



## XII SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE

### **DESENVOLVIMENTO DE UM MODELO PARA PROCESSAMENTO E ANÁLISE DE DADOS HIDROLÓGICOS NO AMBIENTE R**

*Artur Moises Gonçalves Lourenço<sup>1</sup>; Camilo Allyson Simões de Farias<sup>2</sup> & Carlos Oliveira Galvão<sup>3</sup>*

**RESUMO** – O R é uma linguagem e ambiente para computação estatística e gráfica de licença livre que vem ganhando espaço na análise e modelagem hidrológica. Muitos pacotes na área de engenharia de recursos hídricos estão sendo desenvolvidos e disponibilizados por hidrólogos neste ambiente. Neste trabalho, é apresentado o desenvolvimento do pacote “HidroWebFix”, para ser utilizado no pré-processamento e análise preliminar de dados hidrológicos disponibilizados pela Agência Nacional de Águas e Agência Executiva de Gestão das Águas do Estado da Paraíba. Esse pacote, juntamente com outras bibliotecas, possibilitou a realização rápida do trabalho oneroso de pré-processamento dos dados, gerando relatórios, formatos versáteis para análise e é um exemplo de como bibliotecas desenvolvidas e compartilhadas pelos próprios usuários possibilitam a análise e modelagem eficiente no ambiente. Com base nos resultados, conclui-se o grande potencial da ferramenta e espera-se que o pacote apresentado sirva de auxílio e estímulo à análise e modelagem hidrológica no R, fortalecendo ainda mais a comunidade no Brasil.

**ABSTRACT** – R is a language and environment for statistical and graphics computing, of free license that has gained space for the hydrologic analysis and modeling. Many packages in the area of water resources engineering have been developed and made available by hydrologists using this environment. This work presents the development of package "HidroWebFix" to be used in pre-processing and preliminary analysis of hydrological data available from the National Water Agency and Executive Agency of Water Management in the State of Paraíba. This package, along with other libraries, enabled rapid delivery of costly work of the pre-processing data, generating reports, versatile formats for analysis and is an example of how developed and shared libraries by the users

---

<sup>1</sup> Mestrando do Programa de Pós-graduação em Engenharia Civil e Ambiental, Universidade Federal de Campina Grande, Departamento de Engenharia Civil, Rua Aprígio Veloso, nº 882, Bairro Universitário, 58429-900, Campina Grande - PB, Fone: (83) 2101-1000, e-mail: eng.artur.goncalves@gmail.com

<sup>2</sup> Professor Adjunto, Universidade Federal de Campina Grande, Unidade Acadêmica de Ciências e Tecnologia Ambiental, Rua Jairo Vieira Feitosa, 1770, Pereiros, 58840-000, Pombal - PB, Fone: (83) 3431-4000, e-mail: camiloallyson@yahoo.com.br

<sup>3</sup> Professor Associado, Universidade Federal de Campina Grande, Unidade Acadêmica de Engenharia Civil, Rua Aprígio Veloso, 882, Bairro Universitário, 58429-900, Campina Grande - PB, Fone: (83) 2101-1000, e-mail: carlos@dec.ufcg.edu.br

enable efficient analysis and modeling in R environment . Based on the results, we conclude the great potential of the tool and it is expected that the package serves as a stimulus and aid the analysis and hydrologic modeling in R, further strengthening the community in Brazil.

**Palavras-chave:** software livre, HidroWeb, AESA.

## 1 - INTRODUÇÃO

O *R* é uma linguagem e ambiente multiplataforma para computação estatística e gráfica. Este *software* é livre e teve a sua origem nos anos 1990s com o projeto piloto intitulado *R Project*. A linguagem *R* pode ser considerada uma implementação diferente da conhecida linguagem *S* a qual foi desenvolvida no *Bell Laboratories* em 1970 (Venables e Smith, 2014).

Pode-se dizer que o *R* é a versão *Open Source* da linguagem *S*, tendo algumas diferenças importantes, porem todo código escrito em *S* poder ser compilado em ambiente *R*. Nele é fornecido uma ampla variedade de técnicas estatística e gráficas, como modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, *clustering*, etc (Venables e Smith, 2014).

Alguns usuários consideram o *R* como um sistema estatístico, mas para os desenvolvedores do software, deve-se pensar nele como um ambiente no qual as técnicas estatísticas são implementadas, sendo assim, diferentes técnicas podem ser implementadas se utilizando dos recursos do ambiente e abarcando diversas áreas do conhecimento.

Com o apoio de uma comunidade crescente de usuários com diferentes formações, as funcionalidades do *R* vem se expandido. A distribuição original do *R* possui cerca de oito *packages*/pacotes ou bibliotecas, que são conjunto de funções e rotinas com diferentes utilidades. No repositório oficial do *R*, o CRAN (*Comprehensive R Archive Network*), estão disponíveis diversos pacotes desenvolvidos pela comunidade que vão de ferramentas para análise de dados obtidos da área de medicina e saúde, como é o caso do “*fmsb*”, até conjunto de funções para trabalhar com sistemas inteligentes como é o caso dos pacotes “*GA*” e “*neuralnet*”, o download dos pacotes pode ser realizado no site oficial do CRAN ou instalados diretamente pelo *console* do *R*.

Recentemente vários pacotes na área de engenharia de recursos hídricos vem sendo desenvolvidos e disponibilizados por hidrólogos de diversas instituições para trabalhar com dados hidrológicos no *R*. Muitas destas ferramentas vão além das técnicas utilizadas na hidrologia estatística, são ferramentas para gestão e gerenciamento de recursos hídricos como modelos hidrológicos, sistemas para operação de reservatórios, análise de aquíferos, estudo de mudanças climáticas,

modelagem de contaminantes na água e solo que estão disponibilizados gratuitamente no CRAN, fazendo o R uma grande ferramenta para pesquisa na área de recursos hídricos.

Este trabalho tem como objetivo apresentar o desenvolvimento de um pacote em ambiente R, o “HidroWebFix”, para realizar pré-processamento e análise preliminar de dados hidrológicos disponibilizados pelo sistema nacional de informações hidrológicas, o HidroWeb, e pela Agência Executiva de Gestão das Águas do Estado da Paraíba - AESA.

## 2 - MATERIAL E MÉTODOS

### 2.1 Pacotes para manipulação de dados hidrológicos no R

Apesar dos sistemas de informação sobre recursos hídricos terem como objetivo facilitar o acesso aos dados hidrológicos, algumas vezes esses sistemas passam por dificuldades de implantação, ou ainda, o formato que estes dados são disponibilizados dificulta e torna muito oneroso o trabalho de processamento destas informações.

Os pacotes dedicados a manipulação de séries temporais dão um grande auxílio para lidar com séries hidrológicas, a biblioteca “zoo” trabalha as séries de forma indexada facilitando o acesso e manipulação entre intervalos temporais.

Zambrano-Bigiarini e Bellin (2010) relataram o trabalho oneroso para realizar o pré-processamento de 30 anos de dados da rede hidrométrica italiana e desenvolveram o pacote “hydroTSM” para realizar esta tarefa. O pacote fornece funções de gerenciamento, análise, interpolação e plotagem para modelagem hidrológica. Algumas bibliotecas no R funcionam de forma integrada, são dependentes, como é o caso da “hydroTSM”, que se utiliza da versatilidade da “zoo” para trabalhar com séries temporais.

Dave Hutchinson do *Environment Canada* desenvolveu o pacote “HYDATA” para facilitar o acesso aos dados hidrológicos da rede hidrométrica canadense.

Outro pacote para trabalhar com séries hidrológicas é o “hydrosanity”, desenvolvido por Andrew (2007), que fornece uma interface gráfica de usuário para explorar séries temporais hidrológicas. Ele é projetado para trabalhar com os dados hidrológicos superficiais (principalmente séries temporais de chuva e vazão). Existem funções para importar dados a partir de um banco de dados ou de arquivos, sumarizar e visualizar conjunto de dados de diversas maneiras, estimar chuva real, preencher falhas em séries de precipitação e estimar a relação chuva-vazão.

## 2.2 Desenvolvimento do pacote “HidroWebFix”

O sistema de informações hidrológicas brasileiro HidroWeb, que teve início de operação em ano 2002 após a criação da Agência Nacional das Águas - ANA, centralizou todas as informações nacionais de estações hidrométricas em um único banco de dados que vem sendo alimentado até os dias atuais.

No estado da Paraíba, especificamente na época da criação da HidroWeb, quatro instituições/órgãos eram responsáveis por operar as estações de monitoramento hidrológico espalhadas no território paraibano: a Companhia de Águas e Esgotos da Paraíba - CAGEPA, Departamento Nacional de Obras Contra Secas - DNOCS, Superintendência do Desenvolvimento do Nordeste - SUDENE e a empresa de instrumentação hidrológica HOBECO Ltda. Atualmente a maioria dos postos hidrométricos ainda é operada por estas instituições e pela AESA, havendo alguns operados pela ANA (ANA, 2014). Nessa transição e integração de informações algumas estações tiveram dados perdidos ou estão com suas séries alocadas de forma errada, dados repetidos, entre outros problemas.

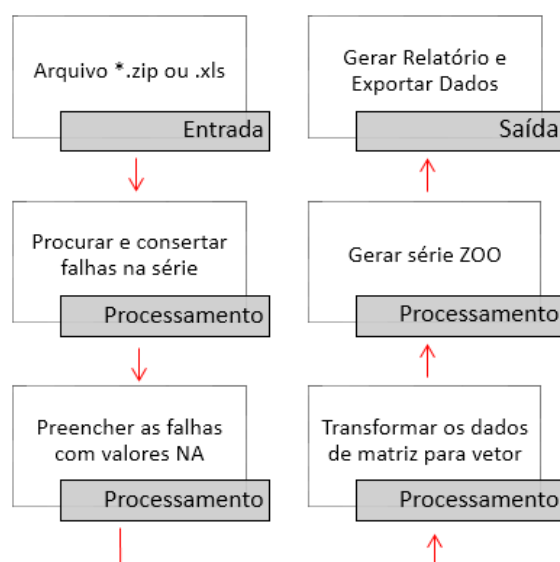
A HidroWeb fornece os dados on-line mediante consulta de formulário, sendo gerado um arquivo compactado .ZIP com um único arquivo interno que pode ser no formato .TXT ou .MDB, dependendo da escolha do usuário, com a série temporal da variável hidrológica escolhida.

O arquivo possui um cabeçalho com informações sobre a estação e legendas e os dados são separados em colunas e linhas por ponto-e-vírgula. O dados de precipitação e vazão, de interesse neste trabalho, são diários com o mês correspondente em linhas e os dias em colunas formando uma matriz.

A AESA disponibiliza os dados em formato semelhante (.XLS).

O objetivo do pacote desenvolvido “HidroWebFix” é realizar o pré-processamento dos arquivos gerados para séries de precipitação e vazão, gerando relatórios e um produto final com as séries indexadas pela data, facilitando a manipulação destas séries no R e possibilitando a exportação para outros formatos.

A Figura 1 apresenta o fluxograma do pré-processamento realizado pelo pacote “HidroWebFix”.



**Figura 1** - Fluxograma do processamento realizado pelo HidroWebFix.

O arquivo compactado gerado pela HidroWeb ou .XLS cedido pela AESA serve como entrada. É realizada, então, uma busca por falhas e outros problemas na série hidrológica do posto fluviométrico ou pluviométrico. As falhas são preenchidas por valores NA (do inglês, *Not Available*), que é que a designação do R para um valor não disponível. Dessa forma é mais fácil computar dias com falhas em cada mês, ano, etc.

Os dados em formato de matriz são transformados em um único vetor com toda a série em sequência ordenada pela data, que é geralmente um formato de entrada utilizado em muitos modelos hidrológicos. Esta etapa é opcional e pode ser definida nos argumentos da função.

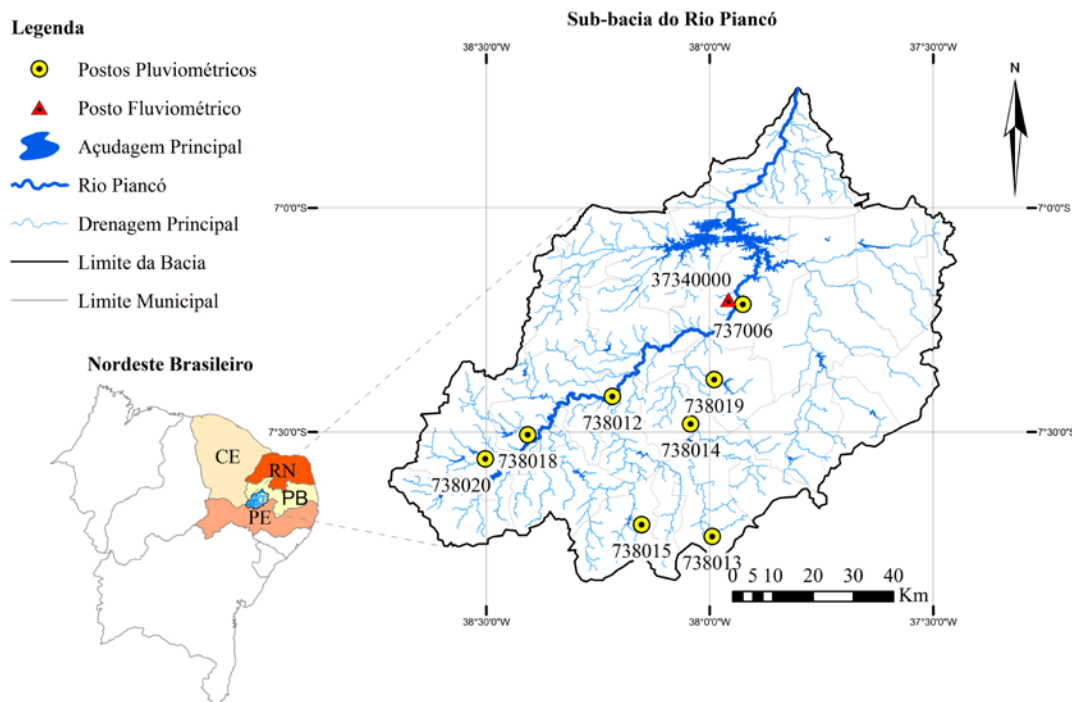
A série é convertida para a classe ZOO para melhor manipulação no R. Em seguida, é gerado relatório em um arquivo PDF, apresentando análises com gráficos e textos, comprimento da série e o número de falhas que podem ser expressos em dias ou meses por ano. Ao final a série pode ser exportada para formatos como TXT, XLS, RDATA, entre outros.

O pacote “HidroWebFix” tem como bibliotecas dependentes, ou seja, que dão suporte, as “zoo”, “hydroTSM” e “lattice”.

### 2.3 Aplicação na bacia hidrográfica do rio Piancó

Foram escolhidas estações hidrométricas localizadas na Bacia do Rio Piancó, uma sub-bacia da Bacia Hidrográfica do Rio Piranhas-Açu, que engloba parte dos estados da Paraíba e Rio Grande do Norte. Esta bacia está localizada no extremo sudoeste da Paraíba, totalmente inserida na mesorregião do sertão paraibano, e é a maior em extensão territorial do estado, com 9.242,76 km<sup>2</sup> e um total de 271.255 habitantes (PERH/PB, 2006).

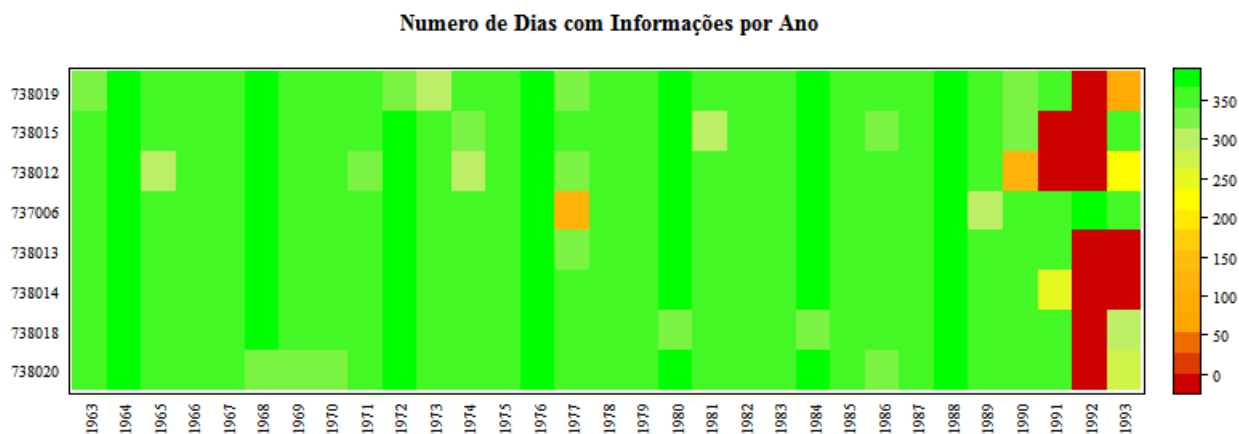
As séries históricas de vazão e precipitação coletadas do website HidroWeb (ANA, 2014) e cedidas pela AESA foram processadas pelo pacote “HidroWebFix”. A Figura 2 apresenta a localização da bacia hidrográfica e dos postos utilizados.



**Figura 2** - Sub-bacia do Rio Piancó e postos utilizados.

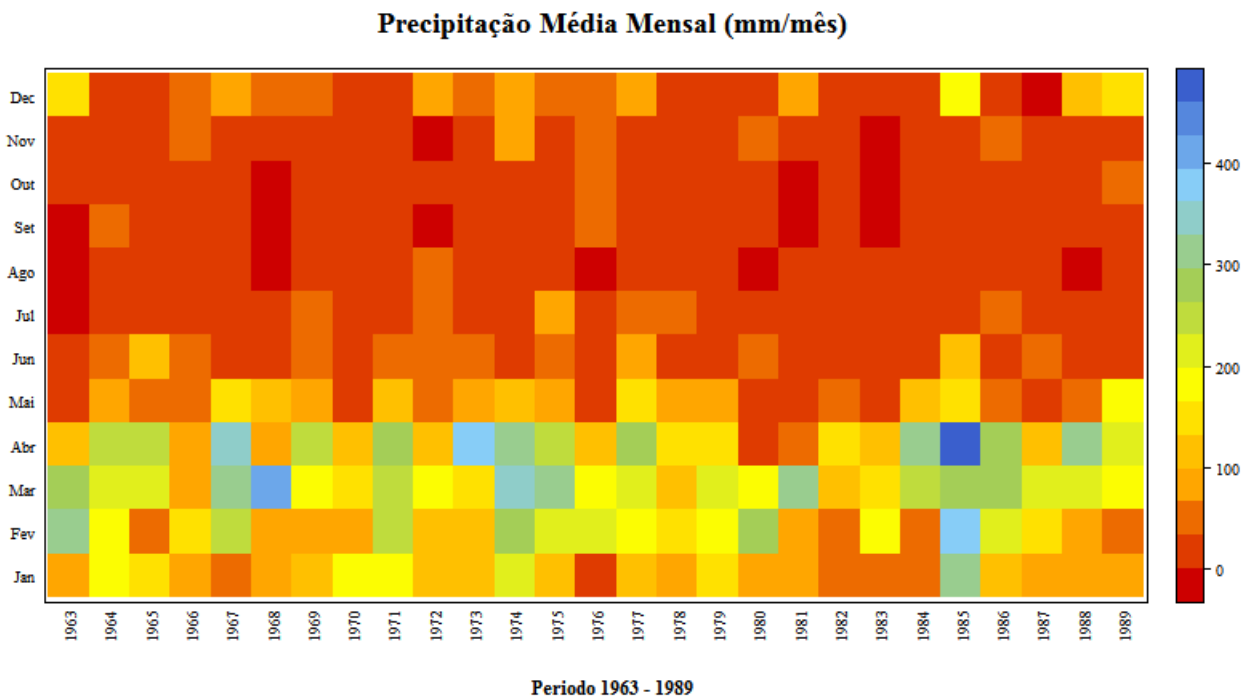
### 3 - RESULTADOS E DISCUSSÃO

Na Figura 3 apresenta-se gráfico contido no relatório gerado após o pré-processamento das séries pelo “HidroWebFix”. O gráfico do tipo matriz identifica o número de dias com informação por ano para a série escolhida para cada posto utilizado. Este tipo de gráfico auxilia visualmente a escolha de séries identificando as falhas e sua quantidade, possibilitando uma visão geral sobre a qualidade dos postos.



**Figura 3** - Número de dias com informações por ano para as estações utilizadas.

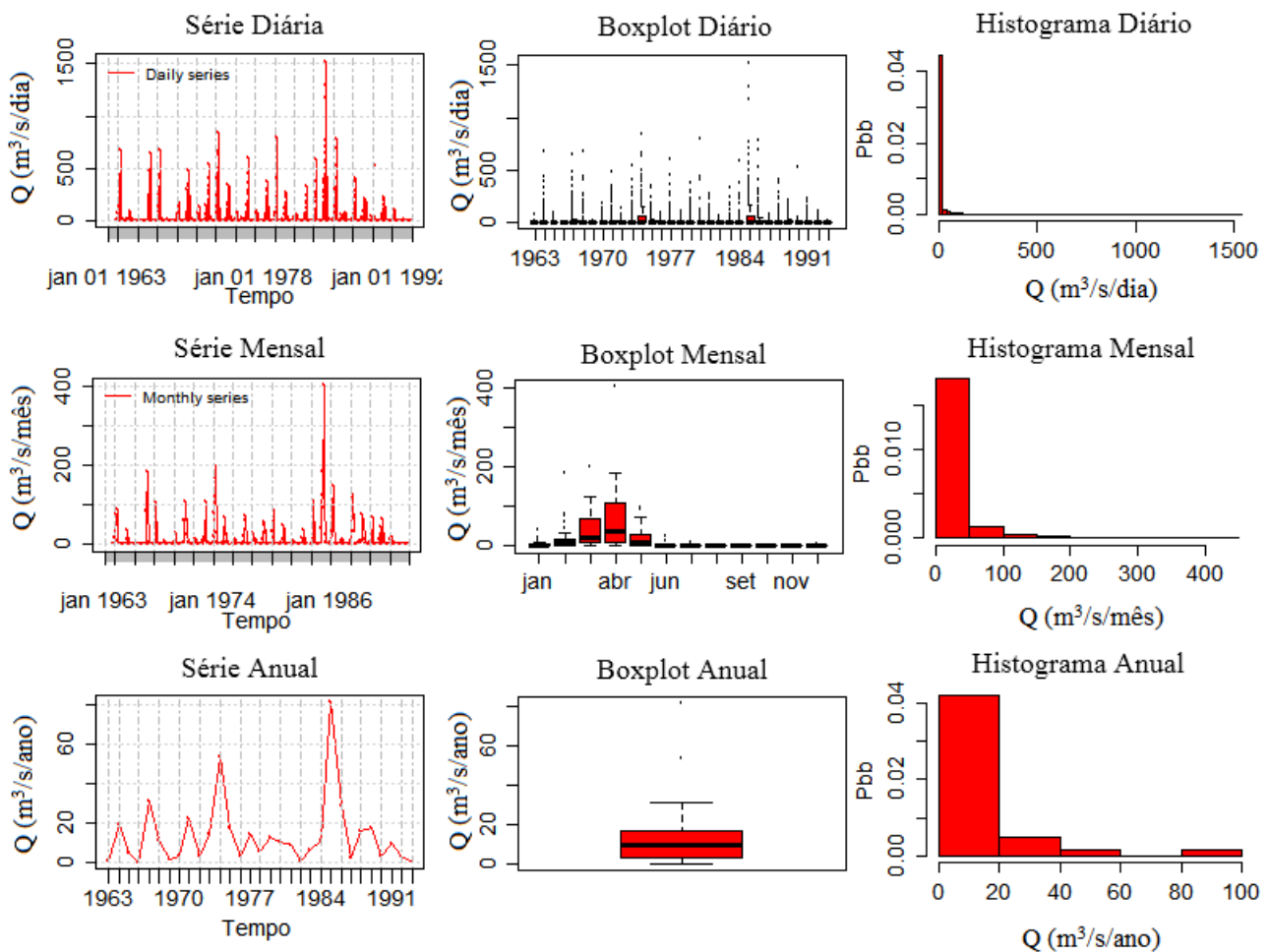
Na Figura 4 apresenta-se também um gráfico do tipo matriz onde é representada a precipitação média mensal dos postos utilizados no período de 1963 à 1989.



**Figura 4** – Precipitação média mensal dos postos utilizados no período de 1963 a 1989.

O gráfico se mostra útil para identificar os meses úmidos e secos, sendo possível perceber as características pluviométricas da região, neste caso com duas estações bem definidas com chuvas de janeiro a maio e período seco até dezembro, sendo grande parte das chuvas concentrada entre os meses de março e abril.

Na Figura 5 apresenta-se uma análise preliminar dos dados do posto fluviométrico Piancó utilizando a função hydroplot do pacote “hydroTSM”.



**Figura 5** – Análise preliminar do posto fluviométrico Piancó.

#### 4 – CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentado o ambiente R como linguagem de programação e ferramenta para realizar análise e modelagem hidrológica. O ambiente, por ter licença GNU GPL (do inglês, *GNU General Public License*), torna-se alternativa viável a softwares proprietários, possibilitando o suporte à pesquisas e a elaboração de estudos e projetos.

A comunidade é crescente e as aplicações em recursos hídricos têm sido bem disseminadas devido à linguagem ser relativamente simples e eficiente. A grande diversidade de bibliotecas disponibilizadas auxiliam desde o trabalho de pré-processamento de dados até modelagem e criação de aplicações destinadas ao usuário final.

Com uso do pacote “HidroWebFix”, foi possível otimizar o oneroso trabalho de pré-tratamento de dados hidrológicos gerando relatórios de texto e gráficos, corrigindo erros e alocando as séries em formatos versáteis para manipulação no R e exportação.



Constata-se, portanto, o grande potencial da ferramenta e espera-se que o pacote apresentado, que será breve disponibilizado no CRAN, sirva de auxílio e estímulo à análise e modelagem hidrológica no R, fortalecendo ainda mais a comunidade no Brasil.

## AGRADECIMENTOS

Os autores agradecem à CAPES, à FINEP e ao CNPq pelo apoio financeiro (Processo n.º 477838/2013-8) e bolsa de estudos do primeiro autor.

## BIBLIOGRAFIA

AESA (2006). Agência Executiva de Gestão das Águas do Estado da Paraíba: “*Plano Estadual de Recursos Hídricos da Paraíba*”, Brasil.

AESA (2014). Agência Executiva de Gestão das Águas do Estado da Paraíba. “*Meteorologia: Chuvas*”. Disponível em <http://goo.gl/R60SyZ>, acessado em 02/04/2014.

ANA (2014). Agência Nacional de Águas. “*HidroWeb: Sistema de Informações sobre Recursos Hídricos*”. Disponível em <http://hidroweb.ana.gov.br/>, acessado em 02/04/2014.

ANDREWS, F. (2007). “*Hydrosanity: a starting point for hydrological analysis*”. Proceedings of MODSIM 2007 - International Congress on Modelling and Simulation, Christchurch, New Zealand, 10-13 December 2007. Modelling and Simulation Society of Australia and New Zealand Inc.

VENABLES, W. N.; SMITH D. M. (2014). “*An Introduction to R: A Programming Environment for Data Analysis and Graphics*”. R Core Team. Apr. pp 105.

ZAMBRANO-BIGIARINI, M.; BELLIN, A. (2010). “*R: a statistical environment for hydrological analysis*”, EGU General Assembly 2010, held 2-7 May, 2010 in Vienna, Austria, p. 13008, vol. 12, pp. 13008.