

## **DETERMINAÇÃO DOS DIFERENTES PADRÕES DE PRECIPITAÇÃO NO ESTADO DO RIO GRANDE DO NORTE ATRAVÉS DE ANÁLISE DE COMPONENTES PRINCIPAIS**

*Adelena Gonçalves Maia<sup>1\*</sup>; Paulo Sérgio Lucio<sup>2</sup>*

**Resumo** – O objetivo deste trabalho é utilizar a Análise de Componentes Principais como ferramenta para a criação de novas variáveis capazes de agregar informação às variáveis originais (precipitação média mensal) quanto ao regime pluviométrico do estado do Rio Grande do Norte. Duas componentes principais foram geradas, capazes de explicar mais de 90% da variância dos dados originais. A CP1 captou informação referente à precipitação média anual, que não fazia parte dos dados originais; e a CP2 captou informação da variabilidade temporal da precipitação ao longo do ano associada à precipitação anual, sendo a CP2 capaz de diferenciar os dois padrões de precipitação existentes no estado. A análise de componentes principais foi uma importante ferramenta de análise de dados multivariados, sendo capaz não só de reduzir o número de variáveis analisadas, como de criar novas variáveis com informações antes inexistentes.

**Palavras-Chave** – regime pluvial, análise multivariada.

## **DETERMINATION OF THE DIFFERENTS PRECIPITATION PATTERNS FROM THE RIO GRANDE DO NORTE STATE BY PRINCIPAL COMPONENTS ANALYSIS**

**Abstract** – The objective of this work is to use the principal component analysis as a tool for creating new variables capable of adding information to the original variables (average monthly rainfall) as the rainfall patterns of the Rio Grande do Norte state. Two principal components were generated, able to explain 90% of the variance of the original variables. The PC1 captured information regarding the average annual rainfall, which was not part of the original data, and PC2 captured information about temporal variability of rainfall throughout the year associated with annual rainfall. The PC2 is able to differentiate the two precipitation patterns existing in the state. The principal components analysis was an important tool for multivariate data analysis, being able not only to reduce the number of variables, such as creating new variables with information previously nonexistent.

**Keywords** – rainfall pattern, multivariate analysis.

<sup>1</sup> Professora Adjunta do Curso de Engenharia Ambiental da Universidade Federal do Rio Grande do Norte (UFRN). UFRN/CT/LARHISA, Cx. Postal 1524, Campus Universitário Lagoa Nova, CEP: 59072-970, Natal – RN; adelenam@gmail.com

<sup>2</sup> Professor Associado do Departamento de Estatística da UFRN. UFRN/CCET/PPGCC. Cx. Postal 1524, Campus Universitário Lagoa Nova, CEP: 59072-970, Natal – RN; pslucio@ccet.ufrn.br

\* Autor Correspondente

## **INTRODUÇÃO**

O estado do Rio Grande do Norte tem uma grande variabilidade espacial e temporal da precipitação, com valores de precipitação anual variando de aproximadamente 500 a 1700 mm. A distribuição desta precipitação ao longo do ano também tem relevantes variações. Nas regiões do litoral e agreste do estado as Perturbações Ondulatórias dos Alísios (POA) são determinantes na ocorrência de chuvas, atuando especialmente de maio a agosto (NOBRE & MOLION, 1998). Já nas regiões central e oeste do estado as ocorrências de chuvas estão relacionadas principalmente ao deslocamento da Zona de Convergência Intertropical (ZCIT), sendo as chuvas concentradas no período de fevereiro a maio.

A Análise de Componentes Principais (ACP) constitui um poderoso método de investigação do comportamento espacial e temporal de variáveis meteorológicas (DE e MAZUNDAR, 1999), tendo sido utilizada em diversos trabalhos para a identificação de padrões climáticos, como em Serra et al. (1999), ou especificamente padrões de precipitação (SNEYERS, 1989; DE e MAZUNDAR, 1999). A ACP também pode ser utilizada associada a técnicas de agrupamento para a determinação de áreas com precipitação pluvial homogênea (ARAÚJO e SOUSA, 2012).

O objetivo deste trabalho é utilizar a ACP como ferramenta para a criação de novas variáveis capazes de agregar informação às variáveis originais (precipitação média mensal) quanto ao regime pluvial do estado do Rio Grande do Norte.

## **MATERIAIS E MÉTODOS**

### **Área de Estudo**

O estudo foi realizado no estado do Rio Grande do Norte, onde 40 estações homogeneamente distribuídas foram utilizadas (figura 1), com séries de 48 anos de dados de precipitação diária, abrangendo o período de 1963 até 2010. Os dados pluviométricos foram cedidos pela Empresa de Pesquisa Agropecuária do Rio Grande do Norte (EMPARN).

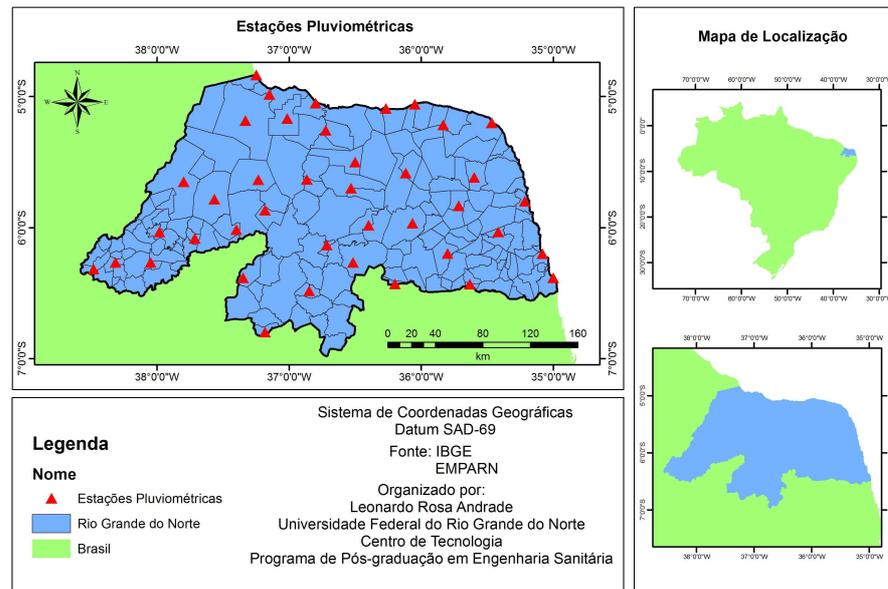


Figura 1 – Localização das estações pluviométricas.

## Análise de Componentes Principais

A análise de componentes principais (ACP) é utilizada principalmente com o objetivo de reduzir o número de variáveis características, através da mudança das variáveis originais por um conjunto de componentes principais ortogonais que irão tornar os cálculos subsequentes mais fáceis (SERRA et al., 1999). A ACP transforma um sistema de eixos originais em um novo sistema que tem duas importantes propriedades: os novos eixos principais são ortogonais e as componentes principais são arranjadas em ordem decrescente de importância, de modo que cada uma delas explique a máxima quantidade da variância das observações, não explicada pelas componentes anteriores (Assis<sup>3</sup> apud Araújo et.al, 2012).

As componentes principais ( $y_1, y_2, y_3, \dots, y_q$ ) são geradas através das combinações lineares das variáveis originais ( $x_1, x_2, x_3, \dots, x_q$ ), conforme apresentado nas equações a seguir:

$$y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1q}x_q \quad (1)$$

$$y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2q}x_q \quad (2)$$

...

$$y_q = a_{q1}x_1 + a_{q2}x_2 + \dots + a_{qq}x_q \quad (3)$$

<sup>3</sup> ASSIS NETO, F.; ARRUDA, H. V.; PEREIRA, A. R. *Aplicação de estatística à climatologia - Teoria e pratica*. Editora e Gráfica Universitária (UFPel), 1996. 161p.

Os pesos (valores de “a”) utilizados são os elementos dos autovetores da matriz de correlação dos dados originais. Os dados originais utilizados nesta pesquisa foram as doze precipitações médias mensais das 40 estações estudadas.

O software utilizado para o desenvolvimento a geração das componentes principais foi o Programa R - The R Project for Statistical Computing (R DEVELOPMENT CORE TEAM, 2012).

## RESULTADOS E DISCUSSÕES

### Análise Exploratória dos Dados

Foram determinadas as precipitações médias mensais para as localidades estudadas e realizada uma análise exploratório dos dados através do gráfico boxplot das 40 estações estudadas (figura 2a) e do gráfico de precipitações médias mensais (figura 2b) de apenas doze localidades distribuídas no estado (figura 2c).

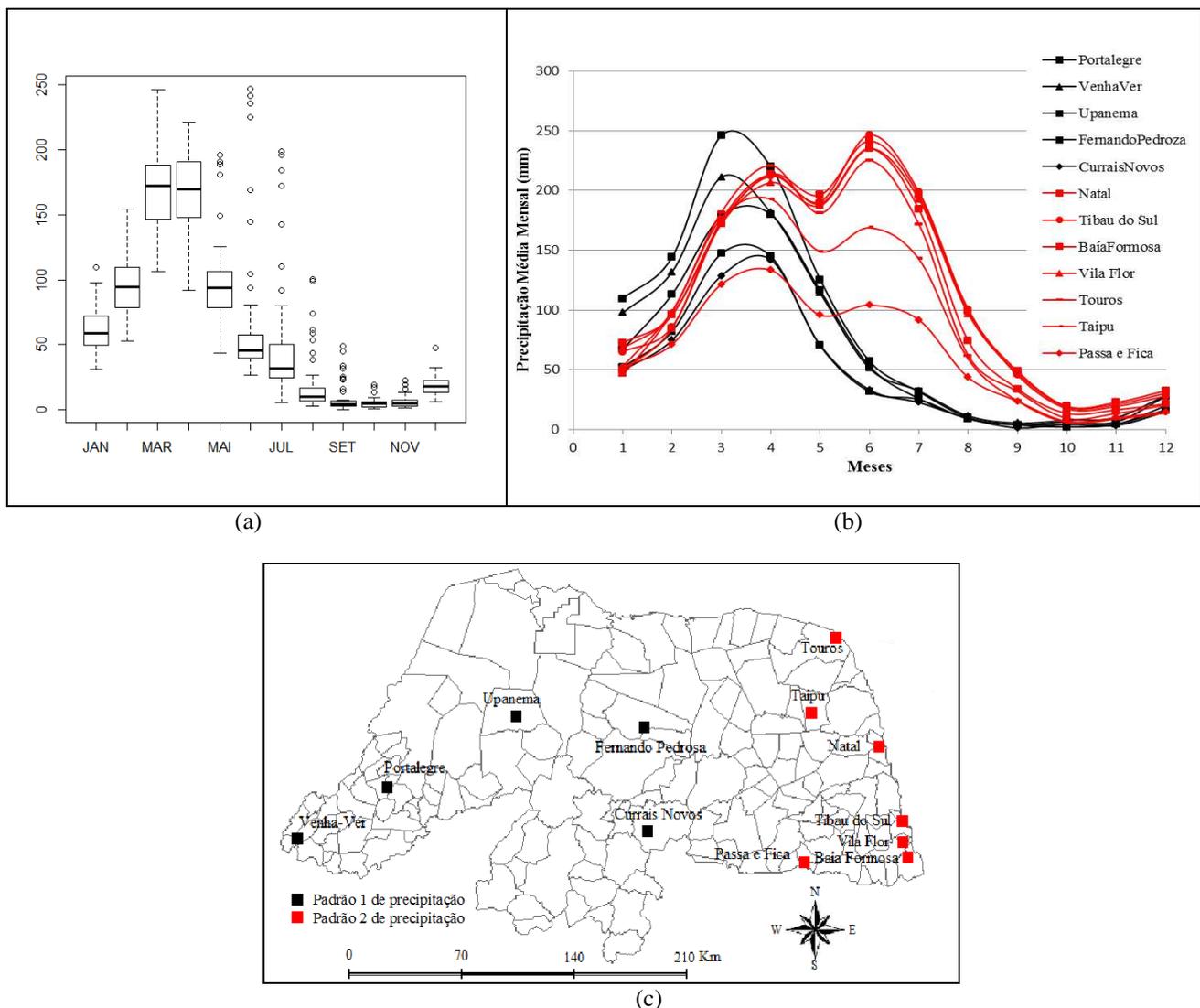


Figura 2- Análise exploratória dos dados através do gráfico boxplot (a) e do gráfico de precipitação média mensal (b) de alguns municípios do estado (c).

Através da análise do gráfico boxplot (figura 2a) dos dados de precipitação média mensal das 40 estações estudadas percebe-se que, em média, os meses de maior precipitação concentram-se no período de março a abril, no entanto no período de maio a dezembro existem diversos outliers que sugerem que existe uma variabilidade espacial do período úmido na área de estudo. O gráfico das precipitações médias de algumas localidades foi plotado (figura 2b) para a investigação desta variabilidade espacial. Na figura 2b podemos identificar dois padrões de precipitação no estado, o primeiro padrão (Padrão 1), em preto, refere-se ao padrão com apenas 1 pico de precipitação, identificado no oeste do estado (Venha Ver, Portalegre e Upanema) e na região central (Fernando Pedroza e Currais Novos), e o segundo padrão (padrão 2), em vermelho, apresenta dois picos de precipitação, encontrado na região leste do estado, no litoral (Natal, Tibau do Sul, Baía Formosa, Vila Flor, Touros) e em uma região um pouco afastada do litoral, conhecida como agreste (Taipu e Passa e Fica).

### Componentes Principais

As componentes principais são combinações lineares das variáveis originais e quantificam a máxima variação dos dados originais. A seguir (figura 3) é apresentada a variância explicada pelas doze componentes principais geradas. A primeira componente explica 57,3% da variância dos dados originais, a segunda explica 33,8% e a terceira 5,6%. A primeira e a segunda componentes explicam mais de 90% dos dados e através da análise dos resultados foram consideradas suficientes para a identificações dos diferentes padrões de precipitação existentes no estado.

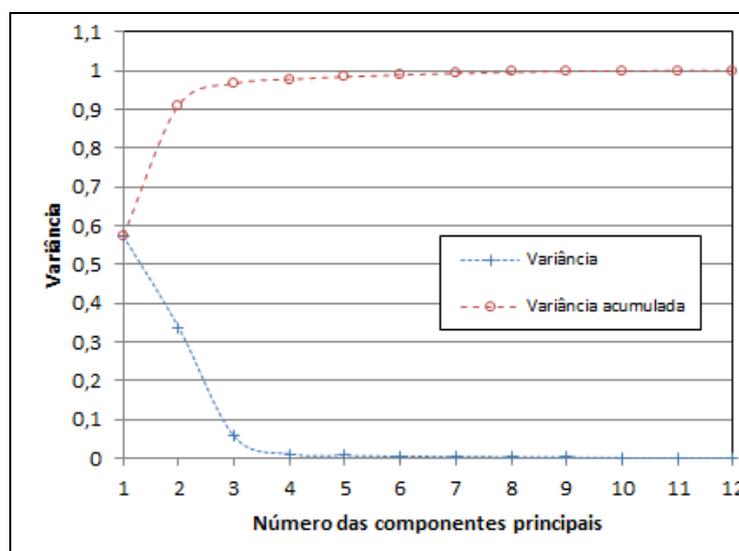


Figura 3- Variância total e variância acumulada das componentes principais.

Os elementos dos dois primeiros autovetores da matriz de correlação dos dados originais são apresentados na figura 4.

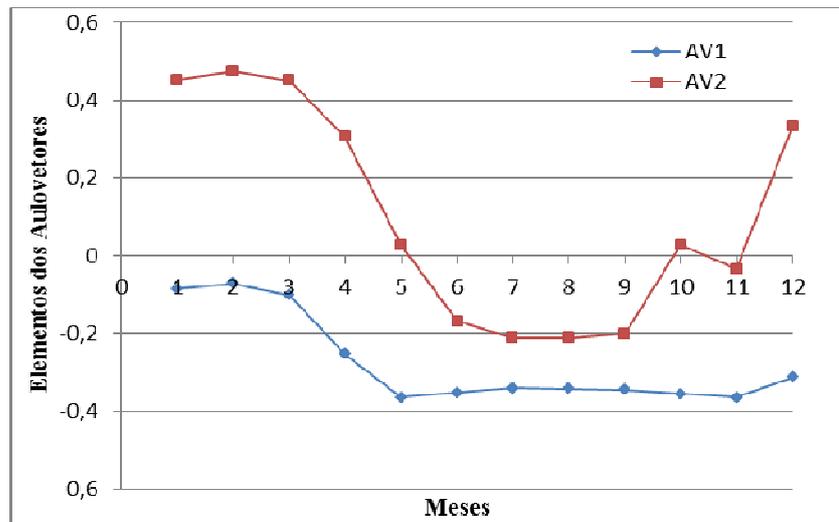


Figura 4- Valores dos elementos dos autovetores 1 e 2

A primeira componente principal será então gerada pela combinação linear dos dados de precipitação média mensal, cujos pesos (valores de “a”, equação 1) para cada mês são apresentados na figura 4. Os valores dos autovetores serão positivos ou negativos por conta do método de resolução do autovetor, mas podemos multiplicar todos os valores de um autovetor por “-1”, sem alterar a interpretação dos resultados, pois o que importa é a diferença entre estes pesos, que serão utilizados em cada mês para a geração da componente principal.

O primeiro autovetor apresenta valores de “-0,07” a “-0,36”, ou seja, todos os valores com o mesmo sinal e maior peso para os meses de maio a dezembro, sendo assim todos os meses contribuirão positivamente (apesar de terem valores negativos) para a geração desta componente. A figura 5 apresenta os valores da CPI e da precipitação média anual no estado.

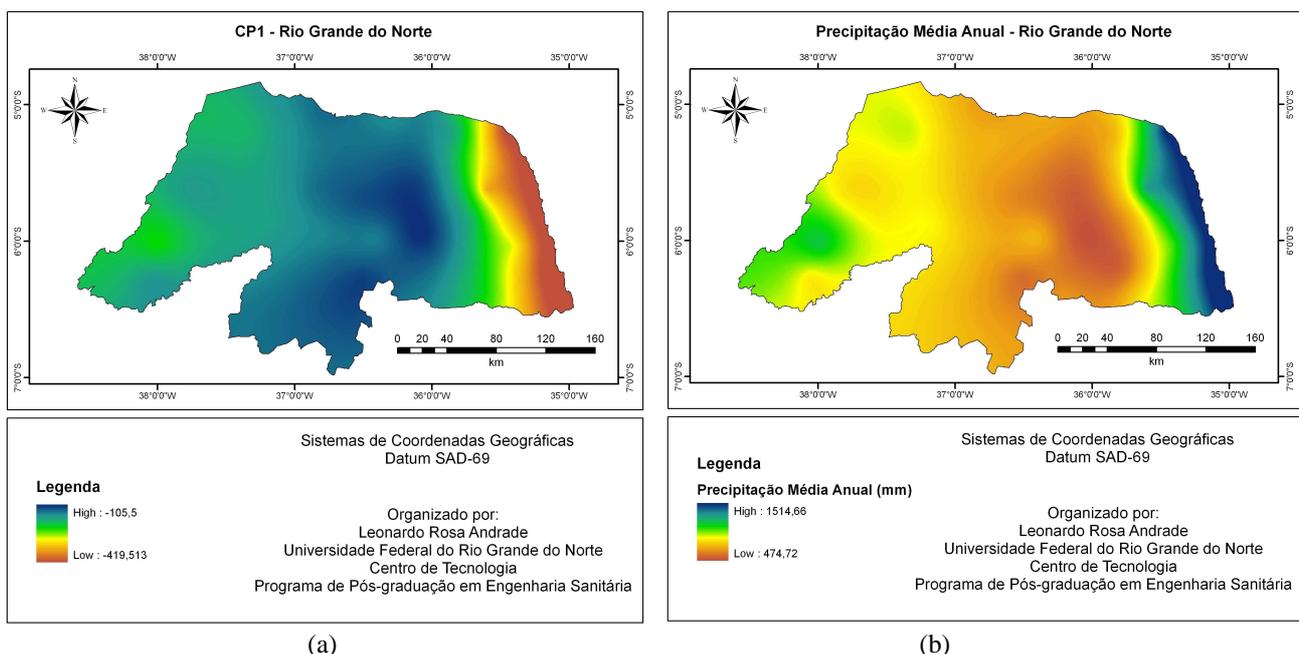


Figura 5 – Mapa com a distribuição espacial da CP1 e da Precipitação média anual

Observa-se que os dois mapas (figura 5) apresentam o mesmo padrão, ou seja a CP1 (primeira componente principal), que representa a maior variabilidade dos dados, pode ser utilizada como representativa da variabilidade da precipitação média anual no estado.

O segundo autovetor (figura 4) apresenta valores positivos dos meses de dezembro a abril, bem discrepantes dos valores negativos dos meses de junho a setembro. O gráfico dos elementos do autovetor 2 (figura 4) segue a mesma tendência da variação da mediana das precipitações mensais apresentada na figura 2a. A figura 6 apresenta os resultados do CP2 distribuídas no estado.

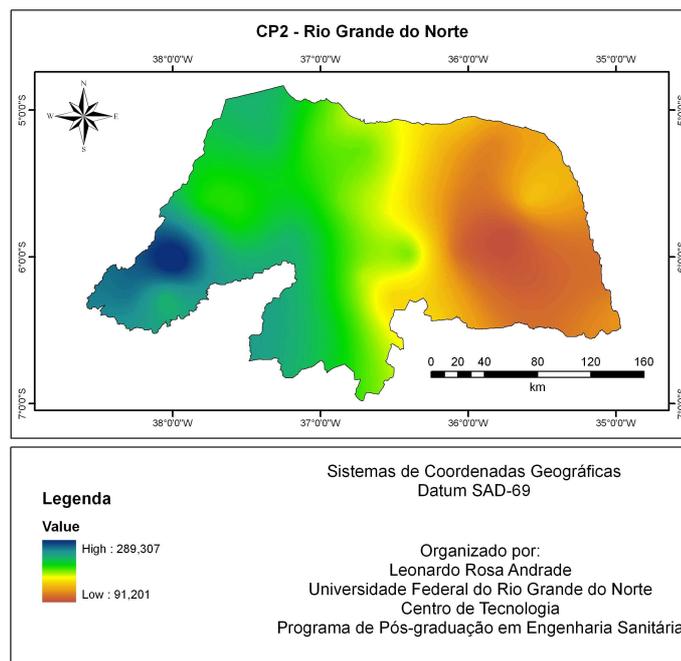


Figura 6 - Mapa com a distribuição espacial da CP2

Observa-se que os valores de CP2 foram maiores para as regiões que se enquadravam no Padrão 1 de precipitação (1 pico, identificado na figura 2b), nestas regiões, leste e centro do estado, foi possível diferenciar as áreas que apresentavam maior valor de precipitação total anual, como a região de Portalegre, das áreas que apresentavam menor precipitação anual, como Currais Novos. As regiões pertencentes ao padrão 2 (leste do estado) apresentaram menores valores de CP2, sendo possível também diferenciar regiões com maior precipitação anual, litoral, das regiões localizadas no agreste, que apresentam menor precipitação anual.

## CONCLUSÕES

O estudo identificou dois padrões de precipitação no estado do Rio Grande do Norte, o primeiro com o pico em Março, característico da região central e oeste do estado; e o segundo com dois picos, Abril e Junho, característico da região leste e agreste.

A análise dos dados através de componentes principais foi capaz de gerar duas novas variáveis (CP1 e CP2) originadas das doze precipitações médias mensais do ano. A CP1 captou informação referente à precipitação média anual, que não fazia parte dos dados originais, e a CP2

captou informação da variabilidade temporal da precipitação ao longo do ano, sendo a CP2 capaz de diferenciar os dois padrões de precipitação existentes.

As componentes principais geradas podem ser úteis no planejamento e gerenciamento de atividades diretamente relacionadas à disponibilidade hídrica, como a atividade agrícola. A CP2 também pode ser utilizada como variável independente de equações empíricas, como de dimensionamento de cisternas, para a inserção da informação da variabilidade temporal da precipitação ao longo do ano.

A análise de componentes principais foi uma importante ferramenta de análise de dados multivariados, sendo capaz não só de reduzir o número de variáveis analisadas, como de criar novas variáveis com informações antes inexistentes.

## REFERÊNCIAS

ARAÚJO, S.; SOUSA, S. (2012). Identificação de regiões pluviometricamente homogêneas no nordeste do Brasil usando análise multivariada. *Revista Brasileira de Climatologia*. 8 (10), pp. 136-152.

DE, U.S.; MAZUNDAR, A.B. (1999). Principal components analysis of rainfall and associated synoptic models of the southwest monsoon over India. *Theoretical and applied climatology*. 64, pp. 213-228.

NOBRE, C.A.; MOLION, L.C. The climatology of droughts and drought predictions. In: PARRY, M.L., CARTER, T.T., KOIJN, N.T. *The impact of climate variations on agriculture*. Dordrecht, Netherlands: KluwerAcademic, v.2, p.115-134, 1988.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

SERRA, C.; FERNANDEZ MILLS, G.; LANA, X. (1999). Winter synoptic weather types in Catalona (NE Spain) and their linkage with minimum temperature anomalies. *International Journal of Climatolog*. 19, pp. 1675-1695.

SNEYERS, R.; VANDIEPENBEECK, M.; VANLIERDE, R. (1989). Principal Component Analysis of Belgian Rainfall. *Theoretical and applied climatology*. 39, pp. 199-204.