

# MODELO DE PREVISÕES DE VAZÕES PARA O NORDESTE SETENTRIONAL UTILIZANDO ANÁLISE DAS COMPONENTES PRINCIPAIS

*Diego de Castro Maia Ribeiro<sup>1</sup>, Marcos Abílio Medeiros de Sabóia<sup>2</sup> & Francisco de Assis de Souza Filho<sup>3</sup>*

**RESUMO** --- O artigo apresenta os resultados de um modelo de previsões de vazões sazonais e interanuais para o sistema de reservatórios do nordeste setentrional, utilizando série histórica de 1931-1998, com as suas respectivas garantias. A metodologia apresentada utiliza a análise das componentes principais (ACP) do sistema de 18 reservatórios e sob essas componentes é realizada uma regressão linear sobre os índices climáticos. Utilizando-se os parâmetros desta regressão aplica-se a metodologia dos k-vizinhos mais próximos para o cálculo da distribuição de probabilidade das vazões afluentes. A aplicação da técnica dos k-vizinhos é apresentada para três reservatórios: Armando Ribeiro, Coremas e Orós. Os valores referentes à correlação existente entre os valores observados foram inferiores aos obtidos para o modelo de previsão do reservatório Orós de forma isolada, havendo espaço para aperfeiçoamento e melhores resultados.

**ABSTRACT** --- *The article presents the results of a model that contains predictions of seasonal and interannual flow to a system of reservoirs in northeastern North, using historical series from 1931 to 1998, with their respective securities. The methodology presented uses the analysis of principal components (PCA) system of 18 reservoirs and under these components is done a linear regression performed on the climatic indices. Using the parameters of the regression is applied the methodology of the k-nearest neighbors to calculate the probability distribution of flows of tributaries. The application of the k-neighbors is given to three reservoirs: Armando Ribeiro, Coremas and Oros. The values for the correlation between the values observed were lower than those obtained from the model to predict the reservoir Oros in isolation, letting space for improvement and better results.*

**Palavras-chave:** Previsão de Vazões, Análise das Componentes Principais, k-vizinhos.

---

<sup>1</sup> Mestrando em Engenharia Civil (Recursos Hídricos) - E-mail: diegocmr@gmail.com

<sup>2</sup> Mestrando em Engenharia Civil (Recursos Hídricos) - E-mail: marcos\_abilio@hotmail.com

<sup>3</sup> Professor do Departamento de Engenharia Hidráulica e Ambiental da Universidade Federal do Ceará ( UFC) - E-mail: assis@ufc.br  
Universidade Federal do Ceará - UFC, Campus do Pici - Bloco 713 - CEP 60455-900 - Fortaleza – CE

## **INTRODUÇÃO**

O Nordeste semi-árido é uma região pobre em volume de escoamento de água dos rios. Essa situação pode ser explicada em razão da variabilidade temporal das precipitações e das características geológicas dominantes, onde há predominância de solos rasos baseados sobre rochas cristalinas e conseqüentemente baixas trocas de água entre o rio e o solo adjacente. O resultado é a existência de densa rede de rios intermitentes. A maior exceção é o Rio São Francisco. Este grande rio, porém, nasce na Serra da Canastra, em Minas Gerais, e só após centenas de quilômetros de percurso entra na região Nordeste.

As regiões semi-áridas, como o Nordeste do Brasil, são particularmente vulneráveis a flutuações do clima e seus impactos nos suprimentos de água. Modelos de previsão de vazões em rios para horizonte de alguns meses ou ano tornam-se muito interessante com vistas a operação mais eficiente dos suprimentos de água e o processo de alocação de água entre usos e usuários conflitantes.

A previsão de vazão pode ser realizada utilizando-se a modelagem dinâmica ou estatística. Um outro procedimento também utilizado consiste no uso de Modelos de Circulação Gerais (GCMs) do oceano e da atmosfera, seguido de "downscaling" que usa modelos regionais de previsão de clima ou aproximações estatísticas e seguido por modelos chuva-deflúvio concentrados ou distribuídos. Esta é uma direção de pesquisa útil. Neste momento, ainda estão sendo avaliados assuntos relacionados à propagação de incertezas ao longo da metodologia (Block et al., 2009), processos estão sendo parametrizados e capacidade de previsão final está sendo avaliada.

Uma alternativa a esta metodologia é o desenvolvimento direto de previsões estatísticas para os aportes de água e as demandas a partir de uma adequada seleção de preditores de clima. Esse artigo pretende aplicar essa linha metodológica para a previsão sazonal de vazões em 18 reservatórios do Nordeste Setentrional sendo a mesma semelhante com a apresentada por Souza Filho e Lall (2002).

## **LOCAL DE APLICAÇÃO**

O estudo de previsão de vazões foi realizado para a região do Nordeste Setentrional. Situado ao Norte do Rio São Francisco, abrange parcialmente os Estados de Pernambuco, Ceará, Paraíba e Rio Grande do Norte. Foram estudados 18 reservatórios principais que estão descritos na Tabela 1; estes são os mesmo reservatórios que serão beneficiados pela transposição do Rio São Francisco.

Tabela 1 - Localização e volume máximo dos reservatórios

Reservatório	Bacia	Estado	Vol. Max. (hm <sup>3</sup> )
Acarape	Metropolitana de Fortaleza	Ceará	33,3
Aracoiaba	Metropolitana de Fortaleza	Ceará	318
Pacoti- Riachao	Metropolitana de Fortaleza	Ceará	420
Pacajus	Metropolitana de Fortaleza	Ceará	240
Armando Ribeiro	Piranhas-Açu	Rio Grande do Norte	2400
Eng. Avidos	Piranhas-Açu	Paraíba	300
Coremas	Rio do Peixe	Paraíba	1358,7
Banabuiú	Banabuiú	Ceará	2134
Castanhao	Médio Jaguaribe	Ceará	4451
Oros	Alto Jaguaribe	Ceará	2100
Atalho	Salgado	Ceará	108,2
Entremontes	---	Pernambuco	339,3
Chapéu	---	Pernambuco	188
Acauã	Rio Paraíba	Paraíba	250
Epitácio Pessoa – Boqueirão	Rio Paraíba	Paraíba	535,7
Barra do Jua	---	Paraíba	71,48
Poço da Cruz	---	Pernambuco	504

## METODOLOGIA

O procedimento utilizado para desenvolver previsões de vazões para os 18 (dezoito) reservatórios no Nordeste Setentrional através do modelo de regressão linear utilizando as Componentes Principais (CP), a série de Dipolo do Atlântico Tropical e NINO3 é descrito a seguir, considerando que as idéias principais do modelo são:

I) As vazões nos 18 locais do Nordeste Setentrional têm alta correlação espacial e, aparentemente, são influenciadas pelo clima de uma maneira semelhante, em particular, pela temperatura da superfície do mar no pacífico e no atlântico tropical. Este fato sugere a possibilidade da utilização de um mesmo modelo para todas as localidades;

II) Enquanto os índices de clima são autocorrelacionados, e tendo distribuição aproximadamente normal, as vazões anuais não parecem ser correlacionadas (pelo menos na análise autoregressiva habitual), sugerindo que um procedimento de previsão apropriado possa ser obtido pela regressão das vazões anuais/mensais e uma seqüência de valores de índices climáticos;

III) O uso de métodos estatísticos tradicionais (métodos paramétricos) para construir um modelo de regressão linear de previsão comum na escala anual e posterior desagregação mensal

pode ser de difícil elaboração devido à grande assimetria das distribuições anuais e mensais, o grande número de vazões zero e a relação não linear entre vazão e os índices climáticos, e entre vazões mensais e anuais;

IV) Os métodos não paramétricos para a regressão e estimativa da densidade de probabilidade podem também ter um sucesso limitado em uma aplicação direta, devido à grande amplitude dos dados de vazão, as dimensões do problema multivariado no espaço e a quantidade de dados disponível;

V) Por conseguinte, desenvolveu-se uma aproximação semi-paramétrica decompondo o problema de estimação em três partes: (1) a transformação das vazões anuais em cada local, com vistas a garantir na regressão destas vazões com os índices num resíduo normal e com variância aproximadamente constante; (2) um modelo de regressão de vetor (neste caso regressão de componente principal) com redução de dimensão com vistas a desenvolver um modelo comum de previsão das vazões anuais utilizando como preditores um conjunto de índices climáticos com certa antecedência; e (3) utilizar este modelo de previsão para a reamostragem não paramétrica do conjunto de dados anuais da série histórica (e conseqüentemente um conjunto de vazões mensais/anuais em todos os locais). Tal reamostragem será construída a partir de valores atuais dos índices climáticos.

A construção do modelo de previsão de vazões segue algumas etapas que serão abordadas a seguir.

Primeiramente, os dados de vazões dos 18 reservatórios foram organizados e analisados utilizando a análise de componentes principais (ACP) que consiste essencialmente em reescrever as coordenadas das amostras em outro sistema de eixo mais conveniente para a análise dos dados. Em outras palavras, as  $n$ -variáveis originais geram, através de suas combinações lineares,  $n$ -componentes principais, cuja principal característica, além da ortogonalidade e de elas serem não correlacionadas entre si, é que são obtidos em ordem decrescente de máxima variância, ou seja, a componente principal 1 detém mais informação estatística do que a componente principal 2 que, por sua vez tem mais informação estatística que a componente principal 3 e assim por diante.

A utilização desse método permite a redução da dimensão dos postos representativos das amostras, tendo em vista que se pode obter em apenas 2 ou 3 das primeiras componentes principais mais que 90% desta informação, apesar de a informação estatística presente nas  $n$ -variáveis originais ser a mesma das  $n$ -componentes principais. O gráfico da componente principal 1 *versus* a componente principal 2 fornece uma janela privilegiada (estatisticamente) para observação dos pontos no espaço  $n$ -dimensional.

A ACP também pode ser usada para julgar a importância das próprias variáveis originais escolhidas, ou seja, as variáveis originais com maior peso (loadings) na combinação linear dos primeiros componentes principais são as mais importantes do ponto de vista estatístico.

Assim, a tarefa realizada com estatística multivariada no presente trabalho consiste na interpretação da distribuição dos pontos no gráfico de componentes principais e na identificação das variáveis originais com maior peso na combinação linear das componentes principais mais importantes. A CP com maior peso foi escolhida para a previsão de vazões.

Após a obtenção da CP, foi realizada a análise da correlação entre a CP e a Temperatura da Superfície do Mar (TSM) no site do NOAA (National Oceanic and Atmospheric Administration) para os meses de Abril-Maio-Junho (AMJ), utilizando a correlação linear com médias mensais/sazonais na atmosfera, de maneira a observar o comportamento da variação da TSM em relação as CP's.

Ainda no NOAA, foram obtidos os preditores considerados para a previsão de julho das vazões anuais (Janeiro-Dezembro) para o ano seguinte que são os valores dos índices do Niño 3 e Dipolo do Atlântico Tropical de (AMJ).

Para a aplicação do modelo de regressão linear entre as CP1 e os índices climáticos, inicialmente padronizou-se a CP1. Após a padronização dos dados, foi aplicado o modelo de regressão linear. Depois da obtenção da equação do modelo de regressão linear, foi aplicado o procedimento dos k-vizinhos de estimativa da função de densidade descritos em Lall e Sharma (1996) e Karlsson e Yakowitz (1987).

O método dos k-vizinhos consiste em realizar uma amostragem das vazões anuais e das vazões mensais que implicitamente reflete a distribuição de probabilidade condicional. A estratégia básica é selecionar os k-vizinhos mais próximos dos preditores no conjunto de dados históricos, estimando-se as probabilidades para nomear a cada um destes vizinhos, bem como os valores para cada um dos valores amostrados.

Foi utilizada uma ferramenta computacional de estatística para realizar todas as operações necessárias à obtenção das componentes principais e agrupamento hierárquico. A ferramenta utilizada foi o software estatístico S-Plus.

## **RESULTADOS DAS ANÁLISES DOS COMPONENTES PRINCIPAIS**

(a) Aplicação da ACP na série de vazões dos reservatórios do Nordeste Setentrional Brasileiro

A análise dos componentes principais (ACP) foi realizada para as vazões afluentes aos rios dos 18 reservatórios de interesse. Foram utilizadas as somas das vazões dos seis primeiros meses de

cada ano para poder alimentar o programa S-Plus com configuração de modelo de escala da covariância matricial.

A Tabela 2 apresenta o percentual de explicação da variância de cada uma das componentes principais. Observa-se que a primeira componente principal explica mais de 85% da variância das vazões nos 18 locais de interesse, isso mostra a similaridade dos regimes pluviais dos reservatórios.

Tabela 2 - Análise das componentes principais

<b>Componentes</b>	<b>Desvio Padrão</b>	<b>Proporção da Variância</b>	<b>Proporção Cumulativa</b>
Comp. 1	1056.782069	0.877891	0.877891
Comp. 2	261.662029	0.053821	0.931711
Comp. 3	173.083743	0.023550	0.955261
Comp. 4	136.064143	0.014553	0.969814
Comp. 5	111.586992	0.009788	0.979602
Comp. 6	84.545797	0.005619	0.985221
Comp. 7	81.184133	0.005181	0.990402
Comp. 8	66.143164	0.003439	0.993841
Comp. 9	51.483830	0.002084	0.995925
Comp. 10	47.219541	0.001753	0.997677
Comp. 11	33.127299	0.000863	0.998540
Comp. 12	26.779676	0.000564	0.999104
Comp. 13	18.528559	0.000270	0.999374
Comp. 14	15.751221	0.000195	0.999569
Comp. 15	14.732887	0.000171	0.999739
Comp. 16	11.694798	0.000108	0.999847
Comp. 17	10.592720	0.000088	0.999935
Comp. 18	9.087632	0.000065	1.000000

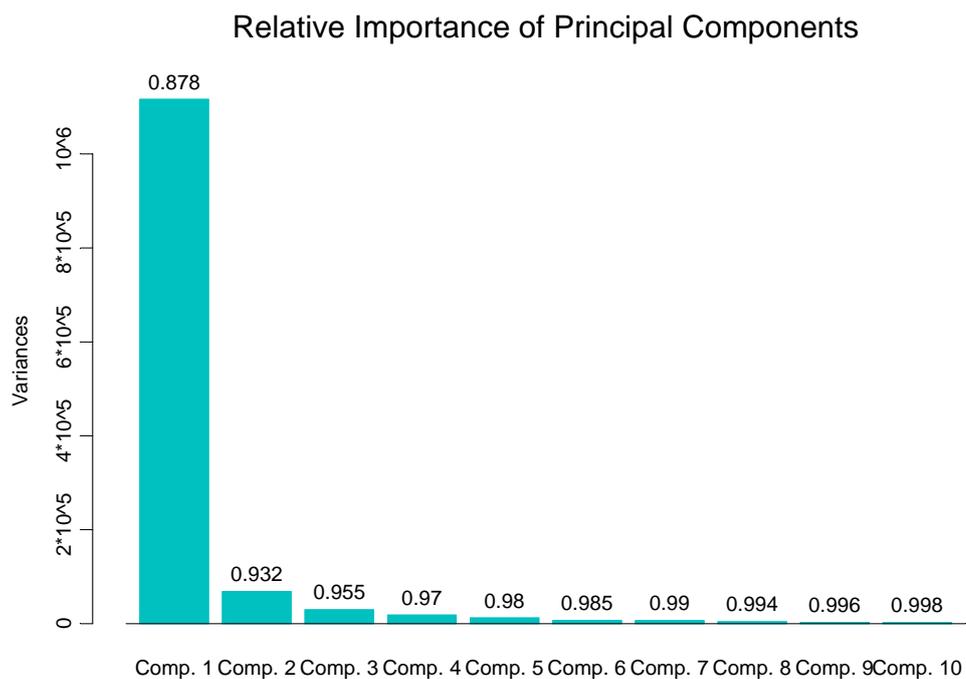


Figura 1 - Percentual acumulado da importância relativa das componentes principais

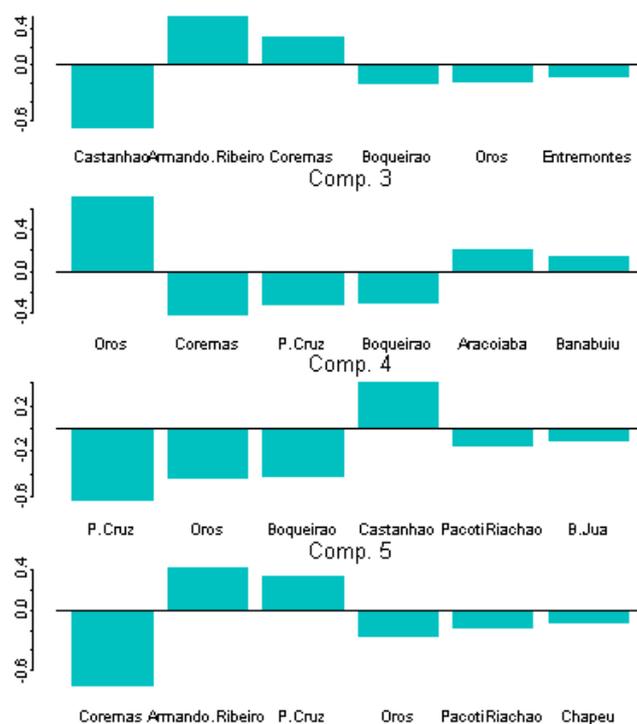


Figura 2 - Pesos na Análise dos componentes principais das afluências aos reservatórios padronizadas

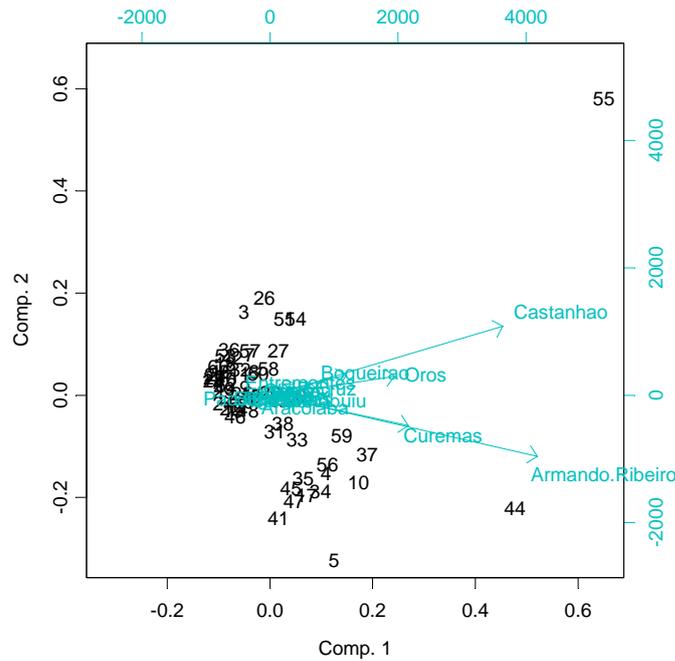


Figura 3 – Componente 1 x Componente 2

(b) Seleção das componentes principais que serão utilizadas na previsão de vazões

O gráfico apresentado na seção anterior mostra o resultado calculado pelo programa, pode-se perceber que a primeira componente é responsável por 87,8% da informação contida nos dados. A componente escolhida para o estudo de previsão de vazões foi a componente 1.

(c) - Correlação da componente principal com a Temperatura da Superfície do Mar (TSM)

Os dados referentes a CP1 foram inseridos no site do NOAA ([www.cdc.noaa.gov/data/correlation/](http://www.cdc.noaa.gov/data/correlation/)) para obter a correlação da CP1 com a TSM. O intervalo de dados utilizados foi restrito ao período de 1948 a 1990, sendo realizado de duas maneiras com índices de LAG 0 e LAG 6. Observe a correlação com a Região do El Nino e do Dipolo Atlântico Norte e Sul.

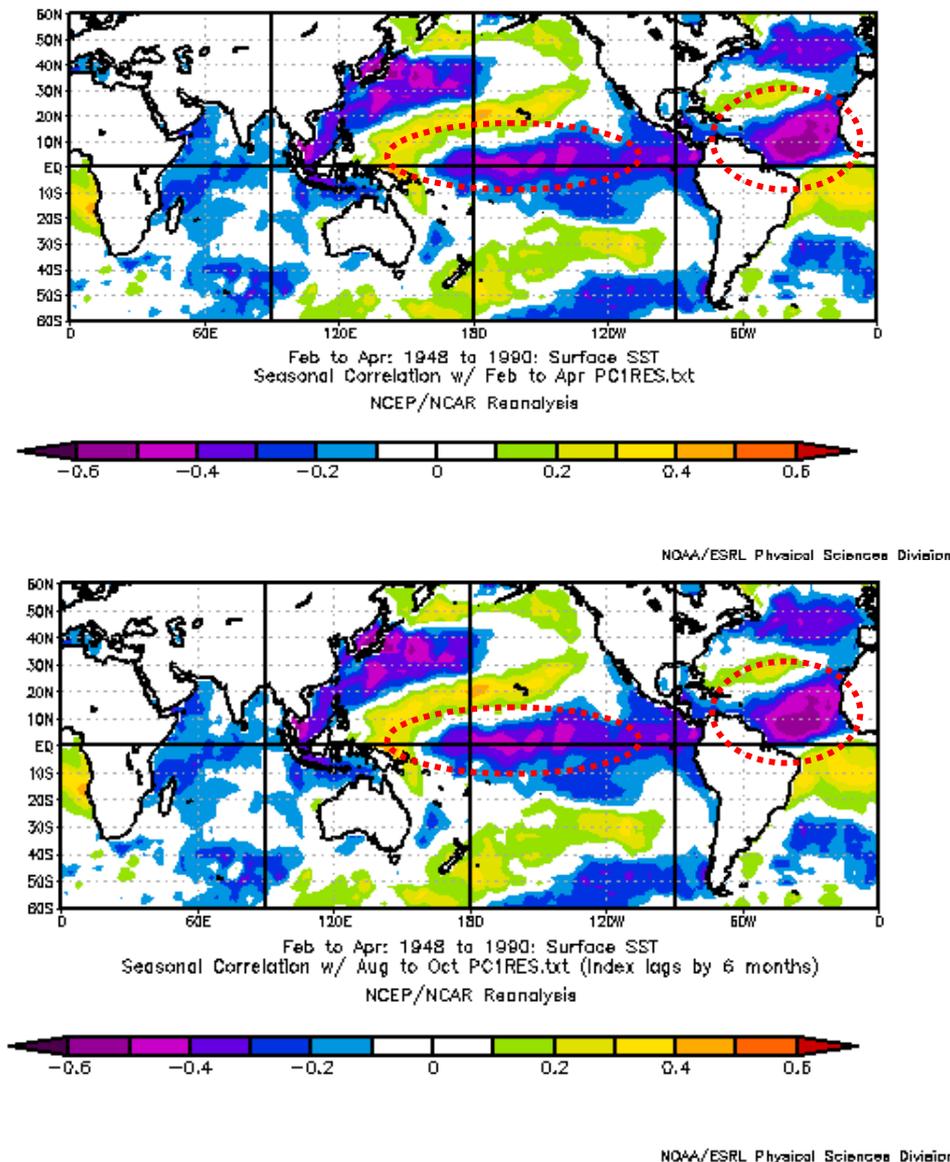


Figura 4 – Correlações da Temperatura da Superfície do Mar (TSM) com as vazões dos reservatórios em questão

#### (d) Obtenção dos Índices Climáticos

Os preditores considerados para a previsão de julho das vazões anuais (Janeiro-Dezembro) para o ano seguinte foram os valores dos índices do Niño 3 e Dipolo do Atlântico Tropical de Abril-Maio-Junho (AMJ). Os índices climáticos do Niño 3 e Dipolo do Atlântico Tropical foram obtidos do site do NOAA (<http://www.cdc.noaa.gov/data/climateindices/list/#TNA>). Esses índices referem-se ao Atlântico Tropical Norte com anomalias mensais de temperatura da superfície do mar situadas entre as seguintes coordenadas 5,5°N a 23,5°N e 15°O a 57,5°O.

(e) Desenvolvimento do modelo de regressão linear entre as CP's e os índices climáticos

Inicialmente, para a aplicação do modelo de regressão linear entre a CP1 e os índices, padronizou-se a CP1, através da seguinte fórmula:

$$CP1(Padronizada)_i = \left( \frac{CP_i - \overline{CP}}{\sigma_{CP}} \right) \quad (1)$$

Após a padronização dos dados aplicou-se o modelo de regressão linear, obtendo os seguintes coeficientes na primeira componente principal:

Tabela 3 - Coeficientes da regressão linear

	Value	Std. Error
(Intercept)	0.0129	0.1128
DIP	1.0563	0.2371
NINO3	-0.1711	0.183

Obtendo como equação final do modelo de regressão linear:

$$\left( \frac{CP_i - \overline{CP}}{\sigma_{CP}} \right) = 1.0563(DIP) - 0.1711(NINO3) + \varepsilon \quad (2)$$

(f) Aplicação da técnica dos k-vizinhos

O procedimento dos k-vizinhos de estimativa da função de densidade descritos em Lall e Sharma (1996) e Karlsson e Yakowitz (1987) é adaptado ao problema atual. Os resultados da aplicação da técnica são mostrados a seguir para três reservatórios: Armando Ribeiro, Coremas e Orós.

## Reservatório: Armando Ribeiro

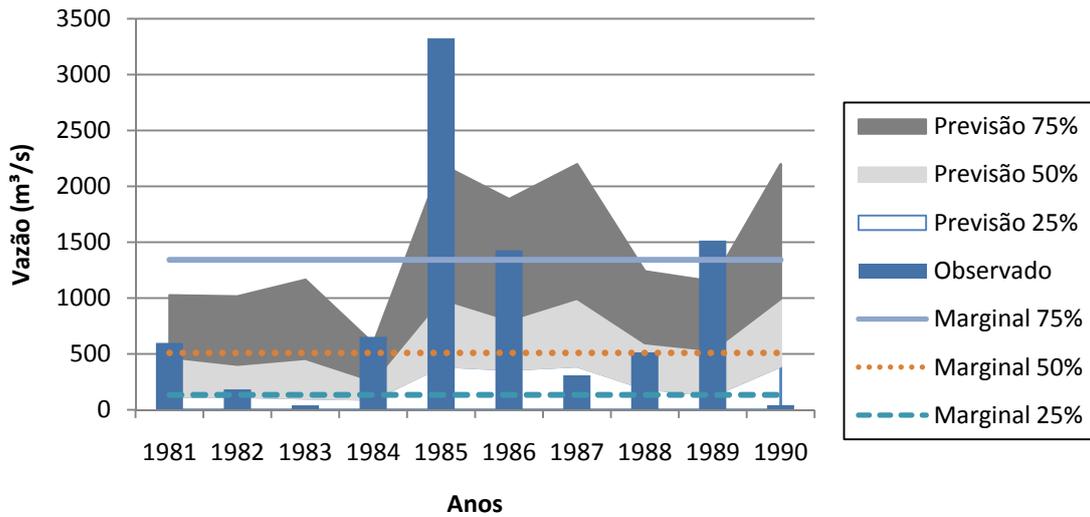


Figura 5 - Vazões dos reservatório Armando Ribeiro

A aplicação do modelo de previsão para o reservatório Armando Ribeiro foi considerada satisfatória para os anos de 1981 - 1986, onde nos três primeiros anos o modelo previa anos com baixa umidade e nos próximos três anos previa anos com muita umidade, o que de fato ocorreu. Porém, nos anos seguintes o modelo não obteve êxito.

## Reservatório: Coremas

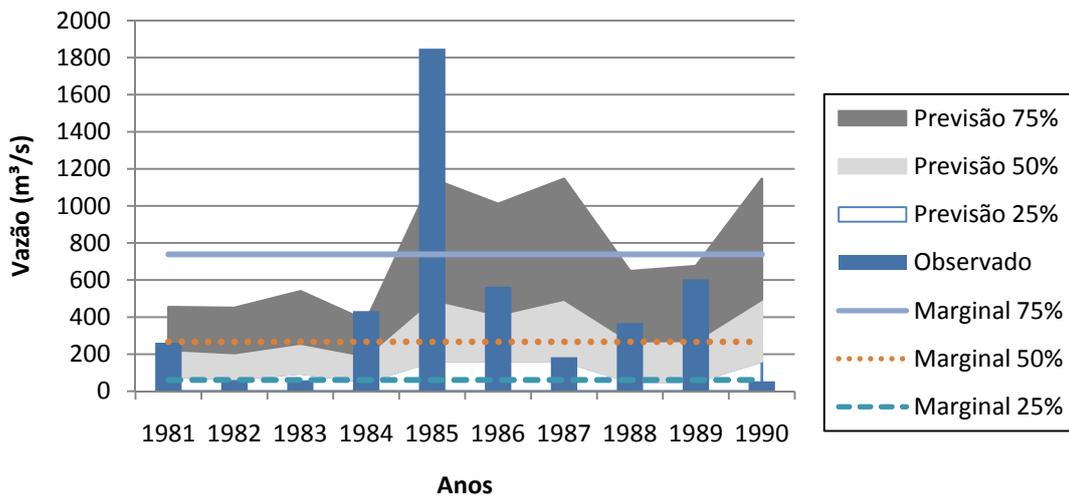


Figura 5 – Vazões do reservatório Coremas

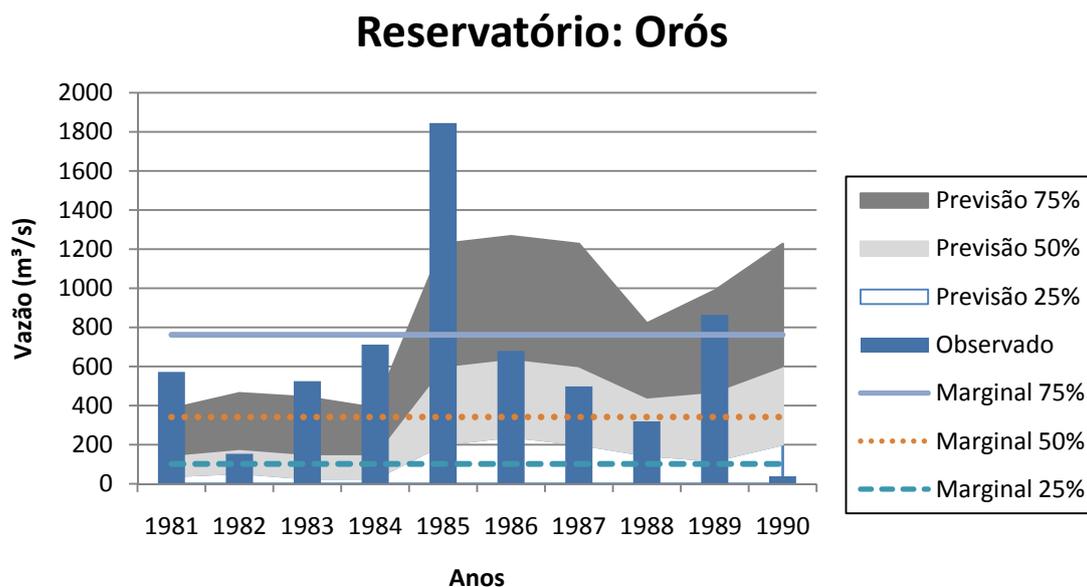


Figura 6 - Vazões do reservatório Orós

Os resultados do modelo aplicado para os reservatórios de Coremas e Orós foram bem semelhantes, sendo considerados satisfatórios para os anos de 1981 - 1989, onde nos três primeiros anos o modelo previa anos com baixa umidade, seguido por anos com mais umidade, e isso realmente ocorreu. Ocorre que no ano de 1990 o modelo falhou ao prever um ano com vazões maiores, e o que se observou foi uma baixa vazão.

(g) Correlação Entre os Valores Observados e Previstos

Após a aplicação da técnica dos k-vizinhos e a obtenção dos resultados, analisou-se a correlação entre os valores observados e os previstos pelo modelo. A tabela a seguir apresenta os valores encontrados.

Tabela 4 - Valores das correlações

Reservatório	Correlação
Armando Ribeiro	0.318
Coremas	0.399
Orós	0.269

Os valores referentes à correlação existente entre os valores observados e os previstos foram baixos, o que demonstra que o modelo pode ser aperfeiçoado para obtenção de melhores resultados.

## **CONCLUSÕES**

A metodologia foi aplicada com sucesso ao sistema dos 18 reservatórios pertencentes ao Nordeste Setentrional utilizando-se a análise das componentes principais por meio das séries históricas de vazões do referido local. Os resultados são promissores, mas ainda podem ser aperfeiçoados para que se possa obter melhores resultados.

Os valores referentes à correlação existente entre os valores observados foram inferiores aos obtidos para o modelo de previsão do reservatório Orós de forma isolada, havendo espaço para aperfeiçoamento e melhores resultados.

Tendo em vista a lacuna existente, pretende-se no futuro a incorporação de uma quantidade maior de preditores na avaliação, de maneira a aperfeiçoar e melhorar os resultados.

## **AGRADECIMENTOS**

Ao Conselho Nacional de Pesquisa (CNPq) pelo financiamento aos estudos que deram origem a este trabalho.

## **BIBLIOGRAFIA**

COGERH, “Plano de Gerenciamento da Bacia do Jaguaribe, Companhia de Gestão dos Recursos Hídricos”, Fortaleza. Ceará. 1999a.

COGERH, “Plano de Gerenciamento das Bacias Metropolitanas, Companhia de Gestão dos Recursos Hídricos”, Fortaleza. Ceará. 1999b.

MINGOTI, S. A. “Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada”. Belo Horizonte: Editora UFMG, 2005. 295p.

SOUZA FILHO, F. A., LALL, U. “Modelo de previsão de vazões sazonais e interanuais”. Revista Brasileira de Recursos Hídricos, 2003