Agrupamento	Correlação Cofenética
Ligação Completa	0.8823474
Ligação Simples	0.8805057
Ligação Média	0.8828067
Ward	0.8710001
Centróide	0.8823919

Este resultado concorda com Barroso e Artes (2003), que afirmam que o método das médias das distâncias produz melhores partições que os métodos de ligação simples e completa.

Observamos na Figura 4, a distribuição dos agrupamentos ou a similaridade entre as cidades onde estão localizadas as Plataformas de Coletas Dados Agrometeorológicos.

Figura 4 -Localização da distribuição das Plataformas de Coletas de Dados das estações agrometeorológicas com a formação dos agrupamentos entre os municípios.



CONCLUSÕES

Os resultados obtidos nos permitiram encontrar um conjunto de variáveis não correlacionadas que maior representassem a explicação do clima no estado de Sergipe eliminando o maior número de variáveis redundantes através da Análise de Componentes Principais. As variáveis encontradas foram umidade relativa do ar, temperatura do ar máxima e direção da

velocidade dos ventos máxima explicando 75,15% das componentes em análise. Quanto aos agrupamentos das PCD's, em relação às variáveis obtivemos três grupos: O primeiro formado por Estância; Segundo formado por: Itabaiana e Japarutuba e Terceiro formado por: Glória, Riachão do Dantas e Poço Redondo. Sendo considerado que para este estudo a melhor técnica de agrupamento hierárquica utilizada foi a da Ligação Média.

REFERÊNCIAS

Alexander L Vet al. 2006. **Global observed changes in daily climate extremes of temperature and precipitation.** Journal of Geophysical Research Atmospheres, v. 111, Issue D5, 16, DOI: 10.1029/2005JD006290.

Andrade N S, Carvalho P S e Russo S L. 2011. **Análise das variações de temperaturas na plataforma de Nossa Senhora da Glória via Box e Jenkins,** Revista Geintec. Sergipe. Universidade Federal de Sergipe. DOI: <https://doi.org/10.7198/geintec.v1i1.6>

Bussab W de O, Miazak E S e Andrade D F de. 1990. **Introdução à análise de agrupamentos.** São Paulo. Associação Brasileira de estatística. 9º Simpósio Nacional de Probabilidade e Estatística, 105 p.

França M S. 2009. **Análise estatística multivariada dos dados de monitoramento de qualidade de água da Bacia do Alto Iguaçu: uma ferramenta para a gestão de recursos hídricos.** Curitiba – PR. Dissertação (Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental) - Universidade Federal do Paraná, 150 p.

Johnson RA and Wichern DW. 1998. **Applied multivariate statistical analysis.** 4 ed. New Jersey: Prentice Hall, 776 p.

Kaiser HF. 1958. **The varimax criterion for analytic rotation in factor analysis.** Psychometrika, 23, p. 187-200.

Marques MAM. 2006. **Aplicação da Análise multivariada no estudo da infraestrutura dos serviços de saúde dos municípios paranaenses.** Curitiba-PR. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas da Universidade Federal do Paraná, 120 p.

Mingoti SA. 2005. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Editora UFMG, Belo Horizonte, 297 p.

Nonato EA, Viola ZGG, Almeida KCB e Schor HHR. 2007. **Tratamento estatístico dos parâmetros da qualidade das águas da bacia do Alto Curso do Rio das Velhas.** Química Nova, Vol. 30, Nº 4, p. 797-804.

Regulamento Técnico da OMM, Vol. I Publ. OMM, nº 49 BD2 Genebra, 1956. **Superintendenciado**

Desenvolvimento do Nordeste, 1990. Dados pluviométricos mensais do Nordeste (Afagoas). Recife, (Série pluviométrica: 7). 116p.

Secretaria de Estado do Meio Ambiente e dos Recursos Hídricos. Disponível em: <<http://www.semarh.se.gov.br/meteorologia/modules/tinyd0/index.php?id=1>>. Acesso em: 16 fev. 2014.

Sistema Integrado de Dados Ambientais. Disponível em: <<http://sinda.crn2.inpe.br/PCD/>>, acesso em 02/02/2014 às 20h37min>. Acesso em: 20 de jan. 2014.

Varella C A A. 2008. **Análise Multivariada Aplicada as Ciências Agrárias**, Pós-Graduação em Agronomia – CPGA_CS, Seropédica – RJ. Disponível em: <http://www.ufrjr.br/institutos/it/deng/varella/analise%20multivariada.htm/>, acesso em 05/01/2014 às 20h37min>. Acesso em: 20 de jan. 2014.