ISSN (*on-line*) 1984-2368

Boletim do Instituto Adolfo Lutz

Bol Inst Adolfo Lutz. 2018: ano 28, número único

Boletim do
INSTITUTO ADOLFO LUTZ

Bol Inst Adolfo Lutz. 2018: ano 28, número único

ISSN (*on-line*) 1984-2368

Carta ao Editor
Avenida Dr. Arnaldo, 355
Cerqueira César – CEP 01246-902
E-mail: bial@ial.sp.gov.br
São Paulo, SP – Brasil
Telefone: (11) 3068-2867
Núcleo de Acervo

Sumário

Azitromicina pó para suspensão oral: perspectivas de utilização de uma alternativa mais rápida para determinação de teor por CLAE-UV

Fernanda Fernandes FARIAS[1]; Ellen Gameiro HILINSKI[2]; Valéria Adriana Pereira MARTINS[1]; Luz Marina TRUJILLO[1]; Adriana Aparecida Buzzo ALMODOVAR[2]

[1]Núcleo de Ensaios Físicos e Químicos em Medicamentos - Centro de Medicamentos, Cosméticos e Saneantes - Instituto Adolfo Lutz

[2]Núcleo de Ensaios Biológicos e de Segurança - Centro de Medicamentos, Cosméticos e Saneantes - Instituto Adolfo Lutz

A azitromicina é um agente antimicrobiano da classe dos macrolídeos, disponível no Brasil desde 1991, foi desenvolvida em laboratório a partir da molécula de eritromicina. Sua nova conformação estrutural permitiu maior tempo de prateleira, maior estabilidade, resistência a meios mais ácidos e melhor difusão tecidual. Possui a capacidade de provocar a inibição da síntese proteica bacteriana através da sua ligação com o RNA ribossomal 23S da subunidade 50S, impedindo a tradução do mRNA e consequente síntese peptídica, sendo principalmente indicada para o tratamento de infecções respiratórias[1].

Para que a infecção seja devidamente combatida, é necessário que a potência do antimicrobiano nas diversas preparações farmacêuticas esteja dentro de parâmetros especificados por compêndios oficiais. Diante deste contexto, o ensaio de teor realizado em laboratório de controle de qualidade é uma ferramenta para evidenciar que a dosagem indicada em rótulo se encontra em conformidade com especificações vigentes[2].

Atualmente, o método oficial preconizado pela Farmacopeia Brasileira[3] para o doseamento de azitromicina na forma de matéria prima, cápsula e pó para suspensão oral é por difusão em ágar. Trata-se de um ensaio microbiológico, no qual se determina a potência ou atividade de um antimicrobiano comparando-se a dose que inibe o crescimento de um microrganismo susceptível em relação à dose de uma substância padrão.

Considerando a Farmacopeia Americana[4], o método de escolha para o teor de azitromicina em pó para suspensão oral é por Cromatografia Líquida de Alta Eficiência (CLAE), por meio do emprego da detecção amperométrica eletroquímica, que utiliza um detector de alto custo, pouco comum em laboratórios de controle de qualidade.

O objetivo deste trabalho foi adaptar métodos desenvolvidos e validados, descritos em literatura científica[5,6], para a determinação de teor de azitromicina por CLAE detecção UV, e verificação destes resultados com os obtidos pela metodologia microbiológica por difusão em ágar, conforme compêndio oficial.

Para a execução deste estudo foram utilizadas unidades de um mesmo lote de azitromicina pó para suspensão oral na concentração de 200 mg/5 mL.

Foi utilizado o cromatógrafo líquido acoplado a um detector UV-Vis (Waters, Milford, EUA), com leitura a 210 nm, coluna empacotada com sílica quimicamente ligada a grupo octadecil (C_{18}) de 5 µm, tamanho 250 x 4,6 mm, mantida a temperatura de 50 °C, fluxo da fase móvel de 1,5 mL/min, sendo a fase móvel uma mistura isocrática constituída de metanol e tampão fosfato de potássio 0,03 M pH 7,5 (80:20). A corrida foi estabelecida em 13 minutos, com tempo de retenção do ativo em aproximadamente 9 minutos (Figura 1), injeção de 30 µL. Foram utilizados reagentes grau HPLC e padrão de trabalho certificado fornecido pelo fabricante do produto de marca. A amostra foi previamente reconstituída segundo indicações da bula do medicamento, obtendo concentração de 1 mg/mL, sendo o diluente a própria fase móvel. Filtrou-se para *vial* com auxílio de cartucho de filtração com poro de 0,45 µm.

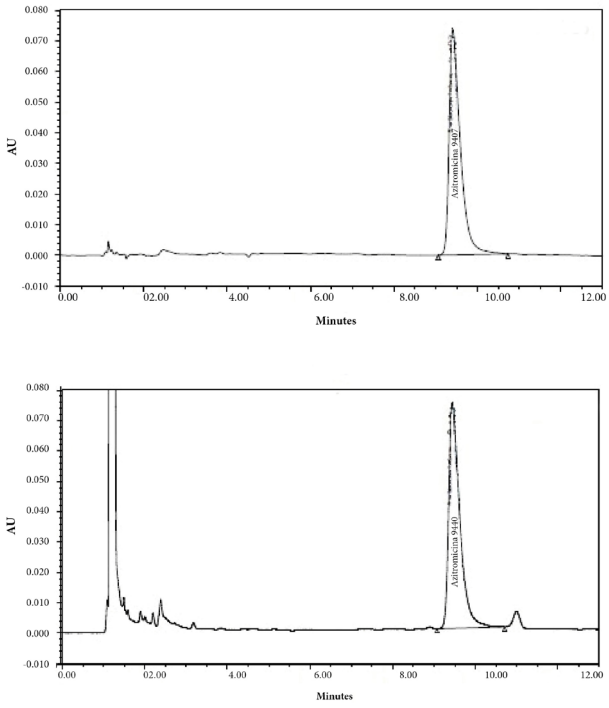

Figura 1. Cromatograma de padrão e amostra de azitromicina pelo método CLAE-UV

O ensaio microbiológico de azitromicina foi realizado de acordo com o preconizado pela Farmacopeia Brasileira[3]. O diâmetro de inibição da zona foi medido após o período de incubação das placas por 16-18 horas a 33,5 ± 1,5 °C, com o auxílio de paquímetro (Figura 2).

A especificação tanto em Farmacopeia Brasileira quanto Americana para a potência de azitromicina em pó para suspensão oral é de 90,0 a 110,0% do teor declarado. O teor obtido de azitromicina por CLAE-UV para a amostra de 200 mg/5 mL foi de 190,7 mg/5 mL (95,4%) e por difusão em ágar 184,5 mg/5 mL (92,3% do teor declarado). A variação entre os resultados obtidos por meio dos dois métodos foi de 3,4%. A análise realizada por cromatografia líquida com detecção por UV foi considerada uma ferramenta adequada para a quantificação do teor de azitromicina por ser uma técnica versátil, sensível e precisa, obtendo-se assim uma boa separação. Além disso, o método desenvolvido apresentou-se como uma alternativa mais rápida, uma vez que o ensaio microbiológico requer incubação mínima de 16 horas para obtenção dos resultados, enquanto o método por CLAE-UV, após preparo de fase móvel e amostra, apresenta corrida com duração de 13 minutos.

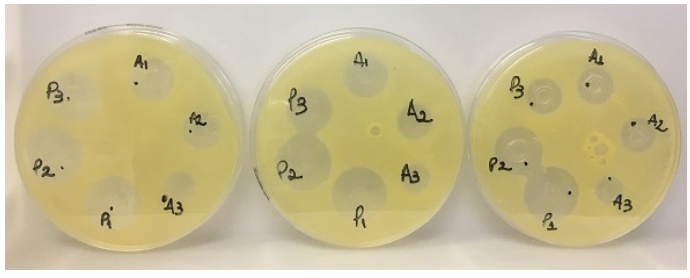

Figura 2. Placas com padrão (P_1, P_2, P_3) e amostra (A_1, A_2, A_3), método oficial microbiológico

Consideramos que este estudo constitui importante ferramenta para a elaboração do plano de validação do método proposto, o qual poderá ser utilizado como alternativa para futuras análises de determinação de teor de azitromicina na forma de pó para suspensão oral.

REFERÊNCIAS

1. Guimarães DO, Momesso LS, Pupo MT. Antibióticos: importância terapêutica e perspectivas para a descoberta e desenvolvimento de novos agentes. Quim Nova. 2010; 33(3): 667-679.

2. Takamune LF, Vieira DCM. Comparação da metodologia para determinação da potência de amoxicilina: método de difusão em ágar e método de espalhamento. Rev Ciênc Farm Básica Apl., 2013;34(4):555-558.

3. BRASIL. Farmacopéia Brasileira. 5.ed. v.1. Brasília: Agência Nacional de Vigilância Sanitária, 2010. Disponível em: [http://www.anvisa.gov.br/hotsite/cd_farmacopeia/pdf/volume1.pdf].

4. The US Pharmacopoeia. 40th. ed. Rockville: The US Pharmacopoeial Convention, 2017.

5. Ghari T, Kobarfard F, Mortazavia SA. Development of a Simple RP-HPLC-UV Method for Determination of Azithromycin in Bulk and Pharmaceutical Dosage forms as an Alternative to the USP Method. Iran J Pharm Res. 2013; 12(Suppl): 57–63.

6. Al-Rimawi F, Kharoaf M. Analysis of azithromycin and its related compounds by RP-HPLC with UV detection. J Chromatogr Sci. 2010;48(2):86-90.

Cálculo de incertezas associadas a medidas de condutividade em amostras de águas

Sergio DOVIDAUSKAS[1], Isaura Akemi OKADA[1]

[1]Núcleo de Ciências Químicas e Bromatológicas – Centro de Laboratório Regional – Instituto Adolfo Lutz de Ribeirão Preto VI

Na medição da condutividade em águas destinadas ao consumo humano no NQBRP (Núcleo de Ciências Químicas e Bromatológicas de Ribeirão Preto), considerou-se que as fontes de incerteza seriam: a condutividade da solução padrão utilizada na calibração do condutivímetro, a resolução do equipamento (condutivímetro marca Metrohm, modelo 912) e a precisão da medição. As medidas foram realizadas em ambiente com controle de temperatura, e o efeito dessa sobre as medições de condutividade não foi considerado como fonte de incerteza uma vez que a cela de condutividade utilizada (marca Metrohm) possui um sensor acoplado que permite a correção automática dos valores de condutividade medidos na temperatura da análise para valores de condutividade a 25ºC. Assim, utilizando a regra da adição/subtração[1], a incerteza padrão combinada da condutividade de uma amostra ($u_{cond\ am}$) pôde ser calculada pela equação 1:

$$u_{cond\ am} = \sqrt{u_{Ccal}^2 + u_{res}^2 + u_{pm}^2} \quad (eq.1)$$

onde:

u_{Ccal} = incerteza padrão associada à condutividade da solução padrão usada na calibração

u_{res} = incerteza padrão associada à resolução do condutivímetro

u_{pm} = incerteza padrão associada à precisão da medição

Existem fornecedores de soluções padrão de condutividade que informam a incerteza padrão associada (u_{Ccal}), mas o NQBRP não dispõe de solução padrão similar com certificação do valor de condutividade. A alternativa adotada foi preparar uma solução 0,001 mol L^{-1} de KCl que apresenta condutividade igual a 146,9 µS cm^{-1} (a 25ºC) segundo o *Standard Methods for the Examination of Water and Wastewater*[2] – essa condutividade é particularmente adequada para a calibração do condutivímetro nas análises de águas destinadas ao consumo humano da região onde o NQBRP atua uma vez que, em estudo recente[3], verificou-se que a mediana e o terceiro quartil de 4.347 amostras analisadas de águas de abastecimento da região apresentaram os valores de 123,1 e 179,8 µS cm^{-1}, respectivamente. A estratégia utilizada para o preparo da solução de calibração envolveu uma diluição na razão 5:500 (pipeta 5,000 ± 0,015 e balão volumétrico de 500,00 ± 0,25 mL) de uma solução 0,1 mol L^{-1} de KCl preparada, por sua vez, dissolvendo-se 0,7455 g do sal de alta pureza (100 ± 0,5%, Sigma-Aldrich/Merck) em água ultrapurificada até 100 mL (balão volumérico de 100,0 ± 0,1 mL). Assim, para calcular u_{Ccal}, foi necessário inicialmente calcular a incerteza padrão combinada da concentração da solução de KCl 0,1 mol L^{-1} (u_{MKCl}) e a incerteza padrão combinada da concentração da solução de KCl 0,001 mol L^{-1} (u_{Mcal}) resultante da diluição da solução concentrada, transformando o resultado final em condutividade (u_{Ccal}), como é descrito a seguir.

A concentração molar da solução inicial de KCl (M_{KCl}) é calculada pela equação 2

$$M_{KCl} = \frac{m_{KCl} * p_{KCl}}{MM_{KCl} * V_{sol\ KCl}} \quad (eq.2)$$

onde:

m_{KCl} = massa de KCl, em g

p_{KCl} = pureza do KCl

MM_{KCl} = massa molar do KCl, em g mol^{-1}

$V_{sol\ KCl}$ = volume da solução de KCl, em L

Pela regra da multiplicação/divisão, a incerteza padrão combinada da concentração (u_{MKCl}) será:

$$u_{M_{KCl}} = M_{KCl} * \sqrt{\left(\frac{u_{m_{KCl}}}{m_{KCl}}\right)^2 + \left(\frac{u_{p_{KCl}}}{p_{KCl}}\right)^2 + \left(\frac{u_{MM_{KCl}}}{MM_{KCl}}\right)^2 + \left(\frac{u_{V_{sol\,KCl}}}{V_{sol\,KCl}}\right)^2} \quad (eq.3)$$

onde:

u_{mKCl} = incerteza padrão associada à massa do KCl

u_{pKCl} = incerteza padrão associada à pureza do KCl

u_{MMKCl} = incerteza padrão associada à massa molar do KCl

$u_{Vsol\,KCl}$ = incerteza padrão associada ao volume da solução preparada

Considerando a incerteza padrão da medida de massa da balança utilizada (marca Mettler Toledo modelo AB204) igual a 0,1 mg, e que o recipiente vazio é inicialmente pesado (tara) originando uma incerteza (s_{mt}) para depois ser adicionado KCl até a massa de 0,7455 g com a respectiva incerteza associada (s_{mp}), temos:

$$u_{m_{KCl}} = \sqrt{(s_{mt})^2 + (s_{mp})^2} \quad (eq.4)$$

$$u_{m_{KCl}} = \sqrt{(0{,}0001)^2 + (0{,}0001)^2} = 1{,}4142 \cdot 10^{-4} g$$

Com relação a u_{pKCl}, o fornecedor do KCl utilizado informa um teor de KCl entre 99,5% e 100,5% (100,0 ± 0,5%, ou 1,000 ± 0,005), sem indicação sobre o fator de abrangência (k) e o nível de confiança adotado. Então, considerando uma distribuição retangular:

$$u_{p_{KCl}} = \frac{a}{\sqrt{3}} \quad (eq.5)$$

$$u_{p_{KCl}} = \frac{0{,}005}{\sqrt{3}} = 2{,}8867 \cdot 10^{-3}$$

Com relação a u_{MMKCl}, segundo dados da *International Union of Pure and Applied Chemistry* (IUPAC)[4], a massa atômica do potássio (K) é 39,0983 ± 0,0001 Da, enquanto cloro (Cl) possui uma massa atômica contida no intervalo de 35,446 a 35,457 Da. Considerou-se a massa atômica do cloro como o ponto médio entre os dois limites (35,4515 Da) e assumiu-se a incerteza dessa massa como a diferença entre os limites e o ponto médio (± 0,0055 Da). Portanto, a massa molar do KCl, envolvendo a soma das massas molares do potássio e do cloro e igual a 74,5498 g mol^{-1}, apresenta uma incerteza (u_{MMKCl}) que foi calculada pela equação 6, utilizando distribuições retangulares para as massas molares dos dois elementos (equação 5):

$$u_K = \frac{0{,}0001}{\sqrt{3}} = 5{,}7735 \cdot 10^{-5}\ g\ mol^{-1}$$

$$u_{Cl} = \frac{0{,}0055}{\sqrt{3}} = 3{,}1754 \cdot 10^{-3}\ g\ mol^{-1}$$

$$u_{MMKCl} = \sqrt{(u_K)^2 + (u_{Cl})^2} \quad (eq.6)$$

$$u_{MMKCl} = 3{,}1759 \cdot 10^{-3}\ g\ mol^{-1}$$

Com relação a $u_{Vsol\,KCl}$, na incerteza declarada no balão volumétrico no qual a solução de KCl 0,1 mol L^{-1} foi preparada (100,0 ± 0,1 mL ou 0,1000 ± 0,0001 L) não haviam informações sobre o fator de abrangência e sobre o nível de confiança adotado. Então, utilizando uma distribuição retangular (equação 5):

$$u_{V_{sol\,KCl}} = \frac{0{,}0001}{\sqrt{3}} = 5{,}8 \cdot 10^{-5}\ L$$

Substituindo na equação 3 os valores e as incertezas calculadas, obteve-se a incerteza padrão combinada da concentração (u_{MKCl}):

$$u_{M_{KCl}} = 0{,}1 * \sqrt{\left(\frac{1{,}4 \cdot 10^{-4}}{0{,}7455}\right)^2 + \left(\frac{2{,}9 \cdot 10^{-3}}{1}\right)^2 + \left(\frac{3{,}2 \cdot 10^{-3}}{74{,}5498}\right)^2 + \left(\frac{5{,}8 \cdot 10^{-5}}{0{,}1}\right)^2}$$

$$u_{M_{KCl}} = 0{,}0003\ mol\ L^{-1}$$

A diluição da solução de KCl 0,1000 ± 0,0006 mol L^{-1} (k = 2; nível de confiança = 95,45%) pode ser expressa pela equação 7:

$$M_{cal} = \frac{M_{KCl_{conc}} \cdot V_{KCl_{conc}}}{V_{KCl_{dil}}} \quad (eq.7)$$

onde:

$M_{KCl\ conc}$ = concentração da solução concentrada (0,1000 ± 0,0006 mol L^{-1})

$V_{KCl\ conc}$ = volume da solução concentrada usada (pipeta de 5,000 ± 0,015 mL)

$V_{KCl\ dil}$ = volume da solução (balão volumétrico de 500,00 ± 0,25 mL)

Para M_{cal} = 0,001 mol L^{-1}, a incerteza padrão combinada (u_{Mcal}) foi calculada através da equação 8:

$$u_{M_{cal}} = M_{cal} \cdot \sqrt{\left(\frac{u_{M_{KCl_{conc}}}}{M_{KCl_{conc}}}\right)^2 + \left(\frac{u_{V_{KCl_{conc}}}}{V_{KCl_{conc}}}\right)^2 + \left(\frac{u_{V_{KCl_{dil}}}}{V_{KCl_{dil}}}\right)^2} \quad (eq.8)$$

onde:

$u_{MKCl\ conc}$ = incerteza padrão associada à concentração da solução de KCl 0,1000 mol L^{-1}

$u_{VKCl\ conc}$ = incerteza padrão associada ao volume usado da solução de KCl 0,1000 mol L^{-1}

$u_{VKCl\ dil}$ = incerteza padrão associada ao volume da solução preparada

Tanto $u_{VKCl\ conc}$ como $u_{VKCl\ dil}$ foram calculados assumindo-se distribuições retangulares (equação 5):

$$u_{V_{KCl_{conc}}} = \frac{0{,}015}{\sqrt{3}} = 8{,}6603 * 10^{-3}\ mL$$

$$u_{V_{KCl_{dil}}} = \frac{0{,}25}{\sqrt{3}} = 0{,}1443\ mL$$

Substituindo os valores na equação 8:

$$u_{M_{cal}} = 0{,}001 * \sqrt{\left(\frac{0{,}0003}{0{,}1000}\right)^2 + \left(\frac{8{,}66 * 10^{-3}}{5{,}000}\right)^2 + \left(\frac{0{,}14}{500{,}00}\right)^2}$$

$$u_{M_{cal}} = 3{,}476 * 10^{-6} \cong 0{,}0000035\ mol\ L^{-1}$$

Uma vez que uma solução de KCl 0,001 mol L^{-1} apresenta a condutividade de 146,9 µS cm^{-1}, considerou-se a transformação da representação da solução de KCl de concentração para condutividade como uma transformação linear, admitindo-se um fator *b* de multiplicação como indicado na equação 9, onde C_{cal} é a condutividade da solução de calibração:

$$M_{cal} \cdot b = C_{cal} \qquad (eq.9)$$

Substituindo os valores de M_{cal} e C_{cal}, é possível calcular *b*:

$$0{,}001 \cdot b = 146{,}9$$

$$b = 146.900\ \mu S\ cm^{-1} mol^{-1} L$$

Multiplicando o valor de *b* pelo valor da incerteza da concentração da solução de calibração (u_{Mcal}), é possível expressar a incerteza padrão combinada da condutividade da solução de calibração (u_{Ccal}):

$$u_{C_{cal}} = u_{M_{cal}} \cdot b = 0{,}0000035 \cdot 146.900$$

$$u_{C_{cal}} \cong 0{,}5\ \mu S\ cm^{-1}$$

Para um fator de abrangência igual a 2 (k = 2) e um nível de confiança igual a 95,45%, a condutividade C_{cal} da solução de calibração foi expressa como 146,9 ± 1,0 µS cm^{-1}.

Uma vez calculado u_{Ccal}, o próximo passo, indicado pela equação 1, foi estimar a incerteza padrão associada à resolução do condutivímetro (u_{res}). O respectivo Manual do Usuário informa uma resolução de 4 algarismos significativos (a escala digital de leitura modifica automaticamente de acordo com a condutividade da solução sendo analisada). Uma vez que, na rotina do NQBRP, esperam-se medidas em torno do valor da condutividade do padrão utilizado (146,9 µS cm^{-1}), considerou-se a resolução igual a 0,1 µS cm^{-1}; assim, considerando novamente distribuição retangular (equação 5):

$$u_{res} = \frac{0{,}1}{\sqrt{3}} = 0{,}05773\ \mu S\ cm^{-1}$$

O último termo da equação 1 está relacionado à incerteza padrão associada à precisão da medição (u_{pm}). Para exemplificar, realizaram-se 3 medições independentes de condutividade em uma amostra de água de abastecimento coletada no NQBRP ("água de torneira"), resultando nas medidas 113,3 , 113,8 e 113,7 µS cm^{-1}, resultando na média 113,6 µS cm^{-1} e no desvio padrão 0,2646 µS cm^{-1}. Nesse caso, u_{pm} será representado por esse desvio padrão e a aplicação da equação 1 resultará em:

$$u_{cond\ am} = \sqrt{(0{,}5141)^2 + (0{,}05773)^2 + (0{,}2646)^2}$$

$$u_{cond\ am} \cong 0{,}6\ \mu S\ cm^{-1}$$

Portanto, o resultado das 3 medições de condutividade para a amostra foi expresso como 113,6 ± 1,2 µS cm^{-1}, considerando k = 2 e nível de confiança 95,45%.

Para estimar a incerteza através da equação 1 de resultados com valores maiores de condutividade e que exigem a calibração com a solução padrão 0,1000 mol L^{-1} (condutividade = 12.890 µS cm^{-1})[2], inicialmente precisamos calcular u_{Ccal} (a incerteza padrão associada à condutividade dessa solução padrão). De forma análoga à transformação linear realizada anteriormente para a solução de KCl 0,001 mol L^{-1}, calculou-se um fator *b* de multiplicação (como indicado na equação 9) entre os valores de concentração e de condutividade:

$$0{,}1 \cdot b = 12.890$$

$$b = 128.900\ \mu S\ cm^{-1} mol^{-1} L$$

Multiplicando o valor de *b* pelo valor da incerteza da concentração da solução de KCl (u_{MKCl}), é possível expressar a incerteza padrão combinada da condutividade da solução 0,1000 de KCl (u_{CKCl}):

$$u_{C_{KCl}} = u_{M_{KCl}} \cdot b = 0{,}0003 \cdot 128.900$$

$$u_{C_{KCl}} \cong 40\ \mu S\ cm^{-1} = 0{,}040\ mS\ cm^{-1}$$

Aplicando a equação 5 para a incerteza associada à resolução do equipamento:

$$u_{res} = \frac{0{,}001}{\sqrt{3}} = 5{,}7735 \cdot 10^{-4}\ mS\ cm^{-1}$$

No cálculo do termo relacionado à precisão de medição da equação 1 (u_{pm}) foi considerado o desvio padrão de cinco medidas independentes de condutividade na amostra: 5,015; 5,016; 5,012; 5,027 e 5,025 mS cm^{-1}; o cálculo do desvio padrão resultou em 0,007 mS cm^{-1}, para uma média igual a 5,019 mS cm^{-1}. Aplicando a equação 1 novamente:

$$u_{cond\ am} = \sqrt{u_{C_{cal}}^2 + u_{res}^2 + u_{pm}^2} \quad (eq.\ 1)$$

$$u_{cond\ am} \cong 0{,}041\ mS\ cm^{-1} = 41\ \mu S\ cm^{-1}$$

Portanto, o resultado da medição de condutividade para essa amostra foi expresso como 5.019 ± 82 µS cm^{-1}, considerando k = 2 e nível de confiança 95,45%.

REFERÊNCIAS

1. Oliveira CCd, Kira CS, Trujillo LM, Carvalho MdFH, Caruso MSF, Silva SAd, Martins VAP. Incerteza de medição em ensaios físico-químicos: uma abordagem prática. São Paulo: SES-SP; 2015. 140 p.
2. APHA, AWWA, WEF. Standard Methods for the Examination of Water and Wastewater. 22nd ed. Rice EW, Baird RB, Eaton AD, Clesceri LS, editors. Washington DC: American Public Health Association, American Water Works Association, Water Environment Federation; 2012.
3. Dovidauskas S, Okada IA, Iha MH, Cavallini ÁG, Okada MM, Briganti RdC, Bergamini AMM, Oliveira MAd. Mapeamento da qualidade da água de abastecimento público no nordeste do Estado de São Paulo (Brasil). Vigil sanit debate. 2017;5(2):53-63.
4. Meija J, Coplen TB, Berglund M, Brand WA, Bièvre PD, Gröning M, Holden NE, Irrgeher J, Loss RD, Walczyk T, Prohaska T. Atomic weights of the elements 2013 (IUPAC Technical Report). Pure Appl Chem. 2016;88(3):265-91.

Breve discussão sobre a importância do pré-tratamento de dados na análise de componentes principais

Sergio DOVIDAUSKAS[1], Isaura Akemi OKADA[1]

[1]Núcleo de Ciências Químicas e Bromatológicas – Centro de Laboratório Regional – Instituto Adolfo Lutz de Ribeirão Preto VI

A análise de componentes principais (ACP) constitui-se em uma ferramenta básica e importante da análise multivariada de dados. É particularmente útil quando se dispõe de dados de muitas variáveis obtidas para um grande número de amostras. A partir de uma matriz de dados de n linhas (amostras) e m colunas (variáveis), procura-se por correlações significativas entre as variáveis de modo a substituir as variáveis originais por outras, as componentes principais – o objetivo é fazer que as correlações permitam que o conjunto de dados possa ser descrito com um menor número de variáveis, ou seja, duas ou três componentes principais. Dessa forma, será possível visualizar padrões e relações entre as amostras e entre as variáveis; em outras palavras, é uma análise exploratória de dados[1].

Uma vez construída a matriz de dados, pode ser necessário (e muito frequentemente o é) o pré-tratamento antes da análise multivariada propriamente dita. Esse pré-tratamento dos dados pode incluir: (i) uma transformação aplicada às linhas da matriz (amostras) como, por exemplo, utilizar técnicas de alisamento e de correção de linha base em espectros, e a normalização ou a mudança para logaritmo; (ii) um pré-processamento aplicado às colunas (variáveis), como a centralização dos dados na média, o escalamento pela variância, o autoescalamento e o escalamento pela amplitude, por exemplo[2]. É particularmente sobre esse pré-processamento aplicado às colunas que lida essa comunicação.

Assim, inicialmente é recomendável que se analise cada coluna (variável) no que diz respeito à distribuição dos dados, ou seja, se essa distribuição pode ser considerada normal ou não. A priori, se a distribuição não for normal, pré-processamentos paramétricos deveriam ser evitados, como a centralização pela média, o escalamento pela variância e o autoescalamento; esse último envolve subtrair a média dos valores de cada elemento da coluna, dividindo-se o resultado pelo respectivo desvio padrão – em outras palavras, para se obter o valor autoescalado (x_a) de uma determinada variável para uma dada amostra, subtrai-se do valor obtido para cada amostra (x) a média obtida para o conjunto de amostras (X_m), dividindo-se o resultado pelo desvio padrão de X_m (s_{Xm}), conforme indica a equação 1 (observe nessa equação a relação direta com a padronização em termos de escores z de uma distribuição normal[3]):

$$x_a = \frac{x - X_m}{s_{X_m}} \quad \textit{equação 1}$$

Não obstante, é importante avaliar o impacto que poderia ser observado nos resultados de uma ACP se pré-processamentos paramétricos fossem aplicados a dados com distribuições assimétricas (ou seja, não normais). Para ilustrar, tomemos como exemplo um estudo realizado em nosso laboratório[4] em que 88 municípios (linhas ou amostras) foram representados pelas respectivas séries em 12 variáveis (colunas). As variáveis consistiam nas medianas de medidas realizadas durante um ano nas águas de abastecimento público dos municípios, nos parâmetros pH, condutividade e concentrações de 10 íons (Li^+, Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, ClO_3^-, NO_3^-, PO_4^{3-} e SO_4^{2-}). Os 88 dados em cada coluna (x) foram centralizados pela mediana da coluna (X_{med}) e escalados pelo respectivo intervalo interquartil (ii), obtendo-se o valor centralizado e escalado x_{medii}, conforme indica a equação 2:

$$x_{medii} = \frac{x - X_{med}}{ii} \quad \textit{equação 2}$$

Esse pré-processamento foi escolhido devido às distribuições tipicamente assimétricas observadas nas variáveis consideradas. A **Figura 1A** traz o gráfico de escores da ACP realizada com o algoritmo NIPALS (*Non-linear Iterative Partial Least-Squares*) e usando-se 4 componentes principais; esse gráfico exibiu 4 grupos: um grupo constituído de apenas um município (Ibitinga); o segundo grupo (em azul) é composto por 4 municípios, enquanto o terceiro apresenta 8 municípios (cor magenta); o grupo mais numeroso (75 municípios, cor verde) situa-se próximo à origem de gráfico CP1/CP2 e foi denominado "grupo típico". O respectivo gráfico de pesos (**Figura 1B**) indica que o grupo em azul apresenta as concentrações de sódio e lítio como variáveis proeminentes, enquanto a concentração de sulfato é a variável mais importante para o grupo em magenta; o grupo em verde não apresenta variáveis proeminentes no modelo estudado (como indica a sua posição no gráfico de escores), enquanto Ibitinga apresentou variáveis físico-químicas incomuns (teores relativamente maiores de sulfato, cloreto, lítio e sódio, além de maiores valores de pH e condutividade), tendo sido estudado individualmente[4]. A formação desses mesmos grupos havia sido também observada quando se efetuou a análise hierárquica de agrupamentos pelo método Ward, utilizando-se os valores centralizados e escalados (x_{medii})[4].

Para comparação, consideremos agora representar cada um dos 88 municípios pela respectiva série de médias (e não medianas) nas mesmas 12 variáveis consideradas anteriormente; em adição, no pré-processamento utilizaremos o autoescalamento da equação 1 em lugar da centralização/escalamento da equação 2. A ACP nas mesmas condições (algoritmo NIPALS usando 4 componentes principais) resultará nos gráficos de escores e de pesos indicados nas **Figuras 1C** e **1D**, respectivamente. Observe-se inicialmente que, ao se passar de um modelo baseado em medianas para um modelo baseado em médias, a variância explicada para 2 componentes principais diminui de 75% (56% em CP1 19% em CP2) para 51% (30% em CP1 e 21% em CP2) – em outras palavras: a qualidade do modelo exibido nas **Figuras 1A** e **1B** (medianas) é melhor que a do modelo indicado nas **Figuras 1C** e **1D** (médias).

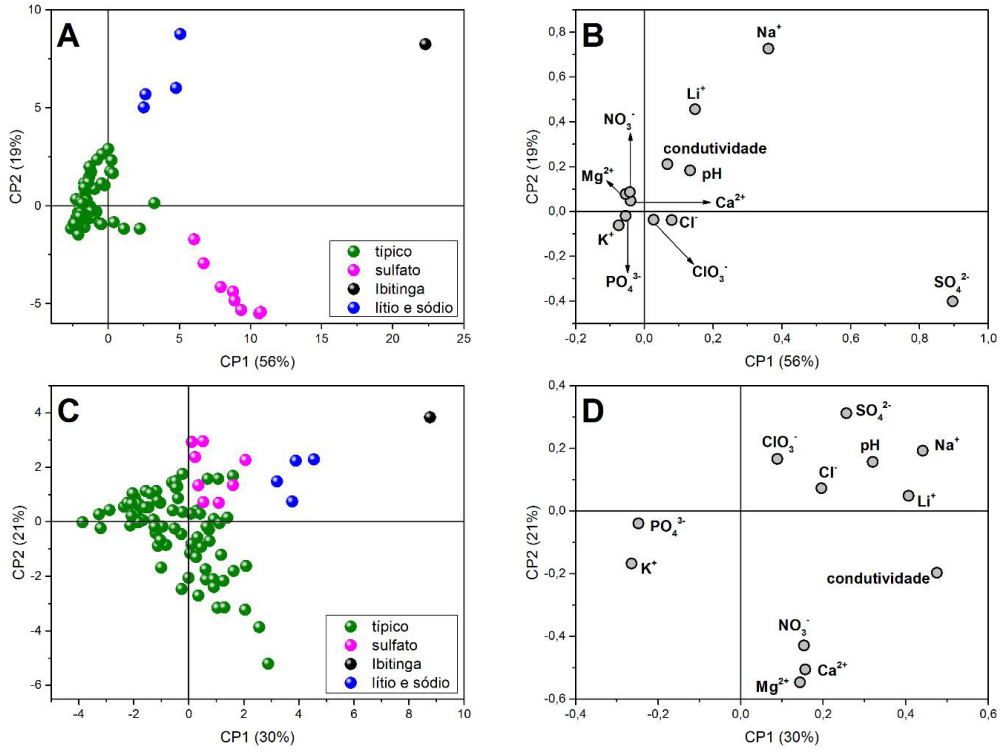

Fig 1. Análise de componentes principais de amostras de águas de abastecimento público de 88 municípios da região nordeste do Estado de São Paulo utilizando 12 variáveis: A/B, medianas; C/D, médias.

Essa queda na qualidade reflete na capacidade do modelo baseado em médias em produzir separações de grupos onde cada grupo inclua municípios com águas de abastecimento de perfis físico-químicos similares. Assim, observe-se que, se ainda é possível visualizar o grupo de apenas um município (Ibitinga), o "grupo do sulfato", claramente visualizado na **Figura 1A**, aparece na **Figura 1C** unido ao "grupo típico", ou seja, não há separação. Em adição, a distância entre o "grupo do cloreto" e o "grupo típico" é diminuída em relação ao modelo baseado em medianas. No que diz respeito à interpretação fornecida pelo gráfico de pesos (**Figura 1D**), observe-se que, se a interpretação para o município de Ibitinga não é muito prejudicada, no caso do grupo do cloreto não fica evidente que essa variável é a responsável pela formação do grupo em azul; situação pior é a do "grupo do sulfato": a análise simultânea dos gráficos das Figuras 1C e 1D não fornece indicações de que esse grupo exista.

Em conclusão: essa comunicação procurou demonstrar que, mesmo quando está se trabalhando com análises mais elaboradas, envolvendo muitas amostras e muitas variáveis, conhecimentos básicos sobre como lidar com os tipos de distribuição que se tem em mãos para as variáveis consideradas individualmente, pode ser a diferença entre a obtenção de um modelo interpretativo ou de um modelo que não fornece informação de qualidade.

AGRADECIMENTO

À Fundação de Amparo à Pesquisa do Estado de São Paulo pelo apoio financeiro (Processo FAPESP nº 2014/10034-2).

REFERÊNCIAS

1. Esbensen, K. H., *Multivariate Data Analysis - In Practice*. CAMO Process AS: Oslo, 2002; p 598.
2. Ferreira, M. M. C., *Quimiometria - Conceitos, Métodos e Aplicações*. Editora da Unicamp: Campinas (SP), 2015; p 495.
3. Moore, D. S.; McCabe, G. P., *Introdução à Prática da Estatística*. 3a ed.; LTC Editora: Rio de Janeiro, 2002; p 536.
4. Dovidauskas, S.; Okada, I. A.; Iha, M. H.; Cavallini, Á. G.; Okada, M. M.; Briganti, R. d. C., Parâmetros físico-químicos incomuns em água de abastecimento público de um município da região nordeste do Estado de São Paulo (Brasil). *Vigil. sanit. debate* **2017,** *5* (1), 106-115.

INSTITUTO ADOLFO LUTZ

CCD
COORDENADORIA DE
CONTROLE DE DOENÇAS

GOVERNO DO ESTADO
DE SÃO PAULO