**ANEXO I**

**PROCEDIMENTOS PARA CONSISTÊNCIA DE DADOS PLUVIOMÉTRICOS**

Gerência de Informações Hidrometeorológicas - GEINF

Superintendência de Gestão da Rede Hidrometeorológica - SGH

Brasília, Setembro de 2014

SUMÁRIO

[1 INTRODUÇÃO 1](#_Toc399757546)

[1.1 Objetivos 2](#_Toc399757547)

[1.2 Objetivos Específicos 2](#_Toc399757548)

[2 Levantamento Bibliográfico 2](#_Toc399757549)

[2.1 Analise Preliminar 3](#_Toc399757550)

[2.2 Métodos de Consistência de Dados Pluviométricos 4](#_Toc399757551)

[2.2.1 Método da Ponderação Regional 5](#_Toc399757552)

[2.2.2 Método da Regressão Linear 6](#_Toc399757553)

[2.2.3 Método da Dupla Massa 7](#_Toc399757554)

[2.2.4 Método do Vetor Regional 8](#_Toc399757555)

[2.2.5 Modelo Multivariado Sazonal Mensal Auto-Regressivo de Ordem 1 - SMMAR(1) 9](#_Toc399757556)

[2.2.6 Método de Redes Neurais Artificiais 11](#_Toc399757557)

[2.2.7 Técnicas de Agrupamento (clusters) 11](#_Toc399757558)

[2.2.8 Componentes Principais 12](#_Toc399757559)

[2.2.9 HidroPlu 12](#_Toc399757560)

[3 SISTEMÁTICAS DE PROCEDImentos de consistência DE DADOS PLUVIOMÉTRICOS UTILIZADOS EM ESTUDOS JÁ REALIZADOS 14](#_Toc399757561)

[3.1 Organização Meteorológica Mundial - OMM (WMO) 14](#_Toc399757562)

[3.2 ORSTOM / IRD 15](#_Toc399757563)

[3.3 U.S. Geological Survey - USGS 16](#_Toc399757564)

[3.4 Agência Nacional de Energia Elétrica - ANEEL 16](#_Toc399757565)

[3.5 Operador Nacional do Sistema Elétrico - ONS 16](#_Toc399757566)

[3.6 Serviço Geológico do Brasil - CPRM 17](#_Toc399757567)

[3.7 Agência Nacional de Águas - ANA 20](#_Toc399757568)

[4 CONSIDERAÇÕES FINAIS 25](#_Toc399757569)

[5 REFERÊNCIAS BIBLIOGRÁFICAS 26](#_Toc399757570)

# INTRODUÇÃO

Os dados e informações gerados pela rede de monitoramento são imprescindíveis para as tomadas de decisão relacionadas ao uso, gestão ou preservação dos recursos hídricos. Com o objetivo de auxiliar a implantação da Política Nacional de Recursos Hídricos e fortalecer o Sistema Nacional de Informações sobre Recursos Hídricos (SNIRH), a partir do ano de 2000, a ANA tem gerenciado o monitoramento hidrometeorológico, em escala nacional, por meio de um conjunto de estações distribuídas estrategicamente por todo país. Isto é feito com o objetivo de levantar os dados básicos necessários ao conhecimento das características quantitativas e qualitativas de nossos rios, assim como os índices pluviométricos, com suas distribuições no espaço e no tempo.

O monitoramento hidrológico no País iniciou, de maneira organizada, em 1920, concentrado na região Sudeste, mais especificamente na bacia do rio Paraíba do Sul, em apoio aos projetos de geração de energia. Os avanços da rede de estações hidrometeorológicas nos anos seguintes ocorreram de forma tímida, para atender às necessidades da época. Somente a partir da década de 70, sob a gestão do Departamento Nacional de Águas e Energia Elétrica (DNAEE), a rede hidrometeorológica nacional começou a ser implantada de forma mais consistente nas diversas regiões do País, com destaque à região Norte. Com a extinção do DNAEE e a criação da Agência Nacional de Energia Elétrica (ANEEL), pela Lei nº 9.427, de 26/12/1996, a gestão da rede hidrometeorológica nacional passou a ser exercida pela ANEEL, assim permanecendo até o ano de 2000, quando a atribuição passou a ser exercida pela Agência Nacional de Águas (ANA). Atualmente, por intermédio da Superintendência de Gestão da Rede Hidrometeorológica (SGH), a ANA gerencia diretamente 4.529 estações, sendo 2.699 estações pluviométricas e 1.830 estações fluviométricas.

Atualmente a SGH/ANA é responsável pelas atividades relativas à alimentação e qualificação da base de dados hidrometeorológicos, sedimentométricos e de qualidade da água, provenientes das redes de monitoramento sob responsabilidade da ANA, bem como pelo controle da quantidade e qualidade dos dados provenientes da rede sob responsabilidade da ANA.

O armazenamento e gerenciamento das informações geradas em toda a Rede Hidrometeorológica Nacional é realizado a partir do sistema HIDRO, composto por um banco de dados e um aplicativo para manipulação dos dados. O sistema HIDRO permite o gerenciamento da base de dados hidrometeorológica (armazenada de forma centralizada em um banco de dados relacional), a entrada de dados por parte das entidades que operam uma Rede Hidrometeorológica, o cálculo de funções hidrometeorológicas básicas e a visualização dos dados armazenados em diversos formatos (gráficos, imagens, etc.).

Em relação a qualificação dos dados hidrometeorológicos disponibilizados no sistema HIDRO ressalta-se que, desde o início da década de 70 com o extinto Departamento Nacional de Águas e Energia Elétrica – DNAEE, foram realizados diversos estudos onde foram aplicadas diferentes metodologias de análise de consistência de dados pluviométricos e fluviométricos. Atualmente, a responsabilidade pela qualificação dos dados também é uma tarefa desempenhada pela SGH/ANA, que desde 2001 vem efetuando a análise de consistência dos dados pluviométricos e fluviométricos existentes no banco de dados Hidro.

Em relação aos dados fluviométricos a SGH desenvolveu e tem utilizado, desde 2011, o Sistema para Análise de Dados Hidrológicos (SiADH), ferramenta computacional que tem por objetivo facilitar e estruturar a análise de dados de cotas e vazões que permitem ao usuário identificar inconsistências e possíveis incoerências nos dados hidrológicos armazenados em banco de dados no formato Hidro 1.2.

No que concerne a consistência de dados pluviométricos, a SGH/ANA utiliza como ferramenta de apoio o aplicativo HidroPlu, desenvolvido pela Universidade Federal de Viçosa – UFV, em 1998, com base na metodologia utilizada pelo sistema PROHD (desenvolvido pelo DNAEE na década de 80). Embora o HidroPlu seja uma ferramenta de análise de dados pluviométricos bastante útil, o mesmo não tem a função de sistematizar os procedimentos de análise dos dados, como é o caso do SiADH. Desta forma, este documento tem por objetivo fazer um levantamento das técnicas e métodos de consistência de dados pluviométricos, visando dar subsídio para a elaboração de um documento de sistematização de processos e metodologias de consistência de dados pluviométricos da ANA, e o consequente desenvolvimento de ferramentas de análise de dados pluviométrico nos moldes do SiADH.

## Objetivos

O objetivo deste documento é fazer um levantamento das técnicas e métodos de consistência de dados pluviométricos, visando dar subsídio para a elaboração de um documento de sistematização de processos e metodologias de consistência de dados pluviométricos da ANA

## Objetivos Específicos

Os objetivos específicos do presente trabalho são:

* Levantamento bibliográficos sobre técnicas e metodologias de análise de consistência de dados pluviométricos;
* Análise crítica referente as técnicas e metodologias de análise de consistência de dados pluviométricos.

# Levantamento Bibliográfico

Para a garantia da qualidade dos dados, diversas exigências devem ser cumpridas pelos responsáveis pela coleta de informações, antes mesmo da utilização ou divulgação das mesmas. Trata-se das atividades de pré-processamento, que incluem a consistência dos dados coletados, o preenchimento e a extensão das séries históricas. Esta etapa dos trabalhos pode ser desenvolvida manualmente ou, por um sistema de informações hidrológicas criado em um ambiente de banco de dados relacional, dado que o nível de tecnologia da informação disponível representa hoje uma base ideal para sistemas otimizados de aquisição, avaliação dos dados e divulgação dos conjuntos de dados hidrológicos (RHA, 2009).

Segundo Tucci (2017), o objetivo de uma estação pluviométrica é obter uma série ininterrupta de precipitação ao longo do tempo, que servirá para se conhecer o regime climatológico de determinado local ou região, estudos específicos sobre intensidade de precipitação, utilização dos dados para modelagem hidrológica, entre outros. A incorporação de dados pluviométricos deve atender o critério de confiabilidade, considerando a sua importância no planejamento e na gestão de recursos hídricos.

Com base na problemática acima decorre que os dados de precipitação, para serem úteis aos diversos tipos de usos deverão ser submetidos a controles mínimos de qualidade, correções e eventual preenchimento de falhas, processo usualmente conhecido como análise de consistência. De um modo geral a análise de consistência pode ser dividida em diferentes etapas, sendo que uma das formas de classificação é apresentada a seguir:

* Analise Preliminar, que visa identificar, registrar e eliminar erros grosseiros contidos nos dados coletados de forma manual ou automática.
* Identificação de regiões homogêneas.
* Analise de inconsistências.
* Correção de dados
* Preenchimento de Falhas

## Analise Preliminar

Na análise preliminar, normalmente são verificadas irregularidades no preenchimento dos boletins. Em seguida os dados diários digitados são avaliados e os totais mensais são comparados com os das estações de apoio. As estações de apoio devem ser selecionadas a partir das estações circunvizinhas à estação em estudo.

No caso de dados coletados automaticamente, por meio de Plataformas Automáticas de Coleta de Dados - PCDs, também deve-se aplicar a análise preliminar dos dados. Esta análise pode ser aplicada tanto de forma manual como automática. Basicamente a análise preliminar automática consta em submeter o dado coletado à filtros e limites pré-estabelecidos de acordo com as características das estações. Uma vez identificado um problema nos dados, a melhor forma de A melhor forma de se conseguir identificar os problemas ocorridos com PCDs é a partir da análise da ficha de inspeção de campo. Nas fichas podem ser checadas as condições gerais do equipamento no momento da visita, realizada a comparação entre os valores lidos pelo equipamento e o observado no local, procedimentos de calibração, entre outros. Além disso, todos os dados provenientes de PCDs devem ser submetidos a filtros automáticos de qualificação, como por exemplo faixa de valores aceitáveis ou fisicamente possíveis de acontecer.

Segue abaixo uma lista dos principais pontos a serem verificados em uma análise preliminar de consistência de dados pluviométricos:

*No boletim pluviométrico:*

* Preenchimento incorreto do cabeçalho
* Preenchimento incorreto das leituras de acordo com o número de dias do mês
* Mudança de caligrafia
* Ilegibilidade das anotações
* Falta de decimal ou mais de um decimal

*Após a digitação dos dados:*

* Erros de digitação
* Repetição de valores iguais, diferentes de zero
* Valores múltiplos de provetas
* Leituras de totais diários muito altos
* Totais mensais discordantes dos valores observados nos apoios

*Dados Automáticos*

* Inserção do código incorreto na PCD ou arquivo de dados com nome trocado, acarretando a inserção equivocada dos dados na base.
* *Datalogger* com relógio interno não ajustado, implicando na inserção de dados com registro de hora incorreta.
* Pluviômetro não calibrado ou entupido (no caso do pluviômetro de báscula), implicando na medição incorreta dos dados.
* Utilização de unidades erradas na importação dos dados. Como exemplo, alguns pluviômetros estão ajustados para o sistema de polegadas, ou registram a quantidade de eventos (basculadas) e não diretamente a precipitação.
* Carga baixa na bateria da PCD, que pode causar a leitura incorreta do pluviômetro.
* Erros durante a transmissão dos dados (para o caso de PCDs telemétricas).

## Métodos de Consistência de Dados Pluviométricos

Segundo Bertoni & Tucci (2013), o “objetivo de um posto de medição de chuvas é o de obter uma série ininterrupta de precipitações ao longo dos anos (ou o estudo da variação das intensidades de chuva ao longo das tormentas) ”. Em qualquer caso, pode ocorrer a existência de períodos sem informações ou com falhas nas observações, devido a problemas com os aparelhos de registro e/ou com o operador do posto. Os dados coletados devem ser submetidos a uma análise antes de serem utilizados.

A análise de consistência dos dados pluviométricos deve ter como objetivo a identificação e correção de erros, bem como o preenchimento de falhas das séries pluviométricas. A condição espacial da precipitação sugere sempre a necessidade de analisar os dados de conjuntos de estações de medição pluviométricas próximas para permitir o preenchimento de lacunas nos registros ou a substituição de dados observados e considerados errôneos (ANA, 2012).

É desejável que o analista tenha conhecimento do regime climático, do sistema de circulação geral e demais processos geradores das chuvas, da orografia, da existência de microclimas e demais fatores que possam influenciar na ocorrência das chuvas na região em estudo. Em muitas ocasiões, totais pluviométricos bastante diferenciados entre estações próximas podem ser explicados por diferenças de altitude, pela localização das estações a barlavento ou sotavento, pela ocorrência de chuvas convectivas etc. A qualidade do observador e a existência de aparelhos registradores são as variáveis mais importantes para decidir pela substituição ou não de um dado duvidoso (DNAEE, 1982; Bertoni & Tucci, 2013).

Na análise preliminar, inicialmente, devem ser verificadas irregularidades na recepção dos dados pelos equipamentos, em seguida deve se proceder à avaliação dos dados diários e dos totais mensais, comparando-os com os das estações de apoio. As estações a serem analisadas deve selecionadas e separadas em dois grupos: estações principais e estações secundárias (ANA, 2012).

As estações principais são os locais a partir de cujos registros os resultados do estudo hidrológico em questão são obtidos e são identificadas a partir de algumas características consideradas básicas, como: localização adequada na bacia para os propósitos do estudo; grande extensão do período de observação (em geral, maior ou igual a 25 anos); poucos períodos de interrupção; presumivelmente bem operada com pouca ou nenhuma alteração na instalação (DNAEE, 1982; Bertoni & Tucci, 2013).

As estações secundárias têm como finalidade principal verificar a existência de erros de observação ou transcrição nos dados das estações principais. Entretanto, em função dos resultados da análise de consistência, poderá uma estação secundária ser levada à condição de principal pela qualidade dos seus registros e das suas condições naturais. De uma forma geral, as estações secundárias ou de apoio, devem ser selecionadas tendo como base, dentre outros, os seguintes aspectos: maiores coeficientes de correlação (no mínimo igual a 0,8); menores distâncias entre as estações (quando possível, distâncias máximas de 200 km); períodos de dados coincidentes; e altitudes semelhantes entre as estações (ANA, 2012).

Dentre as etapas que compõem a consistência dos dados, a delimitação de regiões homogêneas é considerada a mais complexa e sujeita a subjetividades. Uma região é dita homogênea se existem evidências suficientes de que as diferentes estações do grupo possuem a mesma distribuição de frequências. Dentre as diversas técnicas empregadas podem ser citadas: análise de componentes principais; análise de agrupamento (cluster analysis); análise hierárquica, dentre outras (Everitt, 1974; Braga & Targino, 1996; Ward, 1963).

### Método da Ponderação Regional

Segundo Bertoni & Tucci (2013), o método da ponderação regional é um método simplificado normalmente utilizado para o preenchimento de séries mensais ou anuais de precipitações, visando à homogeneização do período de informações e à análise estatística das precipitações, sendo também utilizado para a extensão de séries pluviométricas. O método pode ser classificado segundo o tipo de estatística utilizado na ponderação do peso das estações: médias ou correlação. Abaixo é apresentada uma breve descrição do método conforme descrito em RHA (2009).

*Ponderação por Médias*

Considera-se um grupo de *N* postos, *X1, X2,......., XN-1* e *Y* (pelo menos 3 (três) estações com no mínimo 10 (dez) anos comuns de observação), este último representando a série dependente do grupo homogêneo. O valor proposto para preenchimento ou extensão da série do posto “*Y*”, em determinado mês, é definido pela seguinte equação:



onde,

*y* = total mensal precipitado, estimado (preenchido ou estendido) para o posto “*Y*”, no referido mês;

 = total médio precipitado na estação “*Y*”, no mês em referência, correspondente ao período comum de observação;

 = total médio precipitado para a estação “*Xi*” do grupo homogêneo, no mês em referência, correspondente ao período comum de observação;

*xi* = total mensal observado na estação “*Xi*”, no mês em que o total de precipitação na estação ”*Y*” deve ser preenchido ou estendido.

*Ponderação por Correlação*

O método de ponderação regional por correlação consiste em estabelecer regressões lineares entre o posto com dados a serem preenchidos ou estendidos (posto *Y*), e cada um dos postos vizinhos (postos *Xi, i = 1,2,...,N-1*). Para o período comum de observação, determina-se para cada mês do ano o correspondente vetor de correlação, composto por *N-1* valores determinados pela respectiva correlação linear (*ri*) entre a série da estação *Y* com as correspondentes séries das estações *Xi*. Ou seja, obtêm 12 vetores de correlação, um para cada mês do ano, de forma a levar em consideração os eventuais efeitos de sazonalidade da série em análise. Para um mês específico qualquer, o método é representado pela seguinte equação:



onde,

*y* = total mensal precipitado, estimado (preenchido ou estendido) para o posto “*Y*”, no referido mês;

 = total médio precipitado na estação “*Y*”, no mês em referência, correspondente ao período comum de observação;

sy = desvio padrão do total precipitado na estação “*Y*”, no mês em referência, correspondente ao período comum de observação;

 = total médio precipitado para a estação “*Xi*” do grupo homogêneo, no mês em referência, correspondente ao período comum de observação;

*sxi* = desvio padrão do total precipitado na estação “*Xi*”, do grupo homogêneo, no mês em referência, correspondente ao período comum de observação;

*xi* = total mensal observado na estação “*Xi*”, no mês em que o total de precipitação na estação ”*Y*” deve ser preenchido ou estendido;

*ri* = correlação linear entre a série de total precipitado na estação “*Y*” e a correspondente série na estação “*Xi*”, considerando o período comum de observação no mês em referência.

### Método da Regressão Linear

O método de regressão linear consiste em utilizar regressões lineares simples ou múltiplas para o preenchimento de falhas (Bertoni & Tucci, 2013). Considerando que exista um relacionamento funcional entre os valores *Y* e *X*, responsável pelo aspecto do diagrama, essa função deverá explicar parcela significativa da variação de *Y* com *X* (o valor médio de uma das variáveis em função da outra). Contudo, uma parcela da variação permanece não explicada e deve ser atribuída ao acaso (variação residual). Admitindo ser uma reta a linha teórica de regressão, a função entre *X* e *Y* é a seguinte:

*Y = α + βX + e*

onde, *Y* é a variável dependente, *X* é a variável independente, *α* e *β* são os coeficientes do modelo e “*e”* representa os erros ou resíduos da regressão.

Os coeficientes *α* e *β* da reta teórica são estimados através dos dados observados fornecidos pela amostra. Um dos procedimentos objetivos mais adequados é a aplicação do método dos mínimos quadrados, segundo o qual a reta a ser adotada deverá ser aquela que torna mínima a soma dos quadrados dos erros ou resíduos da regressão. (Naguettinni & Pinto, 2007).

Na regressão múltipla estuda-se o comportamento de uma variável dependente *Y* em função de duas ou mais variáveis independentes *Xj, j = 1, ..., p*. Portanto, um possível modelo para avaliar essa relação pode ser dado por

*Yi =* *β0 + β1Xi1+ β2Xi2+...+ βpXip + ei , i = 1, ..., n*

onde, *n* é o número de observações, *Yi* é a observação da variável dependente para o *i-ésimo* indivíduo, *Xi = (Xi1, Xi2, ..., Xip)’* é um vetor de observações das variáveis independentes para o *i-ésimo* indivíduo, *β = (β0, β1, β2, ..., βp)’* é um vetor de coeficientes de regressão (parâmetros) e *ei* é um componente de erro aleatório. Assume-se que esses erros são independentes e seguem distribuição normal com média zero e variância desconhecida *σ2*.

De maneira análoga à regressão linear simples, os coeficientes desconhecidos *βi* podem ser estimados pela minimização do somatório dos erros quadráticos, (Naguettinni & Pinto, 2007).

Na regressão linear simples (diagrama de dispersão), as precipitações do posto com falhas e de um posto vizinho são correlacionadas. Na regressão linear múltipla as informações pluviométricas do posto com falhas são correlacionadas com as correspondentes observações de vários postos vizinhos (Bertoni & Tucci, 2013).

### Método da Dupla Massa

O método da dupla massa (ou curva dupla acumulativa) é a técnica de análise de dados pluviométricos mais utilizada no Brasil. A metodologia foi desenvolvida pelo Serviço Geológico dos Estados Unidos - USGS (Searcy, 1960), e pode ser utilizada tanto para consistência de séries mensais quanto anuais. Basicamente, o método consiste em:

* Selecionar as estações da região a ser analisada (estações de apoio);
* Acumular, para cada estação, os valores mensais (ou anuais) de precipitação;
* Plotar, em um gráfico cartesiano, os valores acumulados correspondentes à estação em análise (nas ordenadas) contra os valores acumulados de outra estação confiável (estação de referência) adotada como base de comparação (nas abscissas). Também pode ser utilizada como base de comparação, ao invés de uma única estação de apoio, a média das estações selecionadas.

Por esse método é possível verificar o grau de homogeneidade dos dados disponíveis na estação em análise com relação às observações registradas nas estações vizinhas. Se os valores da estação a consistir são proporcionais aos observados na base de comparação, os pontos devem-se alinhar segundo uma única reta. A declividade da reta determina o fator de proporcionalidade entre ambas as séries. Com isso, pode-se verificar a existência de anormalidades na estação pluviométrica, tais como: mudança de local ou das condições do aparelho ou modificação no método de observação. As possíveis situações podem ser identificadas a partir da análise do gráfico de dupla massa (HidroPlu, 2000):

* Mudança na declividade, determinando duas ou mais retas: Constitui o exemplo típico derivado da presença de erros sistemáticos, mudança nas condições de observação ou a existência de uma causa física real, como alterações climáticas no local provocadas pela presença de reservatórios artificiais, por exemplo. Para se considerar a existência de mudança na declividade, é prática comum exigir a ocorrência de pelo menos cinco pontos sucessivos alinhados segundo a nova tendência.
* Alinhamento dos pontos em retas paralelas: ocorre quando existem erros de transcrição de um ou mais dados ou pela presença de anos extremos em uma das séries plotadas. A ocorrência de alinhamentos, segundo duas ou mais retas aproximadamente horizontais (ou verticais), pode ser a evidência da comparação de postos com diferentes regimes pluviométricos;
* Distribuição errática dos pontos: geralmente é resultado da comparação de postos com diferentes regimes pluviométricos, sendo incorreta toda associação que se deseje fazer entre os dados dos postos plotados.

Para corrigir os valores correspondentes ao posto em análise, existem duas possibilidades: corrigir os valores mais antigos para a situação atual ou corrigir os valores mais recentes para a condição antiga. Sendo o mais utilizado o primeiro dos casos.

### Método do Vetor Regional

O método do Vetor Regional, desenvolvido por Hiez (1977 e 1978), e constitui uma forma de realizar análise de consistências e preenchimentos de falhas de dados pluviométricos em níveis mensais e anuais. O vetor regional é definido como “uma série cronológica, sintética, de índices pluviométricos anuais (ou mensais), oriundos da extração por um método de máxima verossimilhança da informação contida nos dados de um conjunto de estações agrupadas regionalmente”.

Esses índices, conforme indica Heinz (1987), são únicos para toda região e estão relacionados às alturas precipitadas em cada posto por meio de coeficientes apropriados, de modo que a estimativa da altura precipitada (anual ou mensal), no ano “*i*”, no posto “*j*”, resulta *Pij = Li Cj*. Onde vetor coluna “*Li*” é o vetor regional e o vetor linha “*Cj*” é o vetor de coeficientes.

O sistema de *n x m* equações, que surge quando *Li (i=1,...n)* e *Cj (j=1,...m)* forem as incógnitas, é, em geral, incompatível, exigindo que, para a estimativa tanto do vetor regional quanto do vetor de coeficientes, se lance mão do método dos mínimos quadrados. Define-se então uma matriz de erros *Eij = Li x Cj x Pij*, devendo-se minimizar a soma dos quadrados dos seus elementos. Segue-se um processo iterativo, partindo-se duma estimativa inicial para o vetor regional, utilizando-se, como exemplo, a média aritmética das precipitações registradas a cada ano *i* nas *m* estações. Com essa estimativa inicial, aplica-se para cada *j* o método dos mínimos quadrados (Heinz, 1987). Uma vez determinado o vetor regional e os coeficientes das estações, calcula-se os erros absolutos (*Eij*) e os relativos (*εij*) dados por *εij = Eij / Li Cj*, e os seus valores acumulados dados por *Σki=1 εij* e *Σki=j Eij,* com *1≤ k ≤ n*.

A análise visual de um gráfico mostrando os erros simples ou acumulados em função do tempo permite detectar erros sistemáticos, erros grosseiros, e anomalias climáticas locais, e estimar valores para preenchimento de falhas nas observações.

Segundo Heinz (1987), o método é adequado apenas para totais anuais ou mensais. No caso mensal, é conveniente usar um conjunto de 12 vetores regionais, um para cada mês, ou pelo menos vetores diferenciados em função de cada estação do ano. Para análise de precipitações diárias, não há notícia da utilização desse processo e acredita-se mesmo que as fortes variações espaciais e temporais das chuvas impeçam essa aplicação.

Em relação ao tratamento de dados pluviométricos em nível mensal observa-se uma dificuldade adicional, posto que os desvios mensais não seguem, em geral, uma distribuição de probabilidades de tipo normal. Assim, adota-se apenas a compatibilização da série mensal consolidada com sua correspondente anual.

Além da simplicidade, o procedimento descrito apresenta a vantagem de corrigir principalmente os meses com maior número de dias chuvosos, o que é coerente com a idéia de proporcionalidade entre o número de leituras efetuadas pelo operador ao longo do mês e a probabilidade de leituras errôneas. No caso de meses sem nenhuma informação, seus totais poderão ser preenchidos pelos valores calculados com base no Vetor Regional mensal. Tais valores são totalizados anualmente calculando-se, em seguida, o percentual de contribuição de cada um destes sobre esta soma (DNAEE, 1982; Bertoni & Tucci, 2013).

### Modelo Multivariado Sazonal Mensal Auto-Regressivo de Ordem 1 - SMMAR(1)

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. Os métodos de estatística multivariada são utilizados em situações nas quais diversas variáveis são medidas simultaneamente, em cada elemento amostral. Para ser considerada verdadeiramente como análise multivariada, as variáveis envolvidas devem ser aleatórias e inter-relacionadas de maneira que seus diferentes efeitos não podem ser significativamente interpretados de forma separada (Hair. et al, 1998).

A regressão linear múltipla é uma técnica multivariada cuja finalidade principal é obter uma relação matemática entre uma das variáveis (a variável dependente) e o restante das variáveis que descrevem o sistema (variáveis independentes). Sua principal aplicação é produzir valores para a variável dependente quando se têm as variáveis independentes. Desta forma a regressão linear múltipla é utilizada na predição de resultados (RHA, 2009).

A RHA (2009) utilizou, em um estudo contratado pela Agencia Nacional de Águas, um Modelo Multivariado Sazonal Mensal Auto-Regressivo de Ordem 1, denominado SMMAR(1), para verificação da consistência e correção dos dados pluviométricos. Este modelo baseia-se na teoria da distribuição multivariada normal, tornando independentes variáveis normalmente distribuídas. Em síntese, para um dado grupo de estações (séries históricas mensais de precipitação) o modelo reproduz automaticamente todas as equações de regressão e aplica, em cada caso, a equação de regressão possível em função da disponibilidade de dados no mês em referência e do mês anterior. O ajuste é feito, um para cada mês, preservando as características sazonais da série em análise, previamente normalizada pela transformação Box-Cox (RHA, 2009).

A transformação Box e Cox, sobre a série pluviométrica original (*Xt*), baseia-se na determinação do valor do parâmetro *λ*, na função



sendo que é um parâmetro que pode ser estimado impondo-se a condição de que o coeficiente de assimetria da série transformada (*Yt*) resulte igual a zero, *Ca(Yt)=0* (correspondente ao valor do coeficiente de assimetria da distribuição normal). No estudo contratado pela ANA a equação *Ca(Yt)=0* foi solucionada usando-se o método do recozimento simulado para estimar que minimiza |*Ca*(*Yt*)|.

O modelo SMMAR(1) pode ser aplicado através do uso da seguinte relação:

*y(i) =m(i)+v(i), i=1,...,k,*

sendo *y(i)* a transformação Box-Cox do total precipitado mensal, para o local *i*, pertencente a um conjunto (considerado homogêneo) com séries temporais (pluviométricas) para n locais; *k* o número de locais (séries de um determinado mês) com valores a serem estimados; *n\*k* o número de locais com séries do mês anterior (valores considerados conhecidos); *m(i)* a média da série de *Y(i).*

*O termo*  *v(i)* é determinado pela expressão:

*v(i)=[zl(i+1,i)v(i+1)...l(n,i)v(n)]/l(i,i), i=k,...,1,*

sendo *z* um número aleatório normal com média zero e variância unitária; *l(i,j)*, com *i=1,...,n* e *j=1,...,i*, são elementos de uma matriz triangular inferior obtida pela fatoração de Cholesky da inversa da matriz de covariâncias de *Yt*; os valores de *v(i)*, para *i=k+1,...,n*, devem ser conhecidos.

Em resumo, o modelo SMMAR(1) deverá ser implementado para cada mês do ano, para as precipitações transformadas pela relação de Box e Cox. Desta forma, para cada modelo necessita-se de uma matriz de covariâncias de dimensão igual ao dobro do número de estações consideradas, para cada mês do ano.

### Método de Redes Neurais Artificiais

Uma rede neural artificial (RNA) é um grupo interligado de neurônios artificiais que tem a propriedade natural de armazenar conhecimento experimental e torná-lo disponível para uso. Uma arquitetura de RNA ótima pode ser considerado como o que proporcionou melhores resultados em termos de minimização do erro, enquanto mantem uma estrutura simples e compacta. Há duas questões importantes relativas à implementação de redes neurais artificiais, ou seja, especificando a arquitetura de rede (o número de nós e camadas na rede) e encontrar os valores ideais para os pesos de conexão (seleção de um algoritmo de treinamento). No processo de especificação do tamanho da rede, um número insuficiente de nós escondidos provoca dificuldades na aprendizagem de dados enquanto que um número excessivo de nós escondidos pode levar a tempo desnecessário de treinamento com uma melhora marginal no resultado de treinamento, bem como tornar a estimativa de um conjunto adequado de pesos de interligação mais difícil (Hung et al, 2009).

O aprendizado em Redes Neurais Artificiais – RNA consiste em, a partir de treinamento sistemático, ajustar pesos (a força de uma conexão) de uma forma sistemática. Independentemente do ajuste de peso, o treinamento da rede neural pode ocorrer de duas maneiras: supervisionado ou não supervisionado.

No treinamento supervisionado, os dados de entrada e saída são fornecidos para a rede em um esforço para ensiná-lo a produzir os vetores de saída desejados. No treinamento não supervisionado, os dados são simplesmente inseridos na rede sem qualquer intervenção humana ou informações adicionais. O treinamento é obtido a partir da formação de construções internas que capturam regularidades em seus vetores de entrada. Em cada método de treinamento o aprendizado é concluído quando a rede atinge um determinado critério de estabilidade. Inicialmente, os pesos de conexão são definidos como pequenos valores aleatórios. Os dados de entrada são então fornecidos à rede, fazendo-os passar por alterações de estado que, consequentemente, introduz alterações nos valores dos pesos. A estabilidade é alcançada quando não há mais alterações de peso. De fato, uma rede neural aprende através de adaptações que são determinadas pelas regras de aprendizagem da metodologia utilizada (Michaelides et al., 2001).

Assim como no método de regressão linear, o emprego de redes neurais artificiais pode também ser útil no preenchimento de falhas e análise de séries pluviométricas (Freitas, 1998; Billib & Freitas, 1996; Depiné et al.,2014; Michaelides et al., 2001; Ghuge, 2013).

### Técnicas de Agrupamento (clusters)

Uma das formas de realizar o agrupamento de estações com comportamentos pluviométricos homogêneos é utilizando a técnica de análise de agrupamento (cluster). Dentre as técnicas de agrupamento mais utilizados para análise de precipitação destaca-se o algoritmo hierárquico k-means. Este algoritmo leva em consideração a similaridade geométrica entre as séries acumuladas de precipitação das diferentes estações. O classificador k-means (Jain et al., 1999) divide o conjunto de dados em n classes (clusters) por meio de um algoritmo de partição no qual as séries de chuva acumulada são aglomeradas iterativamente em classes pelo critério de distância mínima. A medida de distância utilizada é a separação angular entre as séries. Primeiramente é especificada, de forma aleatória, uma série média (centroide) para cada uma das n classes. Cada membro do conjunto de treinamento é alocado à classe em função da menor distância em relação à série de chuva média, formando assim o primeiro conjunto de decisões. Para cada classe, um novo centroide é calculado a partir da média das séries pertencentes à classe, e as séries são submetidas a uma nova classificação, baseado neste novo centroide, podendo ser realocados, segundo a nova configuração. O algoritmo termina quando um número de iterações pré-definido pelo usuário for atingido, ou, quando o número de espectros que mudam de classes alcance um determinado valor limite. Uma função objetiva é utilizada para calcular a diferença média ∆µ dos espectros com relação aos respectivos centroides em duas iterações consecutivas.

Entre os estudos de agrupamento de regiões homogêneas de precipitação pode-se citar Matualla et al. (2003); Munõz & Rodrigues (2004); Ahmad et al. (2013) e Conde et al. (2010).

### Componentes Principais

A técnica denominada de Análise de Componentes Principais, popularmente chamada de PCA, é uma técnica exploratória de sintetização ou de simplificação da estrutura de variabilidade dos dados e constitui-se em um método da Estatística Multivariada. Seu objetivo principal é o de explicar a estrutura de variância e covariância de um vetor aleatório, composto por p-variáveis aleatórias, através da construção de combinações lineares das variáveis originais. Estas combinações lineares são chamadas de componentes principais e são não correlacionadas entre si. A informação contida nas *p-variáveis* originais é substituída pela informação contida em k (k < p) componentes principais não correlacionadas. Desta forma, o sistema de variabilidade do vetor aleatório composto das p-variáveis originais é aproximado pelo sistema de variabilidade do vetor aleatório que contém as k componentes principais. A qualidade da aproximação depende do número de componentes mantidas no sistema e pode ser medida através da avaliação da proporção da variância total explicada por essas. Quando a distribuição de probabilidades do vetor aleatório em estudo é normal p-variada, as componentes principais, além de não correlacionadas, são independentes e têm distribuição normal. No entanto, a suposição de normalidade não é um requisito necessário para aplicação da técnica de análise de componentes principais (RHA, 2009).

O uso mais comum da técnica de componentes principais em consistência de dados hidrometeorológicos está associado a identificação de grupos de estações que possuem características de homogeneidade. Existem na literatura vários estudos que utilizam componentes principais com este propósito (RHA, 2009; Braga et al. 2012; Mills, 1995; Baeriswyl & Rebetez).

### HidroPlu

O programa HidroPlu baseia-se na formulação matemática desenvolvida por Holanda & Oliveira (1979), com o objetivo de auxiliar na análise de consistência de dados pluviométricos. O programa visa, em princípio, a análise e consistência dos totais pluviométricos mensais ou anuais da série de observações realizadas.

O HidroPlu utiliza o método da ponderação regional com base em regressões lineares para realizar as críticas e sugestões de correções e preenchimento de falhas. Este método consiste em estabelecer regressões lineares entre os postos com dados a serem preenchidos, *Px*, e cada um dos postos vizinhos, *P1, P2, ..., PN*. De cada uma das regressões lineares efetuadas obtém-se o coeficiente de correlação.

Na sequência são apresentadas algumas etapas e procedimentos a serem implementados para a realização da análise de consistência com a utilização do HidroPlu.

*Escolha do período de estudo*

O usuário deverá escolher o período que será utilizado para a realização dos cálculos, sendo que diversos parâmetros sofrerão influência do período escolhido, podendo-se citar os coeficientes de correlação, as críticas aos dados e até mesmo o valor estimado para um determinado dado criticado.

*Escolha do tipo de processamento*

O processamento pode ser feito em nível mensal ou anual, sendo o processamento mensal mais utilizado.

*Escolha do tipo de média de teste*

Os testes podem ser feitos com base em médias de longo período ou mensais. Utilizando-se a média de longo período a crítica é feita comparando-se os dados de todos os meses das séries referentes à estação em análise e a cada estação de apoio.

*Escolha das estações de apoio*

Devem ser escolhidas no mínimo uma e no máximo nove estações de apoio, sendo que, preferencialmente, o número de estações deve ser suficiente para que se possuam estações em torno de toda a circunferência próxima da estação em análise. A escolha das estações de apoio é baseada nos seguintes critérios:

* Situação geográfica das estações - São calculadas, com base na latitude e longitude de cada estação, as distâncias entre a estação em análise e as estações de apoio.
* Período de observação - O período de observação das estações também é disponibilizado para que se possa comparar o período de funcionamento das estações de apoio com o da estação em análise.
* Coeficiente de correlação - Um outro parâmetro para a escolha das estações de apoio é o coeficiente de correlação. Este coeficiente procura determinar o grau de relacionamento entre duas variáveis (no caso específico, o total precipitado da estação em análise e de cada estação de apoio).

*Escolha do número de estações envolvidas nas críticas*

Normalmente, mesmo que o usuário utilize nove estações de apoio para fazer uma determinada consistência, escolhem-se apenas algumas delas para serem utilizadas no processamento das críticas. As demais são utilizadas apenas para o cálculo do valor a ser estimado pelo programa.

*Parâmetros de calibragem*

Três parâmetros usados no HidroPlu (G1, G2 e G5) permitem a calibração do modelo. Esses parâmetros têm influência direta na averiguação da qualidade dos dados.

Para a execução do programa deve-se indicar os valores dos parâmetros G1 e G2, que correspondem aos parâmetros de calibragem para homogeneização de dados em primeira e segunda críticas, respectivamente. Pode-se também indicar o valor de G5, que corresponde ao parâmetro de calibragem correspondente à variação proporcional entre o valor de dados testado e o da sua aceitabilidade.

# SISTEMÁTICAS DE PROCEDImentos de consistência DE DADOS PLUVIOMÉTRICOS UTILIZADOS EM ESTUDOS JÁ REALIZADOS

A Organização Meteorológica Mundial – OMM tem por atribuição, entre outras, promover a padronização de observações meteorológicas e hidrológicas, assegurando a uniformidade de publicações das observações e estatísticas. Historicamente, pode-se citar a França e os Estados Unidos como países que tiveram forte influência nos estudos e metodologias de consistência de dados adotadas no Brasil, com destaque para os trabalhos desenvolvidos pela IRD (antiga ORSTON) na França, e USGS nos USA. Em nível nacional destaca-se a atuação de entidades como ANEEL, ONS, CPRM e ANA na realização ou contratação de estudos de análise de consistência de dados pluviométricos.

## Organização Meteorológica Mundial - OMM (WMO)

Segundo a WMO (2008), atualmente os dados hidrometeorológicos são coletados e registrados de muitas formas, variando desde leituras manuais, utilizando medidores simples, a uma multiplicidade plataformas automáticas de coleta de dados de dados, tipos de transmissão e sistemas de armazenamento de dados. Com o desenvolvimento acelerado da tecnologia, se tornou ainda mais importante que o sistema de processamento e controle de qualidade dos dados seja bem organizado e compreendido pelos usuários envolvidos tanto na coleta, processamento e uso dos dados. Neste sentido, o Guia de Práticas Hidrológicas (WMO, 2008) e o Guia de Práticas Meteorológicas (WMO, 2011) apresentam recomendações referentes a práticas e procedimentos que devem ser adotados para assegurar a padronização e qualidade das informações hidrometeorológicas coletadas e disponibilizadas (WMO, 2008).

Para a OMM o hidrólogo tem o dever de ser conservador no que concerne a realização de qualquer correção de dados, e sugere-se o uso de critérios rigorosos para alterar ou adicionar valores de dados. Isso deve ser feito sempre usando pressupostos com base em evidências, e não em suposições. Qualquer alteração efetuada nos dados deve ser registrada de forma que outros usuários possam ter conhecimento do que foi feito e o motivo da atualização.

Para os dados registrados em *datalogger*, verificações preliminares da série temporal e comparação com leituras manuais devem ser incluído como parte dos procedimentos realizados em campo. Para dados provenientes de telemetria, recomenda-se que seja implementado um processo de verificação automatizada. Contudo, este tipo de validação só será capaz de verificar alguns aspectos dos dados (ex. faixa, pontos, etapas ou falha de dado), e o usuário deve ser informado destas limitações do controle automatizado de qualidade.

Os procedimentos de validação de dados podem incluir desde verificações simples, complexas e possivelmente automatizados que está sendo executada em várias etapas no caminho de processamento de dados e arquivamento. Alguns também podem ser realizadas em saídas de dados e análises estatísticas por um usuário de dados informados.

Embora as técnicas de validação de dados estejam se tornando mais úteis e poderosas, deve-se reconhecer que elas nunca poderão ser totalmente automatizadas, chegando ao ponto em que o hidrólogo não precise mais verificar os dados. De fato, para obter o melhor desempenho, o hidrólogo necessita ajustar constantemente os valores limites no programa, e terá de exercer um julgamento informado e considerado sobre a possibilidade de aceitar, rejeitar ou valores de dados corretos sinalizados pelos programas. Os valores mais extremos podem revelar-se corretos e, nesse caso, são de vital importância para todos os aplicativos de dados hidrológicos.

Não há dúvida de que a verificação visual das séries temporais de dados plotados por pessoal experiente é uma técnica muito rápida e eficaz para a detecção de anomalias de dados. Por esta razão, a maioria dos sistemas de validação de dados deve incorporar ferramentas de visualização gráfica dos dados, e a comparação com dados de outras estações adjacentes. Segundo a OMM os seguintes itens devem ser verificados:

* Limites esperados de variáveis;
* Variação máxima esperada em uma variável entre observações sucessivas;
* A diferença máxima esperada em variáveis entre estações adjacentes.

## ORSTOM / IRD

Na década de 80, a ORSTOM desenvolveu os pacotes computacionais para hidrologia, Hydrom e Pluviom, bastante conhecidos mundialmente. Também se pode mencionar o Vetor Regional (consistência e preenchimento de series pelo método de índices) e diferentes módulos para ajuste de distribuições de probabilidade e funções de digitação de gráficos de níveis de rios e totais pluviométricos. Na sua linha de sucessão a IRD (French Research Institute for Development) iniciou em 2000 o desenvolvimento do software livre HYDRACCESS, com base no Microsoft Office. Este programa computacional pode gerenciar séries cronológicas de níveis de água, vazão, qualidade de água, precipitação e dados meteorológicos. Neste software a hidrologia é considerada de forma ortodoxa, aplicando métodos e procedimentos consagrados para o tratamento de séries hidrológicas.

## U.S. Geological Survey - USGS

A análise da dupla massa, largamente utilizado no Brasil, foi desenvolvida pelo U.S. Geological Survey como técnica de ajuste de registros hidrológicos inconsistentes. Esta técnica está descrita no Manual de Hidrologia do USGS (Double-Mass Curves) (Searcy, 1960). Este mesmo manual sugere o preenchimento de falhas em séries históricas pelo método da interpolação linear, usando dados de três estações adjacentes. Outros trabalhos do USGS (Water-Resources Investigation Report), publicados entre 1999 e 2008, sugerem a aplicação da análise de tendências, regressão linear, técnicas da curva de dupla massa e curva de frequência acumulativa em trabalhos de consistência hidrológica. Hidrogramas simultâneos também são mencionados tanto para preenchimento como para consistência de séries hidrológicas (RHA, 2009).

## Agência Nacional de Energia Elétrica - ANEEL

Entre os anos de 1998 a 2000 a Agência Nacional de Energia Elétrica - Aneel foi responsável pela coordenação da Rede Hidrometeorológica Nacional, atualmente sob a coordenação da ANA. Neste período foram contratados vários estudos de análise de consistência de dados pluviométricos, entre os quais pode-se citar IPH-UFRS (1999), FUB-UnB (1999) e CERPCH-EFEI (1999). Todos os estudos coordenados pela Aneel foram baseados no uso do Programa de Homogeneização de Dados - PROHD” para crítica, correção de dados e preenchimento de falhas. Após a consistência dos dados mensais, usando o PROHD, os dados diários foram compatibilizados.

## Operador Nacional do Sistema Elétrico - ONS

O Operador Nacional do Sistema Elétrico - ONS realizou vários contratos para consistência de dados pluviométricos em importantes bacias do país. Entre os estudos contratados pela ONS podem-se citar os relatórios de consistências de dados pluviométricos realizados pela Hicon (2003), Lactec (2003), Hicon (2008) e Enerconsult-Hidrosistem-Internave (2008).

No estudo realizado na bacia do rio Paraíba do Sul, pelo consorcio Enerconsult-Hidrosistem-Internave, foi aplicado o método do vetor regional às séries de totais precipitados mensais, que compõem cada região homogênea, para identificar e corrigir as principais inconsistências, bem como a complementação das falhas. O método da Curva Dupla-Acumulativa foi aplicado, nas respectivas séries de totais precipitados anuais que compõem cada região homogênea, para identificação de inconsistências quanto às tendências.

No estudo do Paranapanema, pelo Lactec, foi utilizado o Método da Ponderação Regional por médias para identificação e preenchimento de falhas, seguido da aplicação da curva duplo acumulativa para análise de consistência. Nos casos cuja aplicação das metodologias acima não foi suficiente para remover as inconsistências realizou-se o traçado dos hietogramas, mensais e anuais, para detectar as leituras possivelmente incorretas, e aplicou-se o método de regressões lineares simples e múltipla para a correção destas leituras.

Nos estudos realizados pela Hicon Engenharia, nas bacias do Parnaíba, Jequitinhonha e Doce, e na bacia do rio Grande, a metodologia utilizada para a análise de consistência das séries pluviométricas incluiu a elaboração das curvas duplo-acumulativas para identificação de inconsistências, e a aplicação de regressão linear, com as series de maior coeficiente de determinação, para correção dos dados e preenchimento de falhas.

## Serviço Geológico do Brasil - CPRM

O Serviço Geológico do Brasil - CPRM também realiza estudos de consistência pluviométrica da rede Hidrometeorológica Nacional. Os procedimentos da análise de consistência utilizados pela CPRM estão sendo documentados em um manual de consistência de dados pluviométricos (CPRM, 2014), cujas principais pontos são listados a seguir:

* Realização de uma análise preliminar onde, inicialmente, são verificadas irregularidades no preenchimento dos boletins, em seguida os dados diários digitados são avaliados e os totais mensais são comparados com os das estações de apoio.
* Analise visual comparando, em um mapa, os totais anuais das estações em estudo com as médias dos totais anuais de longo termo destas estações. Quando disponível utiliza-se um mapa de isoietas para comparação. Analisa-se a proporcionalidade entre os totais anuais em análise e os verificados nas médias de longo termo. A ocorrência de discrepâncias nesta proporcionalidade pode ser um indicativo de inconsistência nas leituras.
* O HidroPlu é utilizado para avaliar os totais mensais, para cada estação. Sempre que o total mensal observado for rejeitado e substituído por um valor estimado, é realizada a compatibilização dos dados diários de modo a ficarem coerentes com o total mensal adotado. Para tal é deletado os dados diários considerados duvidosos. Quando não é possível determinar os dias em que ocorreram as leituras incorretas, todo o mês é deletado. Em caso de situações excepcionais ou de chuvas extraordinárias, admite-se a manutenção deste tipo de identificação, a critério do analista.
* Após correção e preenchimentos dos totais mensais, e compatibilização dos dados diários, procede-se à análise dos totais anuais. Essa análise pode ser realizada através da análise dos índices pluviométricos apresentados pelo HidroPlu e pela análise do mapa da bacia em que se encontram plotadas as médias dos totais anuais de longo termo e dos totais anuais consistidos do ano em estudo.
* Por último, recomenda-se a aplicação do HidroPlu para análise das curvas duplo-acumulativas, e avaliar a consistência e homogeneidade das séries.

Segue abaixo o procedimento detalhada realizado pelo Serviço Geológico do Brasil – CPRM:

*Análise Preliminar*

* Na análise preliminar, inicialmente são verificadas irregularidades no preenchimento dos boletins, em seguida os dados diários digitados são avaliados e os totais mensais são comparados com os das estações de apoio. Como estações de apoio devem ser selecionadas as estações mais próximas e circunvizinhas à estação em estudo.
* Nos boletins podem ser verificadas as seguintes anomalias:
* preenchimento incorreto do cabeçalho (código, nome, mês e ano)
* preenchimento incorreto das leituras de acordo com o número de dias do mês
* mudança de caligrafia
* ilegibilidade das anotações
* falta de decimal ou mais de um decimal
* Após a digitação, devem ser verificados:
* erros de digitação
* repetição de valores iguais, diferentes de zero
* valores múltiplos de provetas
* leituras de totais diários significativos (por exemplo, maiores do que 100mm)
* totais mensais discordantes dos valores observados nos apoios
* Na Tabela 1 são apresentadas as principais ocorrências, os procedimentos adotados e normas de identificação no banco de dados. Após a conclusão dos estudos preliminares, as ocorrências encontradas e os procedimentos adotados para cada estação, devem ser anotados em uma memória técnica e as irregularidades repassadas para as equipes de campo.

**Tabela 1** - Análise Preliminar de Dados Pluviométricos



*Análise dos Dados*

* Para que se adquira um conhecimento preliminar do regime pluviométrico da bacia em análise, sugere-se que sejam plotadas em um mapa as médias dos totais anuais de longo termo de todas as estações em estudo ou que se disponha de um mapa de isoietas de precipitação média anual.
* Uma primeira verificação pode ser realizada plotando-se nesse mesmo mapa os totais do ano em análise e comparando-os com os de longo termo. O esperado é que a proporcionalidade entre os totais anuais em análise seja semelhante à verificada nas médias de longo termo. A ocorrência de discrepâncias nesta proporcionalidade pode ser um indicativo de inconsistência nas leituras.
* Em seguida, os totais mensais são avaliados, para cada estação, segundo critérios pertinentes, como por exemplo, os adotados no HidroPlu. Estes critérios geralmente são estabelecidos através do estudo do comportamento da série em análise em relação às séries históricas das estações de apoio.
* No caso do HidroPlu, intervalos de aceitação são determinados em função dos coeficientes de correlação e da proporcionalidade entre as médias históricas da estação em análise e seus apoios. Os totais mensais verificados fora destes limites são criticados e novos valores são sugeridos para substituí-los.
* No caso da adoção deste aplicativo, recomenda-se a utilização de toda a série histórica, porém, caso a estação possua poucos anos de dados, que sejam utilizadas séries históricas com, no mínimo, 5 anos de período comum. Deve-se estar atento quanto à qualidade das séries históricas das estações de apoio e da existência de boa correlação entre as elas, o que pode ser verificado através dos coeficientes de correlação. As médias de longo período fornecem indicativos da pluviosidade das estações comparadas, permitindo a utilização de estações com diferentes altitudes. A ordem dos apoios dentro do aplicativo deve obedecer a ordem decrescente do coeficiente de correlação, ou seja o primeiro apoio deve ser a estação que apresente o maior coeficiente de correlação.
* Na Tabela 2 são relacionados os critérios utilizados para aceitação dos valores sugeridos pelo HidroPlu na metodologia de consistência de dados pluviométricos utilizada pela CPRM.
* Ao final da consolidação, espera-se que sejam dirimidas todas as dúvidas encontradas na análise preliminar e eliminadas as identificações de dados duvidosos do arquivo de dados consistidos. Entretanto, em situações excepcionais ou de chuvas extraordinárias, admite-se a manutenção deste tipo de identificação, a critério do analista.
* Após correção e preenchimentos dos totais mensais e compatibilização dos dados diários, procede-se à análise dos totais anuais. Essa análise pode ser realizada através da análise dos índices pluviométricos apresentados pelo HidroPlu e pela análise do mapa da bacia em que se encontram plotadas as médias dos totais anuais de longo termo e dos totais anuais consistidos do ano em estudo.
* Após o procedimento de correção do dado e preenchimento de falhas, realiza-se a análise de curvas duplo-acumulativas para avaliar a consistência e homogeneidade da série.

**Tabela 2** – Critérios para Aceitação de Correção e Preenchimento de Falhas.



## Agência Nacional de Águas - ANA

Conforme mencionado anteriormente, a partir de 2000 a ANA passou a gerir a Rede Hidrometeorológica Nacional - RHN, que atualmente conta com cerca de 4.500 estações sendo: 2.700 estações pluviométricas e 1.800 estações fluviométricas. Entre as atribuições da ANA, em relação à RHN, ressalta-se o controle de qualidade dos dados hidrometeorológicos de estações de responsabilidade da ANA.

Desde 2001 a ANA passou a adotar os procedimentos de consistência de dados pluviométricos utilizados pela ANEEL, destacando a utilização da ferramenta HidroPlu. Em 2009, a ANA realizou a contratação de um projeto com objetivo de qualificar as séries mensais de precipitação e vazão no país, além de reconstituir séries de vazões naturais em estações fluviométricas selecionadas (RHA, 2011).

Neste projeto foi realizada a análise de consistência pluviométrica de 2.549 estações, distribuídas nas doze regiões hidrográficas brasileiras, na base mensal e no período de 1931 a 2007. O objetivo desta análise de consistência pluviométrica foi verificar a homogeneidade dos dados, de forma a corrigir eventuais anomalias ocorridas na estação pluviométrica, como por exemplo, erros aleatórios, sistemáticos e/ou grosseiros.

O primeiro passo no processo da consistência dos dados pluviométricos consistiu na identificação de similaridade regional entre as estações, a partir da formação dos denominados “Grupos Homogêneos”, obtidos a partir da análise de componentes principais e conveniência geográfica (critério da distância) do conjunto de estações pluviométricas selecionadas.

A partir destes grupos foram construídas curvas duplo-acumulativas para a avaliação de eventual inconsistência. Esta análise aplicou-se aos dados disponíveis das séries pluviométricas selecionadas para o projeto, no período de janeiro de 1931 a dezembro de 2007. As séries resultantes desta análise, através das curvas duplo acumulativas, serviram de insumo à etapa seguinte de preenchimento e extensão de séries, ocasião em que as alterações propostas foram validadas ou revistas.

Foi realizado o preenchimento e extensão das séries pluviométricas de cerca de 2.500 estações, distribuídas, ao longo das doze regiões hidrográficas brasileiras. Para o preenchimento e extensão das séries hidrológicas, foram utilizados três modelos (Modelo Multivariado Sazonal Mensal Auto-Regressivo de Ordem 1- SMMAR[1], Ponderação Regional por Médias - PRM e Ponderação Regional por Correlação - PRC) para a geração de séries sintéticas e a seleção do melhor ajuste pela comparação da performance com o modelo perfeito, que fornece o valor gerado igual ao valor observado em condições naturais. As estatísticas e índices produzidos nesta etapa serviram de insumo à “Qualificação das Séries Pluviométricas”.

Tendo por base o estudo de consistência e preenchimento desenvolvido sobre as séries hidrológicas, foram estabelecidos índices que permitem qualificar, por nota, as séries pluviométricas analisadas. Entre os índices utilizados para qualificação das séries pluviométricas, tem-se o índice que descreve a qualidade geral das séries, denominado Índice Q, o qual considera o número de falhas, comprimento da série, além dos métodos utilizados para a consistência, preenchimento e extensão. Outro índice de qualificação considerado, Índice E, refere-se à avaliação da estacionariedade das séries considerando a tendência e a ocorrência de saltos (*shifts*) na média.

A Tabela 3, a seguir, apresenta um resumo dos principais critérios utilizados pela RHA Engenharia na execução do estudo de análise de consistência pluviométrica.

**Tabela 3** – Resumo dos principais critérios utilizados no estudo de análise de consistência pluviométrica.

|  |  |
| --- | --- |
| **Análise Preliminar** | * Identificação de erros grosseiros, pela simples observação visual e gráfica, na escala diária, para cada uma das séries selecionadas. * Foram considerados valores incompatíveis (negativos ou muito elevados) ou, eventualmente, distorções significativas observadas quando da comparação com as demais séries; * A ausência de leituras dos pluviômetros em finais de semana e feriados, que acarretam o acúmulo da precipitação na leitura do dia útil seguinte, ou ainda devido à ausência do operador a estimativa por parte deste de valores para determinados dias da amostragem. |
| **Identificação de Grupos Homogêneos** | Identificação de similaridade regional entre as estações, a partir da formação dos denominados “Grupos Homogêneos”, obtidos a partir da análise de componentes principais e conveniência geográfica (critério da distância) do conjunto de estações pluviométricas selecionadas.  Para a identificação das regiões homogêneas foi levando em conta tanto a estatísticas locais e características locais:   * A primeira, consistindo de uma delimitação preliminar, baseada na análise de componentes principais do conjunto de estações pluviométricas selecionadas para o Projeto, identificando as estações longas designadas “principais” dos grupos. * A segunda, consistindo no agrupamento das estações restantes às estações “principais” pelo critério de proximidade.   Para formação dos Grupos Homogêneos, partiu-se da premissa de alocar obrigatoriamente uma série longa dentro de cada grupo, garantindo assim a formação da base necessária para o processo de extensão das demais séries do grupo, requisito imposto pela ANA.   * Selecionou-se “p” séries longas a partir do período definido pela ANA, para cada Região Hidrográfica, para preenchimento e extensão das séries pluviométricas. * Para o conjunto de “p” séries consideradas longas, para cada Região Hidrográfica, aplicaram-se a técnica de Análise de Componentes Principais de modo a identificar “k” séries (k < p) com poder de explicar, no mínimo, 95% da variância total das precipitações. * Desta forma “(p-k)” estações de séries longas são descartadas, e as “k” séries que restam constituem cada uma a estação dita “principal” de cada Grupo Homogêneo que será formado na Região Hidrográfica. * Em cada Região Hidrográfica, para formação dos “k” grupos de estações pluviométricas homogêneas, realizou-se o agrupamento das demais estações (não longas) e das estações descartadas pelo processo de análise de componentes principais, pelo critério da menor proximidade de cada uma delas em relação às “k” estações (longas) selecionadas (distância geográfica). Assim, cada estação não-longa foi direcionada ao grupo da estação longa mais próxima. * Para realizar o agrupamento buscou-se formar “w” grupos homogêneos, onde “w” representasse cerca de 10% do número de estações da base de dados para a respectiva Região Hidrográfica. Para tal, foram selecionadas aproximadamente 20% do total de séries existente na base de cada Região Hidrográfica para série longa. Para se chegar a esta quantidade variou-se o parâmetro denominado PEL, que está relacionado a quantidade de falhas admitidas, no período selecionado, para que a estação fosse considerada longa. * Ressalta-se que os critérios acima foram definidos empiricamente, com base na sensibilidade adquirida com a manipulação da base de dados. |

|  |  |
| --- | --- |
| *... continuação da Tabela 3* | |
| **Verificação de Inconsistências e Correções**  **(Curva Duplo Acumulativa)** | A partir destes grupos foram construídas curvas duplo-acumulativas (CDA) para a avaliação e correções de eventuais inconsistências.  Esta análise foi aplicada a todos os dados disponíveis das séries pluviométricas selecionadas para o Projeto no período de janeiro de 1931 a dezembro de 2007. Com base nas seguintes considerações:   * A escala adotada para consistência foi mensal; * Os subgrupos de estações homogêneas, sobre as quais é aplicada a CDA, são extraídos de grupos maiores de estações homogêneas, obtidos da Análise de Componentes Principais aplicada sobre todas as séries pluviométricas consideradas em cada Região Hidrográfica; * Os subgrupos de estações homogêneas, sobre as quais é aplicada a CDA, foram formados de maneira empírica buscando sempre cobrir o maior período de observações de cada estação do grupo homogêneo, com o menor número de subgrupos; * Cada estação da Região Hidrográfica considerada pertence a apenas um grupo homogêneo, porém pode participar de mais de um subgrupo homogêneo; * Uma sequência exagerada de valores nulos, quando comparados com as demais observações correspondentes das estações do subgrupo, é substituída por falhas para posterior consistência.   O processo de análise da CDA segue basicamente os procedimentos descritos a seguir:   * Observa-se o aspecto da curva CDA gerada e o correspondente coeficiente de determinação; * Caso a curva CDA sugira alguma alteração, função da existência de tendências distintas na relação correspondente, verifica-se complementarmente a regularidade dos dados observados na própria estação, através da acumulação dos próprios totais mensais contra o tempo. * Caso pelas duas avaliações confirme-se a possibilidade de alteração das precipitações na estação em análise em determinado período, conforme as premissas apresentadas, a mesma é efetuada sendo proporcional a relação entre o coeficiente angular da reta representativa da tendência a ser mantida na CDA e o coeficiente angular da reta representativa da tendência a ser alterada na CDA. |
| **Preenchimento e extensão das séries pluviométricas** | A base de dados utilizada para preenchimento e extensão das séries refere-se aos dados mensais consistidos das 2.549 estações selecionadas. As séries resultantes da análise de consistência foram complementadas nos períodos não consistidos (sem períodos comuns com outras estações do mesmo subgrupo) com os dados brutos do banco de dados Hidro.   * Para o preenchimento e extensão das séries de dados pluviométricos, bem como para a consistência de determinado valor, foram adotados três modelos, características comuns relativamente à consideração da sazonalidade das precipitações e ao ajuste com base nos períodos comuns de dados das mesmas estações, todas pertencentes ao mesmo grupo homogêneo, baseados em estatísticas das séries envolvidas, quais sejam: * Modelo Multivariado Sazonal Mensal Auto-Regressivo de Ordem 1 \* SMMAR(1) * Modelo da Ponderação Regional por Médias * Modelo da Ponderação Regional por Correlação * Para a extensão e preenchimento das séries e consistência de determinado valor, o programa adota prioritariamente os valores propostos pelo modelo de Ponderação Regional por Correlação, por ter sido este o modelo que melhor reproduziu a série original. Na inexistência de valores propostos pelo modelo “PRC”, adota-se os valores propostos pelo modelo “SMMAR(1)” e, na impossibilidades matemáticas do SMMAR(1) gerar algum valor pontualmente, adotando o programa nestas situações o valor nulo. Aproximadamente 1,8% do total previsto de estações para o projeto e que por consequência não foram preenchidas nem estendidas nesta etapa. * O subgrupo de “N” estações a serem preenchidas/ estendidas (estação principal e N-1 estações de apoio) é selecionado a partir dos Grupos Homogêneos definidos para análise de consistência. |

|  |  |
| --- | --- |
| *... continuação da Tabela 3* | |
| **Preenchimento e extensão das séries pluviométricas** | * A partir dos Grupos Homogêneos foram formados subgrupos, teoricamente limitados a vinte estações, buscando-se garantir um número mínimo de meses em comum entre as séries observadas; * Para aumentar o período de observação comum entre algumas estações, considerando o limite imposto ao aplicativo computacional desenvolvido, foram eventualmente utilizadas estações longas, externas ao Grupo Homogêneo considerado, porém com a maior proximidade possível a este; * Os modelos utilizados (SMMAR, Ponderação Regional por Médias e Ponderação Regional por Correlação) são ajustados sobre as séries consistidas pelo critério da CDM; * O modelo “SMMAR(1)” é validado pelos modelos de ponderação regional, mês a mês, através de diferenças absolutas e relativas entre os correspondentes valores gerados; * Além do preenchimento e extensão das séries, nesta etapa também é realizada a consistência de dados considerados fortemente suspeitos; * A identificação de um dado suspeito é realizada com base em diferenças absolutas e relativas entre os valores propostos pelos modelos e o correspondente valor registrado na série “SC\_1” em análise; Um valor mensal da série “SC\_1” é dito suspeito e será alterado, se as diferenças absolutas deste com os correspondentes valores propostos pelos três modelos superarem os valores pré-estabelecidos. * Para a alteração de um dado suspeito ou para o preenchimento e a extensão da série em análise é adotado prioritariamente os valores propostos pelo modelo “PRC”, padronizando desta forma o método de geração de dados para a série em análise; No caso de inexistência de valor proposto pela série “PRC” e havendo necessidade de proposição de algum valor, adota-se o valor fornecido pelo modelo “SMMAR(1)”; |
| **Índices de qualificação das séries** | Foram estabelecidos índices que permitem qualificar, por nota, as séries pluviométricas analisadas. No estudo em questão foi estabelecido um *Índice Q*, para descrever a qualidade geral das séries pluviométricas, sendo calculado através da média ponderada entre os *Índices Q-1 (r*elacionado à qualidade da série bruta do banco de dados da ANA) e *QSCP (*relacionado à qualidade da série consistida, preenchida e estendida pelo estudo).  O *Índice Q-1* é calculado através da média ponderada de quatro subíndices relacionados a aspectos da série pluviométrica analisada:   * Comprimento da série pluviométrica; * Número de falhas da série pluviométrica; * Coeficiente de determinação (R2) das curvas duplo-acumuladas das respectivas séries de dados brutos; * Número de dados modificados através da curva duplo-acumulada.   O *Índice Q-SCP* é calculado através da combinação de 4 subíndices, os quais estão descritos a seguir.   * *Índice C-SM* – relaciona o número de dados consistidos através do modelo SMMAR com o grau de ajuste deste modelo; * *Índice C-PRC* - que relaciona o número de dados consistidos através do modelo PRC com o grau de ajuste deste modelo; * *Índice P/E-SM* – relaciona o número de dados preenchidos/estendidos através do modelo SMMAR com o grau de ajuste deste modelo; * *Índice P/E-PRC* - relaciona o número de dados preenchidos/estendidos através do modelo PRC com o grau de ajuste deste modelo.   Após o cálculo da nota representada pelo Índice Q, atribui-se conceito para cada estação pluviométrica, quais sejam: ótimo, bom, razoável e ruim. |

# CONSIDERAÇÕES FINAIS

Este documento teve como o objetivo fazer um levantamento das técnicas e métodos de consistência de dados pluviométricos, visando dar subsídio para a elaboração de um manual de sistematização de processos e metodologias de consistência de dados pluviométricos que possa vir a ser adotado pela ANA.

Uma das constatações das análises de estudos anteriores é que existe uma variação muito grande nas técnicas utilizadas e nos procedimentos metodológicos adotados para a análise de consistência, correção e preenchimento de falhas nas séries de dados pluviométricos, mesmo em diferentes estudos coordenados por uma mesma entidade.

Além disso, verificou-se que a grande maioria das metodologias utilizadas são baseadas na adoção de critérios subjetivos que, apesar de bem documentados, são difíceis de serem replicados de forma sistemática em novos estudos. Isto ocorre mesmo em estudos que tiveram como premissa a sistematização e automação dos procedimentos aplicados, como é o caso do estudo realizado pela RHA Engenharia para a ANA (RHA, 2011), onde em várias etapas do estudo houve a necessidade de se adotar critérios baseados em análise empírica e experiência dos analistas.

Verificou-se ainda que, apesar da hidrometria estar passando por um acelerado processo de modernização, os estudos realizados no Brasil ainda não têm dado ênfase aos resultados destes novos tipos de tecnologias no que se refere a análise de consistência e qualificação de dados hidrometeorológicos. Em particular pode-se citar os dados coletados por Plataformas Automático de Coleta de Dados – PCD, equipamentos que usam sensores automáticos para coletar dados de nível e precipitação (entre outras variáveis) e armazenam as informações em meio digital (*dataloggers*) e, em alguns casos, transmitem os dados em tempo quase-real (telemetria). Esta forma de coletar e registrar dados hidrometeorológicos possui inúmeras vantagens em relação aos métodos convencionais (manual), conforme alertado pela WMO (2008) também traz novas dificuldades e desafios que deverão ser adequadamente abordados.

Com base na problemática acima e visando a elaboração de um manual, e ferramentas computacionais, de sistematização de processos e metodologias de consistência de dados pluviométricos que possa vir a ser adotado pela ANA, sugere-se que, como continuidade deste trabalho, seja realizado um estudo de comparação de metodologias de análise de consistência, correção e preenchimento de falhas e sua implantação em ferramentas automatizadas.

# REFERÊNCIAS BIBLIOGRÁFICAS

Ahmad, N.H.; Othman, I.R.; Deni, S.M. Hierachical Cluster Approach for Regionalization of Peninsular Malaysia based on the Precipitation Amount. Journal of Physics: Conference Series 423 (2013) 012018.

ANA. Agencia Nacional de Águas. Diretrizes e Análises Recomendadas para a Consistência de Dados Pluviométricos. Brasília, 2012, 18p.

Baeriswyl, P. A.; Rebetez, M. - Regionalization of Precipitation in Switzerland by Means of Principal Component Analysis. Theor. Appl. Climatol. 58, 31-41 (1997).

Bertoni, J.C. & Tucci, C.E.M. Precipitação. In.: Hidrologia: Ciência e Aplicação, Org. Carlos E. M. Tucci, 4ª ed., 5. Reimpr., Porto Alegre: Ed. Universidade/UFRGS: ABRH, 2013. 943 p.

Billib, M. H. A. & Freitas, M. A. S.: 1996: Drought Forecasting and Management for Northeast-Brazil by Statistics, Neuro-fuzzy Systems Analysis and Stochastic Simulation. In: Conference on Water Resources & Environment Research: towards the 21st Century, 1996, Kyoto. Symposium Sustainability of Water Resources Under Increasing Uncertainty. Kyoto: IAHS Publication.

Braga, C.B.; Macedo, M.J.H.; Silva, B.B; Braga, R.C. - Aplicação dos Componentes Principais na Simulação e Consistência de Séries Temporais. Revista de Geografia (UFPE), V. 29, No. 1, 2012. P. 113-125.

Braga, C.C. & Targino, A.C.L. 1996. Simulação e Consistência de Séries Temporais de Precipitação para o Estado de Pernambuco. IX Congresso Brasileiro de Meteorologia, Campos do Jordão, SP, Anais, vol. 1, 312-315.

CERPCH-EFEI - Análise de Consistência de Dados Pluviométricos – Bacia do Rio Paraná/Sub-bacia-61 (Cabeceiras), Itajubá, 1999. CERPCH-EFEI. Agência Nacional de Energia Elétrica.

Conde, F.C., Ramos, A.M., Santos, L.A.R.; Ferreira, D.B. Identificação de Regiões Homogêneas de Precipitação Acumulada Mensal do Distrito Federal via Análise de Cluster. Anais do XVI CBMET, 2010, Belém-PA.

CPRM – Serviço Geológico do Brasil. Dados Pluviométricos. 2014. (em construção).

Depiné, H.; Castro, N.M.R.; Pinheiro, A.; Pedrollo, O. Preenchimento de Falhas de Dados Horários de Precipitação Utilizando Redes Neurais Artificiais. RBRH – Revista Brasileira de Recursos Hídricos Volume 19 n.1 –Jan/Mar 2014, 51-63.

DNAEE - Departamento Nacional de Águas e Energia Elétrica. Sistemática para Análise de Consistência de Dados Fluviométricos. Brasília: DNAEE, 1982. 35p.

Enerconsult-Hidrosistem-Internave – Relatório de Análise dos Dados Pluviométricos da Bacia do rio Paraíba do Sul e Ribeirão das Lajes, Contrato nº GPD-CT-185/06-2, Consórcio Enerconsult-Hidrosistem-Internave. Operador Nacional do Sistema Elétrico – ONS. 2008.

Everitt, B. 1974: Cluster Analysis. Heinemann Educational Books, London, 135p.

Freitas, M. A. S. Neurocomputação Aplicada, Ed. Gráfica UFPI, 1ª ed., Teresina, 1998, 60p.

FUB - Análise de Consistência de Dados Pluviométricos – SubBacia-65, Brasília, 1999. Fundação Universidade de Brasília - FUB. Agência Nacional de Energia Elétrica.

Ghuge, H.K.; Regulwar, D.G. Artificial Neural Network Method for Estimation of Missing Data. International Journal of Advanced Technology in Civil Engineering, ISSN: 2231 –5721, Volume-2, Issue-1, 2013.

Hair, J.F.; Tatham, R.L.; Anderson, R.E.; Black, W. Multivariate Data Analysis. Prentice-Hall, Inc. 5th Edition, 1998.

Heinz, D. F., Capítulo “Informações Hidrológicas”, in Modelos para Gerenciamento de Recursos Hídricos, ABRH, São Paulo, 1987.

HICON – Consistência de Vazões Naturais na Bacia do Rio Grande – Relatório de Análise dos Dados Pluviométricos. HICON Engenharia Ltda. Operador Nacional do Sistema Elétrico – ONS. 2003.

HICON – Estudos de Consistência e Reconstituição de Séries de Vazões Naturais nas bacias dos rios Parnaíba, Jequitinhonha e Doce, Relatório Final, vol. I, HICON Engenharia Ltda. Operador Nacional do Sistema Elétrico – ONS. 2008.

HidroPlu. Programa de Homogeneização de Dados Pluviométrico. Universidade Federal de Viçosa – UFV. Agência Nacional de Energia Elétrica – ANEEL. 2000. 38 p.

Hiez, G. L´homogenéité des données pluviométriques. Cahiers ORSTOM. Série Hydologie. Paris, 1977, v.14, n.2, p.129-172.

Hiez, G. Processamento dos dados pluviométricos do nordeste: homogeneização dos dados – métodos do vetor regional. Recife: SUDENE, 1978.

Holanda, C.V.M, Oliveira, E. Programa para Homogeneização de Dados – PROHD. In: Simpósio de Hidrologia, 3, 1979, Brasília. **Anais.** Porto Alegre: Associação Brasileira de Recursos Hídricos, p. 810-845, 1979.

Hung, N.Q., Babel, M.S., Weesakul, S., Tripathi, N.K. An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrol. Earth Syst. Sci., 13, 1413–1425, 2009.

IPH - Consistência de Dados Hidrológicos da Bacia do hidrográfica do Alto Uruguai, Porto Alegre. 1999. Instituto de Pesquisas Hidráulicas – IPH. Agência Nacional de Energia Elétrica.

Jain, A.K.; Murty, M.N.; Flynn, P.J. Data Clustering: A Review. ACM Computing Surveys, 1999. v. 31, n. 3, p. 254-323.

LACTEC – Estudos de Consistência e Reconstituição de Séries de Vazões Naturais na Bacia do Rio Paranapanema, Contrato ONS DPP Nº 040/2003 - Bacia do Rio Paranapanema, Relatório Final, LACTEC-CEHPAR. Operador Nacional do Sistema Elétrico – ONS. 2003.

Matulla, C., Penlap, E.K.; Haas, P.; Formayer, H. Comparative Analysis of Spatial and Seasonal Variability: Austrian Precipitation During the 20th Century. Int. J. Climatol. 23: 1577–1588 (2003).

Michaelides, S.C.H.R.; Pattichis, C.S.; Kleovoulou, G. Classification of Rainfall Variability by Using Artificial Neural Networks. Int. J. Climatol. 21: 1401–1414 (2001).

Mills, G. F. - Principal Component Analysis of Precipitation and Rainfall Regionalization in Spain. Theor. Appl. Climatol. 50, 169-183 (1995).

Munõz-Díaz, D.; Rodrigo, F.S. Spatio-temporal patterns of seasonal rainfall in Spain (1912–2000) using cluster and principal component analysis: comparison. Annales Geophysicae (2004) 22: 1435–1448.

Naghettini, M. & Pinto, E.J.A. Hidrologia Estatística. Belo Horizonte. CPRM, 2007.

RHA. Recursos Hídricos e Ambientais. Qualificação de Dados Hidrológicos e Reconstituição de Vazões Naturais no País. Relatório Final do Contrato N° 016/ANA/2009. Agência Nacional de Águas. 2011. 442 p.

Searcy, J. K. and Hardison, C. H.: 1960, 'Double-Mass Curves', U.S.G.S. Water Supply Paper 154 l-B, 66 p.

WARD, J.H. :1963: Hierarchical grouping to optimize an objective function. Journal of American Association, 58:236-244p.

WMO. Guide to Climatological Practices. WMO-Nº. 100. Edição 2011. 117 p.

WMO. Guide to Hidrological Practices. Volume I - Hydrology – From Measurement to Hydrological Information, WMO-Nº. 168. Edição 2008. 296 p.