

## IV-177 – ORGANIZAÇÃO E SISTEMATIZAÇÃO DE UM BANCO DE MONITORAMENTO DA QUALIDADE DAS ÁGUAS: ESTUDO DE CASO DA BACIA HIDROGRÁFICA DO RIO DAS VELHAS/MG

**Anna Flávia Almeida Perini<sup>(1)</sup>**

Graduando em Engenharia Ambiental - UFMG

**Giovanna Moura Calazans**

Mestre em Saneamento, Meio Ambiente e Recursos Hídricos - UFMG

**Sílvia Maria Correia**

Professora do Departamento de Engenharia Ambiental e Sanitária - UFMG

**Endereço<sup>(1)</sup>:** Universidade Federal de Minas Gerais – Escola de Engenharia - Av. Antônio Carlos, 6627 – Campus Pampulha – Belo Horizonte – Minas Gerais - CEP: 31. 270-901 - Brasil - Tel: +55 (31) 3409 - 3645 - Fax: +55 (31) 3409-1879 - e-mail: [annaflaviaperini@gmail.com](mailto:annaflaviaperini@gmail.com)

### RESUMO

Visando levantar, conhecer e divulgar as condições da qualidade das águas superficiais e propor ações de planejamento e controle de usos dos recursos hídricos, no estado de Minas Gerais, o Instituto Mineiro de Gestão das Águas (IGAM) vem monitorando cursos d'águas desde 1997, por meio do Projeto Águas de Minas. O monitoramento de diversos parâmetros de qualidade da água em diferentes pontos de monitoramento gera uma matriz complexa de dados, necessitando de aplicação de métodos estatísticos mais aprofundados para sua correta interpretação. Todavia, a preparação do banco de dados corresponde a uma etapa básica para a aplicação destes métodos, e diversos são os problemas encontrados na estruturação deste banco de dados. Por isso, este trabalho teve como objetivo apresentar a metodologia utilizada para organização e sistematização do banco de dados de monitoramento das águas superficiais da bacia hidrográfica do rio das Velhas, destacando as principais dificuldades encontradas. Neste estudo foram analisadas 46 estações de monitoramento, com dados de uma série histórica de 16 anos (1997 a 2013). O banco de dados abrange registros de 63 parâmetros físico-químicos e biológicos, com frequência de amostragem variando de mensal a semestral. A avaliação dos parâmetros mensurados, realizada em cada estação de amostragem separadamente, consistiu no cálculo de dados faltantes e de dados censurados e na análise do período e frequência de monitoramento. Foram descritas as principais dificuldades encontradas para elaborar um banco de dados consolidado e adequado para aplicação de técnicas estatísticas mais aprofundadas, como, a presença de *outliers*; elevado percentual de dados censurados e de dados faltantes; ausência de padronização da frequência de monitoramento dos parâmetros. A partir dos resultados encontrados, espera-se auxiliar outros trabalhos de tratamento estatístico de dados ambientais, especialmente os dados de monitoramento de qualidade de água, bem como orientar a otimização de redes de monitoramento novas e as existentes, evidenciando a importância de uma rede padronizada.

**PALAVRAS-CHAVE:** Banco de dados de monitoramento, Qualidade das águas superficiais, Estações de monitoramento, Bacia do rio das Velhas.

### INTRODUÇÃO

A ação antrópica, principalmente a implantação progressiva de atividades econômicas e o adensamento populacional de forma desordenada, ocasionam crescentes problemas para os recursos hídricos, provocando reflexos no regime hidrológico, na qualidade e na quantidade das águas superficiais.

Para acompanhar a evolução destes impactos no estado de Minas Gerais, o Instituto Mineiro de Gestão das Águas (IGAM) vem monitorando cursos d'águas desde 1997, por meio do Projeto Águas de Minas, visando levantar, conhecer e divulgar as condições da qualidade das águas superficiais e propor ações de planejamento e controle de usos dos recursos hídricos.

A análise desses dados, em geral, não é efetuada pelos órgãos geradores no nível de aprofundamento que seria necessário para uma quantificação mais precisa da qualidade das águas. Usualmente, os longos programas de monitoramento da qualidade da água, que resultam na determinação de um grande número de variáveis, são muitas vezes subutilizados ou mesmo de difícil interpretação. Essa situação tem sido percebida em muitos órgãos ambientais, instituições de pesquisa e empresas que lidam com grande número de dados de natureza diversa, o que requer a utilização de métodos adequados que possibilitem a sua máxima exploração. Conforme observado por Vidal (2000), na maioria dos casos, tais estudos restringem-se a simples comparação com os padrões ambientais. Vega *et al.* (1998) enfatizam a crescente utilização de análises estatísticas multivariadas e exploratórias no tratamento de dados ambientais, como uma forma de reduzir a dimensionalidade do problema e facilitar a interpretação sem perda de informações importantes. Além desses, outros métodos são comumente utilizados na fase de aprofundamento, como os testes de hipóteses e análises de séries temporais.

Todavia, a aplicação de métodos estatísticos mais aprofundados demanda uma preparação inicial do banco de dados. Nesta etapa, são encontrados problemas frequentes como a falta de sistematização dos dados, o que dificulta a disponibilização de informações importantes sobre a bacia hidrográfica monitorada. Assim, diversos programas dispendiosos de monitoramento tendem a fornecer informações superficiais sobre a qualidade e a situação ecológica dos sistemas estudados, não possibilitando a identificação de possíveis fontes e fatores que influenciam a qualidade dos sistemas aquáticos.

Portanto, o objetivo deste trabalho é apresentar a metodologia utilizada para organização e sistematização do banco de dados de monitoramento das águas superficiais da bacia hidrográfica do rio das Velhas, destacando as principais dificuldades encontradas, visando orientar futuras pesquisas relacionadas ao tratamento de dados ambientais e extrair informações confiáveis e precisas dos dados analisados.

## **MATERIAIS E MÉTODOS**

### **ÁREA DE ESTUDO**

Localizada na região central de Minas Gerais, a bacia do rio das Velhas, possuiu sua nascente principal na cachoeira das Andorinhas, município de Ouro Preto, desaguando no rio São Francisco em barra do Guaicuí, distrito de Várzea da Palma, correspondendo à Unidade de Planejamento e Gestão de Recursos Hídricos SF5, conforme divisão realizada pelo IGAM (IGAM, 2010). Ao todo, a Bacia possui uma área de drenagem correspondente a 29.173 Km<sup>2</sup>, abrangendo 51 municípios (CBH Rio das Velhas, 2014).

A bacia do rio das Velhas é dividida em três regiões: alto, médio e baixo curso. O alto curso da bacia é uma das regiões mais preocupantes quanto à qualidade das águas superficiais pois compreende toda a região do Quadrilátero Ferrífero, e engloba ainda à Região Metropolitana de Belo Horizonte (RMBH). Dessa maneira, além de apresentar o maior índice populacional da Bacia, o alto curso apresenta também os maiores focos de poluição, principalmente pelas atividades mineradoras e pelos esgotos industriais e domésticos não tratados. Diferente do alto curso, o médio e o baixo curso exibem alto predomínio das atividades agrícolas e pecuárias, além de possuírem uma menor taxa populacional.

Neste estudo foram analisadas 46 estações de monitoramento da bacia do rio das Velhas, coordenadas pelo IGAM, com dados de uma série histórica de 16 anos (1997 a 2013). A distribuição geográfica das estações de monitoramento analisadas está representada na Figura 1. O banco de dados abrange registros de 63 parâmetros físico-químicos e biológicos, com frequência de amostragem variando de mensal a semestral.

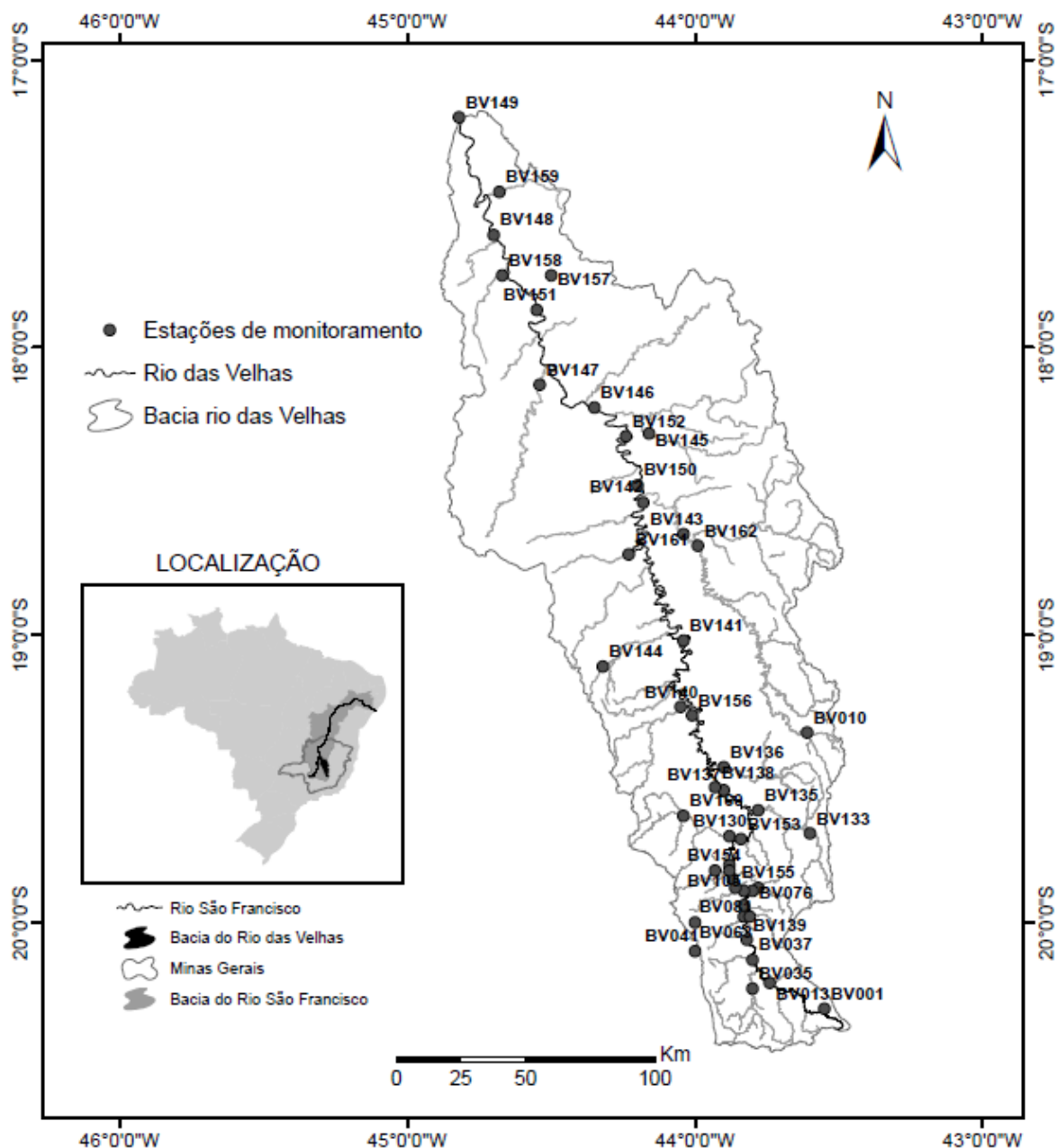


Figura 1. Localização geográfica dos pontos de monitoramento na bacia do rio das Velhas.

A avaliação dos parâmetros mensurados, realizada em cada estação de amostragem separadamente, consistiu no cálculo de dados faltantes e de dados censurados e na análise do período e frequência de monitoramento. Foram descritas as principais dificuldades encontradas para elaborar um banco de dados consolidado e adequado para aplicação de técnicas estatísticas mais aprofundadas.

## RESULTADOS

O banco de dados de monitoramento da qualidade das águas superficiais da bacia do rio das Velhas, que abrange o período de 1997 a 2013, apresenta medições de 63 parâmetros, 3.045 coletas, 46 estações de amostragem, resultando em 107.681 observações válidas. Para a organização e sistematização do banco de dados, as principais dificuldades encontradas foram: presença de *outliers*; elevado percentual de dados censurados e de dados faltantes; ausência de padronização da frequência de monitoramento dos parâmetros. Cada dificuldade será comentada detalhadamente a seguir.

## PRESENÇA DE OUTLIER

*Outliers* são observações atípicas ou infrequentes, que parecem não seguir a distribuição característica do restante dos dados. Eles podem refletir propriedades reais do fenômeno em estudo ou serem consequências de erros de medição ou outras anomalias que não podem ser modeladas. Todos os dados foram analisados segundo o método exploratório proposto por Naghettini e Pinto (2007), que considera os quartis amostrais e a amplitude interquartis, para identificação dos *outliers*. No entanto, somente em caso de inconsistência comprovada estes dados foram excluídos dos bancos de dados, já que eles podem ocorrer naturalmente em dados ambientais.

## ELEVADA PORCENTAGEM DE DADOS CENSURADOS E DE DADOS FALTANTES

*Valores censurados* são frequentes em dados de monitoramento na área ambiental e ocorrem quando as concentrações de uma amostra estão muito perto de zero, podendo ficar abaixo do limite de detecção (LD) dos métodos analíticos. Em alguns casos, os valores censurados são aqueles que ultrapassam o valor máximo do limite de detecção do método, casos observados, por exemplo, para alguns parâmetros microbiológicos, como os coliformes termotolerantes. A presença de dados censurados pode interferir no cálculo de estatísticas descritivas, levando a estimadores enviesados de médias, variâncias e de outros parâmetros da população, além de dificultar a utilização de testes de diferenças entre grupos e desenvolvimento de modelos de regressão. A prática mais comumente adotada em estudos ambientais para tratamento de dados censurados é o da substituição destes por um valor correspondente ao limite de detecção ou de uma fração deste. Gibbons e Coleman (2001) recomendam os métodos de substituição apenas quando o conjunto de dados detectados for igual ou superior a 80%, ou seja, menos 20% de dados censurados. Assim, para conjuntos de dados que apresentem um percentual elevado de observações abaixo do limite de detecção, a substituição dos dados censurados deve ser evitada. Para estes casos, existem outras alternativas que podem ser selecionadas e a escolha correta do método a ser utilizado depende tanto do grau de censura, que interfere diretamente nos resultados, quanto do tipo de aplicação (estatística descritiva, intervalos de confiança; testes de hipóteses, ajuste de distribuições de probabilidade, correlações, análises de regressão e tendências). Dependendo do método utilizado no seu tratamento, os resultados podem sofrer alterações consideráveis, tendo sua interpretação prejudicada.

Dos 107.681 dados coletados nas 46 estações, 21.36% (23.001) foram considerados dados censurados. Esta situação se torna mais abrangente quando os parâmetros são analisados separadamente. Por exemplo, os parâmetros de alumínio dissolvido, boro total, cádmio total, cobre dissolvido e cromo hexavalente apresentaram mais de 90% de dados censurados em todas as estações monitoradas.

Outra situação que se destacou na análise do banco de dados de monitoramento da bacia do rio das Velhas é que alguns parâmetros, como o mercúrio e o sulfeto, apresentaram 100% de dados censurados para todas as estações. Esta situação se justifica pela baixa sensibilidade do método analítico, que possui um alto limite de detecção. E por coincidência, esse limite de detecção também é maior que o valor máximo permitido pela Deliberação Normativa Conjunta COPAM/CERH nº 01/2008. Desse modo, não há como identificar tendências de aumento ou diminuição da concentração desses parâmetros e nem identificar as verdadeiras violações à legislação. Esse fato torna questionável a mensuração desses parâmetros na bacia do rio das Velhas.

Em geral, são considerados *dados faltantes* aqueles valores não computados quando da realização de uma coleta ou quando houver cancelamento de uma coleta em uma data inicialmente prevista. O cálculo de dados faltantes foi dificultado pela mudança de frequência de amostragem (mensal/trimestral/semestral) adotada na bacia.

Os dados faltantes também influenciam negativamente na aplicação de técnicas estatísticas. A alta porcentagem destes pode levar a conclusões não fidedignas da qualidade da água. Por esse motivo, adotou-se que parâmetros que apresentaram mais de 60% de dados faltantes seriam excluídos do banco de dados consolidado. Os dados faltantes também causam prejuízos para a aplicação de técnicas multivariadas. Quando há uma observação faltante em uma estação de amostragem, em determinada data, é necessário remover as observações de todas as estações para aquele parâmetro, naquela data de amostragem, diminuindo a fidelidade das informações geradas.

## **AUSÊNCIA DE PADRONIZAÇÃO DA FREQUÊNCIA DE MONITORAMENTO DOS PARÂMETROS**

O monitoramento dos parâmetros variou de uma estação para outra e também em função da localização de cada estação. Alguns parâmetros são monitorados semestral, trimestral ou mensalmente, mas é usual encontrar oscilações na frequência do monitoramento de um mesmo parâmetro, como citado anteriormente.

O problema das oscilações de frequências é que não se tem conhecimento dos motivos que levaram à sua ocorrência. É muito comum encontrar um parâmetro que começa a ser monitorado semestralmente passando ser trimestralmente e às vezes voltando para o semestral novamente, em um curto período. Em alguns casos podem ser percebidas estas oscilações envolvendo os três tipos de frequência, como é o caso, por exemplo, do nitrato, que na maioria das estações seu monitoramento é encontrado nos três tipos de frequência, mas em outras estações há a presença do monitoramento com apenas duas frequências em relação ao período estudado.

Esta falta de padronização na frequência acarreta inúmeros problemas para o pesquisador, principalmente no momento em que se deseja utilizar os métodos estatísticos na avaliação de comparação das estações ou dos parâmetros monitorados, que podem apresentar números muito diferentes de dados. Dentre os problemas, está a redução do número de parâmetros analisados, diminuindo também o período estudado e os números das estações. Dessa maneira, perde-se também a precisão da análise e/ou a identificação de tendências de elevação ou redução da concentração.

Nos dados avaliados percebeu-se que algumas estações tiveram um período de monitoramento menor que outras. Esse fato ocorreu devido à criação de novas estações nos anos de 2000, 2005 e 2008, período em que já havia o monitoramento de outras estações na região abordada. Foi o que aconteceu com as estações BV001, BV010, BV041, BV080, BV081, BV085, BV133, BV136, BV138, BV144, BV145, BV150, BV151, BV157, BV158, BV159, BV160, BV161 e BV162, criadas posteriormente.

Além das estações, alguns parâmetros possuem períodos de monitoramentos diferentes, como no caso da Amônia que deixou de ser monitorada em outubro de 2010, assim como o fluoreto ionizado e o pH de laboratório, que tiveram o monitoramento encerrado nos anos de 2004 e 2007, respectivamente. Diferente dos outros parâmetros, assim que o monitoramento do cianeto total se encerrou, iniciou-se o monitoramento do cianeto livre, em julho de 2009. Este último foi justificado pela mudança nas diretrizes da legislação ambiental pertinente.

A padronização das frequências de monitoramento dos parâmetros e dos métodos analíticos também é uma preocupação nacional, visto que esse problema impede a comparação da qualidade da água entre os cursos d'água do próprio estado e entre estados da federação (ANA, 2014).

## **CONCLUSÕES/RECOMENDAÇÕES**

Este trabalho identificou que os principais problemas encontrados na estruturação do banco de dados de monitoramento da qualidade das águas superficiais da bacia do rio das Velhas estão relacionados à alta porcentagem de dados faltantes e de dados censurados e à não padronização da frequência e período de amostragem entre os parâmetros e entre as estações de monitoramento.

A partir dos resultados encontrados, espera-se auxiliar outros trabalhos de tratamento estatístico de dados ambientais, especialmente os dados de monitoramento de qualidade de água, bem como orientar a otimização de redes de monitoramento novas e as existentes, evidenciando a importância de uma rede padronizada.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

1. AGÊNCIA NACIONAL DE ÁGUAS - ANA. *O que significa monitorar a qualidade das águas?* Disponível em: <[http://pnqa.ana.gov.br/rede/rede\\_monitoramento.aspx](http://pnqa.ana.gov.br/rede/rede_monitoramento.aspx)>. Acesso em 27 de agosto de 2014.
2. COMITÊ DA BACIA HIDROGRÁFICA DO RIO DAS VELHAS – CBH. Disponível em: <<http://www.cbhvelhas.org.br/>>. Belo Horizonte, Minas Gerais, Brasil. Acesso em 26 de agosto de 2014.

3. GIBBONS, R.D. and COLEMAN, D.E. Statistical methods for detection and qualification of environmental contamination. New York, Wiley, 2001.
4. INSTITUTO MINERO DE GESTÃO DAS ÁGUAS - IGAM. Monitoramento da qualidade das águas superficiais da bacia do rio São Francisco e seus afluentes em 2009. Belo Horizonte, Brazil, 2010.
5. MINAS GERAIS. Deliberação normativa n. 01 maio de 2008. Belo Horizonte: Conselho de Política Ambiental de Minas Gerais/Conselho Estadual de Recursos Hídricos – COPAM/CERH, 2008.
6. NAGHETTINI, M.; PINTO, E. J. DE A. Hidrologia estatística. Belo Horizonte MG, CRPM, 2007. 552P.
7. ROCHA, L. Disposição de resíduos sólidos numa voçoroca e seus impactos sobre as águas: Um estudo de caso em Uberlândia/MG. 2006. 169 f. Dissertação de Mestrado. Universidade Federal de Uberlândia, Uberlândia, 2006.
8. VEGA, M.; PARDO, R.; BARRADO, E.; DEBAN, L., (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research V. 32, N.12, P. 3581-3592.
9. VIDAL, M., LÓPEZ A. SANTOALLA, M.C, et al., 2000. Factor analysis for the study of water resources contamination due to the use of livestock slurries as fertilizer. Agricultural Water Management, v. 45, pp. 1-15.