

Aplicação de análise multivariada no estudo da frequência de amostragem e do número de estações de monitoramento de qualidade da água

Multivariate analysis applied for study of the sampling frequency and the number of sampling stations in water quality monitoring

Kelly Prado Maia¹ , Gilmare Antônia da Silva² , Marcelo Libânio^{1*} 

RESUMO

O cerne do presente trabalho consistiu em aplicar ferramentas de análise exploratória multivariada objetivando avaliar o número de estações de monitoramento de qualidade da água e a frequência de amostragem. Para tal, utilizou-se banco de dados disponibilizado pelo Instituto Mineiro de Gestão das Águas (IGAM) referente à Bacia do Rio das Velhas, na região central mais populosa de Minas Gerais. Foram utilizadas as técnicas de análise das componentes principais (ACP) e a rede neural de Kohonen, que culminaram na significativa redução da frequência de amostragem, em alguns casos de mensal para anual ou semestral, e na redução do número de estações de monitoramento de 36 para 33. Os resultados permitem abrir a possibilidade do emprego dos métodos utilizados como ferramentas de gestão de recursos hídricos de bacias hidrográficas visando à otimização dos programas de monitoramento de qualidade de água.

Palavras-chave: monitoramento de qualidade da água; análise multivariada; análise de componentes principais; rede neural de Kohonen.

ABSTRACT

The core of this work consisted of applying multivariate exploratory analysis tools to evaluate the number of water quality monitoring stations and the sampling frequency. In such way, the database provided by the Minas Gerais Institute of Water Management (IGAM) on the river basin of Rio das Velhas, in the most populous central region of Minas Gerais state, was used. The Principal Components Analysis and the Kohonen neural network techniques were applied, resulting in a significant reduction in sampling frequency, in some cases from monthly to annual or semi-annual, and in the reduction of the number of monitoring stations from 36 to 33. The results open the possibility of using these methods as watershed water resources management tools aimed at optimization of water quality monitoring programs.

Keywords: water quality monitoring; multivariate analysis; principal component analysis; Kohonen neural network.

INTRODUÇÃO E RELEVÂNCIA

Monitoramento de qualidade de água

O monitoramento da qualidade da água surgiu da necessidade de assegurar a viabilidade dos diversos usos, influenciando de forma direta a gestão dos recursos hídricos e a busca por soluções de conflitos pelo uso. Nesse contexto, o monitoramento da qualidade da água pode ser definido como o processo de amostragem, de determinação de parâmetros de interesse e de armazenamento de dados contemplando as características físicas, químicas e biológicas das águas naturais. Os resultados obtidos constituem o pilar para o enquadramento dos corpos d'água em classes,

um dos instrumentos de gestão estabelecidos pela Política Nacional de Recursos Hídricos. À finalidade principal de avaliar a qualidade da água, bem como a adequação desta para a utilização requerida, soma-se o norteamento para implementação de projetos específicos voltados à identificação de determinados problemas. Dessa forma, o programa de monitoramento da qualidade das águas pode se classificar conforme o uso que se pretende dar às informações levantadas, como controle ou planejamento.

A implementação dos programas de monitoramento da qualidade da água torna-se progressivamente mais dispendiosa aos cofres públicos à medida que crescem as dimensões da bacia hidrográfica, a frequência de coleta de amostras e o número de estações de amostragem. Os recursos

¹Universidade Federal de Minas Gerais - Belo Horizonte (MG), Brasil.

²Universidade Federal de Ouro Preto - Ouro Preto (MG), Brasil.

*Autor correspondente: marcelo.libanio@gmail.com

Recebido: 20/02/2017 - Aceito: 01/08/2018 - Reg. ABES: 175743

despendidos predominantemente envolvem, como insumos mais significativos, a coleta, o armazenamento e o transporte das amostras, além da realização das análises dos parâmetros de interesse. A tais custos podem ser acrescentados a capacitação da equipe de profissionais envolvida — de campo e de escritório — e a infraestrutura de informática necessária aos adequados armazenamento e processamento dos dados.

A definição das estações de amostragem para monitoramento da qualidade da água pauta-se na premissa de implantação de uma estação por área de 1.000 km², coadunando-se com a tendência verificada nos países membros da Comunidade Econômica Europeia (IGAM, 2018). Faz-se imperioso o adequado conhecimento das reais necessidades dos usuários da bacia hidrográfica, bem como da capacidade de renovação das fontes naturais. O monitoramento da qualidade da água proporciona, em termos qualitativos, tanto o conhecimento quanto a identificação das relações causa-efeito das ações antrópicas sobre a qualidade dos corpos hídricos. Sua complexidade vincula-se à natureza dinâmica dos ecossistemas, que, por sua vez, sofrem interferência direta das atividades antrópicas na bacia hidrográfica e dos fenômenos naturais.

Índice de Qualidade da Água

Com alguma frequência, o monitoramento da qualidade da água traduz-se pela determinação de indicadores que se prestam a tornar as informações mais inteligíveis à população em geral. Entre eles, assume papel de relevo o Índice de Qualidade da Água (IQA).

A partir da primeira apresentação formal desenvolvida por Horton em 1965, Brown *et al.* (1970) desenvolveram o IQA em pesquisa financiada pela National Sanitation Foundation (NSF). Utilizou-se a técnica Delphi, com painel composto inicialmente de 142 especialistas para a definição dos parâmetros de interesse. Para cada parâmetro, foram atribuídos pesos e, posteriormente, elaboraram-se curvas que representassem a variação da qualidade da água produzida pelas várias possíveis medidas do parâmetro. Os nove parâmetros que o compõem e seus respectivos pesos foram: oxigênio dissolvido (OD) — 17%, coliformes termotolerantes — 15%, pH — 12%, demanda bioquímica de oxigênio (DBO) — 10%, fósforo total — 10%, nitrato — 10%, temperatura — 10%, turbidez — 8% e sólidos totais — 8%. A presença de agrotóxicos, metais-traço (metais pesados) e elementos tóxicos resultaria em valor nulo do IQA.

A primeira formulação do IQA foi definida como a somatória das notas pelos respectivos pesos de cada um dos nove parâmetros. Posteriormente, Landwehr e Deininger (1976) propuseram forma multiplicativa para o IQA, devido fundamentalmente à possibilidade de eventual valor muito baixo de um dos parâmetros tornar-se menos evidente pela formulação anterior. Na forma multiplicativa, os pesos tornam-se potências dos pontos obtidos para cada parâmetro de qualidade da água. Com algumas adaptações, o IQA tem sido utilizado pelos órgãos de gestão de recursos hídricos em diversos países do mundo.

A partir do ano de 1975, iniciou-se, no Brasil, a aplicação do IQA por meio da Companhia Ambiental do Estado de São Paulo (CETESB). Posteriormente, nas décadas seguintes, em diversos estados brasileiros, também se passou a adotar esse índice para avaliar a qualidade da água. Desde então, tem sido o principal indicador aplicado para esse fim no país.

Desde o ano de 1997, o responsável pelo monitoramento da qualidade das águas superficiais e subterrâneas em Minas Gerais é o Instituto Mineiro de Gestão das Águas (IGAM), por meio do Projeto Águas de Minas. Em funcionamento desde 1997, o referido programa tem disponibilizado uma diversidade de dados históricos acerca da qualidade das águas, essenciais à definição de estratégias e tomadas de decisão a estes relacionados, bem como à avaliação da efetividade do Planejamento e Gestão Integrada dos Recursos Hídricos e para as atividades executadas pelos comitês e pelas agências de bacias do Estado.

Atualmente, o IGAM adota o IQA, a Contaminação por Tóxicos (CT), o Índice de Estado Trófico e os Testes Ecotoxicológicos como indicadores para refletir a situação ambiental dos corpos hídricos nas Unidades de Planejamento de Recursos Hídricos (UPGRH) de Minas Gerais de maneira acessível aos não técnicos (IGAM, 2016).

Análise multivariada

A análise estatística multivariada constitui ramo da estatística que envolve o estudo em que as amostras e as variáveis são analisadas simultaneamente, permitindo extrair resultados que a análise univariada não é capaz de mostrar. Isso porque, no estudo univariado, é investigada uma variável por vez, não sendo levada em consideração a interação entre elas. A utilização da análise multivariada permite, entre outros, reduzir a representação dimensional de dados, arranjando-os em estrutura que facilita a visualização de todo o conjunto dos resultados, com a manutenção das informações relevantes e eliminação das irrelevantes e/ou redundantes.

O estudo desenvolvido neste trabalho pautou-se em análise exploratória multivariada, sendo utilizadas técnicas como a análise das componentes principais (ACP) e a rede neural artificial de Kohonen, além do cálculo da correlação de Pearson.

A ACP é um método da estatística multivariada que ajuda na elaboração de hipóteses gerais e é capaz de separar a informação importante da redundante. O principal objetivo de uma ACP é reduzir a dimensionalidade do conjunto de dados, preservando ao mesmo tempo o máximo da informação (RENCHER; CHRISTENSEN, 2012).

A utilização desse método permite extrair as principais contribuições ortogonais (componentes principais — CP) que explicam a maior parte da variância da matriz de dados. Ele é resultado da combinação linear das variáveis originais formando as novas componentes principais, determinadas pela variância explicada por cada uma dessas componentes. A primeira componente principal (CP1) explica a maior variância dos dados, seguida da segunda (CP2), da terceira (CP3) e assim sucessivamente. Dessa forma, a avaliação do comportamento

dos dados é obtida elaborando-se o gráfico CP1 *versus* CP2 e/ou CP3 e, posteriormente, avaliando a conjugação da CP1 com as demais, em busca de se obter a explicação dos dados.

No entanto, essas manobras dificultam a análise gráfica por se tratar de representação limitada (bi ou tridimensional, no caso da análise de duas em duas ou de três em três componentes principais, respectivamente), considerando a possibilidade de visualização de no máximo três componentes principais simultaneamente. Dessa forma, fica impossibilitada a avaliação da variabilidade dos dados em única representação gráfica para um número de componentes superior a três (BARROS NETO; SCARMINIO; BRUNS, 2006).

Para a análise exploratória, é de fundamental importância considerar que os pontos (*scores*) fornecem a composição das componentes principais em relação aos objetos (amostras), enquanto os pesos (*loadings*) apresentam essa mesma composição em relação às variáveis. Como as componentes são ortogonais, é possível examinar as relações entre os objetos por meio dos gráficos dos pontos projetados nas primeiras componentes e entre as variáveis pelos gráficos dos pesos. Esses gráficos apresentam informações úteis sobre tendências subjacentes e características da estrutura dos dados, tais como agrupamentos de objetos ou variáveis e caracterização dos *outliers* (amostras com comportamento anômalo que não se encaixam no modelo). O estudo conjunto dos pesos e pontos permite estimar a influência de cada variável em cada objeto (SENA *et al.*, 2000).

A ACP é uma ferramenta de análise estatística utilizada em diversos estudos relacionados à avaliação de monitoramentos ambientais. Essa análise, associada ao sistema neural de inferência Fuzzy, foi utilizada para a predição da qualidade da água do rio Brahmani, na Índia (SAHOO; PATRA; KHATUA, 2015). Em vertente semelhante, no sul da China, objetivando reduzir custos do monitoramento, um índice de qualidade da água modificado foi desenvolvido utilizando a ACP e análises de correlações dos parâmetros de qualidade da água avaliados nos períodos seco e chuvoso, durante os anos de 2011 e 2012 (SUN *et al.*, 2016).

Ouyang (2005) aplicou a ACP e a análise fatorial principal (AFP) para estudar a eficácia da rede de monitoramento da qualidade das águas superficiais de 22 estações de monitoramento do rio St. Johns, na Flórida, avaliando, nesse estudo, quais estações são as mais importantes para verificar as alterações da qualidade da água ao longo do ano. Os resultados mostraram que três estações de monitoramento foram identificadas como menos relevantes para explicar a variação anual do conjunto de dados e, portanto, poderiam ser tratadas como não principais.

As redes neurais artificiais são uma tecnologia de processamento de informações que se inspira no funcionamento dos neurônios humanos. Um neurônio é formado por um corpo celular, um axônio e os dendritos, ramificações arbóreas. Os axônios são as linhas de transmissão, enquanto os dendritos são zonas receptivas. O pequeno espaço entre o fim do axônio e o dendrito é conhecido como sinapse, que tem como papel fundamental a memorização da informação (ROBERTS, 2007).

As unidades de processamento das redes neurais artificiais não se diferenciam de acordo com o tipo de rede neural. Em todos os tipos, incluindo a de Kohonen, um neurônio artificial simula o comportamento do neurônio biológico. Esse neurônio artificial possui várias entradas, que correspondem às conexões sinápticas com outras unidades similares, e uma saída, cujo valor depende diretamente do somatório ponderado de todas as saídas dos outros neurônios a este conectado. Apresentadas a uma série de dados, considerados como os estímulos nervosos, as redes neurais os organizam e aprendem a reconhecer padrões nos dados de entrada. Se recebem informações sobre a magnitude e direção dos erros de análise, as redes neurais aperfeiçoam o conhecimento, melhorando as respostas seguintes (FRANCISCO *et al.*, 2008).

A rede neural de Kohonen constitui algoritmo desenvolvido em 1982 por Teuvo Kohonen, sendo considerada de fácil interpretação e aplicação, apresentando a capacidade de efetuar a organização dimensional de dados complexos em grupos conforme suas relações. Trata-se de uma técnica que somente requer os parâmetros de entrada, apresentando-se como ferramenta viável para questões em que os padrões sejam desconhecidos ou indeterminados (KOWALSKI *et al.*, 2013). Essa rede é, visualmente, um mapa auto-organizável que tem a capacidade de reduzir a dimensão de determinado grupo de dados não supervisionados. A visualização e interpretação da rede de Kohonen vêm principalmente do fato de que o usuário pode assumir que todas as amostras posicionadas em um mesmo neurônio são consideradas semelhantes entre si, de acordo com o aspecto avaliado.

Outro ponto relevante da rede de Kohonen consiste na possibilidade de formação de conjuntos de amostras com as mesmas características avaliadas, devido à localização em neurônios próximos (neurônios vizinhos). Essa facilidade de visualização da distribuição dos dados é uma de suas principais características, resultando em método mais *amigável* de interpretação dos dados. Tais suposições de interpretação são baseadas no modo de processamento dessa rede, que considera que, quanto mais distante uma amostra for alocada em relação a outra, maiores diferenças podem ser apontadas entre elas. A rede de Kohonen usualmente consiste em única camada de neurônios tipicamente arranjados em um espaço bidimensional. Outros arranjos com maiores dimensões são possíveis, entretanto menos comuns (COSTA; ANDRADE NETTO, 2007).

O fato da rede de Kohonen elucidar a relação entre as amostras em um espaço bidimensional (um plano) torna este o principal aspecto vantajoso em relação a outras rotinas de análise exploratória, como a ACP, para a qual, na maioria dos casos, é necessário trabalhar em espaços multidimensionais fornecidos pelas componentes principais. Esse processo pode resultar em prática bastante morosa, tornando difícil a análise dos dados.

A qualidade da água do Lago de Chilika, localizado na costa leste da Índia, avaliada por meio da ACP e da rede neural de Kohonen, mostrou que a salinidade era o parâmetro mais impactante na qualidade da água. Tal constatação emergiu a partir da avaliação de dez parâmetros

de qualidade da água, monitorados mensalmente em 30 pontos de amostragem, durante o período entre 1999 e 2009 (KIM *et al.*, 2016). Mustonen *et al.* (2008) realizaram estudo utilizando mapas auto-organizáveis para explorar elevado número de variáveis de qualidade da água, com o objetivo de identificar alterações anormais. O uso dessa ferramenta tornou possível identificar os responsáveis pelas tais alterações. Belkhir *et al.* (2018) utilizou mapas para simular os valores de oxigênio dissolvido no rio Yamuna, a jusante da cidade de Mathura, na Índia, e também na classificação e caracterização da qualidade de águas subterrâneas, mostrando que essa ferramenta pode ser empregada com sucesso para classificar e caracterizar a água subterrânea em termos de qualidade.

Estudos realizados no Brasil em amostras coletadas de 12 diferentes pontos de um corpo d'água, durante as estações seca e chuvosa, demonstraram a influência da matéria orgânica na distribuição, mobilidade e biodisponibilidade de alumínio, arsênio e ferro em águas naturais de uma área de mineração. A determinação da influência ou relação entre as variáveis foi verificada por meio da aplicação da rede neural de Kohonen, que também mostrou que a sazonalidade não impactou a maioria das amostras (GONTIJO *et al.*, 2016).

Por fim, uma vez que as técnicas de análise exploratória multivariada apontam relações entre as variáveis, tais interações podem ser corroboradas por meio do cálculo do coeficiente de correlação de Pearson. Tal ferramenta de análise estatística refere-se a uma medida de associação de força quanto ao grau de relação que pode haver entre duas variáveis (HAIR *et al.*, 2005). A correlação é uma forma adotada para medir tanto a direção quanto o grau da relação linear possível entre duas variáveis quantitativas. Assim sendo, a correlação de Pearson pode ser vista como um coeficiente usado para mensurar a associação linear entre variáveis (MOORE, 2007).

Nesse cenário, o objetivo desta pesquisa consistiu em investigar a perspectiva da redução do número de estações de amostragem de qualidade da água por meio de técnicas de análise exploratória multivariada. Adicionalmente, avaliou-se a possibilidade da redução da frequência de coleta de amostras para o monitoramento da qualidade da água, bem como o comportamento dos dados em relação à sazonalidade.

METODOLOGIA

Banco de dados de qualidade da água da Bacia do Rio das Velhas

Para a consecução dos objetivos, utilizaram-se dados secundários de monitoramento da qualidade da água da Bacia do Rio das Velhas (MG), disponibilizados pelo IGAM, abrangendo o período de janeiro de 2000 a dezembro de 2010. Os dados utilizados fundamentam o cálculo do IQA nos pontos de monitoramento.

A bacia estudada constitui-se uma das principais do Rio São Francisco, abrange 51 municípios mineiros, área de 29.173 km² e população da ordem

de 4,5 milhões de habitantes (24% do estado). Dividida em trechos de acordo com os cursos Alto (9,8% da área), Médio (45,0%) e Baixo (45,2%), responde por 42% do produto interno bruto (PIB) de Minas Gerais. No Alto Rio das Velhas, onde está localizada a Região Metropolitana de Belo Horizonte, prevalecem os usos dos recursos hídricos para abastecimento humano, indústria e mineração. Os trechos Médio e Baixo Rio das Velhas concentram principalmente atividades de irrigação e dessedentação de animais.

A rede básica de monitoramento da qualidade da água operada pelo IGAM dispunha de 553 estações de amostragem, distribuídas em 17 bacias hidrográficas de todo o Estado. Especificamente na Bacia do Rio das Velhas, do primeiro ano considerado neste trabalho (janeiro de 2000) a 2004, existiam 29 estações em operação. Em 2005, acrescentaram-se quatro e, em 2008, outras três, totalizando as estudadas 36 estações de amostragem apresentadas na Figura 1 (IGAM, 2016).

O período de análise escolhido neste trabalho baseou-se nas características do banco de dados. Os dados originais abarcavam monitoramentos entre 1997 e 2015. No entanto, o número de estações e a frequência de amostragem entre 1997 e 1999 mostraram-se inferiores quando comparado ao período de 2000 a 2010. Como houve modificação na sistemática da determinação do IQA a partir de 2013 — com a substituição do parâmetro Coliformes Termotolerantes por *Escherichia coli* —, a consecução do objetivo central do trabalho fiou-se nos dados referentes ao período 2000 a 2010, abrangendo, dessa forma, dados de monitoramento ao longo de 10 anos.

As estações de amostragem da bacia vinculam-se às classes de enquadramento consoante à Resolução nº 357 do Conselho Nacional do Meio Ambiente (CONAMA, 2005), com sete pontos da Classe 1, 21 pontos da Classe 2 e 8 da Classe 3, os últimos localizados na Região Metropolitana de Belo Horizonte.

Neste estudo, foram utilizados os parâmetros analisados pelo IGAM para o monitoramento da qualidade das águas do Rio das Velhas por meio do cálculo do IQA. As análises físico-químicas e bacteriológicas foram realizadas segundo o *Standard Methods for the Examination of Water and Wastewater* (APHA; AWWA; WEF, 2005). Esse monitoramento contempla diferentes frequências de coleta (trimestral, bimestral e mensal) nas distintas estações de amostragem ao longo do período abordado.

Para a avaliação dos dados, realizou-se a análise do conjunto original formado por matriz composta de 10 colunas, representadas pelos resultados de análise de cada parâmetro que compõe o IQA e o resultado obtido para o próprio IQA, e 1.834 linhas, que correspondem aos resultados de análise de cada parâmetro e o respectivo IQA. Após a análise do conjunto, foi necessário excluir todas as linhas nas quais pelo menos um dos parâmetros avaliados não possuía o devido resultado de análise. Com a exclusão das linhas sem resultados completos, uma nova matriz foi gerada, tratada como matriz final de dados originais, formada por 1.710 linhas e 10 colunas.

Por fim, foram estabelecidos os códigos que denominaram cada conjunto de dados analisados. Essa nomenclatura foi criada para subsidiar a

localização dos resultados de cada linha que forma a matriz final de dados originais durante o desenvolvimento da análise exploratória. O código foi composto da letra *S* (amostra ou *sample*), seguida do número que indica a ordem dessa amostragem no ano (primeira, segunda e assim sucessivamente) e dos dois últimos algarismos do respectivo ano, quando a

amostragem foi realizada. Cada código foi representado pelo resultado da análise dos nove parâmetros do IQA e o IQA resultante desses valores, em uma determinada estação de monitoramento da Bacia do Rio das Velhas.

A matriz de dados originais abrangeu todos os resultados de monitoramento para cada parâmetro e o IQA resultante desses valores. Esta foi a base para a aplicação da análise exploratória. Em uma planilha eletrônica, a matriz de dados originais foi mantida como referência para identificação de cada ponto de amostragem com a inclusão da data de coleta e a respectiva estação, além do código atribuído a cada linha da matriz do conjunto de dados.

Para viabilizar o processamento e a obtenção de informação dos dados, a matriz final de dados originais foi desmembrada em cada ano e esta, por sua vez, em cada mês do ano. A grande dimensão dos dados originais demandou a aplicação de exaustiva análise exploratória multivariada, visando a identificar o comportamento e as possíveis relações existentes neles.

Análise exploratória multivariada

Análise das componentes principais

Para aplicação da ACP, a matriz obtida para cada mês do ano foi avaliada no ambiente computacional utilizando o *software* Matlab, da MathWorks®. Durante a execução do programa, a matriz de dados era carregada e os dados eram pré-processados por autoescalamento. A aplicação do autoescalamento visa a minimizar a influência de uma variável dominante. Durante o processamento por meio do autoescalamento, subtraíam-se de cada elemento de uma coluna da matriz de dados o valor médio da respectiva coluna, e o resultado era dividido por seu desvio padrão.

Logo após, gerava-se o modelo inicial que indicava as variâncias relativas a cada uma das componentes principais. Ao gerar o modelo, o menor número de componentes principais que explicasse a máxima variância dos dados (acima de 99%) era selecionado e, posteriormente, geravam-se os gráficos de pontos e pesos que informavam, respectivamente, sobre os comportamentos das amostras e das variáveis.

O uso da ACP prestou-se também como indicador para a aplicação da análise de Kohonen. Pelo número de componentes principais necessário para explicar a maior variabilidade existente na matriz de dados, a aplicação da rede neural de Kohonen se tornava vantajosa caso esse número se mostrasse acima de três.

Rede neural de Kohonen

A rede neural de Kohonen foi aplicada com o auxílio de um algoritmo de rede neural artificial disponível em domínio público (<http://www.cis.hut.fi/projects/somtoolbox/>). Essa ferramenta possibilitou a execução da rede gerando como saída os mapas de Kohonen. Esses mapas conduziram ao agrupamento de dados similares e ao afastamento daqueles considerados distintos, de acordo com as variáveis investigadas.

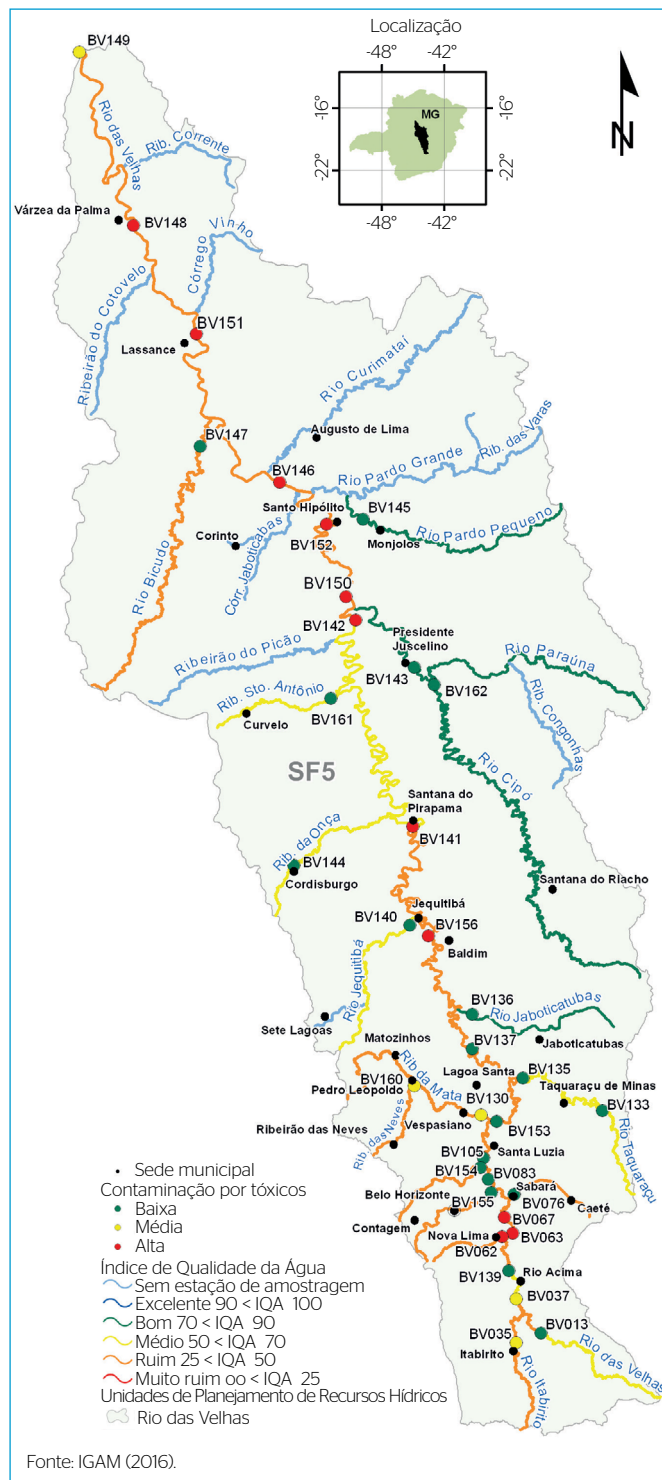


Figura 1 - Distribuição geográfica das 36 estações de amostragem para o monitoramento da qualidade da água da Bacia do Rio das Velhas.

Os mapas de Kohonen foram criados e inicializados linearmente. A rede de Kohonen foi treinada com os dados usando o algoritmo de treinamento em batelada; a função de vizinhança usada no treinamento foi a gaussiana, a estrutura foi hexagonal e a forma do mapa, planar. Para definir a arquitetura de trabalho, foram testados diferentes tamanhos de rede.

As matrizes carregadas inicialmente correspondem às utilizadas para a realização da ACP, ou seja, uma matriz para cada mês de cada ano avaliado, sendo as colunas representadas pelos parâmetros de qualidade da água e o conseqüente valor do IQA.

As matrizes mensais foram submetidas à rede de Kohonen pelo menos cinco vezes: para a arquitetura ideal, para duas arquiteturas abaixo e para duas acima dessa arquitetura, conforme exemplificado na Tabela 1 para o ano de 2010.

Dessa forma, foi escolhida a arquitetura que apresentou o melhor agrupamento entre os conjuntos de dados, a partir da qual foram selecionados novos conjuntos, um para cada mês analisado no ano. O melhor agrupamento do conjunto de dados dá-se pela localização de um ou mais conjunto de dados em um mesmo neurônio ou em neurônios vizinhos. Assim, o agrupamento é definido como a agregação de conjuntos de dados em neurônios próximos, circundados por neurônios vazios.

A seleção dos conjuntos de dados para cada mês do ano gerou uma nova matriz para cada ano, denominada matriz de dados anual. Esta foi submetida a um novo tratamento, seguindo os mesmos critérios adotados para a mensal, como indicado na Tabela 1. A melhor arquitetura resultante foi tratada com a seleção de outros conjuntos de dados, gerando a matriz de dados anual reduzida.

Essa matriz de dados anual reduzida de cada ano, obtida na última análise de Kohonen para cada ano, foi utilizada em nova etapa para verificação do comportamento dos dados nos períodos predominantemente

seco (abril a setembro) e chuvoso (outubro a março). Na Tabela 2, apresentam-se o número de conjuntos de dados para cada ano avaliado e as arquiteturas utilizadas para a análise de Kohonen.

Posteriormente, foi gerada a matriz final reduzida, composta da junção de todas as matrizes de dados anuais. Essa matriz possibilitou a realização da avaliação global de redução da frequência e do número de estações de monitoramento da qualidade da água. Para tal, os dados originais suprimidos da matriz inicial foram considerados como não relevantes, sendo a matriz final representativa da originalidade dos conjuntos de dados avaliados.

Correlação de Pearson

O cálculo dos coeficientes da correlação de Pearson foi realizado com o objetivo de avaliar a manutenção da intensidade da correlação entre as variáveis que formaram uma matriz antes e depois da realização da análise exploratória multivariada e, posteriormente, para avaliar a correlação existente entre os parâmetros que compõem o IQA e o próprio IQA.

A primeira avaliação de correlação deu-se com a totalidade dos dados que formavam a matriz original. Em seguida, a matriz de dados originais de cada ano também foi avaliada para a determinação da correlação. Esta foi importante para avaliar a manutenção da intensidade de correlação entre os dados após a realização da primeira análise de Kohonen, que gerou a matriz de dados anual. Dessa forma, avaliou-se a intensidade de correlação entre os dados de cada ano, comparando a intensidade de correlação obtida para a matriz de dados originais de cada ano e a de dados anual, também para cada ano. Não se realizou a avaliação da matriz de dados mensais, tendo a avaliação sido efetuada ano a ano, e não mês a mês.

Mantida a intensidade da correlação entre os dados da matriz original de cada ano e a de dados anual, prosseguiu-se com a aplicação da rede neural de Kohonen nesta última, gerando a versão reduzida para cada ano. A intensidade da correlação entre os dados dessa matriz foi também analisada e comparada à matriz original de cada ano e à de dados anual.

Tabela 1 - Exemplo de arquiteturas avaliadas para cada matriz mensal durante o tratamento dos dados pela rede neural de Kohonen para o ano de 2010.

| Ano | Meses | Arquitetura avaliada | | | | | |
|------|-----------|----------------------|-------|-------|-------|-------|-------|
| | | Abaixo | Ideal | Acima | | | |
| 2010 | Janeiro | 4 × 4 | 5 × 5 | 6 × 6 | 7 × 7 | 8 × 8 | -- |
| | Fevereiro | 2 × 2 | 3 × 3 | 4 × 4 | 5 × 5 | 6 × 6 | -- |
| | Março | 2 × 2 | 3 × 3 | 4 × 4 | 5 × 5 | 6 × 6 | -- |
| | Abril | 4 × 4 | 5 × 5 | 6 × 6 | 7 × 7 | 8 × 8 | 9 × 9 |
| | Maio | 2 × 2 | 3 × 3 | 4 × 4 | 5 × 5 | 6 × 6 | -- |
| | Junho | 2 × 2 | 3 × 3 | 4 × 4 | 5 × 5 | 6 × 6 | -- |
| | Julho | 4 × 4 | 5 × 5 | 6 × 6 | 7 × 7 | 8 × 8 | 9 × 9 |
| | Agosto | 2 × 2 | 3 × 3 | 4 × 4 | 5 × 5 | 6 × 6 | -- |
| | Setembro | 2 × 2 | 3 × 3 | 4 × 4 | 5 × 5 | 6 × 6 | -- |
| | Outubro | 4 × 4 | 5 × 5 | 6 × 6 | 7 × 7 | 8 × 8 | -- |
| | Novembro | 2 × 2 | 3 × 3 | 4 × 4 | 5 × 5 | 6 × 6 | -- |
| | Dezembro | 2 × 2 | 3 × 3 | 4 × 4 | 5 × 5 | 6 × 6 | -- |

Tabela 2 - Arquiteturas avaliadas para cada matriz anual reduzida durante o tratamento pela rede neural de Kohonen.

| Ano | Número de conjuntos de dados | Arquiteturas avaliadas |
|------|------------------------------|------------------------|
| 2000 | 28 | De 3 × 3 a 10 × 10 |
| 2001 | 41 | De 4 × 4 a 12 × 12 |
| 2002 | 27 | De 3 × 3 a 9 × 9 |
| 2003 | 39 | De 4 × 4 a 11 × 11 |
| 2004 | 31 | De 4 × 4 a 11 × 11 |
| 2005 | 37 | De 4 × 4 a 12 × 12 |
| 2006 | 32 | De 3 × 3 a 11 × 11 |
| 2007 | 31 | De 3 × 3 a 11 × 11 |
| 2008 | 78 | De 6 × 6 a 17 × 17 |
| 2009 | 89 | De 7 × 7 a 16 × 16 |
| 2010 | 91 | De 7 × 7 a 17 × 17 |

A última etapa da análise de correlação contemplou a comparação entre a intensidade de correlação obtida para a matriz final de dados originais e para a matriz final reduzida, gerada com a junção de todas as matrizes de dados reduzidas de cada ano. Nessa avaliação, o princípio norteador foi a manutenção da intensidade de correlação que, se verificada, indicava que a redução do conjunto de dados obtida pela análise exploratória multivariada não influenciava a representatividade dos resultados originais do monitoramento de qualidade de água na Bacia do Rio das Velhas.

RESULTADOS E DISCUSSÃO

Análise exploratória dos dados

A inspeção visual dos dados originais fez-se necessária, conforme mencionado, porque alguns conjuntos não possuíam todos os resultados, como ocorreu, por exemplo, para *Coliformes termotolerantes*. A análise exploratória dos dados com a retirada dos resultados inconsistentes resultou em um banco de dados de 1.710 amostras, pontos de coleta com 10 variáveis, *i.e.*, o IQA e os 9 parâmetros que o integram, tendo sido descartados 124 conjuntos de resultados. Além de equalizar os dados, a análise exploratória multivariada também permitiu a avaliação mais acurada da frequência de monitoramento e sua variação ao longo dos anos.

Essa avaliação tornou-se ainda mais relevante, visto que a ausência do resultado de análise para quaisquer dos parâmetros necessários ao cálculo do IQA impossibilita a determinação desse indicador. Essa análise também se pautou na manutenção de pelo menos uma campanha de amostragem para o período seco e outra para o chuvoso em cada ano do período amostrado.

Para o ciclo compreendido entre 2000 e 2004, foram avaliadas 29 estações de monitoramento, em diferentes meses do ano. Entre 2000 e 2001, o monitoramento ocorreu nos meses de janeiro, maio, julho e novembro; em 2003, em janeiro, abril, julho, setembro e outubro; e em 2004 e 2005, em janeiro, abril, julho e outubro. A partir de 2005, foram monitoradas 33 estações até 2007. A partir de 2008, foram monitoradas 36 estações e as campanhas de monitoramento passaram a ter frequência mensal.

Análise multivariada

Análise das componentes principais

Neste estudo, o número de componentes principais foi determinado considerando aquele para o qual não há uma alteração significativa da variância acumulada (valores menores que 1,0). A ACP apontou que, para todos os meses avaliados, à exceção de fevereiro de 2006 e fevereiro de 2008, são necessárias mais que três componentes principais

para explicar a maior variância dos dados, considerando o número para o qual não há alteração significativa da variância. A necessidade de várias componentes principais — para explorar maior porcentagem das informações relevantes dos dados — gerou espaço multidimensional que dificultou a inspeção visual das distribuições espaciais. Adicionalmente, a organização dos dados no plano dos gráficos não permitiu avaliação detalhada de seu agrupamento, devido ao número de componentes principais avaliadas.

A fim de exemplificar a necessidade do grande número de componentes principais para explicar o comportamento dos dados, serão apresentados os resultados da ACP para as amostras de 2000, o ano com o segundo menor número de conjuntos de dados, conforme denotado na Tabela 2. Para esse ano, foram necessárias nove componentes principais para explicar 99,73% da variância dos dados. A ACP apresentaria atratividade de utilização se houvesse possibilidade de redução do número de variáveis em poucas componentes principais, com a menor perda de informação possível. Com isso, variáveis poderiam ser desconsideradas quando sua contribuição fosse pequena, mediante exames visuais em dispersões gráficas. Esse resultado culminou na aplicação subsequente da rede neural de Kohonen.

Rede neural de Kohonen

A rede neural de Kohonen foi aplicada a fim de atingir três objetivos:

- proporcionar a redução da dimensão dos dados originais;
- avaliar a possibilidade de redução do número de estações de monitoramento;
- avaliar a possibilidade de redução da frequência de amostragem ao longo da Bacia do Rio das Velhas.

Conforme ressaltado, na análise de Kohonen, a estrutura dos gráficos baseia-se em arquiteturas formadas por neurônios, nos quais dados semelhantes se distribuem próximos uns dos outros ou até mesmo dentro de um único neurônio. Assim, foram avaliadas várias arquiteturas em cada etapa da análise de Kohonen e, entre estas, as que apresentaram as distribuições dos dados, formando novos agrupamentos de dados semelhantes, foram as escolhidas para a análise exploratória.

Após a definição das melhores arquiteturas, prosseguiu-se a análise de Kohonen para avaliação da redução dos conjuntos de dados, inspecionando os agrupamentos formados para cada mês de cada ano. Nessa etapa, foram selecionados indivíduos de cada agrupamento, os quais formaram uma nova matriz para cada ano. Nessa fase, o conjunto de dados reduziu-se de 1.710 para 850, da ordem de 49 ou 53%, ao se considerar o número inicial (1.824). As reduções obtidas para cada mês do ano de 2010 estão mostradas na Tabela 3.

Na Figura 2, ilustram-se a definição da melhor arquitetura e a seleção dos agrupamentos de dados para a matriz do mês de janeiro de 2010, na qual foi avaliado o comportamento das amostras. A avaliação

Tabela 3 - Redução dos conjuntos de dados obtida com a rede neural de Kohonen para cada mês de 2010.

| Meses | Número inicial de amostras dos conjuntos de dados | Número de amostras após a aplicação da rede neural de Kohonen |
|-----------|---|---|
| Janeiro | 36 | 18 |
| Fevereiro | 19 | 9 |
| Março | 13 | 7 |
| Abril | 36 | 17 |
| Maio | 19 | 12 |
| Junho | 19 | 9 |
| Julho | 36 | 19 |
| Agosto | 19 | 9 |
| Setembro | 18 | 9 |
| Outubro | 36 | 19 |
| Novembro | 19 | 9 |
| Dezembro | 19 | 11 |

do comportamento das amostras se deu em conjunto com a avaliação do comportamento das variáveis (Figura 3). Dessa forma, a interpretação dos dados só se torna possível com o uso conjunto dos mapas de distribuição das amostras e dos mapas de distribuição das variáveis.

Na Figura 2, pode ser observada uma arquitetura de 8 × 8 neurônios, escolhida por apresentar a melhor distribuição entre os grupos formados, proporcionando a melhor interpretação de seus comportamentos. Nessa arquitetura, os dados foram reunidos em 12 grupos, nos quais as circulações internas representam as amostras escolhidas para representar o agrupamento, partindo da premissa de que quanto mais próximas as amostras estiverem entre si, maior a semelhança entre elas em função das variáveis inspecionadas.

Sobrepondo o mapa de distribuição das amostras ao de distribuição individual das variáveis, é possível analisar os grupos apresentados na Figura 2, em função do mapa dos valores das variáveis (Figura 3). Percebe-se que as variáveis *DBO*, *nitrato*, *fosfato* e *pH* se relacionam com influência mais

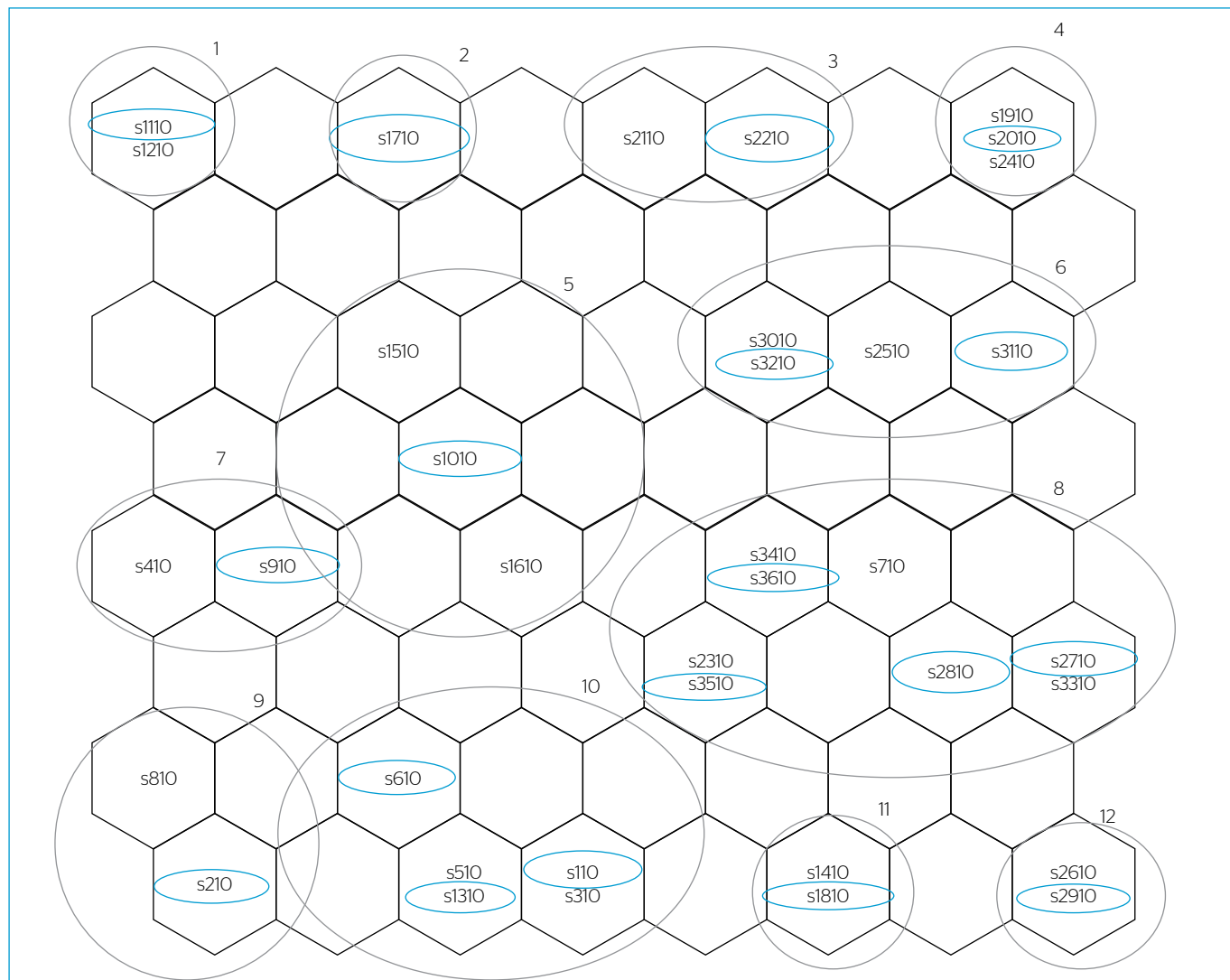


Figura 2 - Perfil geral de distribuição das amostras de janeiro de 2010, demonstrando as amostras do novo conjunto demarcadas com pequenas elipses.

acentuada e a *temperatura* com menor influência, para os conjuntos 1 e 2, sendo que a influência para o grupo 2 é menor quando comparada ao grupo 1.

O parâmetro *nitrato* apresentou forte influência no grupo 3, juntamente com *pH* e *temperatura*, estes com menor influência. O grupo 4 apresentou influência acentuada para os parâmetros *nitrato*, *sólidos totais*, *temperatura* e *turbidez*. As amostras do grupo 5 não foram intensamente influenciadas por nenhuma das variáveis, ao passo que as do grupo 6 o foram intensamente pela *temperatura*.

As variáveis *OD* e *pH* apresentaram influência para o grupo 7, o que determinou a sua formação, assim como dos grupos 8, 9, 10, 11 e 12. O grupo 8 foi notadamente influenciado pela *temperatura* com maior influência e com menor pelo *OD*, seguido do *pH*.

Os grupos 9, 10, 11 e 12 apresentaram grande influência do parâmetro *OD*, e o grupo 11 apresenta também grande influência para o IQA. Para os dados de janeiro de 2010, o conjunto de dados s1410 e s1810 apresentou o melhor valor de IQA.

As matrizes anuais geradas após a análise exploratória de Kohonen também foram submetidas a novo tratamento pela rotina dos mapas auto-organizáveis para possível seleção de agrupamentos. As reduções obtidas para cada ano são apresentadas na Tabela 4.

A redução do número de conjuntos de dados mostrou-se significativa, considerando que, após a realização das duas etapas da rede de Kohonen, o conjunto de dados total foi reduzido de 1.710 para 524. Essa redução foi gradual e dependente das duas etapas da rede neural

de Kohonen realizadas, as análises mensal para cada ano e anual. A aplicação da rede neural de Kohonen resultou em redução total de 71,4% do número original do conjunto de dados de monitoramento da bacia do Rio das Velhas. Esse resultado abre perspectiva de aplicação da metodologia para otimização dos programas de monitoramento de qualidade de água pelos órgãos estaduais gestores dos recursos hídricos.

Tabela 4 - Redução do número de conjuntos de dados obtida com a rede neural de Kohonen para os dados de monitoramento da qualidade da água de 2000 a 2010.

| Ano | Número de conjunto de dados após a aplicação da rede neural de Kohonen mês a mês | Número de conjuntos de dados após a aplicação da rede neural de Kohonen ano a ano |
|------|--|---|
| 2000 | 62 | 28 |
| 2001 | 65 | 41 |
| 2002 | 52 | 27 |
| 2003 | 56 | 39 |
| 2004 | 52 | 31 |
| 2005 | 56 | 37 |
| 2006 | 50 | 32 |
| 2007 | 51 | 31 |
| 2008 | 113 | 78 |
| 2009 | 145 | 89 |
| 2010 | 148 | 91 |

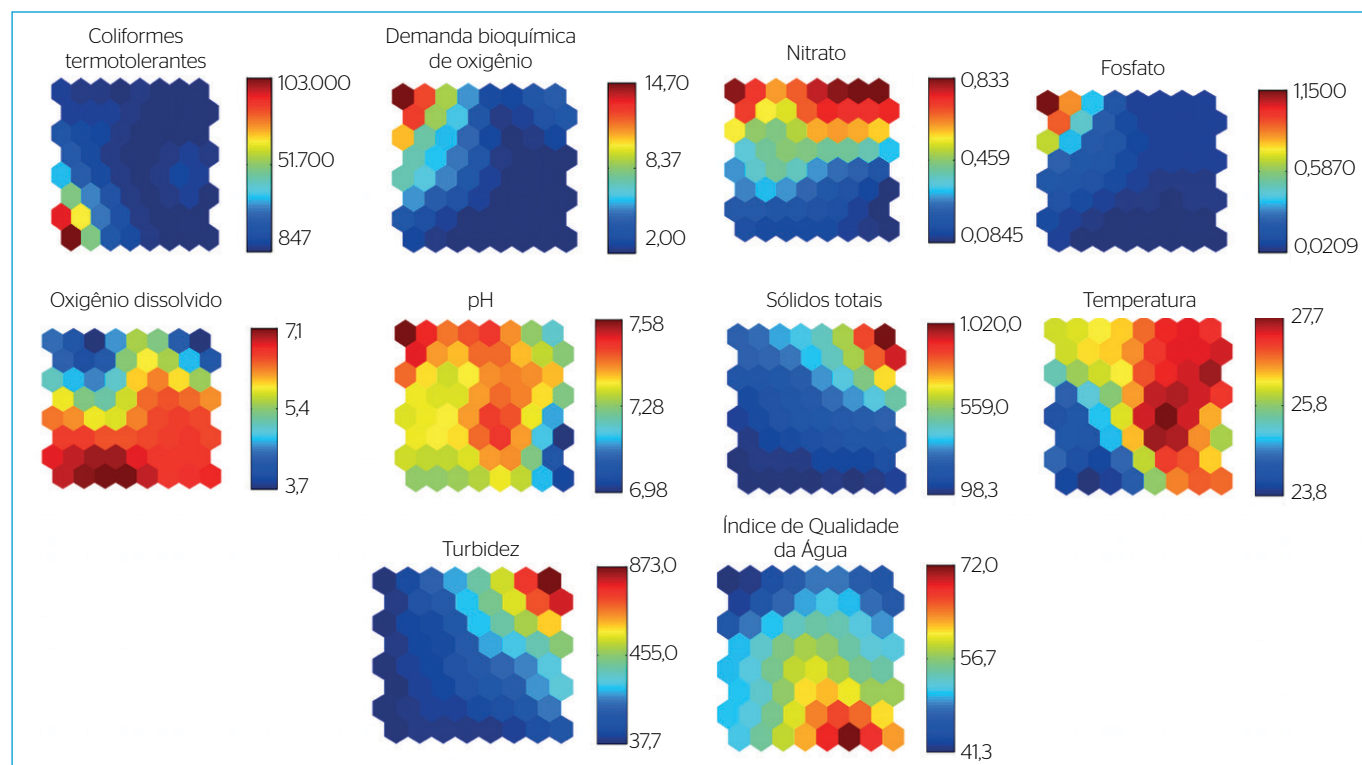


Figura 3 - Mapas de distribuição individual das variáveis analisadas em janeiro de 2010 para os parâmetros integrantes do Índice de Qualidade da Água e o próprio índice.

Correlação de Pearson

A análise de correlação de Pearson foi realizada para avaliar se a redução no número de conjuntos de dados obtida com a rede neural de Kohonen não foi prejudicial quanto às informações fornecidas por esses dados, garantindo que a representatividade da base reduzida de dados se mantivesse quando comparada à base original.

Na Figura 4, são comparadas a matriz de dados original e a final reduzida. Nessa comparação, os resultados obtidos para a intensidade de correlação de dados originais e para a intensidade de correlação da matriz final reduzida mantiveram-se os mesmos, evidenciando, mais uma vez, que a redução de dados por meio da rede de Kohonen não excluiu aqueles relevantes para o estudo em desenvolvimento.

É sabido que a correlação de Pearson evidencia a relação entre os parâmetros, o que indica que, ao analisar sempre os mesmos parâmetros, essa correlação seria mantida. No entanto, há que se considerar que a intensidade dessa correlação pode sofrer mudanças, o que justificou a aplicação da correlação de Pearson para averiguar a possibilidade de redução do número de conjunto de dados sem perda de informações relevantes.

Influência da sazonalidade

Avaliou-se o comportamento dos dados em relação aos períodos seco e chuvoso por meio de análise exploratória multivariada com a rede neural de Kohonen para a matriz final reduzida de cada ano. Para essa avaliação, as arquiteturas foram escolhidas segundo os mesmos critérios seguidos para

as demais análises de Kohonen desenvolvidas neste trabalho. Na Tabela 5, apresentam-se as arquiteturas avaliadas e o número de conjuntos de dados dos períodos seco e chuvoso para cada ano avaliado entre 2000 e 2010.

Na Figura 5, exemplifica-se a avaliação do comportamento dos conjuntos de dados ao longo dos períodos seco e chuvoso para o ano de 2010. Na Figura 6, é apresentado o comportamento das variáveis para o mesmo ano em relação aos períodos seco e chuvoso.

Para 2010, os dados agruparam-se em 37 grupos, em arquitetura formada por 15 × 15 neurônios. Nesse ano, também é observada a separação dos dados de acordo com os períodos seco e chuvoso. Os dados do período chuvoso distribuíram-se na extremidade esquerda do primeiro quadrante superior e parte dos primeiro e segundo quadrantes inferiores. Os dados do período seco distribuíram-se ao longo do segundo quadrante superior e nos primeiro e segundo quadrantes inferiores.

Avaliação da redução do número de estações de monitoramento e/ou frequência de amostragem

Essa etapa foi realizada para a matriz final, na qual os dados relativos às datas de amostragem e as estações monitoradas foram considerados para interpretação dos resultados da análise exploratória. Assim, para cada ano, foram comparadas as estações, a frequência de monitoramento realizada pelo IGAM e as informações mantidas na matriz final.

A aplicação da rede neural de Kohonen apontou que o monitoramento da qualidade da água da Bacia do Rio das Velhas — no que tange à determinação dos parâmetros que integram o IQA — poderia e poderá ser realizado com menor número de estações e menor frequência de amostragem. Na Tabela 6, apresenta-se o número de estações de monitoramento calcado no IQA ao longo do período amostrado obtido neste trabalho.

Após a rede de Kohonen, associou-se a redução do número de estações à menor frequência de amostragem recomendada para a Bacia

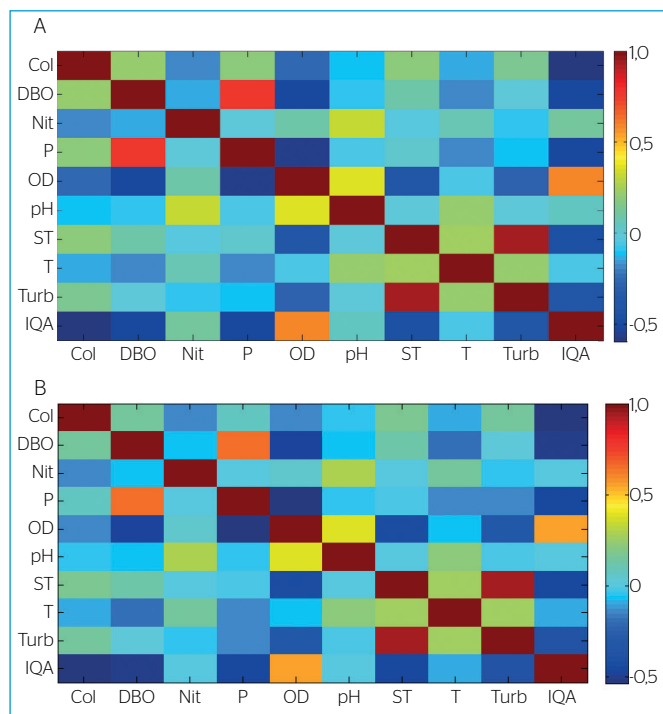


Figura 4 - Gráficos de correlação para a matriz de dados originais (A) e para a final reduzida (B) para o ano de 2010, composta dos resultados de monitoramento de qualidade da água, para os parâmetros do Índice de Qualidade da Água e o próprio índice.

Tabela 5 - Conjuntos de dados para o estudo da sazonalidade do período amostrado e a arquitetura utilizada na rede neural de Kohonen.

| Ano | Número de conjuntos de dados para o período seco da matriz anual reduzida | Número de conjuntos de dados para o período chuvoso da matriz reduzida | Arquitetura utilizada |
|------|---|--|-----------------------|
| 2000 | 13 | 15 | 8 × 8 |
| 2001 | 16 | 25 | 10 × 10 |
| 2002 | 15 | 12 | 7 × 7 |
| 2003 | 23 | 16 | 9 × 9 |
| 2004 | 19 | 12 | 9 × 9 |
| 2005 | 19 | 18 | 10 × 10 |
| 2006 | 19 | 13 | 9 × 9 |
| 2007 | 20 | 11 | 9 × 9 |
| 2008 | 40 | 38 | 15 × 15 |
| 2009 | 44 | 45 | 14 × 14 |
| 2010 | 52 | 39 | 15 × 15 |

do Rio das Velhas. Nesse contexto, novamente exemplificando apenas para o ano de 2010, cuja frequência de monitoramento foi a mesma de 2009, tem-se: sete estações podem ser monitoradas anualmente (062, 140, 142, 144, 145, 155 e 160); 12, semestralmente (013, 035, 037, 076, 133, 136, 139, 147, 148, 150, 152 e 161); três, três vezes ao ano (105, 130 e 137); sete, trimestralmente (063, 067, 146, 151, 153, 156 e 162); duas, cinco vezes ao ano (083 e 149); uma, com frequência bimestral (141); e uma, sete vezes ao ano (138).

Cabe salientar que, em todos os anos avaliados, a frequência de amostragem foi alterada, sinalizando para redução da frequência de amostragem nos anos subsequentes. Por outro lado, comparando-se as estações eliminadas e as mantidas, conclui-se que cada estação foi mantida pelo menos uma vez ao longo do período avaliado.

A avaliação da redução dos custos com o monitoramento da qualidade da água traz relevante potencial para ampliação da rede de monitoramento de qualidade de água em Minas Gerais, para o desenvolvimento de ferramentas mais avançadas para essa prática ou mesmo para a ampliação do espectro dos parâmetros monitorados. Nessa vertente, a perspectiva de redução do custo do programa de monitoramento de qualidade da água da Bacia do Rio das Velhas — e de outras bacias hidrográficas do Estado — centrar-se-á mais significativamente na coleta e no transporte das amostras do que na possível redução do número de parâmetros que integram o IQA.

Por fim, cabe ressaltar que a aplicação das metodologias apresentadas neste trabalho a esta e a outras bacias hidrográficas — com provável redução dos custos de monitoramento da qualidade da

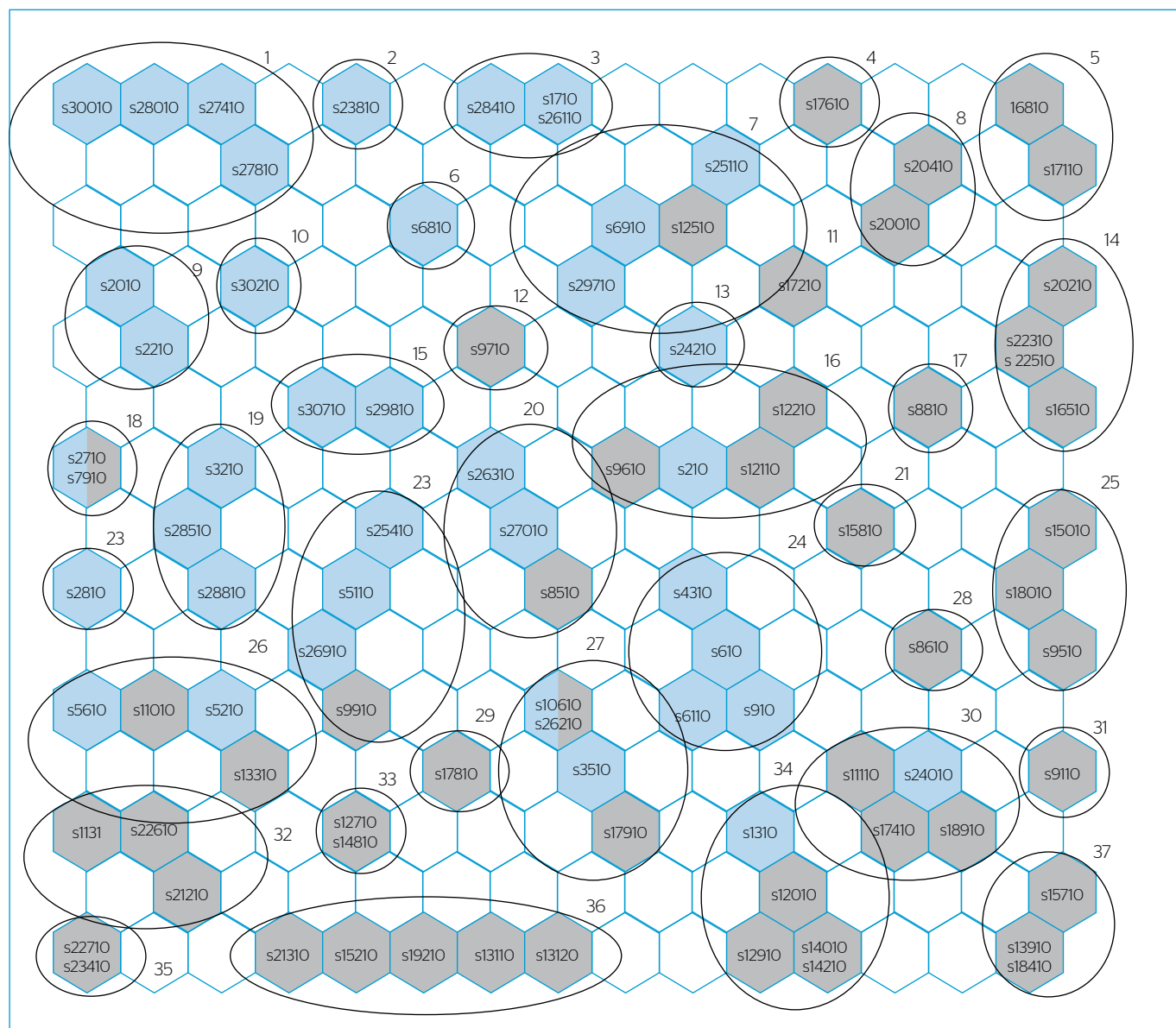


Figura 5 - Distribuição das amostras da matriz reduzida do ano de 2010 em relação aos períodos seco (cor cinza) e chuvoso (cor azul).

água — não exige o órgão gestor de análise acurada dos predominantes usos dos recursos hídricos. Essa ressalva fia-se na perspectiva de que a redução da frequência de monitoramento e do número de estações de amostragem também reduz a possibilidade de identificação mais expedita dos lançamentos pontuais, acidentais ou não, de cargas poluidoras.

CONCLUSÕES

A partir dos resultados auferidos, é possível concluir que:

- há perspectiva de redução do número de estações e da frequência de monitoramento da qualidade dos cursos d'água pelos órgãos gestores estaduais, mantendo-se a fidedignidade das informações, por meio da aplicação da rede neural de Kohonen. A menor frequência de amostragem oriunda da rede de Kohonen para a Bacia do Rio das Velhas resultou na possibilidade de realização das coletas em intervalos de até 50 dias;
- a rede neural de Kohonen mostrou-se também adequada para avaliar o comportamento dos dados ao longo dos anos, segundo os períodos seco e chuvoso. Os dados avaliados que correspondem ao mesmo período tendem a se posicionar na rede neural de forma que se crie um grupo e, no fim da análise, obtenha-se a clara separação dos dados em função da sazonalidade. Essa constatação abre a possibilidade de avaliação de distintas frequências de monitoramento nos períodos de estiagem e de chuva, assim

como a possibilidade de verificação da influência dos parâmetros avaliados em função da sazonalidade;

- a ACP é uma ferramenta muito apropriada para análise exploratória multivariada, mas mostrou-se de menor praticidade para casos semelhantes. Ela pode tornar a interpretação dos resultados laboriosa quando o número de componentes principais conduz a um espaço multidimensional que torna onerosa a inspeção visual das distribuições espaciais.

Tabela 6 - Número de estações de monitoramento da qualidade da água da Bacia do Rio das Velhas em operação e necessário após a aplicação da rede neural de Kohonen.

| Ano | Estações de monitoramento em operação | Estações de monitoramento apontadas neste trabalho |
|------|---------------------------------------|--|
| 2000 | 29 | 19 |
| 2001 | 29 | 25 |
| 2002 | 29 | 20 |
| 2003 | 29 | 22 |
| 2004 | 29 | 22 |
| 2005 | 33 | 23 |
| 2006 | 33 | 22 |
| 2007 | 33 | 22 |
| 2008 | 36 | 33 |
| 2009 | 36 | 35 |
| 2010 | 36 | 33 |

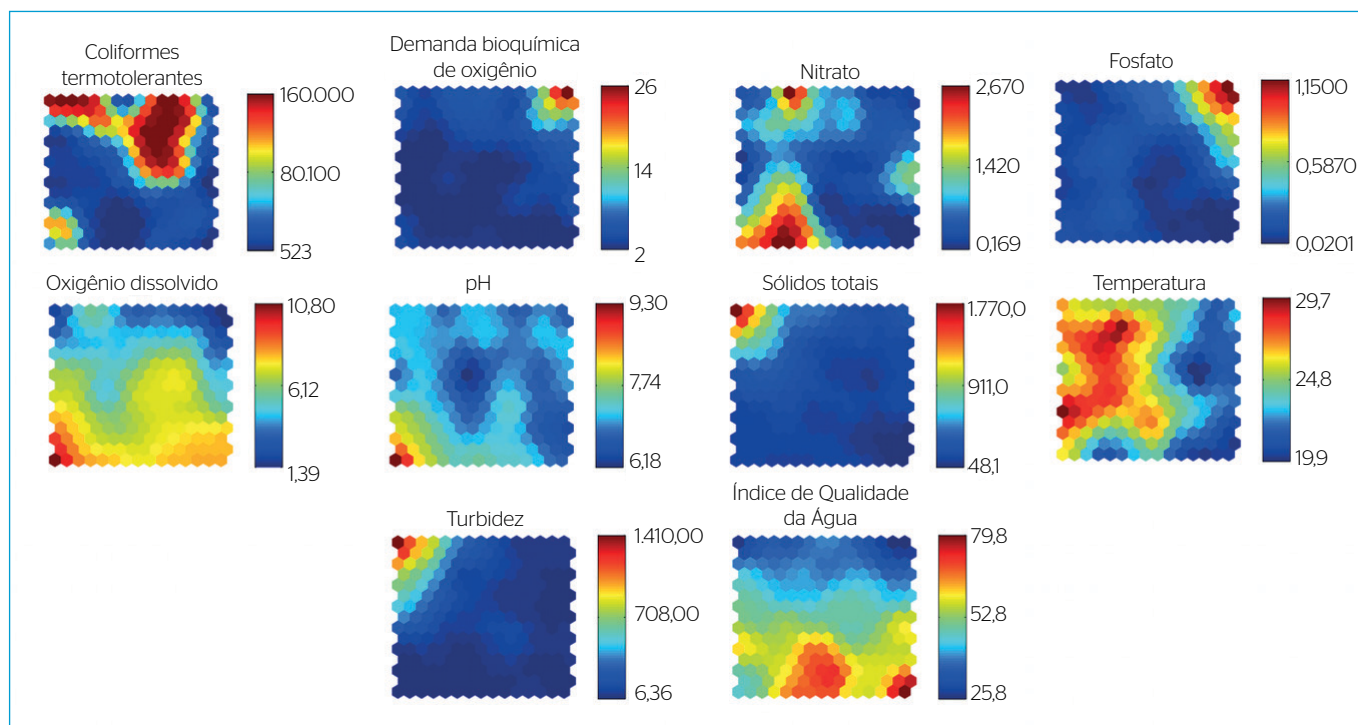


Figura 6 - Mapas de distribuição individual das variáveis analisadas na matriz reduzida no ano de 2010.

REFERÊNCIAS

- AMERICAN PUBLIC HEALTH ASSOCIATION (APHA); AMERICAN WATER WORKS ASSOCIATION (AWWA); WATER ENVIRONMENT FEDERATION (WEF). (2005) *Standard methods for the examination of water and wastewater*. 21. ed. Washington, D.C.: American Public Health Association.
- BARROS NETO, B. de; SCARMINIO, I.S.; BRUNS, R.E. (2006) 25 anos de quimiometria no Brasil. *Química Nova*, v. 29, n. 6, p. 1401-1406. <http://dx.doi.org/10.1590/S0100-40422006000600042>
- BRASIL. (2005) Conselho Nacional de Meio Ambiente. Resolução nº 357. 18 p.
- BROWN, R.M.; MCCLELLAND, N.I.; DEINIGER, R.A.; TOZER, R.G. (1970) A water quality index – do we dare? *Water & Sewage Works*, Chicago, v. 117, n. 10, p. 339-343.
- COSTA, J.A.F.; ANDRADE NETTO, M.L. (2007) Segmentação de mapas auto-organizáveis com espaço de saída. *Controle e Automação*, v. 18, n. 2, p. 150. <http://dx.doi.org/10.1590/S0103-17592007000200002>
- FRANCISCO, E.; WHIGHAM, P. A.; ARANHA FILHO, F.; ZAMBALDI, F.; PONCHIO, M. (2008). A Consumer Income Predicting Model Based on Survey Data: An Analysis Using Geographically Weighted Regression (GWR). *Latin American Advances in Consumer Research*. v. 2.
- GONTIJO, E.S.J.; WATANABE, C.H.; MONTEIRO, A.S.C.; TONELLO, P.S.; SILVA, G.A.; FRIESE, K.; ROESER, H.M.P.; ROSA, A.H. (2016) Distribution and bioavailability of arsenic in natural waters of a mining area studied by ultrafiltration and diffusive gradients in thin films. *Chemosphere*, Oxford, v. 164, p. 290-298. <https://doi.org/10.1016/j.chemosphere.2016.08.107>
- HAIR, J.J.F.; Anderson, R.E.; Tatham, R.L.; Black, W.C. (2005) *Análise multivariada de dados*. 5. ed. Porto Alegre: Bookman. 593 p.
- INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS (IGAM). (2016) *Qualidade das Águas Superficiais de Minas Gerais em 2015*. Resumo Executivo. IGAM. 75 p. Disponível em: <<http://200.198.57.118:8080/jspui/handle/123456789/322>>. Acesso em: 22 ago. 2016.
- INSTITUTO MINEIRO DE GESTÃO DAS ÁGUAS (IGAM). (2018) *Monitoramento da qualidade das águas*. IGAM. Disponível em: <<http://www.igam.mg.gov.br/monitoramento-da-qualidade-das-aguas2>>. Acesso em: 9 maio 2018.
- KIM, J.Y.; BHATTA, K.; RASTOGI, G.; MUDULI, P.R.; DO, Y.; KIM, D.K.; PATTAIK, A.K.; JOO, G.J. (2016) Application of multivariate analysis to determine spatial and temporal changes in water quality after new channel construction in the Chilika Lagoon. *Ecological Engineering*, v. 90, p. 314-319. <https://doi.org/10.1016/j.ecoleng.2016.01.053>
- KOWALSKI, C.H.; SILVA, G.A.; GODOY, H.T.; POPPI, R.J.; AUGUSTO, F. (2013) Application of Kohonen neural network for evaluation of the contamination of Brazilian breast milk with polychlorinated biphenyls. *Talanta*, v. 116, p. 315-321. <https://doi.org/10.1016/j.talanta.2013.05.033>
- LANDWEHR, J.M.; DEININGER, R.A. (1976) A comparison of several water quality indexes. *Journal of Water Pollution Control Federation*, v. 48, n. 5, p. 954-958.
- MOORE, D.S. (2007) *The Basic Practice of Statistics*. Nova York: Freeman.
- MUSTONEN, S.M.; TISSARI, S.; HUIKKO L.; KOLEHMAINEN, M.; LEHTOLA, M.J.; HIRVONEN, A. (2008) Evaluating online data of water quality changes in a pilot drinking water distribution system with multivariate data exploration methods. *Water Research*, v. 42, n. 10-11, p. 2421-2430. <https://doi.org/10.1016/j.watres.2008.01.015>
- OUYANG, Y. (2005) Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*, v. 39, n. 12, p. 2621-2635. <https://doi.org/10.1016/j.watres.2005.04.024>
- RENCHER, A.C.; CHRISTENSEN, W.F. (2012) *Methods of Multivariate Analysis*. 3. ed. Nova York: John Wiley & Sons.
- ROBERTS, M. (2007) *Schizophrenia daily news blog: New neurobiology knowledge may lead to schizophrenia research and treatments*. Disponível em: <<http://www.schizophrenia.com/sznews/archives/005490.html>>. Acesso em: 27 abr. 2018.
- SAHOO, M.M.; PATRA, K.C.; KHATUA, K.K. (2015) Inference of Water Quality Index using ANFIA and PCA. *Aquatic Procedia*, v. 4, p. 1099-1106. <https://doi.org/10.1016/j.aqpro.2015.02.139>
- SENA, M.M. de; POPPI, R.J.; FRIGHETTO, R.S.; VALARINI, P.J. (2000) Avaliação do uso de métodos quimiométricos em análise de solos. *Química Nova*, v. 23, n. 4.
- SUN, W.; XIA, C.; XU, M.; GUO, J.; SUN, G. (2016) Application of modified water quality indices as indicators to assess the spatial and temporal trends of water quality in the Dongjiang River. *Ecological Indicators*, v. 66, p. 306-312. <https://doi.org/10.1016/j.ecolind.2016.01.054>