



Análise do comportamento da precipitação em Taubaté - SP na escala de tempo diária em comparação com os eventos de El Niño e La Niña

doi: 10.4136/ambi-agua.1421

Received: 16 Aug. 2013; Accepted: 02 Dec. 2013

Claudio Miro Maurício da Silva^{1*}; Gilberto Fernando Fisch^{1,2}

¹Programa de Pós-Graduação em Ciências Ambientais - Universidade de Taubaté –UNITAU, SP

²Instituto de Aeronáutica e Espaço, Departamento de Ciência e Tecnologia Aeroespacial (IAE/DCTA), São José dos Campos, SP

*Autor correspondente: e-mail: claudomiro.silva@hotmail.com,
fisch.gilberto@gmail.com

RESUMO

A precipitação é um elemento climático de fundamental importância no âmbito das ciências ambientais. Seu estudo é necessário, principalmente devido ao advento do tema “aquecimento global” e do alerta de muitos pesquisadores quanto ao aumento da temperatura e, conseqüentemente, da precipitação. Este trabalho teve por objetivo utilizar a técnica de análise de agrupamento, uma ferramenta da análise multivariada para agrupar a série temporal de precipitação de 1983 a 2010 (28 anos) em grupos homogêneos internamente. Foi utilizado o Coeficiente de Correlação de Pearson, como função de agrupamento, e o método de Ward como método de agrupamento e análise do comportamento do nível de fusão, para determinar o número final de grupos. A análise de agrupamento foi aplicada na série completa, e para isso os dados foram organizados por dia em forma de matrizes. Os grupos formados foram comparados com grupos de anos em que aconteceram os eventos de El Niño e La Niña, para observar se houve alguma influência desses eventos na climatologia de Taubaté e Região. A formação dos grupos revelou que não há evidências de que esses eventos influenciem no comportamento do clima de Taubaté e Região.

Palavras-chave: precipitação, análise de agrupamento, climatologia, El Niño-La Niña.

Analysis of the behavior of precipitation in Taubaté - SP in the daily time scale compared to El Niño and La Niña events

ABSTRACT

Rainfall is a fundamentally important climatic element in the environmental sciences. It is therefore necessary to study and understand it, especially with the advent of "Global Warming," in which many researchers warn of increasing temperature and as a consequence of precipitation. This study used the technique of cluster analysis, a tool for multivariate analysis, to identify internally homogeneous groups in the precipitation time series from 1983 to 2010 (28 years). We used the Pearson correlation coefficient as a clustering function, and Ward's method to group and analyze the level of fusion to determine the final number of groups. Cluster analysis was applied to the entire series, and for this, the data were organized by days in matrices. The groups formed were compared with groups of years that the events

of El Niño and La Niña occurred to see if these events had any influence on the climatology of the Taubaté region. The analysis of the groups revealed no evidence that the events of El Niño and La Niña influence the climate of Taubaté region.

Keywords: precipitation, cluster analysis, climatology, El Niño-La Niña.

1. INTRODUÇÃO

A precipitação, quando em excesso e em alto índice de intensidade, torna-se um elemento climático que pode trazer sérias consequências para a sociedade, visto que causa inundações e deslizamento de massas, em decorrência da formação de tempestades.

Nas últimas décadas tem-se presenciado um aumento considerável, não só na frequência e intensidade, mas também nos danos e prejuízos causados pelos desastres naturais. Alguns estudos indicam que esse aumento pode estar diretamente vinculado às mudanças climáticas globais. Segundo Horikoshi e Fisch (2007), estudos de simulação climática sobre os possíveis cenários futuros (2010-2099) mostram um aumento na precipitação devido ao aumento na temperatura, e total anual mais elevado, com diminuição durante o verão e leve aumento no inverno. Salienta-se de que esse estudo foi realizado com simulações climáticas efetuadas com um modelo climático global (Hadley Center) com resolução horizontal da ordem de 250 x 250 km.

O relatório do IPCC (Intergovernmental Panel on Climate Change - Painel Intergovernamental Sobre Mudanças Climáticas) de fevereiro de 2007 indica aumento da precipitação nas regiões Sul e Sudeste e secas mais intensas no norte e nordeste brasileiro para o final do século XXI. Sendo assim, é preciso estar atento quanto ao monitoramento do comportamento do clima em geral.

Segundo Ayres (2010), o início do século XXI tem sido marcado por intensas discussões a respeito das mudanças climáticas. Desde o século XVIII, com o advento da Revolução Industrial, muitas são as evidências, inclusive científicas, que cogitam a possibilidade de o planeta Terra estar em alteração climática.

Muito se tem discutido sobre essas mudanças climáticas e suas consequências, que afetarão bilhões de pessoas no mundo. Essas consequências serão das mais diversas, como: alterações no comportamento do regime de chuva, destruição de florestas, desertificação, aumento do nível do mar, aumento da poluição atmosférica, enchentes, secas, mudanças nos padrões de distribuição de doenças infecto-contagiosas, derretimento das calotas polares, entre outras.

Associada à questão do balanço de energia modificado pela variação da composição química da atmosfera, a questão do ciclo hidrológico também é importante e poderá ter modificações substanciais, em um futuro próximo.

A precipitação atmosférica constitui o ramo descendente do ciclo hidrológico, cujo regime é determinado por suas características físicas, topográficas e geológicas, que contribuem diretamente no balanço hídrico. A precipitação pluviométrica é importante devido ao fato de apresentar uma distribuição espacial e temporal com influência na evaporação, a qual reduz o escoamento superficial, retirando grande quantidade de água das superfícies líquidas, incorporando-a novamente à atmosfera (Amorim et al., 2002).

Dessa maneira, Fisch e Valério (2005) reiteram a importância do estudo do ciclo hidrológico na atualidade, podendo-se citar aplicações imediatas nas áreas agrícolas, de produção de energia elétrica e de gerenciamento hídrico para o consumo humano, e também para práticas de lazer e entretenimento.

Segundo Nunes e Calbete (2000), para um aprofundamento do estudo da variabilidade climática do vale do Paraíba paulista é interessante analisar os totais diários de chuva, para

observar qual se coloca como o nível escalar mais adequado para a análise de problemas ambientais.

De acordo com Oliveira (2010), o evento ENOS traz bastantes consequências para o Brasil, principalmente nos setores de agricultura e geração de energia, pois afetam diretamente os índices pluviométricos. Ainda segundo Oliveira (2010), o evento ENOS é um fenômeno oceano-atmosférico interanual que afeta o tempo e o clima em diversos locais do planeta, incluindo o Brasil. É caracterizado pelo Índice de Oscilação Sul (IOS), que é baseado na diferença mensal ou sazonal da pressão ao nível do mar (PNM) entre o Pacífico Central (Tahiti) e o Pacífico Oeste (Darwin/Austrália). A componente oceânica utilizada para o cálculo desse índice é representada pelo evento El Niño (La Niña), fase quente (fria) do fenômeno.

Fisch e Valério (2005), em estudo da variabilidade da precipitação em Taubaté associada aos eventos de El Niño e La Niña, concluíram que não há correspondência direta de eventos de El Niño com um aumento da precipitação e casos de La Niña com a redução.

Para mostrar os acontecimentos dos eventos de El Niño e La Niña, foram extraídas da página da internet do INPE/CPTEC as datas dos eventos e suas intensidades, organizando-as, conforme Figura 1. O que se observa é que ocorrem mais eventos de El Niño do que eventos de La Niña e que não existe regra definida para o revezamento entre eles.

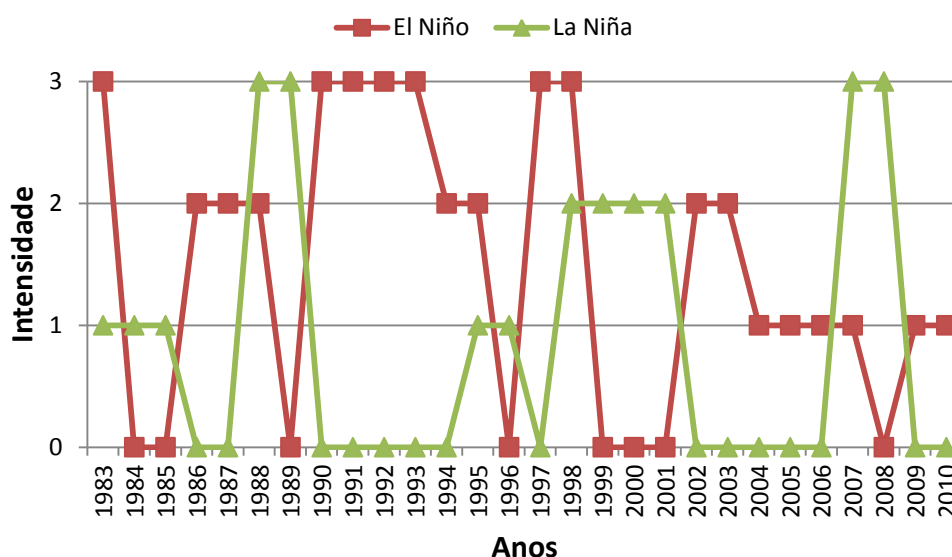


Figura 1. Eventos de El Niño e La Niña entre o período de 1983 e 2010. (Intensidade 0: Não houve evento 1: Fraco 2: Moderado 3: Forte).

Fonte: Adaptado de: http://enos.cptec.inpe.br/tab_elnino.shtml e http://enos.cptec.inpe.br/tab_lanina.shtml

De acordo com Moita Neto (2004), necessita-se de ferramentas estatísticas que apresentem uma visão mais global do fenômeno que aquela possível numa abordagem univariada. A denominação "Análise Multivariada" corresponde a um grande número de métodos e técnicas que utilizam simultaneamente todas as variáveis na interpretação teórica do conjunto de dados obtidos.

Moita Neto (2004) afirma ainda que a realidade é complexa e multivariada, necessitando de uma abordagem estatística adequada. A estatística multivariada permite uma visão global das variáveis e amostras analisadas, sendo um instrumento valioso, numa pesquisa complexa.

Mingoti (2007) define a estatística multivariada como um conjunto de métodos estatísticos utilizados em situações nas quais várias variáveis são medidas simultaneamente

em cada elemento amostral, em geral correlacionadas entre si. Quanto maior o número de variáveis, mais complexa será a análise por métodos comuns de estatística univariada.

Vários trabalhos utilizaram os métodos multivariados.

Braga et al. (1998) dividiram o estado da Bahia em 9 regiões pluviometricamente homogêneas, aplicando a técnica de análise de agrupamento e o método de Ward, e utilizando a distância euclidiana como função de agrupamento para medir a proximidade entre as estações. Para isso, utilizaram dados de 140 estações meteorológicas com 30 anos de observações, sendo feitos totais decendiais da precipitação em que cada mês foi dividido em três decêndios.

Fechine e Galvêncio (2008) determinaram as regiões pluviometricamente homogêneas no âmbito da bacia hidrográfica do rio Brígida-PE, utilizando dados de precipitação mensal de seis estações meteorológicas, de 1965 a 1985. A pesquisa consistiu em determinar o grau de aglomeração entre os dados de precipitação dos municípios daquela bacia. Para tanto, utilizaram a análise de correlações, calculando a distância euclidiana para determinar o nível de agrupamento existente entre as variáveis.

André et al. (2008), por meio da análise hierárquica de agrupamento, a orografia e a proximidade do mar, mostraram que o estado do Rio de Janeiro pode ser dividido, quanto à precipitação, em seis regiões pluviometricamente homogêneas. Para isso, utilizaram médias mensais da precipitação de 48 estações meteorológicas, em um período de 30 anos (1971-2000).

Macedo et al. (2010) utilizaram a análise de agrupamento para estudar a seca no estado da Paraíba, tendo como objetivo subdividir o estado em três subregiões pluviometricamente homogêneas.

Segundo Mingoti (2007), a técnica da análise multivariada teve o seu uso historicamente relacionado com trabalhos na Psicologia, nas Ciências Sociais e nas Ciências Biológicas, mas recentemente tem sido aplicada em um grande universo de áreas diferentes.

Para obter um resultado mais profundo na análise do comportamento da precipitação em Taubaté/SP na escala de tempo diária, neste estudo será feita a utilização do método de “Análise de agrupamento”, uma das técnicas da análise multivariada.

A análise de agrupamento, segundo Mingoti (2007), 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 forem medidas, e de forma que os elementos em grupos diferentes sejam heterogêneos em relação a essas mesmas características.

A análise de agrupamento, devido ao avanço da informática, encontra-se implementada em diversos softwares estatísticos, como SAS, STATISTICA, SPSS, MINITAB, R e IDRISI ANDES, entre muitos outros.

Uma base de dados é formada pelo conjunto de variáveis que são divididas em dois tipos: Quantitativas (Métricas) e Qualitativas (Não Métricas).

Uma variável é quantitativa quando seus valores são expressos em números, ou seja, os indivíduos podem ser identificados como diferenciados em quantia e grau, e pode ser classificada em discreta ou contínua. A variável quantitativa discreta é uma representação numérica cujo valor está contido em um intervalo razoável e pequeno. A variável quantitativa contínua pode assumir magnitudes diferentes em escalas contínuas de medidas, por exemplo, altura e peso. Já uma variável qualitativa é aquela resultante das classificações não quantitativas das observações, podendo ser feitas nas escalas nominal ou ordinal. São atributos, características ou propriedades categóricas que identificam ou descrevem um objeto. A escala nominal fornece o número de ocorrências em cada classe ou categoria da variável em estudo. Logo, os números ou símbolos designados aos objetos não têm

significado quantitativo, como exemplo, o tipo de religião ou partido político de um indivíduo. A escala ordinal refere-se ao próximo nível de precisão em medida. As variáveis podem ser ordenadas com escalas ordinais em relação à quantia do atributo possuída. As subclasses podem ser comparadas com outras em termos de uma relação como “maior que” ou “menor que”. Por exemplo, o nível de satisfação de um consumidor em relação a diversos produtos (Alves, 2007).

Neste trabalho os dados são exclusivamente quantitativos (métricos), tratando de medidas de precipitação expressas em mm (milímetros).

2. MATERIAL E MÉTODOS

2.1. Área de estudo e dados

O local da coleta de dados deste estudo foi o Departamento de Ciências Agrárias da Universidade de Taubaté, que está situado na parte sudeste do município de Taubaté, SP, entre as confluências do rio Una e um dos seus afluentes, o ribeirão Itaim. O rio Una é um dos principais afluentes da margem direita do rio Paraíba do Sul, no Estado de São Paulo. Ocorre, na área do Departamento de Ciências Agrárias da UNITAU, relevo colinoso desenvolvido sobre os sedimentos terciários da Bacia Sedimentar de Taubaté. Os solos são constituídos por espessos latossolos vermelho-amarelos (LVA40), em relevo suavemente ondulado (Oliveira et al., 1999 apud Diniz, 2008).

As coordenadas geográficas do posto meteorológico do Departamento de Ciências Agrárias são 23° 02' S, 45° 30' W e 577 m, localizado na Fazenda Piloto, na estrada municipal Dr. José Luiz Cembranelli, 5000, na zona rural do município.

Todos os dados foram formatados por dia, da seguinte forma: precipitação das 12 horas, como sendo a precipitação ocorrida das 00h01 até às 12h; precipitação das 24 horas, sendo a precipitação ocorrida das 12h01 até 00h e, precipitação total, sendo a soma da precipitação das 24 horas do dia anterior com a precipitação das 12 horas do dia, para compor a precipitação de 1 (um) dia.

A precipitação diária i será considerada como a chuva que ocorre entre 12h do dia i e 12h do dia $i + 1$. As leituras diárias foram efetuadas nos horários preconizados pela OMM (Organização Mundial de Meteorologia), nesse caso, às 9h, 15h e 21h, até o ano de 2008. A partir de 2009 o horário das 15h foi interrompido, por falta de pessoal (observador meteorológico).

Antes de serem submetidos à análise de agrupamento, os dados originais, coletados por meio do pluviômetro convencional Ville de Paris (VP), sofreram correções de +26,8%, em decorrência de subestimação em relação ao pluviômetro automático Tipping Bucket Rain Gauge (TBR), em estudo que contou com a comparação dos dados de precipitação do pluviômetro convencional VP e o automático TBR, no período de novembro de 2009 a fevereiro de 2011.

2.2. Aplicação da técnica de análise de agrupamento na série temporal de dados

A proposta deste estudo vem da necessidade de se conhecer o comportamento da chuva na região, utilizando métodos que levem a utilizar o tratamento da precipitação no menor nível possível de chuva acumulada, nesse caso a chuva acumulada no período de um (1) dia.

Dessa forma, a preparação dos dados de precipitação se deu em forma de matriz em que os dias (1 a 365) representaram as linhas e os anos (1983 a 2010) representaram as colunas. Para os anos bissextos, os valores do dia 29 de fevereiro foram divididos ao meio, e metade foi somada ao dia 28 de fevereiro e a outra metade somada ao dia 01 de março, obtendo-se dessa forma a matriz completa $A_{365 \times 28}$ com toda a série temporal de precipitação.

Os métodos utilizados para o tratamento dos dados utilizando a abordagem da Análise de Agrupamento foram baseados em Mingoti (2007), que propõe uma abordagem clara e prática sobre esse assunto.

Suponha-se que se tenha disponível um conjunto de dados constituídos de n elementos amostrais, tendo-se medidos p variáveis aleatórias em cada um deles. O objetivo é agrupar esses elementos em g grupos. Para cada elemento amostral j , tem-se, portanto, o vetor de medidas X_j definido por:

$$X_j = [X_{1j} \ X_{2j} \ \dots \ X_{pj}], \quad J = 1, 2, \dots, n$$

em que: X_{ij} representa o valor observado da variável i medida no elemento j .

Uma questão importante refere-se ao critério a ser utilizado para se decidir até que ponto dois elementos do conjunto de dados podem ser considerados como semelhantes ou não. Para isso é necessário considerar medidas que descrevam a similaridade entre elementos amostrais de acordo com as características que neles foram medidas. Segundo Mingoti (2007), considerando-se que para cada elemento amostral têm-se informações de p variáveis armazenadas em um vetor, a comparação de diferentes elementos amostrais poderá ser feita por meio de medidas matemáticas (métricas) que possibilitem a comparação de vetores, como as medidas de distância. Assim, podem-se calcular as distâncias entre os vetores de observações dos elementos amostrais e agrupar aqueles de menor distância.

Dessa forma, o conjunto de dados de precipitação é composto pelos elementos amostrais (dias) e variáveis (anos).

De acordo com Alves (2007), existem várias medidas apropriadas para variáveis quantitativas, como: Distância Euclidiana, Distância Generalizada ou Ponderada, Distância de Minkowsky, Coeficiente de Concordância Simples, Coeficiente de concordância Positiva, Coeficiente de Concordância de Jaccard e Distância Euclidiana Média.

De acordo com Naldi (2011), é possível também utilizar o Coeficiente de Correlação de Pearson, o qual será utilizado neste trabalho. Segundo Naldi (2011), para essa medida os objetos são considerados vetores espaciais ou sequências temporais com n_a componentes. O cálculo é feito a partir da média do vetor, dado pela seguinte equação:

$$\bar{X}_i = \frac{1}{n_a} \sum_{z=1}^{n_a} X_i(z)$$

O coeficiente é calculado pela seguinte equação:

$$r_p = \frac{\sum_{m=1}^d (x_{im} - \bar{x}_i) - (x_{ij} - \bar{x}_j)}{\left(\sum_{m=1}^d (x_{im} - \bar{x}_i)^2 \sum_{l=1}^d (x_{jl} - \bar{x}_j)^2 \right)^{\frac{1}{2}}}$$

A distância de Pearson é dada por $S_{ij} = 1 - r_p$, o que faz com que S_{ij} assumam valores entre 0 e 2.

Dessa maneira, os dois elementos amostrais são comparados em cada variável pertencente ao vetor de observações.

O método hierárquico parte do princípio de observar cada elemento do conjunto de dados como um conglomerado isolado, até o ponto de todos serem considerados num único grupo. No estágio inicial, tem-se a partição com a menor dispersão interna (variância igual a zero), e no final ela atinge sua maior dispersão interna, dado que nesse estágio estão contidos todos os

elementos amostrais. Os passos principais para a aplicação da técnica são resumidos por Mingoti (2007). O primeiro passo relata que cada elemento constitui um cluster tamanho “um”. Logo, existem n clusters. No próximo passo, indica-se que apenas um novo conglomerado pode ser formado em cada passo, e que seu tamanho é diminuído a cada um deles. O terceiro é que, se em algum estágio do processo dois elementos amostrais aparecem juntos num cluster, eles aparecerão juntos em todos outros subsequentes (propriedade de hierarquia). O último passo consiste na possibilidade de construir, devido à propriedade de hierarquia (Timm, 2002 apud Mingoti, 2007), um gráfico conhecido como dendrograma, em forma de árvore, no qual o nível de similaridade (ou dissimilaridade) é indicado na escala vertical. No eixo horizontal são apresentados os elementos amostrais na ordem conveniente relacionada à história de agrupamento. As linhas verticais, partindo dos elementos amostrais agrupados, têm altura correspondente ao nível em que os elementos foram considerados semelhantes, isto é, a distância do agrupamento ou o nível de similaridade.

Existem vários métodos de agrupamentos hierárquicos, e a maioria deles está disponível na maioria dos softwares estatísticos, como: Método de Ligação Simples (Single Linkage), Método de Ligação Completa (Complete Linkage), Método da Média das Distâncias (Average Linkage), Método do Centróide (Centroid Method) e Método de Ward. O objetivo é obter grupos heterogêneos com elementos homogêneos entre si.

O método utilizado nesse estudo é o de Ward. Braga et al. (1998), André et al. (2008) e Fachine e Galvíncio (2008) optaram por esse método pela razão de haver apresentado melhores resultados. Alves (2007) acrescenta que esse método é apropriado para variáveis quantitativas, e Mingoti (2007) acrescenta que, para usá-lo, basta que as p variáveis sejam quantitativas e passíveis, portanto, do cálculo da média.

2.3. O Método de Ward

Em 1963, Ward propôs um método de agrupamento fundamentado nos princípios da “Mudança de Variação” entre os grupos e dentro dos grupos que se formam em cada passo do agrupamento. Conforme Naldi (2011), o algoritmo de agrupamento de variância mínima de Ward é um dos métodos de agrupamento hierárquicos mais conhecidos. Assim como outros algoritmos, tem como principal objetivo a redução da variância dos grupos formados. Por ser aglomerativo, o algoritmo é iniciado com grupos de um único objeto e termina com um único grupo contendo todo o conjunto de dados. A cada passo, ele aglomera a dupla de grupos que resultará no menor incremento da variância dos grupos, ou seja, o menor aumento na soma das dissimilaridades entre os objetos do grupo e seu centróide. Esse índice também é minimizado pelo algoritmo K-médias e é dado pela seguinte equação:

$$\min_{c_1, \dots, c_k} obj = \sum_{i=1}^k \sum_{x_j \in C_i} d(X_j, C_i)$$

Para escolher qual dupla será aglomerada, o algoritmo tentará todas as aglomerações de duplas possíveis, calculará o valor da variância para cada aglomeração e selecionará o menor. O processo é repetido até que um único grupo seja formado.

2.4. A escolha do número ideal de grupos

De acordo com Mingoti (2007), uma questão importante é como se deve proceder para escolher o número final de g grupos que define a partição do conjunto de dados analisado, ou, de outra forma, em qual passo k o algoritmo de agrupamento deve ser interrompido. Não existe uma resposta exata, entretanto existem alguns critérios que podem auxiliar na decisão final, como: Análise do comportamento do nível de fusão (distância), Análise do

comportamento do nível de similaridade, Análise da soma de quadrados entre grupos (coeficiente R^2), Estatística Pseudo F, Correlação semiparcial (Método de Ward), Estatística Pseudo T^2 e Estatística CCC (Cubic Clustering Criterium).

Mingoti (2007) ressalta que se for feito o agrupamento dos dados amostrais pelo método de Ward, o critério do coeficiente de Correlação Semiparcial para a escolha do número de grupos será equivalente à aplicação do critério de Análise de Agrupamento do Nível de Fusão (Distância).

Assim sendo, será adotado neste trabalho o critério de Análise do Comportamento do Nível de Fusão (distância) para a escolha do número de grupos.

Então, o critério de Análise do Comportamento do Nível de Fusão se dá da seguinte maneira: à medida que se avança no algoritmo de agrupamento, ou seja, que se passa do estágio k para o estágio $k+1$, a similaridade entre os conglomerados que estão sendo combinados nos respectivos passos vai decrescendo. Consequentemente, a distância entre eles vai aumentando. Desse modo, se for feito um gráfico do passo (ou números de grupos) versus o nível de distância (nível de fusão) do agrupamento de cada estágio do processo, será possível visualizar se há “pontos de salto” relativamente grandes em relação aos demais valores de distância. Esses pontos indicam o momento ideal de parada do algoritmo, isto é, o número de conglomerados final g e a composição final dos grupos. Logo, se a função apresentar vários “pontos de salto”, pode-se delimitar uma região de prováveis valores do número de grupos g que deveriam ser mais bem investigados por algum outro procedimento. Quando n não for muito grande, o dendrograma também poderá ser utilizado como uma ferramenta para a visualização dos pontos de salto das distâncias, ou da perda acentuada de similaridade e dos grupos formados.

3. RESULTADOS E DISCUSSÃO

3.1. Aplicação da técnica de “análise de agrupamento”

O software utilizado foi o MINITAB versão 1.4. O método de agrupamento adotado foi de Ward, e o critério para escolha do número final de grupos foi o de Análise do comportamento do Nível de Fusão, sugerido por Mingoti (2007), e a medida adotada para o cálculo da distância, o Coeficiente de correlação de Pearson.

A análise utilizou a série temporal completa representada pela matriz $C_{365 \times 28}$.

Os grupos formados foram comparados aos anos em que aconteceram os eventos de El Niño e La Niña, como forma de observar a influência desses eventos na formação dos diversos grupos.

3.2. Critério para a divisão dos grupos

A Tabela 1 mostra os passos, os números de grupos no determinado passo, o nível de similaridade e o nível de distância entre os grupos em cada passo.

Tabela 1. Resultado da aplicação do método de Ward de agrupamento na série temporal de 1983 a 2010.

Passos	Número de	Nível	Nível	Grupos	Novo	Número de
	Grupos	Similaridade	Distância	Unidos	Grupo	No novo Grupo
1	27	66,4187	0,67163	18 24	18	2
2	26	62,4847	0,75031	4 9	4	2
3	25	61,2978	0,77404	5 22	5	2
4	24	61,2834	0,77433	11 27	11	2
5	23	60,8301	0,7834	20 26	20	2
6	22	60,0774	0,79845	6 14	6	2
7	21	59,7664	0,80467	18 28	18	3
8	20	59,2026	0,81595	13 16	13	2
9	19	58,9262	0,82148	3 23	3	2
10	18	57,8793	0,84241	12 15	12	2
11	17	57,199	0,85602	4 17	4	3
12	16	56,4655	0,87069	7 11	7	3
13	15	55,3758	0,89248	8 19	8	2
14	14	54,8169	0,90366	3 25	3	3
15	13	53,6339	0,92732	12 21	12	3
16	12	53,3364	0,93327	2 20	2	3
17	11	52,2429	0,95514	1 6	1	3
18	10	51,9034	0,96193	5 12	5	5
19	9	51,7015	0,96597	8 10	8	3
20	8	47,9145	1,04171	13 18	13	5
21	7	46,9251	1,0615	2 5	2	8
22	6	45,9113	1,08177	1 3	1	6
23	5	43,8462	1,12308	4 8	4	6
24	4	41,4741	1,17052	1 13	1	11
25	3	41,1772	1,17646	4 7	4	9
26	2	34,5166	1,30967	1 2	1	19
27	1	32,9107	1,34179	1 4	1	28

Ao final de todos os estágios do algoritmo restará somente um grupo. Nesse momento interrompe-se o processo, gerando assim o número de grupos de seu interesse.

A Figura 2 mostra a formação dos grupos ao final de todos os estágios do algoritmo. É difícil perceber o momento de interromper o algoritmo para a geração de uma quantidade de grupos que seja o mais ideal.

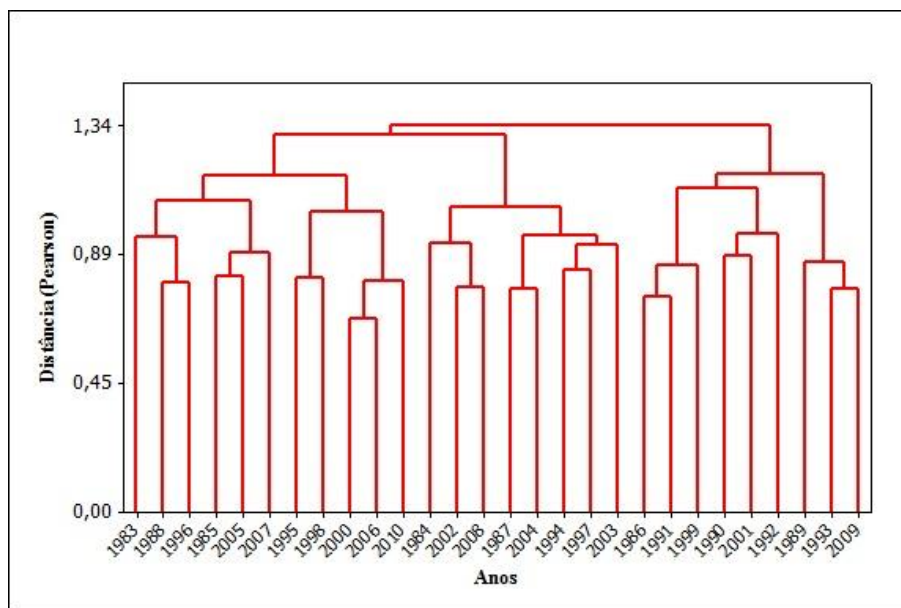


Figura 2. Dendrograma do agrupamento da série temporal de 1983 a 2010.

Utilizando a análise do comportamento do nível de fusão (distância) como critério para a escolha dos grupos, observa-se, na Figura 3, um pequeno salto no passo 19 e um salto maior no passo 25. Tem-se, então uma forte indicação para agrupar os dados da série temporal completa em nove grupos (passo 19) ou em três grupos (passo 25), observando-se a Tabela 1.

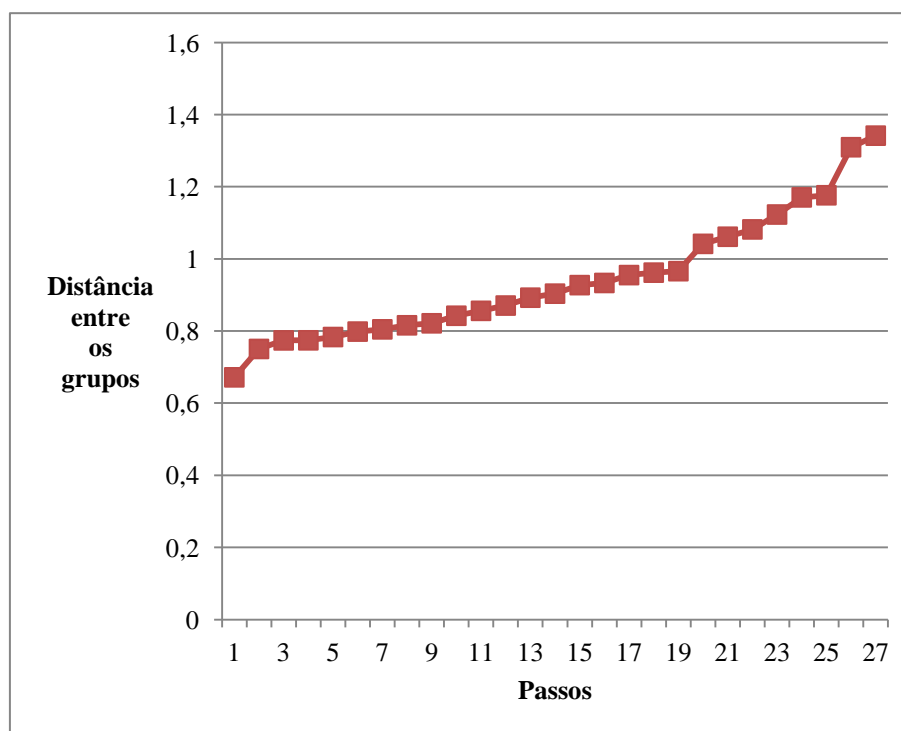


Figura 3. Pontos de salto, passos X nível de fusão aplicado na série temporal de 1983 a 2010.

3.3. Divisão da série temporal completa em grupos

Seguindo a indicação do método, pode-se observar, na Figura 4, a formação dos nove grupos e, na Figura 5, a formação dos três, para a série temporal.

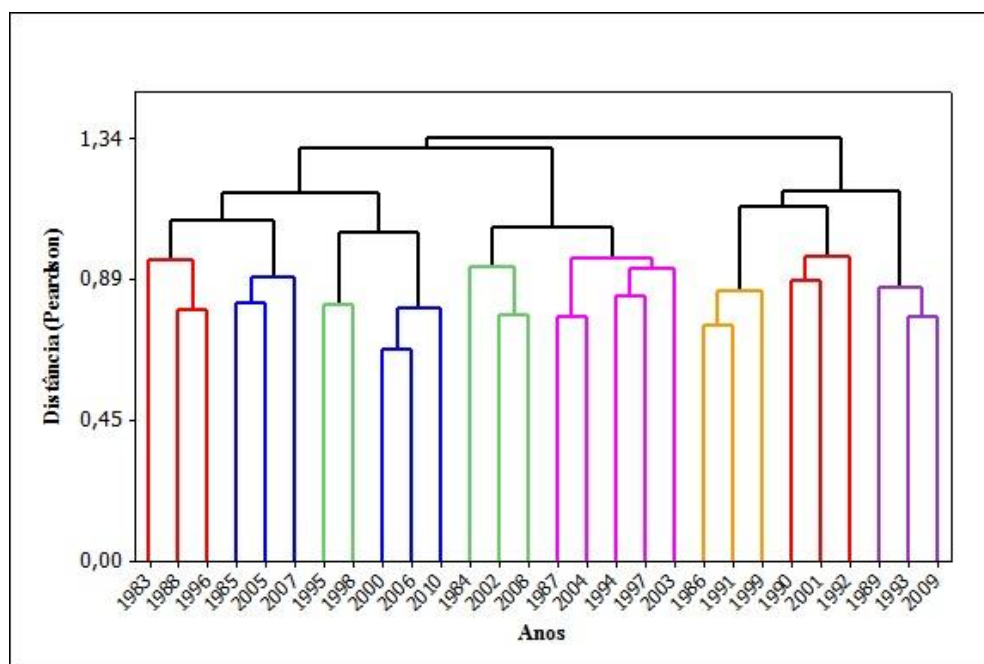


Figura 4. Dendrograma do agrupamento da série temporal de 1983 a 2010, em nove grupos.

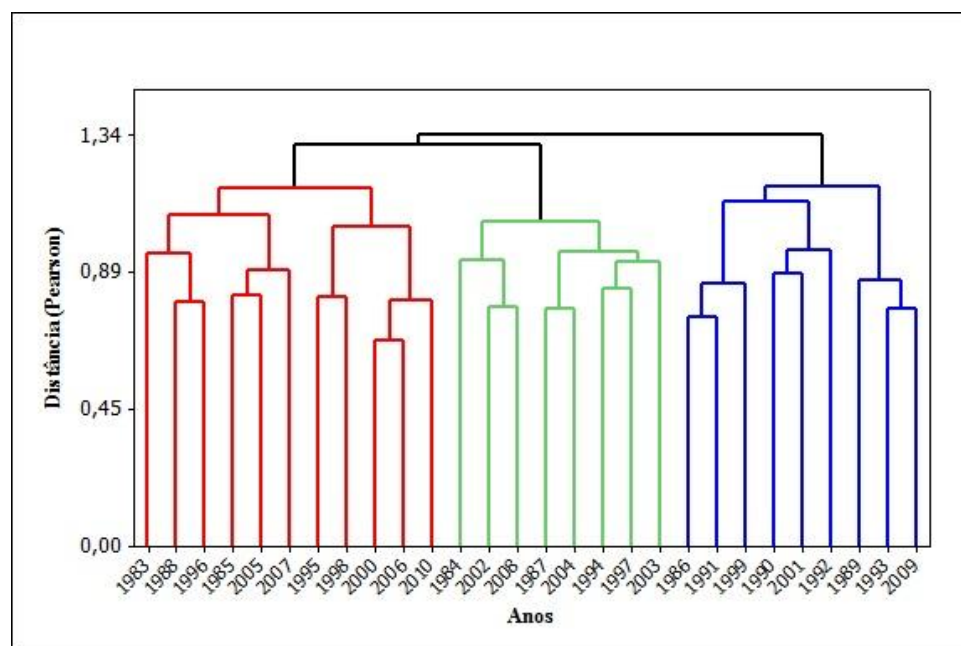


Figura 5. Dendrograma do agrupamento da série temporal de 1983 a 2010, em três grupos.

3.4. Análise da formação dos grupos

Procurou-se, neste estudo, uma resposta para a formação dos diversos grupos após aplicação da análise de agrupamento nos dados da série completa de 1983 a 2010. Tem-se

falado muito, atualmente, sobre os eventos El Niño e La Niña e tem-se procurado observar as relações desses eventos com o comportamento do clima em diversas regiões.

Fisch e Valério (2005), utilizando dados acumulados por mês, fizeram um estudo relacionado com esse tema e não encontraram evidências de que esses eventos influenciasssem o comportamento da precipitação em Taubaté e na região do vale do Paraíba.

Com os dados de precipitação no nível inferior de dados acumulados, nesse caso em um dia, é feito neste estudo uma comparação com esses eventos, podendo dessa maneira observar, com essa técnica utilizada, se há evidências da influência dos eventos de El Niño e La Niña no comportamento da precipitação em Taubaté e região.

A Tabela 2 mostra os anos em que ocorreram os eventos de El Niño e os anos em que ocorreram eventos de La Niña. Alguns anos ocorreram ambos os eventos, como no caso de 1983, 1988, 1995, 1998 e 2007. Na Tabela 5 pode-se observar a formação dos três grupos da série temporal completa, e na Tabela 6 observa-se o percentual do quanto cada grupo foi influenciado pelos eventos de El Niño ou La Niña.

Para obter um resultado favorável à influência do evento na formação do grupo, o ideal é que cada grupo seja responsável por pelo menos um acontecimento de 100% no evento El Niño ou no evento La Niña, em sua formação.

Para efeito de exemplo, observe-se, na Tabela 3 e na Tabela 4, o efeito da influência desses eventos na formação dos grupos ideais. Os percentuais na formação do grupo são obtidos com a divisão da quantidade de elementos encontrados de cada evento pela quantidade de elementos do grupo.

Conforme a Tabela 3, observam-se: a quantidade do grupo 01(Qg01)= 5, a quantidade do grupo 02 (Qg02)= 8 e a quantidade do grupo 03(Qg03)=15. Tem-se também: Elementos do grupo 01 pertencentes a El Niño (G01elniño) = 5, Elementos do grupo 01 pertencentes a La Niña (G01laniña) = 5, Elementos do grupo 02 pertencentes a El Niño (G02elniño)=0, Elementos do grupo 02 pertencentes a La Niña (G02laniña)=8, Elementos do grupo 03 pertencentes a El Niño (G03elniño)=15 e Elementos do grupo 03 pertencentes a La Niña (G03laniña)=0.

Na Tabela 4 observam-se os grupos formados em relação ao evento de El Niño e La Niña: Grupo 01 (El Niño) = G01elniño/Qg01, Grupo 01 (La Niña) = G01laniña/Qg01, Grupo 02 (El Niño) = G02elniño/Qg02, Grupo 02 (La Niña) = G02laniña/Qg02, Grupo 03 (El Niño) = G03elniño/Qg03 e Grupo 03 (La Niña) = G03laniña/Qg03.

Entretanto, é preciso que se obtenha um número máximo de contribuições de 100% de eventos El Niño ou de eventos La Niña na formação dos grupos, para assumir que o grupo sofreu influência desses eventos em sua formação.

Tabela 2. Eventos de El Niño e La Niña no período de 1983 a 2010.

El Niño	La Niña
1983, 1986, 1987, 1988, 1990, 1991, 1992, 1993, 1994, 1995, 1997, 1998, 2002, 2003, 2004, 2005, 2006, 2007, 2009, 2010	1983, 1984, 1985, 1988, 1989, 1995, 1996, 1998, 1999, 2000, 2001, 2007, 2008

Tabela 3. Exemplo da Formação Ideal de grupos sob influência dos eventos de El Niño e La Niña.

Grupo 01	Grupo 02	Grupo 03
1983, 1988, 1995, 1998, 2007	1984, 1985, 1989, 1996, 1999, 2000, 2001, 2008	1986, 1987, 1990, 1991, 1992, 1993, 1994, 1997, 2002, 2003, 2004, 2005, 2006, 2009, 2010

Tabela 4. Exemplo do Percentual da Contribuição dos eventos de El Niño e La Niña na formação dos grupos ideais.

	El Niño	La Niña
Grupo 01	100,0%	100,0%
Grupo 02	0,0%	100,0%
Grupo 03	100,0%	0,0%
Geral	100,0%	

Dessa forma, observando-se a Tabela 5, têm-se os anos que formaram os três respectivos grupos da série temporal, de 1983 a 2010.

Observando-se a Tabela 6, verifica-se que o percentual de 72,7% na contribuição da formação do grupo 01 corresponde a 8 elementos de cada grupo, pertencendo tanto ao grupo dos eventos de El Niño (1983, 1988, 1995, 1998, 2005, 2006, 2007 e 2010) quanto ao grupo dos eventos de La Niña (1983, 1985, 1988, 1995, 1996, 1998, 2000 e 2007).

Dessa forma, observa-se que não se conseguiu em nenhuma formação a participação de 100% na contribuição, tanto do evento de El Niño, quanto no evento de La Niña, na formação dos grupos, considerando-se que os anos em que ocorreram ambos os eventos (1983, 1988, 1995, 1998 e 2007), 100% estão presentes no grupo 01.

Da mesma forma, na formação do grupo 02 os 75% de contribuição dos eventos de El Niño se deu pelo fato de 6 dos 8 elementos do grupo 02 (1987, 1994, 1997, 2002, 2003 e 2004) estarem contidos nos anos dos eventos de El Niño e, por outro lado, somente 2 elementos (anos) do grupo 02 (1984 e 2008) estão contidos nos anos dos eventos de La Niña.

Na formação do Grupo 03, dos 9 elementos do grupo 6 pertencem aos anos de El Niño (1986, 1990, 1991, 1992, 1993 e 2009), e 3, aos anos de La Niña (1989, 1999 e 2001).

No geral, calculando a média dos percentuais da contribuição dos eventos na formação dos grupos em relação ao número de grupos, obtém-se a média de 71,5%, demonstrando-se assim o percentual da influência dos eventos de El Niño e La Niña na formação dos três grupos da série temporal completa de 1983 a 2010.

Essa média de 71,5% foi obtida dividindo-se a soma dos maiores percentuais encontrados em cada grupo por 300, que é a soma máxima dos percentuais dos 3 grupos $((72,7+75+66,7)/300=71,5)$.

Tabela 5. Formação conforme a sugestão dos três grupos para a série temporal de 1983 a 2010.

Grupo 01	Grupo 02	Grupo 03
1983, 1985, 1988, 1995, 1996, 1998, 2000, 2005, 2006, 2007, 2010	1984, 1987, 1994, 1997, 2002, 2003, 2004, 2008	1986, 1989, 1990, 1991, 1992, 1993, 1999, 2001, 2009

Tabela 6. Contribuição dos eventos de El Niño e La Niña na formação dos três Grupos da série temporal de 1983 a 2010.

	El Niño	La Niña
Grupo 01	72,7%	72,7%
Grupo 02	75,0%	25,0%
Grupo 03	66,7%	33,3%
Geral	71,5%	

Como o modelo sugeriu também uma divisão da série temporal completa em nove grupos, pode-se observar, na Tabela 7, a formação desses nove grupos e, na Tabela 8, o

percentual de contribuição dos eventos de El Niño e La Niña na formação desses grupos, de acordo com o mesmo critério adotado para a formação dos três grupos.

Observando-se a Tabela 8, não se pode afirmar com convicção que houve influência dos eventos de El Niño e La Niña na formação dos nove grupos (em 100%), mas ocorreu na formação de três grupos (Grupo 01, Grupo 05 e Grupo 08). Os outros seis grupos ficaram com índices de 66,7%, suficientes para aumentar o índice geral. Assim, calculando-se a média dos percentuais da contribuição dos eventos na formação dos grupos em relação ao número de grupos, obtém-se a média de 77,8%. Esses seis grupos são compostos por 3 elementos (anos) cada, e o modelo conseguiu agrupar 2 elementos (anos), ora sendo de El Niño, ora sendo de La Niña, em um mesmo grupo. O ideal seria se o modelo separasse exatamente anos de El Niño e anos de La Niña em cada grupo, portanto essa média geral de 77,8% demonstra também uma influência pequena dos eventos de El Niño e La Niña na formação dos nove grupos da série temporal completa de 1983 a 2010.

Tabela 7. Formação conforme a sugestão dos nove grupos para a série temporal de 1983 a 2010.

Grupo 01	Grupo 02	Grupo 03	Grupo 04	Grupo 05	Grupo 06	Grupo 07	Grupo 08	Grupo 09
1983	1984	1985	1986	1987	1989	1990	1995	2000
1988	2002	2005	1991	1994	1993	1992	1998	2006
1996	2008	2007	1999	1997	2009	2001		2010
				2003				
				2004				

Tabela 8. Contribuição dos eventos de El Niño e La Niña na formação dos nove Grupos para a série temporal de 1983 a 2010.

	El Niño	La Niña
Grupo 01	66,7%	100,0%
Grupo 02	33,3%	66,7%
Grupo 03	66,7%	66,7%
Grupo 04	66,7%	33,3%
Grupo 05	100,0%	0,0%
Grupo 06	66,7%	33,3%
Grupo 07	66,7%	33,3%
Grupo 08	100,0%	100,0%
Grupo 09	66,7%	33,3%
Geral		77,8%

Mesmo com menos elementos em cada grupo, não foi possível encontrar uma quantidade maior de grupos com a participação de 100% dos eventos em suas formações.

Dessa maneira, fica claro que o regime de precipitação em Taubaté e região tem sua própria característica e que, portanto, não sofre influência dos eventos de El Niño ou dos eventos de La Niña, concordando-se com Fisch e Valério (2005) que concluíram o mesmo usando apenas os totais mensais de precipitação, além de uma série temporal menor.

Grim e Ferraz (1998 apud Fisch e Valério, 2005) já alertavam para o fato de que a resposta dos eventos El Niño e La Niña na região Sudeste podem não ser tão lineares quanto nos extremos Sul e Norte do Brasil, pois a região Sudeste tem alta correlação com as anomalias de temperatura da água do mar do Atlântico Sul.

4. CONCLUSÃO

As divisões dos vários grupos propostos pela técnica de Análise de Agrupamento para a série temporal completa mostram como o valor da precipitação de cada dia tem sua importância própria, quando observada no plano cartesiano.

Os índices percentuais da participação dos eventos de El Niño e La Niña na formação dos grupos, utilizando a Análise de Agrupamento com o método de agrupamento de Ward, a medida de distância, o Coeficiente de Correlação de Pearson e o critério de agrupamento, e a Análise do Comportamento do Nível de Fusão (Distância), não revelaram influência desses eventos na formação dos diversos grupos.

Dessa forma, os diversos grupos gerados representam uma dinâmica própria da climatologia de Taubaté e da região do vale do Paraíba.

Sabendo-se que, quanto ao uso da água, 70% são destinados à agricultura, 20% à indústria e 10% ao uso doméstico, o vale do Paraíba assume importante posição no que se refere ao uso e distribuição da água, uma vez que nessa região ocorrem todas essas situações.

O controle do comportamento da precipitação, principalmente pelo Departamento de Ciências Agrárias da UNITAU – Universidade de Taubaté torna-se mais importante ainda, caso no futuro se concretize, como divulgado pela imprensa, a transposição da água da bacia do Rio Paraíba do Sul para a capital São Paulo.

5. REFERÊNCIAS

- ALVES, L. B. **Tratamentos de dados multivariados por análise de correspondência em dados de idosos de São José dos Campos**. 2007. 116f. Dissertação (Mestrado em Engenharia Elétrica e Mecânica) – Instituto Tecnológico de Aeronáutica, São José dos Campos, 2007.
- AMORIM, R. C. F.; RICIERI, R. P.; FILHO, J. S. V.; JÚNIOR, R. S. S.; GNOATO, E. Determinação do período de retorno da precipitação pluviométrica por meio da distribuição de Gumbel para a região de Cascavel/PR. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 12., 2002, Foz do Iguaçu. **Anais...** Foz do Iguaçu: SBMET, 2002.
- ANDRÉ, R. G. B.; MARQUES, V. S.; PINHEIRO, F. M. A.; FERRAUDO, A. S. Identificação de regiões pluviometricamente homogêneas no Estado do Rio de Janeiro, utilizando-se valores mensais. **Revista Brasileira de Meteorologia**, v. 23, n. 4, p. 501-509, 2008. <http://dx.doi.org/10.1590/S0102-77862008000400009>
- AYRES, A. C. **Variabilidade e desastres naturais da região do Vale do Paraíba/SP: Passado e Futuro**. 2010. 89f. Dissertação (Mestrado em Ciências Ambientais) – Universidade de Taubaté, Taubaté, 2010.
- BRAGA, C. C.; MELO, M. L. D.; MELO, E. C. S. Análise de agrupamento aplicada a distribuição da precipitação no Estado da Bahia. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 10., 1998, Brasília. **Anais...** Brasília: SBMET, 1998. p. 1857-1862.
- DINIZ, H. N. Interferência entre poços tubulares profundos na bacia do rio Una, município de Taubaté. In: CONGRESSO BRASILEIRO DE ÁGUAS SUBTERRÂNEAS, 15., 2008, Natal. **Anais...** Natal: ABAS, 2008.

- FECHINE, J. A. L.; GALVÍNIO, J. D. Agrupamento da precipitação mensal da bacia hidrográfica do Rio Brigida-PE, através da multivariada. **Revista Brasileira de Geografia Física**, v. 01, n. 01, p. 39-46, 2008.
- FISCH, G.; VALÉRIO, M. C. Variabilidade intra e interanual da precipitação em Taubaté - SP associado aos eventos El Niño e La Niña. **Revista Biociências**, Taubaté, v. 11, n. 1/2, p. 19-29, 2005.
- HORIKOSHI, A. S.; FISCH, G. Balanço hídrico atual e simulações para cenários climáticos futuros no Município de Taubaté, SP, Brasil. **Revista Ambiente & Água**, Taubaté, v. 2, n. 2, p. 33-46, 2007. <http://dx.doi.org/10.4136/ami-agua.25>
- INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Centro de Previsão de Tempo e Estudos Climáticos – CPTEC. **Ocorrências de El Niño**. 2011. Disponível em: <http://enos.cptec.inpe.br/tab_elnino.shtml> Acesso em: 05 abr. 2011.
- INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Centro de Previsão de Tempo e Estudos Climáticos – CPTEC. **Ocorrências de La Niña**. 2011. Disponível em: <http://enos.cptec.inpe.br/tab_lanina.shtml> Acesso em: 05 abr. 2011.
- INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE. **Climate Change 2007: the physical science basis**. IPCC: Genebra, 2007. Disponível em: <http://www.ipcc.ch/publications_and_data/ar4/wg1/en/contents.html>. Acesso em: 02 dez. 2009.
- MACEDO, M. J. H.; GUEDES, R. V. S.; SOUZA, F. A. S.; DANTAS, F. R. C. Análise do índice padronizado de precipitação para o estado da Paraíba, Brasil. **Revista Ambiente e Água**, v. 5, n. 1, p. 204-214, 2010. <http://dx.doi.org/10.4136/ambi-agua.130>
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2007. 297 p.
- MOITA NETO, J. M. Estatística multivariada: uma visão didática–metodológica. **Crítica Revista de Filosofia**, maio 2004. Disponível em: <http://criticanarede.com/cien_estatistica.html>. Acesso em: 20 abr. 2011.
- NALDI, M. C. **Técnicas de combinação para agrupamento centralizado e distribuído de dados**. 2011. 245p. Tese (Doutorado em Ciências - Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemática e de Computação, Universidade de São Paulo, São Carlos, 2011.
- NUNES, L. H.; CALBET, N. O. Variabilidade pluviométrica no Vale do Paraíba Paulista. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 11., 2000, Rio de Janeiro. **Anais...** Rio de Janeiro: Sociedade Brasileira de Meteorologia, 2000. 1 CD-ROM.
- OLIVEIRA, V. **Influência do Oceano Atlântico Sul na precipitação do Brasil com ênfase sobre o Rio Grande do Sul**. 2010. 104f. Dissertação (Mestrado em Meteorologia) - Universidade Federal de Alagoas, Maceió, 2010.