



XII SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE

PROPOSTA PRELIMINAR PARA ANÁLISE DE CONSISTÊNCIA DE DADOS DE QUALIDADE DE ÁGUA

Rita de Cássia Cerqueira Condé de Piscocoya¹; Walszon Terllizzie Araújo Lopes^{1,2}; Gabriel Meldau Lemos¹; João Carlos Carvalho¹ & Maria Célia Alencar Machado da Silva¹

RESUMO – A Agência Nacional de Águas – ANA é responsável pelo gerenciamento da Rede Hidrometeorológica Nacional e a sua base de dados relacionada (Hidro). A análise de consistência dos dados fluviométricos existentes no banco de dados Hidro vem sendo realizada pela ANA, desde 2011, com o auxílio do Sistema para Análise de Dados Hidrológicos – SIADH (Lopes *et al.*, 2013). Quanto à análise de consistência de dados de qualidade de água, nenhuma metodologia havia sido implementada pela ANA até então. O objetivo deste trabalho é apresentar uma proposta de análise preliminar de consistência de dados de qualidade de água a ser implementada no Sistema para Gerenciamento de Dados Hidrológicos da ANA (Hidro). Foi selecionada metodologia de análise por critérios para validação dos dados de qualidade de água onde foram determinados intervalos de “Valores Possíveis” e de “Valores Normais” para os parâmetros de qualidade de água do Hidro. Foi realizada análise da representatividade espacial dos dados de cada parâmetro para avaliar se o intervalo do valor normal definido poderia ser implementado no sistema Hidro.

ABSTRACT – The Brazilian National Water Agency – ANA is responsible for managing the National Hydro-Meteorological Network and its related database (Hidro). The streamflow data analysis has been performed by ANA, since 2011, using the Computational System for Hydrological Data Analysis - SiADH (Lopes *et al.*, 2013). Concerning to water quality data analysis, no methodology was implemented by the ANA until then. The aim of this work is to present a proposal for a preliminary analysis of water quality data to be implemented in the ANA’s System for Hydrological Data Management (Hidro). A methodology for that analysis has been selected by criteria for validation of water quality data where “Possible Values Intervals” and “Normal Values Intervals” were determined for many Hidro’s water quality parameters. Analysis of the spatial representativeness of the data for each parameter was performed to assess if the “Normal Values Intervals” could be implemented in the Hidro System.

¹ Especialistas em Recursos Hídricos da Agência Nacional de Águas – ANA, Setor Policial Sul, Área 05, Quadra 03, Bloco L, Brasília-DF. CEP: 70.610-200. Fone: 55-61-2109-5549. Fax: 55-61-2109-5328. (rita.piscocoya@ana.gov.br; walszon@ana.gov.br; gabriel.lemos@ana.gov.br; maria.silva@ana.gov.br; joao.carvalho@ana.gov.br).

² Professor do Instituto de Ensino Superior Planalto – IESPlan/DF, Dep. Eng. Civil. SEPS AV. W5 SUL, EQ 708/907, Brasília-DF. CEP: 70.390-070. Fone: 55-61-3442-6000. Fax: 55-61-3443-2933. (walszon@yahoo.com).

Palavras-Chave – Análise preliminar de consistência; Dados de qualidade de água.

1 – INTRODUÇÃO

A Agência Nacional de Águas – ANA é responsável atualmente por, aproximadamente, 1.600 estações de qualidade de água no território brasileiro, constituindo parte da Rede Hidrometeorológica Nacional, e vem desenvolvendo esforços para agregar a qualidade da água à sua rede quantitativa, implementando o Programa Nacional de Avaliação da Qualidade das Águas – PNQA.

Metodologias de análise de consistência de dados pluviométricos e fluviométricos são utilizadas desde o início da década de 70, pelo extinto Departamento Nacional de Águas e Energia Elétrica - DNAEE, que promoveu o desenvolvimento do primeiro sistema de banco de dados brasileiro destinado a gerenciar as informações coletadas em toda a Rede Hidrometeorológica Nacional, que deu origem ao atual Sistema para Gerenciamento de Dados Hidrológicos (Hidro 1.2).

A análise de consistência dos dados pluviométricos e fluviométricos existentes no banco de dados Hidro vem sendo realizada pela ANA, desde 2011, com o auxílio do Sistema para Análise de Dados Hidrológicos (Lopes *et al.*, 2013). Quanto à análise de consistência de dados de qualidade de água, nenhuma metodologia havia sido implementada pela ANA até então.

O objetivo deste trabalho é apresentar a proposta de análise preliminar de consistência de dados de qualidade de água que foi implementada no sistema Hidro pela ANA.

Palavras-Chave – análise preliminar de consistência, dados de qualidade de água.

2 - MATERIAIS E MÉTODOS

2.1 Levantamento bibliográfico

Foi realizado levantamento de metodologias para análise, processamento e validação de dados de qualidade de água e levantamento de métodos de estatística robusta para a identificação de dados que não se encaixam no padrão estabelecido pela maioria dos dados de qualidade de água.

2.2 Seleção de proposta para consistência preliminar dos dados de qualidade

A proposta utilizada no Projeto ODEANA (Belo, 2006; Vieira, 2006; Vieira, 2008), que realiza uma série de análises por critérios, foi selecionada, com adequações da nomenclatura. A proposta para implementação no Sistema Hidro é apresentada na **Tabela 1**.

Tabela 1. Proposta de critérios para certificação/validação dos dados de qualidade de água.

Critérios	Descrição
Critério 1 – Valor Impossível	Valor físico, químico ou biológico inaceitável por ser absurdo. Intervalo: definido por dois extremos.
Critério 2 – Valor Anormal	Valor possível de muito baixa probabilidade de ocorrência que indicaria uma alteração extraordinária do sistema natural ou do equipamento de monitoramento. A probabilidade é estabelecida através de métodos estatísticos baseados em séries históricas de medições no sistema ou em sistemas análogos. Intervalo: definido por dois extremos.
Critério 3 – Valor Limite Crítico	Valor limite imposto pela Norma Legal vigente. Foram utilizados os valores padrão (valor limite adotado como requisito normativo de um parâmetro de qualidade de água ou efluente) da Resolução do CONAMA nº 357/2005. Intervalo: definido por valor extremo.
Critério 4 – Valor Limite Operacional	Valor estabelecido a partir da norma legal vigente e que permite, em tempo útil, a adoção de medidas de gestão adequadas de modo a evitar que o limite crítico seja ultrapassado. Intervalo: definido por dois extremos.
Critério 5 – Gradiente Anormal	Variação anormal do valor do parâmetro. Valor: definido por valor.

Neste trabalho foram abordados os dois primeiros critérios, que foram implementados no sistema Hidro.

2.2.1 Determinação do intervalo do valor possível

Para determinação dos limites do valor possível (valores mínimo e máximo) foi realizada uma primeira consulta no *Standard Methods for the Examination of Water and Wastewater* (Clesceri, 1998), que trás informações sobre os métodos utilizados para determinação dos parâmetros de qualidade de água, a fim de levantar os valores máximos e mínimos obtidos pelos métodos de análise dos parâmetros. Não sendo obtidos os valores dos limites de detecção dos métodos, foi realizada consulta ao *Handbook of Hydrology* (Maidment, 1992) que possui tabelas dos intervalos e concentrações típicos dos parâmetros de qualidade de água de rios e águas subterrâneas. Por último, foi realizada consulta às tabelas de solubilidade do *Perry's Chemical Engineers' Handbook* que contém tabelas de solubilidade dos compostos inorgânicos e orgânicos em água a várias temperaturas. Esta última consulta foi descartada para a maioria dos parâmetros tendo em vista os altos valores de saturação dos parâmetros em água pura, situação que não ocorre na natureza.

2.2.2 Determinação do intervalo do valor normal

Para determinação do intervalo do valor normal (valores mínimo e máximo) foi realizado tratamento estatístico dos dados de cada parâmetro de qualidade de água presente no Hidro. Foram elaboradas tabelas de distribuição de frequência e respectivos gráficos para melhor visualização da distribuição dos valores. Em seguida, foi determinado intervalo do valor normal, no qual se encontra a maioria dos dados de cada parâmetro.

Foi constatada a necessidade de obtenção de uma metodologia que pudesse ser aplicada de forma automática no Hidro para determinação do intervalo do valor normal. A primeira opção levantada foi ajustar os limites do intervalo do valor normal de forma a se ter, fora do intervalo, apenas 5% dos valores. Tendo em vista, dependendo do número da amostra, o número grande de valores que teriam de ser verificados (numa amostra de 100 dados cerca de cinco, correspondendo ao percentual de 5%), optou-se por considerar o intervalo excluindo apenas 1% dos valores. O ajuste foi feito em planilha eletrônica, não sendo possível encontrar para todos os parâmetros um intervalo em que exatamente 1% dos dados ficasse de fora. Desta forma, optou-se por utilizar os percentis 1% e 99%, e depois incorporando à análise o valor de 0,5% (0,5% e 99,5%) a fim de compará-los. Foram calculados os valores dos percentis, por bacia, para toda a base de dados disponíveis na ANA e para as sub-bacias 58, 64, 62 e 65 (escolhidas por serem as sub-bacias com a maior quantidade de dados de qualidade de água no Hidro). Com a grande quantidade de parâmetros a serem analisados, a análise dos resultados foi focada nos parâmetros principais medidos na rede de monitoramento da ANA (temperatura do ar, temperatura da amostra, pH, turbidez, condutividade elétrica e oxigênio dissolvido - OD).

2.2.2.1 Uso de estatística robusta para definição do valor anormal

A definição do valor anormal, ou seja, do valor possível de muito baixa probabilidade de ocorrência que indicaria uma alteração extraordinária do sistema natural ou do equipamento de monitorização, enquadra-se perfeitamente na definição estatística de *outliers* apresentada por Naghettini e Pinto (2007): “ponto amostral considerado atípico que se desvia significativamente do conjunto de outros pontos”. Considerando então a semelhança dos conceitos, a diversidade de comportamento encontrada nos 138 parâmetros de qualidade de água possíveis de serem armazenados no Hidro - o que dificulta a utilização automática de um único método estatístico para definição dos *outliers* de cada parâmetro – e a necessidade de obtenção de valores automaticamente no sistema, foi feita revisão na literatura a fim de levantar métodos e ferramentas estatísticos para

identificação de *outliers*, estabelecendo limites do valor normal que pudessem ser utilizados no Hidro em todos os parâmetros de qualidade de água. Foram levantados os seguintes métodos estatísticos para identificação de *outliers*: Chauvenet, Dixon-Thompson, Rosner, Log-Pearson Tipo III, Pearson Tipo III, Grubbs, Grubbs e Beck, Hampel, Median Absolute Deviation e Intervalo inter quartis.

Dos métodos de detecção de *outliers* levantados, foram testados nos dados do Hidro, métodos com as seguintes características: que não dependem de algum tipo de distribuição, que não necessitam de conhecimento prévio do número de *outliers* existentes, que fossem aplicados de uma única vez a todos os *outliers*, que não dependessem do tamanho da amostra e de tabela para determinação dos valores críticos a serem comparados e que utilizassem procedimentos resistentes aos *outliers*, ou seja, que utilizassem os conceitos da estatística robusta. A presença de pontos atípicos em uma dada amostra pode afetar drasticamente o ajuste de certa distribuição de probabilidades dos dados, por isso optou-se por utilizar os conceitos da estatística robusta, utilizando, por exemplo, a mediana, que praticamente não sofre influência de *outliers*.

A fim de determinar uma metodologia para replicar em todos os parâmetros de qualidade de água no Hidro, os percentis de 1%, o intervalo inter-quartis, o teste de Hampel, o método de Grubbs e o método MAD (*Median Absolute Deviation*) foram aplicados a todos os dados armazenados no Hidro dos parâmetros principais medidos na rede de monitoramento da ANA, quais sejam: temperatura do ar, temperatura da amostra, pH, turbidez, condutividade elétrica e OD. A partir dos *outliers* (limite inferior e superior), foi estabelecido um intervalo de valor normal para cada parâmetro.

2.3 Representatividade espacial dos parâmetros de qualidade

Após a determinação do intervalo do valor normal dos parâmetros, foi realizada análise da representatividade espacial dos dados de cada parâmetro com a finalidade de determinar se o intervalo do valor normal poderá ser implementado ou não no Hidro, e se necessita ou não de revisão futura. Para esta análise, foi criado um índice de representatividade espacial que avalia a existência de medições de cada parâmetro nas oito bacias, 78 sub-bacias e 3228 estações do Hidro. São calculados valores da representatividade nas bacias, sub-bacias e estações, da seguinte forma:

$$\text{Representatividade nas Bacias (RB)} = \frac{\text{N}^\circ \text{ de Bacias com Medição}}{(\text{N}^\circ \text{ Total de Bacias}=8)} \quad (1)$$

$$\text{Representatividade nas Sub – bacias (RS)} = \frac{\text{N}^\circ \text{ de SubBacias com Medição}}{(\text{N}^\circ \text{ Total de Sub–Bacias}=78)} \quad (2)$$

$$\text{Representatividade nas Estações (RE)} = \frac{\text{N}^\circ \text{ de Estações com Medição}}{(\text{N}^\circ \text{ Total de Estações}=3228)} \quad (3)$$

Os valores individuais da representatividade variam de 0 a 1, recebendo 0 quando não há medição do parâmetro na bacia ou na sub-bacia ou na estação; e 1 quando existe medição em todas as bacias ou sub-bacias ou estações (Tabela 2).

Tabela 2. Variação possível dos valores de representatividade espacial nas bacias, sub-bacias e estações.

Representatividade Espacial	Valor	Observação
Nas Bacias	0 a 1	Valor = 0: Nenhuma bacia possui valores de medições do parâmetro Valor = 1: Existem medições do parâmetro em todas a 8 bacias
Nas Sub-bacias	0 a 1	Valor = 0: Nenhuma subbacia possui valores de medições do parâmetro Valor = 1: Existem medições do parâmetro em todas a 78 sub-bacias
Nas Estações	0 a 1	Valor = 0: Nenhuma estação possui valores de medições do parâmetro Valor = 1: Existem medições do parâmetro em todas a 3228 estações

O Índice de Representatividade Espacial é formado pelo somatório da análise da representatividade espacial nas bacias, sub-bacias e estações e é calculado pela seguinte equação:

$$\text{Índice de Representatividade Espacial (IRE)} = RB + RS + RE \quad (4)$$

O Índice varia de 0 a 3, recebendo 0 quando não há dado de medição do parâmetro em qualquer bacia, sub-bacia e estação; e 3 quando todas as bacias, sub-bacias e estações tem dado de medição do parâmetro. A partir do cálculo do Índice de Representatividade Espacial foi feita uma classificação dos parâmetros: Baixa representatividade espacial: valor menor que 1; Média: valor entre um e dois; Alta: maior que dois. O intervalo do Valor Normal foi implementado apenas para os parâmetros com Índice de Representatividade Espacial médio e alto.

3. RESULTADOS E DISCUSSÃO

3.1 Determinação e implementação do intervalo do valor possível

Por meio da pesquisa bibliográfica foram obtidos os limites do valor possível de 69 dos 138 parâmetros de qualidade de água passíveis de serem armazenados no Hidro, o que corresponde a 50% dos parâmetros (Tabela 3). O intervalo destes 69 parâmetros foram implementados no Hidro de forma a identificar os valores fora dos limites estabelecidos. Quando o usuário do sistema tentar inserir um valor fora do limite estabelecido para o valor possível, uma tela aparecerá com mensagem de que o valor digitado está fora do limite do intervalo considerado possível para o parâmetro. Os valores fora do limite do Valor Possível não poderão ser inseridos no Hidro.

3.2 Determinação e implementação do intervalo do valor normal

3.2.1 Aplicação dos métodos e ferramentas a todos os dados de qualidade de água do Hidro

O resultado da aplicação dos percentis de 1%, o intervalo inter-quartis, o teste de Hampel, o método de Grubbs e o método MAD (Median Absolute Deviation) a todos os dados armazenados no Hidro dos parâmetros principais medidos na rede de monitoramento da ANA (temperatura do ar, temperatura da amostra, pH, turbidez, condutividade elétrica e OD) são apresentados na Tabela 4. Tendo em vista a grande quantidade de dados do banco de dados Hidro, não foi possível aplicar o método de Grubbs e Beck a todos os dados do Hidro para os seis parâmetros selecionados, sendo realizado recorte por estação.

A regra de decisão do método MAD é comparar a estatística de teste com um ponto de corte arbitrário, que pode ser mais conservador ou mais rigoroso, dependendo do valor crítico utilizado (2,5, 4,5 ou 5,0; sendo 2,5 considerado mais conservador, e 4,5 e 5,0 mais rigoroso). Considerando os limites mínimo e máximo que perfazem o intervalo de valor normal, o intervalo obtido pelo método é menor, e com isto, o número de *outliers* é muito grande, variando, dependendo do parâmetro, de 5,8 a 28,9 %. Este valor crítico não se mostra adequado para definição de um valor normal que represente um banco de dados nacional como o Hidro, que tem dados de todo o território brasileiro.

Analisando os resultados da estatística utilizando valores críticos iguais a 4,5 e 5,0 no método MAD, e o teste Humpel que utiliza a mesma metodologia de cálculo e utiliza como valor crítico o valor de 5,06 para um nível de significância de 95%; quanto maior o valor crítico, maior o intervalo obtido entre os valores mínimo e máximo, e menor o número de valores considerados *outliers*. Os intervalos encontrados são maiores, menos conservadores, sendo mais adequados para expressar a variabilidade existente nas diversas regiões brasileiras. Analisando a quantidade de *outliers*, observa-se que os valores obtidos variam de 0,7 a 1,1% para a temperatura do ar, de 0,1 a 1,2% para temperatura da amostra, e de 1,8 a 3,4% para pH. O número de *outliers* obtidos para os parâmetros turbidez, condutividade elétrica e OD, apesar de serem mais baixos que o método MAD

utilizando o valor crítico de 2,5, são ainda muito altos, variando de 11 a 12,3% para a turbidez, 4,7 a 27,1% para condutividade elétrica e de 4,9 a 6,6% para OD. A grande quantidade de *outliers* obtida na análise dos dados de turbidez, condutividade elétrica e OD; e os valores mínimos negativos obtidos para os parâmetros turbidez e condutividade elétrica mostram uma inadequação do uso do método MAD e do teste Humpel para estes três parâmetros.

Para os parâmetros turbidez, condutividade elétrica e OD, a quantidade de *outliers* encontrada para os métodos MAD e Hampel e a ferramenta estatística intervalo inter-quartis foi muito grande, variando de 10,1 a 19,6% para os valores de turbidez; de 4,7 a 29,0% para a condutividade e de 4,9 a 17,2% para OD.

Tabela 3. Intervalo do valor possível (IVP) de 69 dos 138 parâmetros de qualidade de água do Hidro.

Parâmetro	Unidade	IVP	Parâmetro	Unidade	IVP
Temperatura do ar	°C	-20 a 50	Cádmio	mg/l Cd	0 a 1,9
Temperatura da amostra	°C	-10 a 100	Chumbo	mg/l Pb	0 a 0,1
pH	-	0 a 14	Cobre	mg/l Cu	0 a 2
Cor	mg Pt/Co	0 a 500	Cromo hexavalente	mg/l Cr	0 a 1
Turbidez	FTU	0 a 4.000	Manganês	mg/l Mn	0 a 10
Condutividade elétrica	uS/cm a 20°C	0 a 500.000	Mercúrio	mg/l Hg	0 a 0,01
Dureza total	mg/l CaCO ₃	0 a 1.000	Níquel	mg/l Ni	0 a 47
DQO	mg/l O ₂	0 a 100	Zinco	mg/l Zn	0 a 7
DBO	mg/l O ₂	0 a 100	Índice de fenóis	mg/l C ₆ H ₅ OH	0 a 5
OD	mg/l O ₂	0 a 19	Bário	mg/l Ba	0 a 0,37
Sólidos totais	mg/l	0 a 20.000	Berílio	mg/l Be	0 a 1,9
Sólidos fixos	mg/l	0 a 20.000	Boro	mg/l B	0 a 5,1
Sólidos voláteis	mg/l	0 a 20.000	Cobalto	mg/l Co	0 a 2,34
Sólidos em suspensão totais	mg/l	0 a 20.000	Estanho	mg/l Sn	0 a 0,1
Sólidos em suspensão fixos	mg/l	0 a 20.000	Lítio	mg/l Li	0 a 0,5
Sólidos em suspensão voláteis	mg/l	0 a 20.000	Prata	mg/l Ag	0 a 0,18
Sólidos dissolvidos totais	mg/l	0 a 20.000	Selênio	mg/l Se	0 a 1,41
Sólidos dissolvidos fixos	mg/l a 180°C	0 a 20.000	Urânio total	mg/l U	0 a 0,1
Sólidos dissolvidos voláteis	mg/l	0 a 20.000	Vanádio	mg/l V	0 a 4,6
Sólidos sedimentáveis	mg/l	0 a 20.000	Óleos e graxas	mg/l	0 a 150
Detergentes	mg/l LAS	0 a 2.000	Carbono orgânico total	mg/l C	0 a 120
Alcalinidade CO ₃	mg/l	0 a 800	Cromo total	mg/l Cr	1 a 1,4
Alcalinidade HCO ₃	mg/l	0 a 400	Nitrogênio orgânico	mg/l	0 a 100
Cloretos	mg/L	0 a 1.000	Sódio total	mg/l Na	0 a 15.000
Sulfatos	mg/l SO ₄	0 a 4.000	Magnésio total	mg/l Mg	0 a 400
Sulfetos	mg/l S	0 a 20	Sílica dissolvida	mg/l SiO ₂	0 a 9,4
Fluoretos	mg/l F	0 a 25	Potássio total	mg/l K	0 a 14,15
Cianetos	mg/l CN	0 a 100	Cálcio total	mg/l Ca	0 a 500
Nitrogênio total	mg/l N	0 a 200	Ferro total	mg/l Fe	0 a 2.000
Nitrogênio amoniacal	mg/l N	0 a 1.400	Fósforo total	mg/l P	0 a 20
Nitratos	mg/l N	0 a 1.400	Bismuto total	mg/l Bi	0 a 0,1
Nitritos	mg/l N	0 a 100	Nitrogênio total KJELDAHL	mg/l N	0 a 1.500
Alumínio	mg/l Al	0 a 5	IQA		0 a 100
Arsênio	mg/l As	0 a 1,9			

Analisando os valores obtidos pelo uso dos percentis, observa-se que os valores obtidos para os valores mínimos e máximos pelo cálculo dos percentis (1% e 99%) são bastante adequados (com exceção do parâmetro condutividade) e a quantidade de dados a ser analisada (2% dos dados) é adequada para verificação dos dados. O valor máximo da condutividade elétrica definido pelo

percentis é muito alto (1.422,8 $\mu\text{S}/\text{cm}$ a 20°C), o que pode demonstrar a quantidade de dados errados existentes no banco de dados Hidro.

Os valores máximos obtidos pelo intervalo inter-quartis para os parâmetros temperatura do ar, temperatura de amostra, pH e OD são bastante próximos aos valores obtidos pelo percentis, já os valores obtidos para os parâmetros turbidez e condutividade elétrica são muito inferiores aos obtidos pelo cálculo dos percentis, o que confere uma quantidade muito grande de outliers altos, de cerca de 10% para turbidez, e 12% para condutividade elétrica. Os valores mínimos obtidos pelo intervalo inter-quartis para os parâmetros temperatura do ar e da amostra e pH são bastante próximos aos obtidos pelo percentis, entretanto, a quantidade de outliers obtida para o parâmetro pH, de cerca de 5%, é muito alta, o que, se implementado, na prática dificultará o trabalho dos alimentadores do Hidro que deverão verificar 5% dos dados de pH inseridos no sistema, além de não refletir os valores muito baixos e muito altos de pH encontrados em determinadas regiões brasileiras. Da mesma forma que os valores máximos, os valores mínimos obtidos para os parâmetros turbidez e condutividade elétrica pelo intervalo inter-quartis são bastante inferiores ao percentis, sendo valores negativos. Como não existem valores negativos para os parâmetros turbidez e condutividade elétrica, o intervalo inter-quartis não identifica *outliers* inferiores.

3.2.2 Aplicação dos percentis por bacia

Visando verificar as diferenças entre os intervalos de valores normais mínimo e máximo obtidos para todos os valores do Hidro e os obtidos para as oito bacias, foi realizada análise dos percentis por bacia. Foram determinados os valores mínimos e máximo do Hidro das oito bacias para o percentis de 1%, foi determinado se o intervalo (diferença entre valor máximo e mínimo) obtido para a bacia está inserido no intervalo obtido para o Hidro (valor normal máximo da bacia inferior ao valor normal máximo do Hidro, e valor mínimo da bacia superior ao valor normal mínimo do Hidro), e foi calculada a diferença entre o valores mínimos e máximos do Hidro e o valores mínimo e máximo de cada bacia.

A análise para os seis parâmetros avaliados mostra que não podem ser utilizados os valores normais mínimo e máximo determinados para o banco de dados Hidro (todos os dados do território nacional) para análise dos dados da bacia. Isto se deve às diferenças de comportamento dos parâmetros que está associada as variações espaciais e também a possíveis valores errados que estejam presentes no Hidro.

A Tabela 5 apresenta os intervalos de valor normal calculados com a ferramenta percentis para os parâmetros de média e alta representatividade espacial da Bacia 3 que possuem dados disponíveis no banco Hidro.

3.2.3 Aplicação dos métodos e ferramentas aos dados das estações com mais dados em cada uma das bacias

Tendo em vista as diferenças encontradas entre os parâmetros, foi realizada uma análise mais detalhada dos seis parâmetros. Foi selecionada, para cada um dos seis parâmetros analisados e para cada bacia, a estação com maior número de dados para aplicação dos métodos e ferramentas selecionados para identificação de outliers (método MAD, método de Grubbs e Beck, teste Humpel e as ferramentas estatísticas intervalo inter-quartis e percentis).

Dos métodos estatísticos aplicados para detecção de outliers e determinação dos intervalos do Valor Normal (teste de Humpel, método de Grubbs e Beck e método MAD) nenhum se mostrou adequado ao mesmo tempo para os seis parâmetros escolhidos.

Tabela 4. Resultado dos testes de outliers, percentis e intervalo inter-quartis de seis parâmetros de QA.

		Temp Ar (°C)	Temp Amostra (°C)	pH	Turbidez (FTU)	Cond. Eletrica (µS/cm a 20°C)	OD (mg/L O ₂)
PERCENTIL/ESTATÍSTICA	Média	26	23,3	7,1	46,6	208,4	6,5
	Mediana	26	23,7	7	18	62,8	6,9
	Mediana do Resíduo Absoluto	3,4	3,2	0,4	12	28,38	1,1
	Moda	28	22	7	15	50	7,2
	Desvio Padrão	37,8	11,2	19,9	203,7	3955,9	5,2
PERCENTIL	0,01	36	31,6	8,8	490,0	1422,8	10,8
	0,99	10,5	12,6	4,7	1	3,8	0,07
	0,005	37,8	32,2	9,1	776,4	2777,9	12,2
	0,995	0	11,4	4,3	0,6	0,7	0
OUTLIERS TOTAL	Nº Outliers MAD Method 2,5	8716	6202	13538	14210	27405	16155
	Nº Outliers MAD Method 4,5	935	185	3345	8795	25687	6172
	Nº Outliers MAD Method 5,0	648	105	1842	8005	25597	4643
	Nº Outliers Humpel 5,06 95%S	645	105	1820	7972	4484	4627
	% Outliers MAD Method 2,5	10,0	5,8	13,8	19,6	29,0	17,2
	% Outliers MAD Method 4,5	1,1	0,2	3,4	12,1	27,1	6,6
	% Outliers MAD Method 5,0	0,7	0,1	1,9	11,0	27,0	4,9
	% Outliers Humpel 5,06 95%S	0,7	0,1	1,9	11,0	4,7	4,9
nº Dados	86971	107194	98363	72504	94662	94148	
VALOR MÁXIMO	Valor máximo MAD Method 2,5	34,5	31,7	8,0	48,0	133,8	9,7
	Valor máximo MAD Method 4,5	41,3	38,1	8,8	72,0	190,5	11,9
	Valor máximo MAD Method 5,0	43,0	39,7	9,0	78,0	204,7	12,4
	Valor máximo Humpel 5,06 95%S	43,2	39,9	9,0	78,7	206,4	12,5
OUTLIERS SUP	Nº Outliers MAD Method 2,5 SUP	2199	996	5146	14210	32671	2751
	Nº Outliers MAD Method 4,5 SUP	55	39	844	8795	27405	564
	Nº Outliers MAD Method 5,0 SUP	40	27	551	8005	25687	434
	Nº Outliers Humpel 5,06 95%S SUP	39	27	546	7972	25597	432
	% Outliers MAD Method 2,5 SUP	2,5	0,9	5,2	19,6	34,5	2,9
	% Outliers MAD Method 4,5 SUP	0,1	0,0	0,9	12,1	29,0	0,6
	% Outliers MAD Method 5,0 SUP	0,0	0,0	0,6	11,0	27,1	0,5
	% Outliers Humpel 5,06 95%S SUP	0,0	0,0	0,6	11,0	27,0	0,5
VALOR MÍNIMO	Valor mínimo MAD Method 2,5	17,5	15,7	6,0	-12,0	-8,1	4,2
	Valor mínimo MAD Method 4,5	10,7	9,3	5,2	-36,0	-64,9	2,0
	Valor mínimo MAD Method 5,0	9,0	7,7	5,0	-42,0	-79,1	1,4
	Valor mínimo Humpel 5,06 95%S	8,8	7,5	5,0	-42,7	-80,8	1,3
OUTLIERS INF	Nº Outliers MAD Method 2,5 INF	6517	5206	8392	0	4484	13404
	Nº Outliers MAD Method 4,5 INF	880	146	2501	0	0	0
	Nº Outliers MAD Method 5,0 INF	608	78	1291	0	0	0
	Nº Outliers Humpel 5,06 95%S INF	606	78	1274	0	0	0
	% Outliers MAD Method 2,5 INF	7,5	4,9	8,5	0,0	4,7	14,2
	% Outliers MAD Method 4,5 INF	1,0	0,1	2,5	0,0	0,0	0,0
	% Outliers MAD Method 5,0 INF	0,7	0,1	1,3	0,0	0,0	0,0
	% Outliers Humpel 5,06 95%S INF	0,7	0,1	1,3	0,0	0,0	0,0

Tabela 5. Intervalo do Valor Nomal (IVN) calculados para a Bacia 3 (Atlântico,Trecho Norte/Nordeste).

Parâmetro	Unidade	IVN	Parâmetro	Unidade	IVN
Temperatura do ar	°C	23,1 a 38,0	Nitrogênio Amoniacal	mg/L N	0 a 1,24
Temperatura da amostra	°C	22 a 34	Nitratos	mg/L N	0 a 7
pH	-	4,5 a 9,01	Nitritos	mg/L N	0 a 8,46
Cor	mg Pt/Co	4,24 a 500	Cádmio	mg/L Cd	0 a 0,26
Turbidez	FTU	0 a 290	Chumbo	mg/L Pb	0 a 0,1
Condutividade elétrica	uS/cm a 20°C	1,09 a 2900	Cobre	mg/L Cu	0 a 0,06
DQO	mg/L O2	0 a 94	Manganês	mg/L Mn	0,01 a 0,32
DBO	mg/L O2	0 a 59,9	Zinco	mg/L Zn	0 a 0,34
OD	mg/L O2	0 a 10,5	Coliformes Totais	NMP/100 ml	0 a 187
Sólidos totais	mg/L	2,35 a 1594	Coliformes Fecais	NMP/100 ml	0 a 9000
Sólidos em suspensão totais	mg/L	1,2 a 536	Alcalinidade Total	mg/L	2,9 a 122
Sólidos dissolvidos totais	mg/L	0,09 a 2717	Carbono orgânico total	mg/L C	2,1 a 17,7
Cloretos	mg/L	10 a 741	Fósforo total	mg/L P	0 a 3,24
Nitrogênio total	mg/L N	0 a 12	Nitrogênio total KJELDAHL	mg/L N	0 a 3,05
Amônia não Ionizável	mg/L N	0,058 a 42,9			

Comparando os métodos e as ferramentas estatísticas aplicadas (percentis e intervalo inter-quartis), a ferramenta percentis a 1% e 99% se mostrou a mais adequada e foi escolhida para identificação dos Valores Anormais a serem revisados no momento de sua inserção no Hidro. Apesar da ferramenta percentis não ser um método estatístico de hipótese, a revisão de 2% dos dados (obtidos com a aplicação dos percentis a 1% e 99%) reúne duas funções importantes: a aplicação de frequência desejável para revisão dos dados; e revisão dos outliers identificados pelos métodos, uma vez que a quantidade de *outliers* altos e baixos identificados pelos métodos foram identificados também pela ferramenta percentis a 1% e 99%.

3.2.4 Avaliação da representatividade espacial dos dados do HIDRO

Dos 138 parâmetros de qualidade avaliados quanto a representatividade espacial, 77 parâmetros foram classificados com baixa representatividade espacial; 41 com média; e apenas 16 parâmetros com alta representatividade espacial. Como um todo, os parâmetros mais representativos espacialmente foram a temperatura da amostra e pH que ocorrem em 100% das bacias, 100% das sub-bacias e em 99% das estações.

A Tabela 6 apresenta o índice de representatividade espacial dos parâmetros com alta e média representatividade espacial.

4. CONCLUSÕES

O intervalo do Valor Possível de 67 parâmetros de qualidade foi implementado no Hidro de forma a não se permitir a entrada no Hidro de dados fora do limite do intervalo considerado possível para o parâmetro.

Para a determinação do intervalo do Valor Normal, o uso da ferramenta percentis 1% e 99% em cada bacia se mostrou o mais adequado.

O intervalo do Valor Normal foi implementado por Bacia para 61 parâmetros de média e alta representatividade espacial de forma que o usuário do sistema Hidro, ao tentar inserir um valor

fora do limite estabelecido para o Valor Normal, receberá mensagem de alerta de que o valor digitado está fora do intervalo considerado normal para o parâmetro, permitindo a verificação do dado e inclusão do valor correto no sistema Hidro.

Tabela 6 - Índice de Representatividade Espacial dos parâmetros de alta e média representatividade espacial.

Parâmetro	IRP	Parâmetro	IRP	Parâmetro	IRP
Temperatura da Amostra	3,0	Fósforo Total	1,8	Detergentes	1,4
pH	3,0	Chumbo	1,7	Potássio Total	1,4
OD	3,0	Sólidos Dissolvidos Totais	1,7	Manganês	1,3
Condutividade Elétrica	2,9	Ortofosfato Total	1,7	Níquel	1,2
Temperatura do Ar	2,9	Carbono Orgânico Total	1,6	Nitrogênio Total KJELDAHL	1,2
Turbidez	2,8	Índice de Fenóis	1,6	Cromo Total	1,2
DBO	2,3	Fosfato Total	1,6	Arsênio	1,1
Coliformes Fecais	2,3	Mercúrio	1,6	Alumínio	1,1
DQO	2,2	Sulfatos	1,6	Alcalinidade HCO ₃	1,1
Descarga Líquida	2,2	Cádmio	1,6	Sólidos em Suspensão Fixos	1,1
Nitrogênio Amoniacal	2,1	Ferro Total	1,5	Amônia Não Ionizável	1,1
Sólidos Totais	2,1	Cobre	1,5	Bário	1,1
Nitratos	2,1	Nitrogênio Total	1,5	Cianetos	1,1
Nitritos	2,1	Zinco	1,5	Lindano	1,1
Coliformes Totais	2,1	Magnésio Total	1,5	Aldrin	1,1
Sólidos em Suspensão Totais	2,1	Cálcio Total	1,5	Sólidos Fixos	1,1
Cor	2,0	Sódio Total	1,4	Heptacloro	1,1
Cloretos	1,9	Ferro Solúvel	1,4	Dureza	1,1
Alcalinidade Total	1,8	Óleos Graxas	1,4	Nitrogênio Orgânico	1,0
Dureza Total	1,8	Alcalinidade CO ₃	1,4	Oxigênio Meio Ácido	1,0
				Cromo Trivalente	1,0

5. BIBLIOGRAFIA

- BELO, O.; LOURENÇO, A.; SARMENTO, P.; MAGRIÇO, A.; PINHO, J.L.; LIMA, M.; VIEIRA, J. (2006). "AQUA – Um Sistema de Informação para Análise e Validação de Parâmetros de Qualidade da Água em Alqueva". Associação Portuguesa dos Recursos Hídricos. Disponível em: <http://www.engenhariacivil.com/sistema-de-informacao-para-analise-e-validacao-de-parametros-de-qualidade-da-agua-em-alqueva>. Acesso em: 20 mai. 2014.
- CLESCERI, L.S. (Ed.) (1998). *Standard methods for the examination of water and wastewater*. 20th ed. Washington, DC: American Public Health Association.
- NAGHETTINI, M.; PINTO, É.J.A. (2007). **Hidrologia Estatística**. CPRM, 2007. 561p.
- LOPES, W.T.A.; LEMOS, G.M.; SILVA, L.R.S.; SILVA, M.C.A.M.; PISCOYA, R.C.C.C.; GOMES, A.O.; SANTOS, A.G. (2013) "Sistema para Análise de Dados Hidrológicos - SiADH". In Anais do XX Simpósio Brasileiro de Recursos Hídricos, Bento Gonçalves, Out. 2013. Disponível em: https://www.abrh.org.br/SGCv3/UserFiles/Sumarios/385de6bc0406e5837c9279cda917d043_2051bc3789682f344b751f0509c3756a.pdf. Acesso em 6 de junho 2014.
- VIEIRA, J.M.P.; PINHO, J.L.S. (2008) "Sistema de Suporte à Decisão para a Gestão da Qualidade da Água em Alqueva". Engenharia Civil. Nº 33, 2008. Disponível em: <http://www.civil.uminho.pt/revista/n33/Artigo14%20Pag175-186.pdf>. Acesso em 20 de maio 2014.
- VIEIRA, J.M.P.; PINHO, J.L.S., LIMA, M.M.C.L. (2006). "ODEANA – Um Ambiente Hidroinformático de Suporte à Decisão na Gestão da Água numa Bacia Hidrográfica". 8º Congresso da Água, Figueira da Foz, Portugal. 2006. Disponível em: <http://www.engenhariacivil.com/odeana-um-ambiente-hidroinformatico-de-suporte-a-decisao-na-gestao-da-agua-numa-bacia-hidrografica>. Acesso em: 20 de maio de 2014.