

5

Determinação de azaarenos básicos em querosene de aviação por espectrofluorimetria e regressão por mínimos quadrados parciais

5.1.

Fluorescência

A luminescência molecular é a radiação eletromagnética (na região do ultravioleta-visível) emitida quando moléculas que foram previamente excitadas por fótons, sofrem desativação para estados de menor energia (em geral o estado fundamental). Esse processo envolve transição de elétrons de valência.

A desativação radiativa ocorre a partir de estados excitados moleculares que podem ser classificados como singleto ou tripleto. No estado excitado singleto, a direção do spin do elétron promovido para um orbital de maior energia é preservada. No estado excitado tripleto o spin do elétron promovido é invertido posteriormente à excitação [80]. O estado fundamental da maioria das substâncias orgânicas é singleto (S_0) e a absorção de luz promove elétrons da molécula para um dos vários estados excitados singletos ($S_1, S_2, \dots, S_n$), dependendo do comprimento de onda da radiação incidente. A população de moléculas excitadas tende a retornar para o estado fundamental e, para tal, vários processos podem ocorrer. Primeiramente, por meio de uma série de relaxamentos vibracionais (RV) e cruzamentos internos (CI), ambos processos não-radiativos, essa população é levada através dos níveis vibracionais de estados eletrônicos de mesma multiplicidade até atingir o primeiro nível vibracional do estado excitado singleto de menor energia (S_1). Esses processos ocorrem com muita rapidez (10^{-13} a 10^{-11} s) e, por causa disto, a desativação radiativa (fotoluminescência) só pode ocorrer a partir dos níveis vibracionais de menor energia dos estados eletrônicos excitados (regra de Kasha) [81].

A partir de S_1 , a população de moléculas pode seguir três caminhos distintos para retornar ao estado fundamental (Figura 75). Primeiro, se a diferença de energia entre S_1 e S_0 (estado fundamental) não for muito grande e existir possibilidade de sobreposição de níveis vibracionais, a população de moléculas pode ser levada a S_0 por CI e RV (sem a emissão de luz). No entanto, se a diferença energética entre S_1 e S_0 for relativamente grande a desativação

para o estado fundamental se dá com emissão de radiação na forma de fluorescência. A fluorescência ocorre na escala de tempo da ordem de nanosegundos (10^{-7} a 10^{-9} s).

Em algumas situações, a população de moléculas pode trocar de multiplicidade, passando do S_1 para o estado tripleto T_n . Essa transição ocorre em circunstâncias especiais onde há acoplamento spin-orbital relevante, alterando a direção relativa do spin dos elétrons em orbitais de maior energia. Esse processo é conhecido como cruzamento intersistema (CIS) e o resultado final é o acúmulo da população de moléculas no nível vibracional de menor energia de T_1 . A transição radiativa de T_1 para S_0 é chamada fosforescência e tem um tempo de vida longo que pode variar de 10^{-6} a 10^2 s, dependendo da estrutura da molécula e do meio onde se encontra. O espectro de emissão de fosforescência ocorre em maiores comprimentos de onda que os de fluorescência por causa da perda de energia que ocorre na conversão intersistema do estado singlete para o tripleto [81].

Como consequência de ser um processo mais rápido e que envolve estados excitados de mesma multiplicidade, a fluorescência é intrinsecamente um fenômeno luminescente mais comum que a fosforescência, competindo eficientemente com processos de desativação não-radiativos do estado excitado. Isso possibilita a fácil observação de fluorescência na temperatura ambiente e diretamente em soluções líquidas.

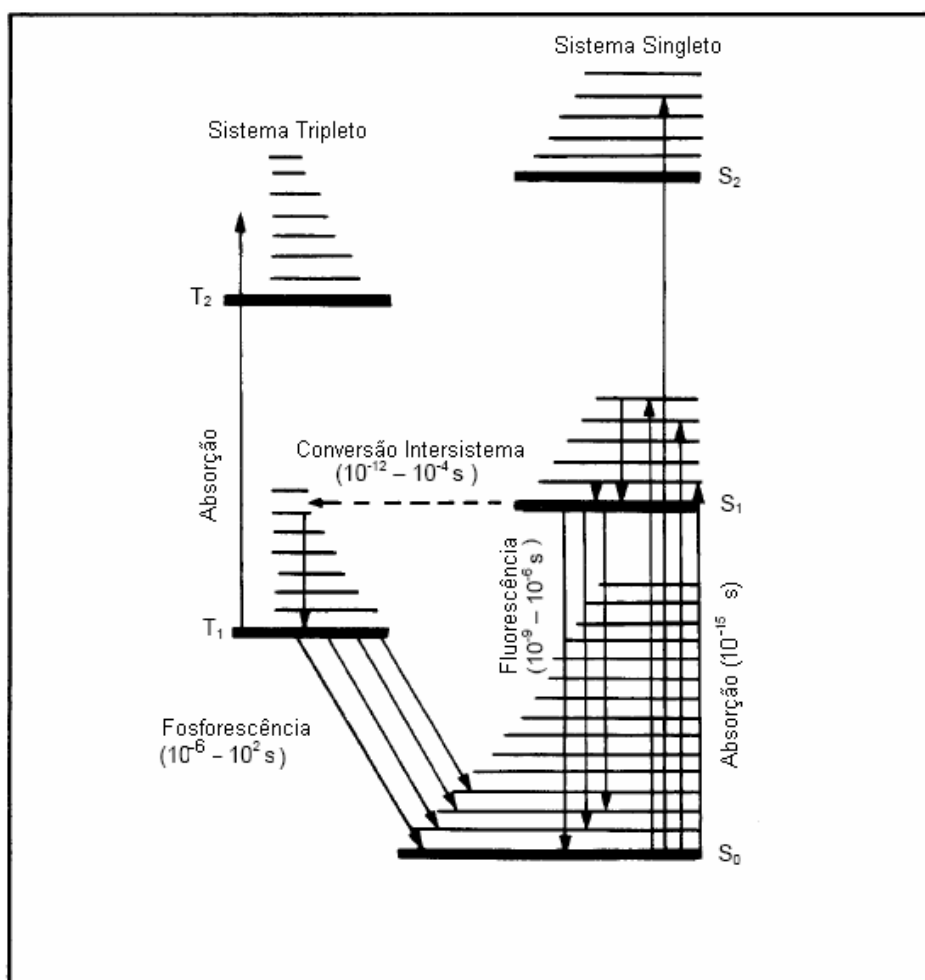


Figura 72 - Diagrama de energia simplificado de uma molécula poliatômica [81].

5.1.1. Fatores que afetam a fluorescência

A eficiência quântica luminescente é a razão entre o número de fótons emitidos por fluorescência (I_L) e o número de fótons absorvidos pela espécie química (I_A), como expresso na Equação 8. Se a eficiência quântica tiver magnitude entre 0,1 e 1, uma molécula será significativamente luminescente.

$$\frac{I_L}{I_A} = \phi_L$$

Equação 8

Para que ocorra a luminescência, uma molécula precisa ter estrutura apropriada e estar em um meio que favoreça a desativação radiativa, sendo estes dois fatores críticos na magnitude da eficiência quântica luminescente (ϕ_L) de uma substância.

Embora seja difícil prever teoricamente se uma molécula exibirá luminescência, é possível, de um modo geral, observar alguns requisitos. Primeiramente, moléculas com estruturas relativamente rígidas e ricas em elétrons π deslocalizados (como no caso das moléculas aromáticas), são potencialmente luminescentes. Estruturas moleculares rígidas têm menor capacidade de perder energia por processos não radiativos. Já os elétrons π em sistemas de ligações conjugadas podem ser facilmente promovidos para orbitais de maior energia com radiação nas regiões do ultravioleta próximo e do visível do espectro eletromagnético [82].

Outros fatores, tais como a temperatura, pH do meio, sistema de solventes e a presença de outras espécies, também afetam as características luminescentes de uma substância, alterando não somente as velocidades dos processos luminescentes e dos processos não-radiativos, mas também a natureza e a energia relativa do estado excitado de menor energia.

Em geral, o aumento da temperatura tem como consequência um aumento na eficiência dos processos de RV e CI na desativação do estado excitado. No entanto, por ser um fenômeno de tempo de vida relativamente curto, esse fator é menos crítico no caso da fluorescência, o que permite fácil observação do fenômeno na temperatura ambiente.

A natureza do sistema de solventes também é fator relevante, pois a sua viscosidade, polaridade e caráter prótico podem afetar significativamente a luminescência. A viscosidade pode diminuir a taxa de colisões bimoleculares desativadoras, pela diminuição da difusão de espécies atenuadoras no meio, em especial o oxigênio. No caso da fluorescência, a presença do oxigênio e de outras espécies atenuadoras não é tão crítica, pois a velocidade de difusão dos atenuadores é menor que a do processo fluorescente propriamente dito. No entanto, esse efeito pode se tornar crítico na medida em que a quantidade do atenuador no meio aumenta. Já a polaridade e o caráter prótico do solvente são importantes, pois afetam a energia relativa do estado excitado [82].

A influência do solvente no espectro eletrônico ocorre devido às interações eletrostáticas. As interações do solvente com as moléculas do soluto podem ser dipolo-dipolo, dipolo-dipolo induzido, dipolo induzido-dipolo induzido ou ligações

de hidrogênio, dependendo das estruturas moleculares do solvente e do soluto. As moléculas do solvente rapidamente se reorientam em torno da molécula luminescente logo após serem promovidas para o estado excitado e antes do retorno para o estado fundamental. Conseqüentemente, a energia relativa do estado excitado após a luminescência pode ser significativamente diferente do que era durante o processo de absorção. Assim, a mudança da polaridade do solvente ou do seu caráter prótico, acarreta em deslocamentos batocrômicos do espectro (em direção ao vermelho) ou deslocamentos hipsocrômicos (em direção ao azul) [83].

O efeito do pH nos sistemas de solventes próticos é relevante na luminescência de moléculas aromáticas contendo grupos funcionais básicos ou ácidos, sendo muito comum observar significativa diferença entre as propriedades luminescentes de moléculas protonadas e não-protonadas. Isso é decorrente dos efeitos de atenuação ou de desativação que os grupos substituintes têm sobre o sistema de elétrons deslocalizados. Assim, os rendimentos e o perfil do decaimento da fluorescência podem variar significativamente com o pH do meio.

5.1.2.

A técnica espectrofluorimétrica

A fluorescência é caracterizada por um procedimento experimental simples e que oferece algumas vantagens, tais como, instrumentação de custo relativamente baixo e excelentes faixas lineares dinâmicas [84]. Na maioria das determinações, a desoxigenação da amostra não é um fator crítico e as medições de fluorescência são rapidamente realizadas mesmo diretamente em solução em temperatura ambiente.

Técnicas luminescentes são relativamente baratas do ponto de vista de aquisição e manutenção do equipamento e de operação. As metodologias são simples e oferecem rapidez nas análises. Além disso, quando adequadamente otimizadas, as metodologias apresentam grande detectabilidade (limites de detecção no nível de ng mL^{-1}) e seletividade.

Métodos fluorimétricos convencionais utilizam a varredura seqüencial das bandas de excitação e de emissão (ou o contrário) com medição de uma em relação à outra usando o comprimento de onda máximo de cada banda. Essa abordagem para detecção de fluorescência, embora contando com a

seletividade da excitação relativa à emissão, pode ser afetada por espécies químicas concomitantes que possuem bandas de emissão e de excitação que parcialmente se sobrepõem as do analito de interesse. Assim, a similaridade e a sobreposição das bandas de fluorescência em temperatura ambiente geralmente interferem no processo de caracterização de componentes puros. Várias técnicas luminescentes têm sido desenvolvidas para melhorar a seletividade de medições fluorescentes, o que inclui a resolução temporal [85, 86], análise das matrizes de excitação-emissão (EEM) da intensidade fluorescente [87, 88], e o estreitamento das bandas fluorescentes com o uso do efeito Shpol'skii [24, 89]. Uma abordagem relativamente simples e rápida para a análise simultânea de misturas de múltiplos componentes é a técnica de varredura sincronizada, que consiste em variar simultaneamente os comprimentos de onda de excitação (λ_{exc}) e de emissão (λ_{em}) ao manter um intervalo constante entre eles ($\Delta\lambda = \lambda_{em} - \lambda_{exc}$). As vantagens da varredura sincronizada incluem a redução da largura da banda espectral, a simplificação do espectro de excitação, a redução da faixa espectral para um dado analito e minimização dos espalhamentos Rayleigh e Raman [84].

5.2.

Análise multivariada dos dados

Com a sofisticação crescente das técnicas instrumentais de análise tornam-se cada vez mais necessários tratamentos de dados mais complexos do ponto de vista matemático e estatístico, a fim de relacionar os sinais analíticos obtidos (intensidades espectrais, por exemplo) com os resultados desejados (concentrações das espécies químicas de interesse, na maioria dos casos).

Muita ênfase tem sido dada aos sistemas multivariados, nos quais muitas variáveis podem ser medidas simultaneamente, ao se analisar uma amostra qualquer. Nesses sistemas, a conversão da resposta instrumental no dado químico de interesse, requer a utilização de técnicas de estatística multivariada e álgebra matricial. Essas técnicas são apontadas, atualmente, como a melhor alternativa para a interpretação de dados e para a aquisição do máximo de informação sobre o sistema.

Desde o início da década de noventa, o número de publicações em química analítica envolvendo calibrações multivariadas vem aumentando acentuadamente e, neste contexto, as aplicações mais frequentes foram as

análises utilizando técnicas espectroscópicas. Isso ocorreu, provavelmente, em decorrência da versatilidade dessas técnicas e do fato de, em alguns casos, permitirem análises não-destrutivas. Recentemente, a aplicação de calibrações multivariadas associadas a técnicas espectroscópicas em análises químicas está ganhando popularidade, através de sua aplicação em análise de misturas de múltiplos componentes nas indústrias de alimentos [90-92], de petróleo [93, 94] e farmacêutica [95-97] e em amostras ambientais [98, 99].

Na maioria dos casos, a sobreposição espectral é um sério problema analítico e requer o uso de pré-tratamentos da amostra e procedimentos de separação das espécies químicas de interesse. Muitas vezes pode ser uma tarefa difícil encontrar uma região espectral seletiva para um determinado analito devido à grande sobreposição espectral comumente observada entre os componentes de uma mesma amostra, principalmente para amostras complexas como querosene de aviação. Conseqüentemente, construir uma calibração univariada, na maioria das vezes, não é possível. Em situações como essa, o uso de calibração multivariada permite a modelagem dos interferentes junto com o analito de interesse. Com isso, a idéia central consiste em substituir a separação física das espécies pela separação quimiométrica de seus sinais [96].

O método de calibração multivariada mais conhecido é a regressão por mínimos quadrados parciais (PLS, do inglês "Partial Least Squares"), que foi proposta inicialmente por Herman Wold [100, 101]. A regressão PLS é uma técnica de análise de dados multivariados utilizada para relacionar uma ou mais variáveis resposta (Y) com diversas variáveis independentes (X), baseada no uso de fatores. Usando como exemplo a espectrofluorimetria, a matriz X seria formada por valores de fluorescência em diversos comprimentos de onda e a matriz Y formada por valores de concentração dos analitos de interesse.

O PLS permite identificar fatores (combinações lineares das variáveis X) que melhor modelam as variáveis dependentes Y. Além disso, admite, com eficiência, trabalhar com conjuntos de dados onde haja variáveis altamente correlacionadas e que apresentam ruído aleatório considerável.

Diversos modelos relacionando propriedades de amostras com dados espectrais já foram desenvolvidos utilizando calibração multivariada, sendo a regressão por PLS uma das ferramentas mais utilizadas. Alguns bons artigos de revisão e tutoriais já foram publicados descrevendo a regressão por PLS e a sua aplicação em química analítica [102, 103]. Um exemplo dessa aplicação pode ser encontrado em DA LUZ [104], que usou PLS para estimar propriedades de gasolinas a partir de espectros de absorção na região do infravermelho.

No texto a seguir, são apresentadas as bases para entender como funciona o PLS e como essa calibração pode ser aplicada para estimar propriedades de amostras a partir de medidas espectrais. É importante ressaltar que no exemplo citado as respostas são obtidas em termos de concentração do analito. No entanto, a resposta pode ser qualquer propriedade física ou química da amostra que se deseja estimar.

5.2.1.

Organização dos dados

Em problemas como os de calibração multivariada, onde o número de objetos e de variáveis é muito grande, torna-se absolutamente indispensável a disposição ordenada dos dados em forma de matriz para tornar mais fácil a sua manipulação.

Dados multivariados são, em geral, organizados em matrizes através de vetores em linha ou coluna, de acordo com a convenção adotada. Além disso, os valores relativos às variáveis independentes (espectros das amostras) e às variáveis dependentes (concentração dos analitos de interesse) são organizados separadamente em duas matrizes.

Na matriz com os dados espectrais (absorvância, intensidade de emissão, etc.), que chamaremos aqui de matriz *fluorescência* para facilitar a explicação, cada espectro é representado como um vetor linha.

F_{11}	F_{12}	F_{13}	F_{14}	...	F_{1w}
F_{21}	F_{22}	F_{23}	F_{24}	...	F_{2w}
F_{31}	F_{32}	F_{33}	F_{34}	...	F_{3w}
...	...	...	...	...	...
F_{s1}	F_{s2}	F_{s3}	F_{s4}	...	F_{sw}

F_{sw} representa a fluorescência da amostra **s** no comprimento de onda **w**, resultando numa matriz com o número de linhas correspondente ao número de amostras e o de colunas ao número de comprimentos de onda.

Já na matriz *concentração*, os valores de concentração dos componentes para cada amostra são representados como vetores coluna. Dessa forma, cada amostra ocupa uma linha da matriz e cada componente uma coluna.

$$\begin{array}{cccc}
 C_{11} & C_{12} & \dots & C_{1c} \\
 C_{21} & C_{22} & \dots & C_{2c} \\
 C_{31} & C_{32} & \dots & C_{3c} \\
 \dots & \dots & \dots & \dots \\
 C_{s1} & C_{s2} & \dots & C_{sc}
 \end{array}$$

C_{sc} representa a concentração do componente **c** na amostra **s**, resultando numa matriz com o número de linhas correspondente ao número de amostras e o de colunas ao de componentes.

Essas matrizes de dados são organizadas em pares de modo que cada matriz *fluorescência* possua uma matriz *concentração* correspondente. Um par de matrizes forma um conjunto de dados, que pode receber diferentes nomes de acordo com sua origem e utilidade.

Um conjunto de dados denominado *conjunto treinamento* ou *calibração* contém medidas de amostras conhecidas e é utilizado para desenvolver a calibração. No exemplo em questão, o *conjunto treinamento* consiste de uma matriz *fluorescência* contendo os espectros obtidos e de uma matriz *concentração* contendo valores determinados por um método de referência confiável e independente.

Para que uma calibração seja válida o *conjunto treinamento* utilizado para construí-la deve conter dados que sejam representativos das amostras reais a serem analisadas. Além disso, como o PLS é uma técnica multivariada, é muito importante que as amostras no *conjunto treinamento* sejam mutuamente independentes [105], ou seja, que a correlação entre os valores de concentração dos analitos seja próxima de zero.

Em termos práticos, isso significa que um *conjunto treinamento* deve: (i) conter todos os componentes esperados; (ii) abranger a faixa de concentração de interesse; (iii) abranger as condições de interesse (temperatura, pH, etc.); (iv) conter amostras mutuamente independentes.

De todos os pré-requisitos, a independência mútua costuma ser a mais difícil de assegurar, principalmente porque diluições ou adições sucessivas não

podem ser utilizadas para o preparo das amostras. Apesar de padrões assim obtidos serem perfeitamente aplicáveis a calibrações univariadas, eles não se aplicam a técnicas multivariadas, pois produzem um sistema inconsistente de equações lineares. O problema é que, com tal procedimento, as concentrações relativas dos vários componentes na amostra não variam e, conseqüentemente, os erros relativos entre as concentrações dos vários componentes também não. As únicas fontes de variação do erro seriam os erros de diluição e o ruído instrumental. As diluições sucessivas conduzem a valores múltiplos de concentração, que levam a indeterminação na resolução das equações simultâneas [105].

Outro conjunto de dados muito importante é o *conjunto validação*, que contém medidas de amostras conhecidas que sejam independentes das amostras usadas no *conjunto treinamento*, utilizado para avaliar o desempenho da calibração. Tratam-se as amostras de validação como se seus valores de concentração não fossem conhecidos e utiliza-se a calibração construída com o *conjunto treinamento* para estimá-los. Comparam-se, então os valores estimados com os valores teóricos (determinados pelo método de referência) para avaliar o desempenho da calibração em amostras realmente desconhecidas.

Há, finalmente, o *conjunto de amostras desconhecidas* ou *amostras-teste* que contém apenas a matriz *fluorescência*. A calibração construída é utilizada para calcular a matriz resultado que contém os valores de concentração preditos.

5.2.2. Princípios de uma técnica multivariada

A calibração multivariada tem como princípio a utilização simultânea de muitas variáveis independentes $x_1, x_2, \dots, x_n$ (por exemplo, valores de fluorescência em vários comprimentos de onda), para quantificar alguma variável dependente y (no caso, a concentração dos analitos).

Quando se trabalha com muitas variáveis, alguns pontos devem ser levados em conta para a obtenção de dados com qualidade e sem redundância de informação [105], entre eles: (i) o número de amostras no conjunto treinamento deve ser igual a pelo menos três vezes o número de componentes presentes na amostra. Ou, no mínimo, igual a três vezes o número de componentes que se deseja estimar; (ii) calibrações satisfatórias são obtidas, em geral, a partir de valores de concentração determinados por métodos de

referência com erro relativo em relação à média inferior a $\pm 5\%$; (iii) para o número de amostras no conjunto validação (quando houver) em geral utiliza-se um número igual a 30% do total de amostras de calibração e validação; (iv) o nível de ruído nos espectros deve ser sempre avaliado para não interferir nos resultados da análise.

Com os espectros e os valores de referência obtidos, a etapa de construção do modelo de calibração é geralmente a mais rápida de todo o processo quando se têm a disposição programas computacionais adequados. É nessa etapa que algumas escolhas quanto ao pré-tratamento dos dados e aos parâmetros utilizados na construção de modelo PLS (a serem discutidos mais tarde) devem ser feitas.

O modelo obtido é testado na etapa de validação, calculando-se o erro entre os valores de concentração teóricos (fornecidos pelo método de referência) e estimados para as amostras de validação. Esse cálculo indica o erro que se pode esperar ao utilizar-se a calibração para estimar a concentração de amostras reais desconhecidas.

A aplicabilidade do modelo obtido deve sempre ser avaliada à medida que novas amostras são analisadas, pois a representatividade inicial das amostras treinamento pode não se manter. Uma maneira de garantir a aplicabilidade do modelo é analisar, a intervalos de tempo apropriados, uma amostra de referência. Em geral, como o desempenho dos instrumentos muda e os sistemas de amostras envelhecem, verifica-se uma deterioração gradual no desempenho da calibração inicial. Uma atualização periódica no conjunto treinamento pode promover a correção da resposta, evitando a deterioração dos resultados das análises.

5.2.3.

Análise dos Componentes Principais (PCA)

A análise dos componentes principais (PCA, do inglês *Principal Component Analysis*) é uma técnica estatística poderosa que pode ser utilizada para redução do número de variáveis e para fornecer uma visão estatisticamente privilegiada do conjunto de dados. A PCA consiste em reescrever as variáveis originais em novas variáveis, denominadas componentes principais (CP), através de uma transformação de coordenadas. Além disso, fornece as ferramentas adequadas para identificar as variáveis mais importantes no espaço dos CP.

Cada CP é uma combinação linear de todas as variáveis originais. Por exemplo, um sistema com oito variáveis, após a transformação, terá oito CP. Cada um destes CP, por sua vez, será escrito como uma combinação linear das oito variáveis originais. Nestas combinações, cada variável terá uma importância ou peso diferente.

Duas são as características dos componentes principais que os tornam mais efetivos que as variáveis originais para a análise do conjunto das amostras [106]. Primeiro, as variáveis podem guardar entre si correlações que são suprimidas nos CP, ou seja, os CP são ortogonais entre si. Desse modo, cada CP traz uma informação estatística diferente dos demais. A segunda característica importante é decorrente do processo matemático de geração de cada componente que maximiza a informação estatística para cada uma das coordenadas que estão sendo criadas. As variáveis originais têm a mesma importância estatística, enquanto que os componentes principais têm importância estatística decrescente. Ou seja, os primeiros CP são tão mais importantes que, na maioria dos casos, podemos desprezar os demais.

Matematicamente falando, a análise dos componentes principais é um método para decompor uma matriz de dados X de posto r (ou "rank" r), como uma soma de matrizes de posto igual a 1, onde posto é um número que expressa a dimensão de uma matriz. Para compreender melhor a descrição matemática da PCA, algumas definições devem ser feitas.

5.2.3.1.

Definições importantes em PCA

O posto de uma matriz é o número de vetores linearmente independentes que compõem uma matriz, ou seja, são os vetores que não podem ser escritos como uma combinação linear de outros vetores que pertençam ao mesmo espaço vetorial. A interpretação química para o posto é o número de espécies químicas distintas contidas nas amostras, desprezando os ruídos aleatórios inerentes às medidas [107].

Quando um operador, representado por uma matriz, é aplicado a um espaço vetorial e o produto dessa operação retorna o próprio espaço vetorial multiplicado por uma constante, tem-se uma equação de autovetores e de autovalores [108].

$$\mathbf{M}\Psi = \Lambda\Psi$$

Equação 9

Onde a matriz $\mathbf{M}$ é o operador que é aplicado no espaço vetorial formado pelos vetores da matriz Ψ resultado em uma constante, Λ , multiplicada pelo próprio espaço vetorial Ψ .

O espaço de fatores nada mais é que um sistema de coordenadas particular que oferece certas vantagens para técnicas multivariadas. Quando se trabalha em um espaço de fatores, ao invés do espaço formado pelos dados originais, faz-se simplesmente uma troca do sistema de coordenadas empregado, sem qualquer modificação nos dados em si.

Há várias razões para o uso de um sistema de coordenadas formado por um espaço de fatores apropriado, ao invés das coordenadas originais [105]: (i) eliminação de problemas causados por dados altamente colineares como um conjunto de espectros muito semelhantes; (ii) remoção de ruído dos dados de forma mais eficiente; (iii) o espaço de fatores pode elucidar quais variáveis x apresentam maior correlação com as variáveis y , quantos componentes estão realmente presentes, ou quais amostras são semelhantes ou diferentes entre si; (iv) redução da dimensionalidade dos dados.

A utilização dos componentes principais (autovetores) para definir um espaço de fatores que englobe os dados, não modifica os dados em si, mas simplesmente encontra um sistema de coordenadas mais conveniente, capaz de remover ruído dos dados sem distorcê-los e de diminuir sua dimensionalidade sem comprometer seu conteúdo de informações.

Cada componente principal tem um autovalor associado a ele. Esse autovalor é igual à soma dos quadrados das projeções (“scores”) dos dados sobre o fator correspondente, que nada mais é que a medida da variância total capturada pelo autovetor. Como cada fator captura o máximo de variância possível, ao fator seguinte resta a variância residual, que se torna cada vez menor a cada fator sucessivo. Conseqüentemente, cada autovalor terá um valor menor que o de seu antecessor.

5.2.3.2.

Descrição matemática da PCA

Para descrever matematicamente a análise dos componentes principais [102], considere que n amostras tiveram seus espectros adquiridos em m comprimentos de onda. Essas informações podem ser arranjadas na forma de uma matriz *fluorescência* $\mathbf{X}$ de dimensões $n \times m$. A PCA é um método de decomposição de uma matriz $\mathbf{X}$ de posto r em um somatório de r matrizes de posto 1, onde posto é o número que expressa a dimensão de uma matriz.

As novas matrizes de posto 1 podem ser escritas como produtos dos vetores chamados “scores” (t_h) e “loadings” (p'_h), calculados par a par, como na Equação 10.

$$\mathbf{X} = t_1 p'_1 + t_2 p'_2 + \dots + t_h p'_h \quad \text{Equação 10}$$

A Figura 73 apresenta a matriz $\mathbf{X}$ decomposta em produtos de matrizes “scores” e “loadings”.

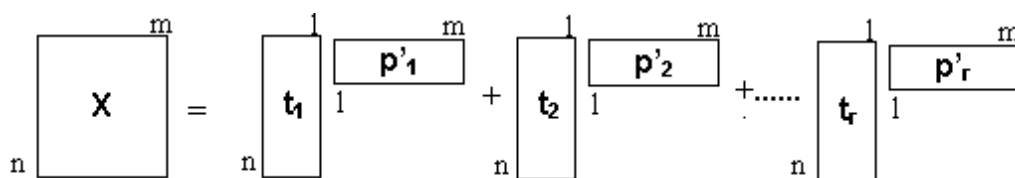


Figura 73 - Representação da matriz de dados $\mathbf{X}$ decomposta em produtos de matrizes de posto igual a 1 [102].

Para ilustrar o significado de t_h e p'_h , na Figura 74 são mostrados, no plano bidimensional, duas variáveis x_1 e x_2 . Na Figura 74A, é mostrado um componente principal que é a reta que aponta para a direção de maior variabilidade das amostras da Figura 74B. Os “scores” t_h são as projeções das amostras na direção do componente principal e os “loadings” p'_h são os cossenos dos ângulos formados entre a componente principal e cada variável.

Em síntese, como já mencionado, a análise dos componentes principais é um método que tem por finalidade básica a redução de dados a partir de combinações lineares das variáveis originais.

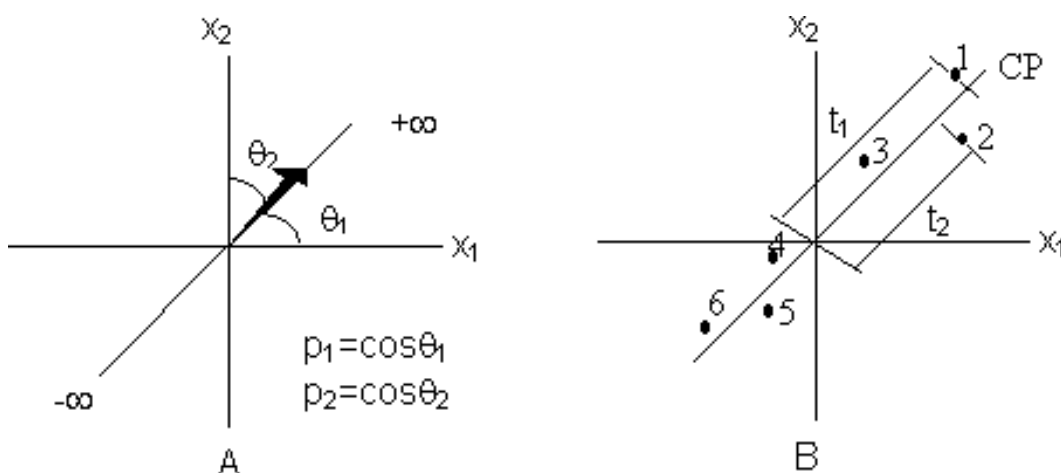


Figura 74 - Um componente principal no caso de duas variáveis: (A) *loadings* são os cosenos dos ângulos do vetor direção; (B) *scores* são as projeções das amostras 1 a 6 na direção do componente principal [102].

5.2.4.

A Regressão por mínimos quadrados parciais

Toda calibração multivariada utiliza modelos matemáticos para estabelecer uma relação entre uma propriedade que possa ser monitorada com alguma outra propriedade de interesse. O método dos mínimos quadrados parciais é um modelo baseado em variáveis latentes (componentes principais ou fatores), onde cada fator é definido como uma combinação linear das variáveis originais das matrizes **X** (variáveis independentes) ou **Y** (variáveis dependentes) [102].

O primeiro componente principal correspondente ao maior autovalor é, por definição, a direção no espaço de **X** que descreve a máxima quantidade de variância das amostras. Quando toda a variância de um conjunto de amostras não puder ser explicada por apenas um componente principal, um segundo componente principal perpendicular ou ortogonal ao primeiro será utilizado, e assim por diante. Após a modelagem, teoricamente, a matriz dos quadrados dos resíduos deverá conter apenas a variância não explicada associada ao ruído.

A importância da ortogonalidade dos componentes principais se dá pelo fato de que somente desta forma pode-se garantir que a nova base formada

resulta de uma combinação de vetores linearmente independentes e, portanto, constituindo um novo espaço vetorial.

A regressão por mínimos quadrados parciais implica em encontrar um conjunto de vetores base (componentes principais) para os dados espectrais e um conjunto separado de vetores base para os dados de concentração e, em seguida, relacioná-los um com o outro. A relação básica entre esses dois conjuntos de vetores é apresentada na Equação 11 [105],

$$Y_f = B_f * X_f \quad \text{Equação 11}$$

onde, Y_f é a projeção dos dados de concentração sobre o “f-ésimo” fator de concentração.

X_f é a projeção dos dados espectrais correspondentes sobre o “f-ésimo” fator espectral.

B_f é a constante de proporcionalidade para o “f-ésimo” par de fatores concentração e espectral.

A idéia geral do PLS é tentar alcançar, tanto quanto possível, a congruência ótima entre cada fator espectral e seu fator concentração correspondente, ou seja, encontrar uma relação perfeitamente linear entre as projeções (*scores*) dos dados espectrais e de concentração sobre os seus respectivos fatores.

No entanto, como o ruído dos dados espectrais é independente do ruído dos dados de concentração, a obtenção daquela relação perfeitamente linear não é possível. A melhor maneira, então, de alcançar uma congruência ótima é utilizar o conceito dos mínimos quadrados. Para isso, os fatores espectrais e de concentração correspondentes sofrem uma rotação até que o ângulo entre eles seja zero [105]. Em outras palavras, o PLS procura por um único vetor, **W**, que represente o melhor compromisso entre os fatores espectral e concentração, ou seja, que maximize a relação linear entre as projeções dos dados espectrais sobre o vetor **W** e as projeções dos dados de concentração correspondentes sobre o mesmo vetor. Cada vetor **W** terá tantos elementos quantos forem os comprimentos de onda nos espectros e, embora **W** seja de fato um fator

abstrato, normalmente seus elementos são chamados de pesos (*loading weights*).

Os fatores **W** são obtidos um a um. Após o primeiro fator **W**₁ ser encontrado, a porção da variância dos dados espectrais capturada por ele é removida dos espectros. Do mesmo modo, a porção da variância dos dados de concentração capturada por **W**₁ é removida. Logo, o próximo fator, **W**₂, é encontrado para os resíduos espectrais e de concentração que não foram capturados por **W**₁. Esse processo continua até que todos os possíveis fatores sejam encontrados.

As projeções dos vetores **W** sobre o plano contendo os dados espectrais são chamadas de cargas espectrais (*spectral loadings*), geralmente designados como variável **P**. Do mesmo modo, as projeções dos vetores **W** sobre o plano contendo os dados de concentração são chamadas de cargas de concentração (*concentration loadings*), designados como variável **Q**.

No caso de a variância espectral ser linearmente correlacionada com a variância dos dados de concentração, os fatores **W** do PLS, e suas correspondentes cargas espectrais, **P**, serão muito semelhantes entre si e também tenderão a ser muito semelhantes aos componentes principais. Assim sendo, no PLS as matrizes X e Y são decompostas simultaneamente em uma soma de h variáveis latentes [102], como nas Equação 12 e Equação 13,

$$X = TP' + E = \sum t_h p'_h + E \quad \text{Equação 12}$$

$$Y = UQ' + F = \sum u_h q'_h + F \quad \text{Equação 13}$$

onde T e U são as matrizes de “scores” das matrizes X e Y, respectivamente; P' e Q' são as matrizes dos “loadings” das matrizes X e Y, respectivamente; e E e F são os resíduos. A correlação entre os dois blocos X e Y é simplesmente uma relação linear obtida pelo coeficiente de regressão linear, tal como descrito na Equação 14, para h variáveis latentes, sendo que os valores de b_h são agrupados na matriz diagonal B, que contém os coeficientes de regressão entre a matriz de “scores” U de Y e a matriz de “scores” T de X. Como já foi mencionado, a melhor relação linear possível entre os “scores” desses dois

blocos é obtida através de pequenas rotações das variáveis latentes dos blocos de X e Y.

A matriz Y pode ser calculada de u_h , através da Equação 15 e a concentração de novas amostras prevista a partir dos novos “scores”, T^* , substituídos na Equação 14.

$$u_h = b_h t_h \quad \text{Equação 14}$$

$$Y = TBQ' + F \quad \text{Equação 15}$$

$$Y = T^*BQ' \quad \text{Equação 16}$$

Nesse processo, um passo crítico é o de estabelecer o número correto de componentes principais a serem utilizados nos modelos de calibração, já que os valores preditos para as propriedades das amostras, calculados a partir desses modelos, dependem diretamente do número de componentes principais utilizados. Poucos fatores podem não ser suficientes para modelar adequadamente o sistema, enquanto muitos fatores podem introduzir ruído à calibração, o que resulta num baixo poder de predição para amostras fora do conjunto calibração [109].

A maioria dos programas para o cálculo de regressões por PLS disponíveis fornece dados para a seleção do número ótimo de componentes principais, construindo o gráfico do erro médio quadrático da predição (RMSEP, do inglês “*root mean square error of prediction*”) versus o número de componentes principais utilizado. O RMSEP é calculado segundo a Equação 17, onde y_i é o valor de referência da concentração, $\hat{y}_i$ é o valor previsto pelo modelo e n é o número de amostras previstas. O número de componentes selecionado é, em geral, aquele que fornece um erro de predição mínimo.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{Equação 17}$$

O cálculo do erro da predição pode ser feito através de um conjunto de amostras independente da calibração, o conjunto validação, ou através de validação cruzada (RMSECV). Na validação cruzada, as mesmas amostras são usadas tanto para construir o modelo quanto para testá-lo. Esse método de validação consiste em deixar algumas amostras de calibração de fora da construção do modelo e então utilizá-las para predição e cálculo dos resíduos. O processo é repetido com outro subconjunto de amostras de calibração até que todas as amostras tenham sido utilizadas para predição. No passo seguinte, todos os resíduos são combinados para computar a variância residual da validação e o valor do RMSEP e uma calibração final é então calculada com todas as amostras. A validação cruzada completa (“full cross validation”, FCV) deixa de fora uma única amostra de cada vez.

5.2.5.

Pré-tratamentos opcionais dos dados

Antes de desenvolver qualquer modelo de calibração multivariada recomenda-se um pré-tratamento dos dados para eliminar amostras anômalas, minimizar ruídos e informações superpostas das espécies de interesse, bem como das espécies interferentes.

Há diversas possíveis maneiras de tratar os dados antes de encontrar os componentes principais e realizar a regressão. As mais utilizadas são:

- i. *Remoção de artefatos ou linearização.* Em espectroscopia, a forma mais comum de remover artefatos é a correção de linha base. Apesar disso, esse tipo de correção deve ser feito com critério para que outros artefatos não sejam introduzidos nos dados ao invés da remoção dos já existentes. Para minimizar o ruído (aplicar “*smooth*”) pode-se utilizar um filtro com média móvel, transformada de Fourier, transformada de Wavelet, Savitsky-Golay, entre outros. A derivada primeira e a derivada segunda são frequentemente utilizadas para melhorar a definição de picos que se encontram sobrepostos em uma mesma região e para correção de linha base.
- ii. *A centralização dos dados em torno da média* é simplesmente a subtração do sinal analítico médio (no presente caso, fluorescência) em cada comprimento de onda, de cada espectro no conjunto de dados [105]. Do

ponto de vista estatístico, a centralização tem como objetivo prevenir que os pontos mais distantes do centro dos dados tenham maior influência que os mais próximos. Dependendo do tipo dos dados e da sua aplicação, a centralização pode ter efeito positivo, negativo ou neutro no desempenho da calibração. Após centrar os dados, a matriz terá valores positivos e negativos, e o novo valor de fluorescência média para todas as amostras em cada comprimento de onda será igual a zero. Da mesma maneira, a matriz concentração também pode ser centralizada em torno da média, componente a componente. A decisão de centrar ou não os dados da matriz fluorescência é independente da mesma decisão para a matriz concentração.

Para ilustrar a centralização em torno da média, na Figura 75A são mostrados os dados antes da centralização e na Figura 75B são mostrados os dados após a centralização. Pode-se verificar que as posições relativas dos dados não foram alteradas, havendo apenas o deslocamento da origem do novo sistema de coordenadas para o centróide dos dados.

- iii. *Escalonar ou ponderar os dados* implica em multiplicar todos os espectros por um diferente fator de escala para cada comprimento de onda, de modo a aumentar ou diminuir a influência sobre a calibração de cada comprimento de onda particular [105]. Um dos tipos mais comuns de escalonamento é o de variância (*variance scaling*), muitas vezes chamado de **padronização**.

Aplicar a padronização significa ajustar o conjunto de dados de modo a igualar a variância de cada variável, ou seja, igualar a influência de cada variável sobre o conjunto de dados. Para isso, a influência de variáveis que contenham informações analíticas úteis pode diminuir enquanto que a de variáveis que contenham principalmente ruído pode aumentar. Fica claro, portanto, que a padronização não deve ser feita a menos que haja uma razão específica para isto. No caso da matriz concentração, por exemplo, a variância de cada componente é ajustada para a unidade dividindo-se o valor de concentração de cada amostra pelo desvio padrão para este componente.

Ao contrário de centrar os dados, padronizá-los modifica as posições relativas dos dados, mas não modifica a posição do centróide do conjunto de dados, como mostrado na Figura 75C. Por outro lado, a decisão de padronizar ou não a matriz concentração também é independente da decisão de padronizar ou não a matriz fluorescência. Uma ilustração do

efeito da padronização aplicada junto com a centralização sobre os dados é mostrada na Figura 75D.

- iv. *Seleção de variáveis.* Esse pré-tratamento permite eliminar os termos que não são relevantes na modelagem, gerando uma submatriz com apenas as variáveis que possuem informação. No presente trabalho, a seleção de variáveis importantes foi feita com auxílio do programa estatístico utilizado ou simplesmente através do conhecimento do comportamento espectral dos respectivos analitos.

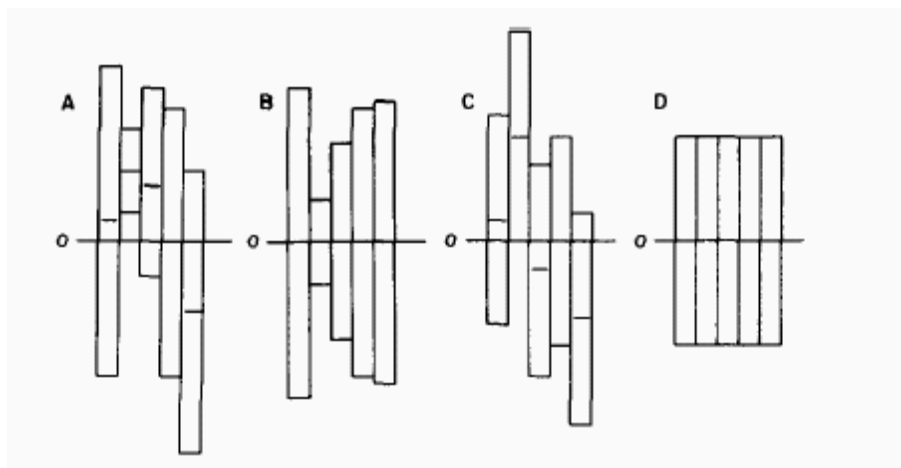


Figura 75 - Pré-processamento dos dados. Os dados para cada variável estão representados por uma barra de variância e seu centro. (A) A maioria dos dados sem tratamento apresentam esse tipo de variação. (B) O resultado após somente a centralização em torno da média. (C) O resultado após somente a padronização. (D) O resultado após centrar e padronizar os dados [102].

5.2.6. Detecção de amostras anômalas

Uma amostra considerada anômala, também chamada “*outlier*”, apresenta-se tão diferente das demais que pode não ser bem descrita pelo modelo ou influenciá-lo demais. Assim sendo, pelo menos um dos componentes principais do modelo pode concentrar-se apenas em descrever como essa amostra é diferente das outras, mesmo que isso seja irrelevante para modelar as características mais importantes presentes nas demais amostras. Usando a análise dos componentes principais, as amostras anômalas detectadas visualmente podem ser confirmadas através do gráfico dos “*scores*” e dos valores dos resíduos e “*leverages*”, em geral combinados num gráfico de influência [110].

- i. O gráfico dos “*scores*” mostra o perfil das amostras em relação a um, dois ou três componentes principais. Com esse gráfico é fácil identificar uma amostra que se situe longe das demais, indicando a possibilidade de ser considerada anômala.
- ii. Os resíduos são uma medida do quanto as amostras ajustam-se ao modelo. Uma amostra com resíduo alto é pobremente descrita pelo modelo, que descreve bem as demais amostras. Esse tipo de amostra é considerado como um elemento estranho em relação ao conjunto de amostras bem descritas pelo modelo, podendo ser considerada uma amostra anômala.
- iii. “*Leverage*” é uma medida de quanto um ponto de um conjunto de dados é extremo, comparado com a maioria. Em PCA ou PLS, “*leverage*” pode ser interpretado como a distância entre o ponto projetado e o centro do modelo. Pontos com alto “*leverage*” têm uma alta influência sobre o modelo e podem ou não ser amostras anômalas. Uma amostra com alto “*leverage*” e com alto resíduo é considerada um caso mais crítico e pode ser detectada usando-se o gráfico de influência.

5.3.

Parâmetros analíticos de mérito

5.3.1.

Sinal analítico líquido

Os métodos analíticos baseados em calibração multivariada são relativamente recentes e sua aceitação e implantação ainda são restritas [96] principalmente devido às exigências de validação pelas agências e organismos reguladores. Essa validação pode ser acessada pela determinação de parâmetros que são conhecidos como figuras de mérito, tais como exatidão, precisão, sensibilidade, e limite de detecção. Na estimativa dos parâmetros analíticos de mérito em calibração multivariada, a parte do sinal que se relaciona unicamente com o analito de interesse é mais importante que o sinal total e é denominada sinal analítico líquido (NAS, do inglês "Net Analyte Signal") [111]. O NAS é definido como a parte do sinal analítico que é ortogonal ao sinal dos interferentes presentes na amostra [112]. A representação geométrica da propriedade de ortogonalidade do NAS [113] é mostrada na Figura 76.

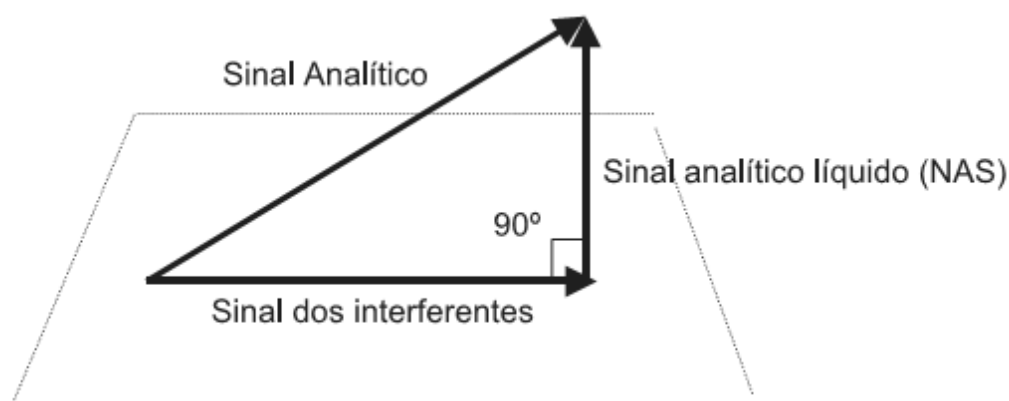


Figura 76 - Representação geométrica do vetor de sinal analítico [97].

O vetor NAS contém os valores para cada amostra e pode ser relacionado com o vetor dos coeficientes de regressão, **b** (Equação 14), em modelos de calibração inversa, tais como o PLS, através da Equação 18, na qual $\|b\|$ corresponde a raiz quadrada da soma dos quadrados de cada elemento no vetor **b** [111],

$$\|NAS\| = \frac{1}{\|b\|} \quad \text{Equação 18}$$

onde o símbolo " $\| \cdot \|$ " indica a norma euclidiana de um vetor. Vários parâmetros analíticos de mérito podem ser calculados como função do NAS e, felizmente, para calculá-los, o NAS não precisa ser explicitamente determinado, mas apenas o escalar associado a ele calculado através da Equação 18 [111].

5.3.2. Exatidão

A exatidão, que expressa o grau de concordância entre o valor estimado e o valor tido como verdadeiro ou de referência, é estimada comumente em aplicações com calibração multivariada, através do RMSEP e do erro padrão relativo da previsão (RSEP, do inglês "Relative Standard Error of Prediction"), calculados de acordo com as Equação 17 e Equação 19, respectivamente.

$$RSEP(\%) = 100 \times \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum (y_i)^2}} \quad \text{Equação 19}$$

No presente trabalho, a exatidão foi determinada com base em 20 soluções-padrão utilizadas como conjunto validação, no qual os valores de concentração dos analitos foram distribuídos em oito níveis, variando entre 13 e 95 $\mu\text{g L}^{-1}$. Para fins de comparação, o mesmo erro padrão relativo foi calculado nas etapas de calibração e validação cruzada.

5.3.3. Precisão

A precisão expressa o grau de concordância entre os resultados de uma série de medições realizadas para uma mesma amostra homogênea em

condições determinadas [97]. No presente trabalho ela foi considerada nos seguintes níveis:

- i. repetitividade, obtida a partir de seis medições de uma mesma amostra realizadas no mesmo dia;
- ii. precisão intermediária entre dias, obtida a partir de seis medições de uma mesma amostra por dia, realizadas em seis dias consecutivos;
- iii. precisão média, obtida a partir do desvio padrão médio de três amostras de concentrações diferentes, correspondentes aos extremos e a um ponto central da calibração, que foram medidas seis vezes em seis dias consecutivos. A precisão média foi calculada de acordo com a seguinte equação:

$$\text{precisão média} = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (\hat{y}_{ij} - \hat{\bar{y}}_i)^2}{n(m-1)}} \quad \text{Equação 20}$$

Onde m é o número de repetições por amostra, n é o número de amostras, $\hat{y}_{ij}$ é o valor estimado para cada repetição e $\hat{\bar{y}}_i$ é a média estimada para cada amostra.

5.3.4. Faixa de reposta linear

A avaliação da linearidade da resposta em modelos de regressão PLS pode ser difícil uma vez que as variáveis são decompostas pelos componentes principais. Assim, uma medida quantitativa para a faixa linear de resposta não corresponderia a uma tarefa simples, ou mesmo possível [97]. Qualitativamente, os gráficos de resíduos e dos *scores* contra a concentração podem indicar se os dados seguem ou não o comportamento linear. Nesse caso, o gráfico de resíduos deve apresentar comportamento aleatório e o gráfico dos *scores*, comportamento linear.

5.3.5.

Ajuste do modelo

O ajuste para um modelo não consiste em uma figura de mérito abordada com frequência em guias e normas oficiais. Entretanto, devido à dificuldade em realizar estimativas para a faixa de resposta linear, esse parâmetro é com frequência determinado em trabalhos que envolvem modelos de calibração multivariada [114]. Este parâmetro pode ser estimado a partir da correlação entre os valores tidos como referência e os valores estimados pelo modelo de calibração multivariada, para a propriedade de interesse. Isso é feito determinando por mínimos quadrados, a reta que melhor se ajusta aos valores de referência e os valores estimados pelo modelo, para as amostras de calibração.

5.3.6.

Sensibilidade

A sensibilidade (*SEN*) é definida como a fração do sinal responsável pelo acréscimo de uma unidade de concentração à propriedade de interesse. Para modelos de calibração multivariada, como PLS, pode ser determinada como [112, 115]:

$$SEN = \frac{\|NAS\|}{\|b\|} = \frac{1}{\|b\|} \quad \text{Equação 21}$$

onde *b* é o vetor dos coeficientes de regressão estimados pelo modelo de regressão PLS.

A sensibilidade analítica (γ), normalmente não abordada em protocolos de validação, apresenta, de forma clara, a sensibilidade do método em função da unidade de concentração que é utilizada [96]. Por analogia à calibração univariada [111], γ é definida como a razão entre a *SEN* e o ruído instrumental (ϵ), de acordo com a Equação 22. O vetor ϵ contém o desvio padrão do sinal analítico em cada comprimento de onda e, neste trabalho, foi estimado a partir de quinze replicatas do branco.

$$\gamma = \frac{SEN}{\|\varepsilon\|} \quad \text{Equação 22}$$

O inverso da sensibilidade analítica, γ^1 , permite estabelecer a menor diferença de concentração entre amostras, que pode ser distinguida pelo método.

5.3.7. Razão sinal ruído

Em calibração univariada, a razão sinal/ruído, S/N (do inglês *signal to noise*), representa o quanto o sinal do analito é maior que o ruído instrumental. No caso da calibração multivariada, essa razão indica o quanto a intensidade do NAS da espécie de interesse está acima do desvio padrão do ruído instrumental [112, 115]. A S/N de uma dada amostra pode ser calculada através da Equação 23 [111], na qual c é a concentração do analito na amostra e ε é o erro instrumental aleatório em r , o espectro da amostra. É importante notar que a razão sinal ruído associada a uma espécie medida em um instrumento de primeira ordem como um espectrômetro, ou seja, do tipo que é capaz de gerar múltiplas medidas ao mesmo tempo para uma mesma amostra, é tanto função da sensibilidade do instrumento para a espécie quanto uma função do grau de similaridade entre o espectro do analito e os das outras espécies presentes na amostra. Quanto maior o grau de interferência espectral sobre o sinal do analito na amostra, menor será a razão sinal ruído [111].

$$S / N = \frac{c}{\|b\| \times \|\varepsilon\|} = \frac{c\|NAS\|}{\|\varepsilon\|} \quad \text{Equação 23}$$

5.3.8. Limites de detecção e de quantificação

Os limites de detecção (LOD) e de quantificação (LOQ) de um procedimento analítico expressam as menores quantidades da espécie de

interesse que podem ser detectadas e determinadas quantitativamente, respectivamente. Para um conjunto de dados que apresenta comportamento homoscedástico (variância constante ao longo da faixa de trabalho), os LOD e LOQ na calibração multivariada podem ser calculados através das Equação 24 e Equação 25 [96, 116]:

$$LOD = \frac{3\|\varepsilon\|}{\|NAS\|} \quad \text{Equação 24}$$

$$LOQ = \frac{10\|\varepsilon\|}{\|NAS\|} \quad \text{Equação 25}$$

5.4.

Procedimentos experimentais específicos para a construção dos modelos de regressão PLS

Os conjuntos de amostras de calibração e de teste (20 misturas para compor o conjunto calibração e 20 misturas para compor o conjunto validação) foram preparados de modo que todas as amostras contivessem A, DBA, 78BQ, 79DMBA, 9ATHA e 9MA, distribuídas em oito níveis de concentração. As concentrações usadas foram entre 10 e 100 $\mu\text{g L}^{-1}$ para o conjunto calibração e entre 13 e 95 $\mu\text{g L}^{-1}$ para o conjunto validação. O planejamento experimental utilizado para a confecção dessas misturas foi baseado em trabalho de FIGUEIREDO [117], com o objetivo de obter amostras mutuamente independentes, ou seja, a composição de cada um dos componentes foi escolhida de modo a variar independentemente da composição dos outros componentes, minimizando a correlação entre os componentes.

Os planejamentos para os conjuntos calibração e validação estão apresentados nas Tabela 14 e Tabela 15. Além disso, nas Figura 77 e Figura 78, o planejamento utilizado está mais claramente ilustrado, através da comparação das composições de alguns dos componentes entre si. Esses gráficos mostram, no plano formado por dois componentes, a distribuição mais ou menos homogênea de seus teores nas misturas.

Tabela 14 – Planejamento experimental para o conjunto calibração ($\mu\text{g L}^{-1}$).

Mistura	9ATHA	A	9MA	78BQ	DBA	79DMBA
C01	25	10	20	50	100	15
C02	25	25	10	20	50	100
C03	100	15	50	30	15	20
C04	75	25	25	10	20	50
C05	30	15	20	100	30	75
C06	75	10	50	25	75	25
C07	100	30	75	75	10	50
C08	10	10	10	10	10	10
C09	30	75	75	10	50	25
C10	15	50	30	15	20	100
C11	15	20	100	30	75	75
C12	50	30	15	20	100	30
C13	20	100	30	75	75	10
C14	20	50	100	15	50	30
C15	10	20	50	100	15	50
C16	25	75	25	25	10	20
C17	50	25	75	25	25	10
C18	10	50	25	75	25	25
C19	50	100	15	50	30	15
C20	75	75	10	50	25	75

Tabela 15 - Planejamento experimental para o conjunto validação ($\mu\text{g L}^{-1}$).

Mistura	9ATHA	A	9MA	78BQ	DBA	79DMBA
T01	28	13	23	45	95	18
T02	28	28	13	23	45	95
T03	95	18	45	35	18	23
T04	70	28	28	13	23	45
T05	35	18	23	95	35	70
T06	70	13	45	28	70	28
T07	95	35	70	70	13	45
T08	13	13	13	13	13	13
T09	35	70	70	13	45	28
T10	18	45	35	18	23	95
T11	18	23	95	35	70	70
T12	45	35	18	23	95	35
T13	23	95	35	70	70	13
T14	23	45	95	18	45	35
T15	13	23	45	95	18	45
T16	28	70	28	28	13	23
T17	45	28	70	28	28	13
T18	13	45	28	70	28	28
T19	45	95	18	45	35	18
T20	70	70	13	45	28	70

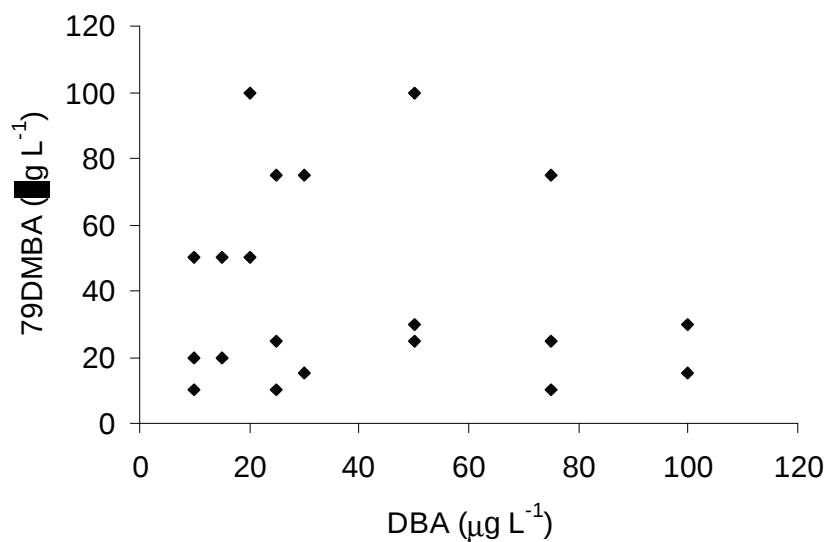


Figura 77 - Planejamento experimental para o conjunto calibração, ilustrando a distribuição das concentrações de DBA e 79DMBA.

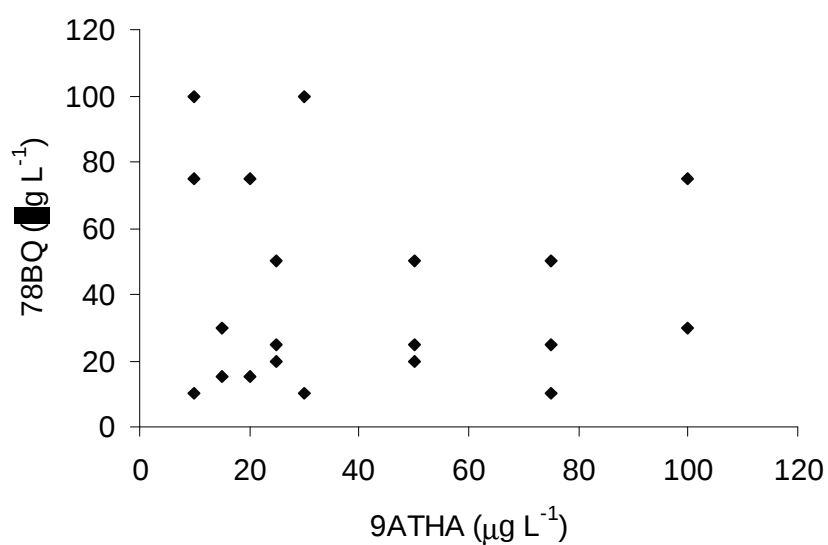


Figura 78 - Planejamento experimental para o conjunto calibração, ilustrando a distribuição das concentrações de 9ATHA e 78BQ.

Para cada analito em estudo, diversos modelos de regressão PLS foram gerados a partir dos espectros de varredura sincronizada (variáveis X) *versus* os valores das respectivas concentrações dos analitos (variáveis Y), variando diferentes parâmetros. Cada modelo obtido a partir do conjunto calibração foi

avaliado por validação cruzada total. O modelo que apresentou um valor de erro de predição (RMSECV) mínimo, de acordo com a variação do número de componentes principais, foi selecionado e aplicado ao conjunto validação para a escolha de um modelo final, com o objetivo de comparar a capacidade dos diferentes modelos em predizer um conjunto independente de amostras.

Os parâmetros testados para o melhor ajuste do modelo, além da seleção do $\Delta\lambda$ e das variáveis espectrais para compor a matriz X, incluíram os seguintes pré-tratamentos: (i) centralização em torno da média; (ii) padronização; (iii) “smooth” (remoção de ruído); (iv) normalização, dividindo-se cada espectro por seu valor máximo; (v) primeira e segunda derivadas; (vi) combinações dos tratamentos acima.

5.5.

Resultados e discussão

5.5.1.

Escolha do solvente para as amostras

Soluções de cada substância pura e misturas-padrão dos seis analitos (A, 9MA, DBA, 9ATHA, 78BQ e 79DMBA) foram preparadas em diferentes solventes, de acordo com o procedimento de extração do QAV (item 2.1.2). Os espectros de emissão e de excitação dessas soluções foram obtidos com o objetivo de identificar os comprimentos de onda de máxima intensidade fluorescente (λ) e a diferença entre os comprimentos de onda de excitação e de emissão ($\Delta\lambda$). Em seguida, foram obtidos, para as mesmas soluções, espectros de varredura sincronizada total (VST) na região de 200 a 450 nm, com $\Delta\lambda_{\text{inicial}}$ de 30 nm, $\Delta\lambda_{\text{final}}$ de 250 nm e incremento de 10 nm, totalizando um conjunto de 23 espectros. Esses espectros, ao serem traçados num mesmo gráfico, formam uma superfície de projeção em três dimensões (intensidade e comprimentos de onda de excitação em função do $\Delta\lambda$) ou um mapa de contorno que podem ser utilizados para selecionar os $\Delta\lambda$ que proporcionem maior seletividade para a análise. O mesmo experimento de VST também foi realizado para os extratos de QAV preparados naqueles mesmos solventes.

As superfícies de projeção e os mapas de contorno obtidos como resultado da VST das misturas padrão e dos extratos de QAV fortificado em diferentes solventes estão apresentados na Figura 79 até a Figura 84.

Com esses resultados, pode-se observar que no caso do solvente metanol:amônia (96:4), o extrato de QAV fortificado ($15 \mu\text{g L}^{-1}$ para 100% de recuperação na extração, Figura 80), além de apresentar sinais fluorescentes na mesma região que a mistura padrão ($20 \mu\text{g L}^{-1}$, Figura 79), apresenta sinais em regiões diferentes e, em ambos os casos, o sinal é muito maior para o extrato que para a mistura padrão, indicando interferência espectral. Sendo assim, esse sistema de solventes foi considerado inadequado para o preparo das amostras para a determinação dos azaarenos básicos em QAV por fluorimetria.

Comparando os mapas de contorno das misturas-padrão preparadas com os outros dois sistemas de solventes (metanol e metanol:HCl ($0,01 \text{ mol L}^{-1}$), apresentados na Figura 81B e na Figura 83B, respectivamente, observa-se que o meio ácido proporciona espectros muito mais ricos em informação, o que, provavelmente, facilitaria a discriminação entre as substâncias que compõem a mistura. Além disso, analisando os mapas de contorno dos extratos de QAV fortificado para os mesmos sistemas de solventes, Figura 82B e Figura 84B, observa-se uma menor interferência espectral em meio ácido, o que torna esse sistema de solventes o mais adequado para a análise. As regiões de interferência estão destacadas nas figuras.

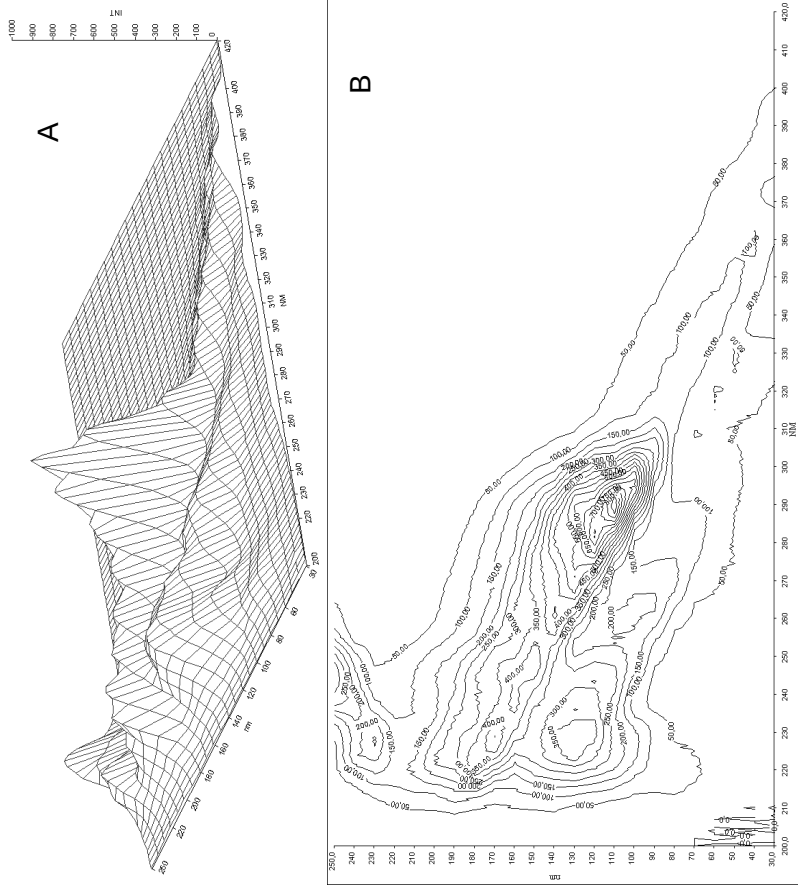


Figura 79 - Superfície de Projeção (A) e Mapa de Contorno (B) da solução padrão 20 $\mu\text{g L}^{-1}$, preparada em Metanol: NH_3 (96:4).

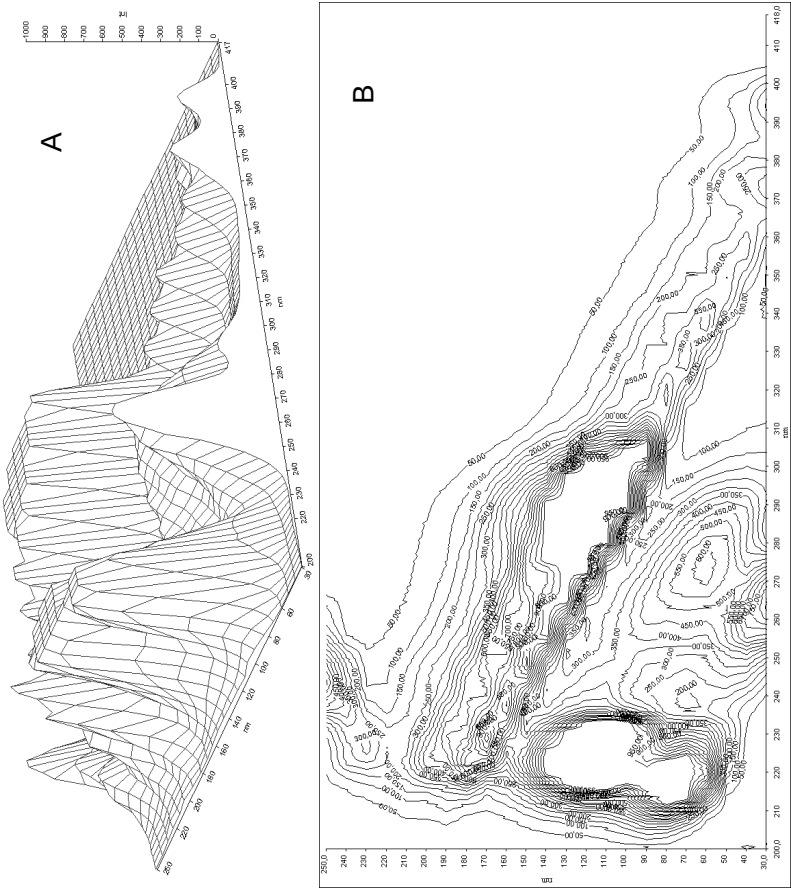


Figura 80 - Superfície de Projeção (A) e Mapa de Contorno (B) do extrato de QAV fortificado ($15 \mu\text{g L}^{-1}$ para recuperação de 100%), preparado em Metanol: NH_3 (96:4).

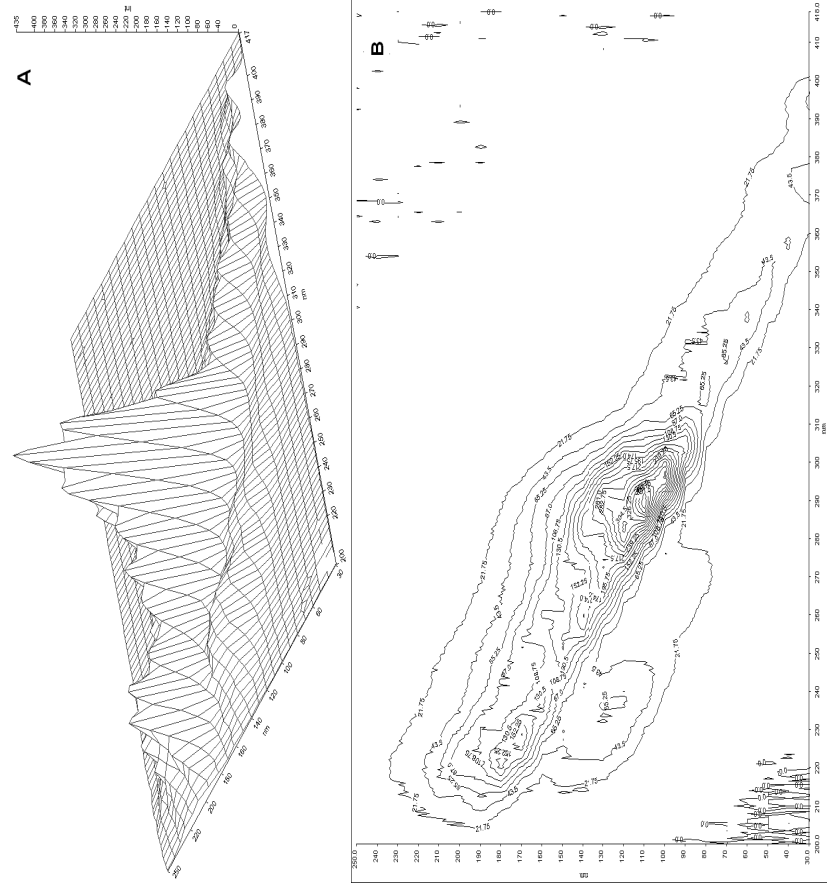


Figura 81 - Superfície de Projeção (A) e Mapa de Contorno (B) da solução padrão 100 µg L⁻¹, preparada em Metanol.

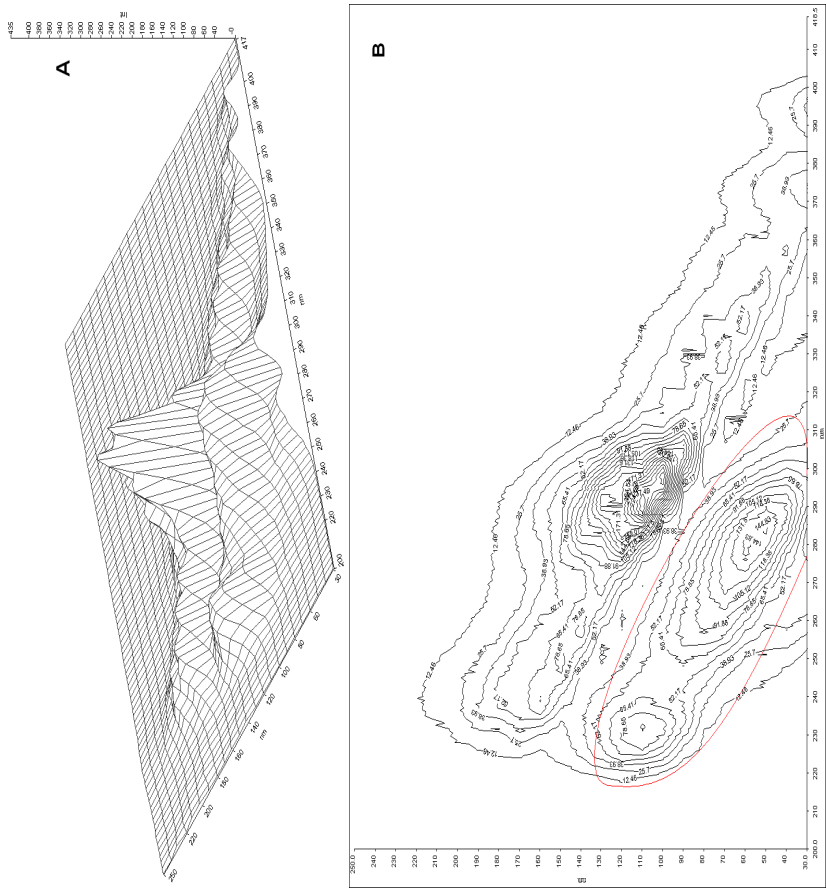


Figura 82 - Superfície de Projeção (A) e Mapa de Contorno (B) do extrato de QAV fortificado (100 µg L⁻¹ para recuperação de 100%), preparado em Metanol.

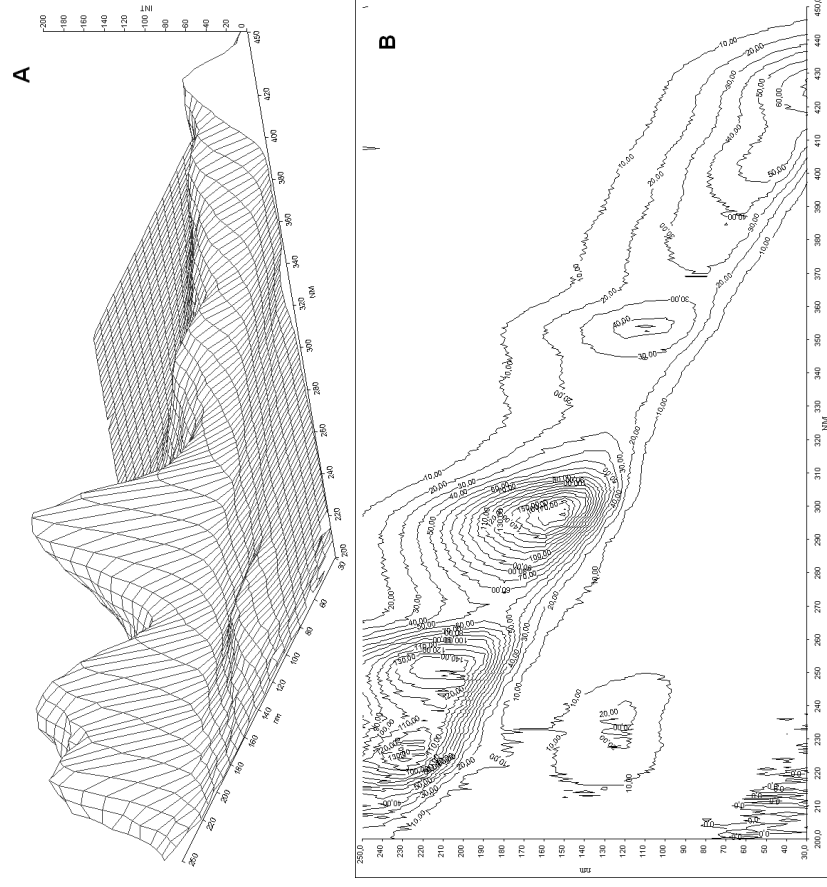


Figura 83 - Superfície de Projeção (A) e Mapa de Contorno (B) da solução padrão $50 \mu\text{g L}^{-1}$, preparada em Metanol: H_2O (70:30), meio ácido $[\text{HCl}] = 10^{-2} \text{ mol L}^{-1}$.

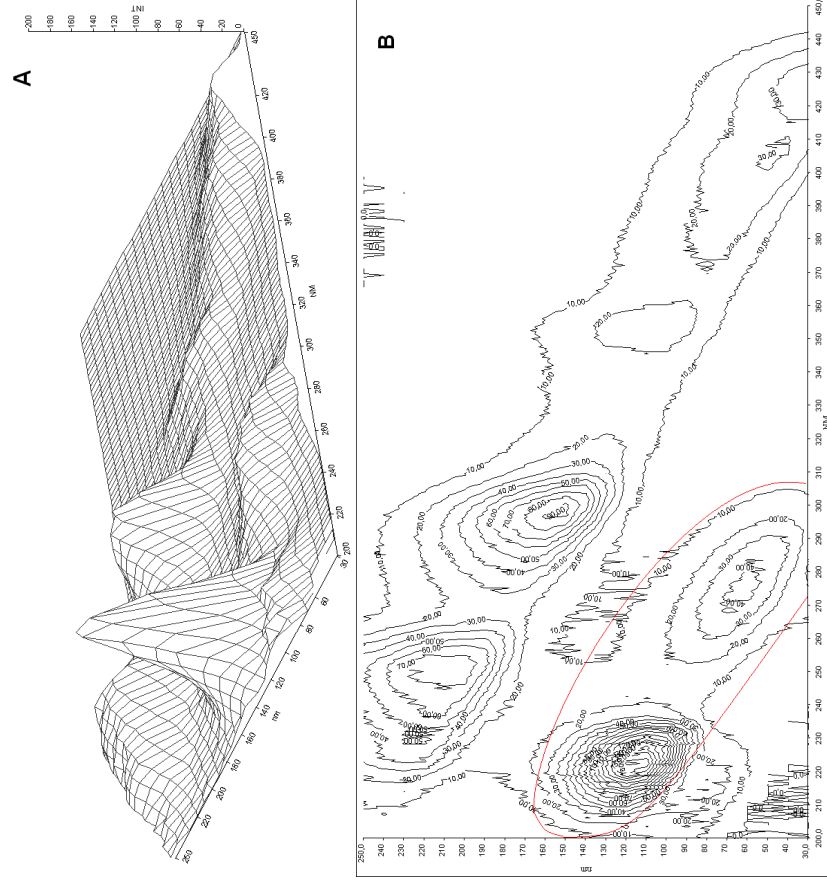


Figura 84 - Superfície de Projeção (A) e Mapa de Contorno (B) do extrato de QAV fortificado ($30 \mu\text{g L}^{-1}$ para recuperação de 100%), preparado em Metanol: H_2O (70:30), meio ácido $[\text{HCl}] = 10^{-2} \text{ mol L}^{-1}$.

5.5.2.

Escolha dos $\Delta\lambda$ para as medidas de varredura sincronizada

Os espectros de excitação e de emissão das substâncias puras foram utilizados para a determinação dos comprimentos de onda máximos de excitação e de emissão e, conseqüentemente, dos valores de $\Delta\lambda$ correspondentes ($\Delta\lambda_{\text{máx}} = \lambda_{\text{exc máx}} - \lambda_{\text{em máx}}$), apresentados na Tabela 16. As medições de varredura sincronizada das soluções-padrão (conjuntos calibração e validação) foram realizadas utilizando os $\Delta\lambda$ 30, 50, 100, 120, 150 e 180 nm, baseados nos resultados apresentados na Tabela 16, e diferentes modelos de regressão PLS foram construídos para cada analito e cada $\Delta\lambda$. Os $\Delta\lambda$ iguais a 30, 120 e 150 nm foram selecionados, pois foram aqueles que proporcionaram os modelos de regressão com os menores valores de RMSECV e RMSEP.

Tabela 16 – Comprimentos de onda máximos para os analitos puros no solvente Metanol:HCl 10^{-2} mol L $^{-1}$)

	λ_{em}^b	λ_{exc}	$\Delta\lambda_{\text{máx}}^b$	$\Delta\lambda_{\text{selecionado}}$
9ATHA	<u>238</u> / 328	360	<u>122</u> / 32	30
A+9MA ^a	<u>251</u> / 353	473	<u>222</u> / 120	150
9MA	<u>252</u> / 352	476	<u>224</u> / 124	120
78BQ	275	430	155	120
DBA	<u>299</u> / 426 / 403	450	<u>151</u> / 24 / 47	30
79DMBA	<u>291</u> / 381	475	<u>184</u> / 94	120

^a Não foi possível determinar a acridina individualmente, mas sim a sua composição com a 9-metilacridina.

^b Os valores sublinhados correspondem aos λ de máxima intensidade fluorescente.

Nas Figura 85, Figura 86 e Figura 87 estão apresentados os espectros de varredura sincronizada obtidos com os $\Delta\lambda$ selecionados, mostrando a correspondência entre os espectros dos analitos puros e os das misturas-padrão, e algumas regiões onde há interferência espectral proveniente de outras substâncias presentes no extrato de QAV, principalmente sobre o sinal da 9ATHA.

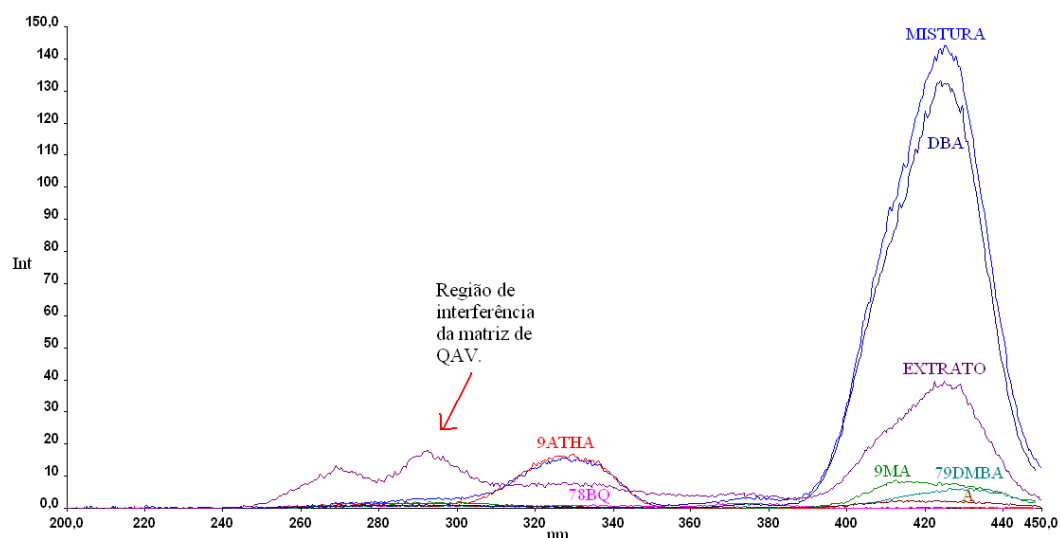


Figura 85 – Espectros de varredura sincronizada ($\Delta\lambda = 30$ nm) das soluções dos analitos puros e de uma mistura dos mesmos, $100 \mu\text{g L}^{-1}$ em Metanol: H_2O (70:30), $\text{HCl } 10^{-2} \text{ mol L}^{-1}$, e do extrato de QAV fortificado, $30 \mu\text{g L}^{-1}$ para recuperação de 100%, no mesmo solvente.

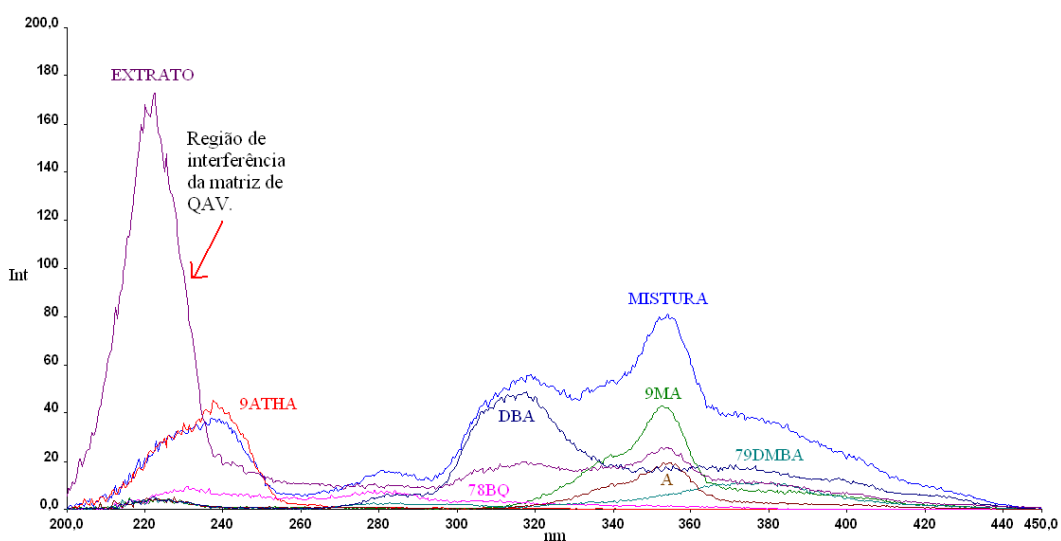


Figura 86 - Espectros de varredura sincronizada ($\Delta\lambda = 120$ nm) das soluções dos analitos puros e de uma mistura dos mesmos, $100 \mu\text{g L}^{-1}$ em Metanol: H_2O (70:30), $\text{HCl } 10^{-2} \text{ mol L}^{-1}$, e do extrato de QAV fortificado, $30 \mu\text{g L}^{-1}$ para recuperação de 100%, no mesmo solvente.

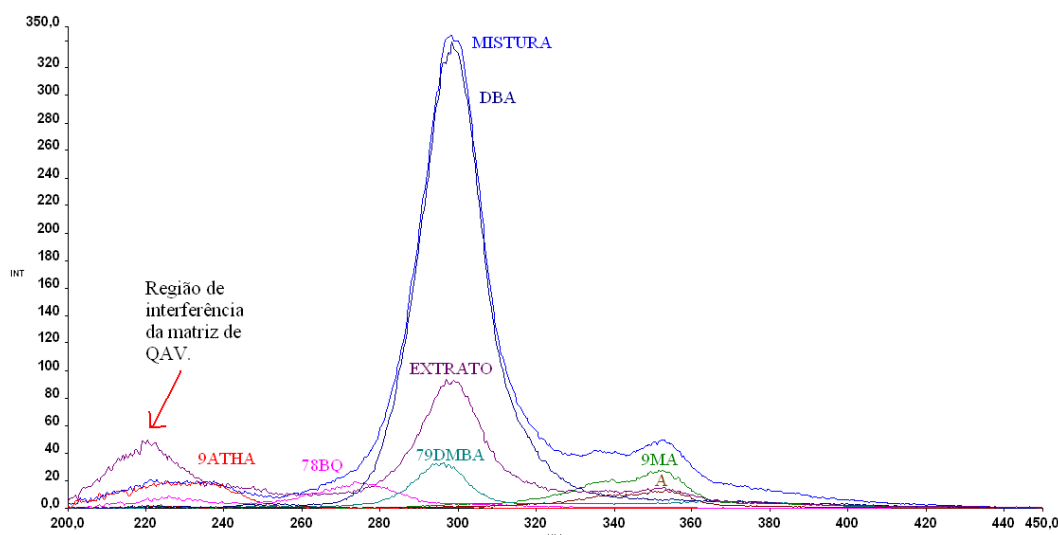


Figura 87 - Espectros de varredura sincronizada ($\Delta\lambda = 150$ nm) das soluções dos analitos puros e de uma mistura dos mesmos, $100 \mu\text{g L}^{-1}$ em Metanol:H₂O (70:30), HCl $10^{-2} \text{ mol L}^{-1}$, e do extrato de QAV fortificado, $30 \mu\text{g L}^{-1}$ para recuperação de 100%, no mesmo solvente.

5.5.3. Seleção de variáveis

A utilização de todas as variáveis X (comprimentos de onda) disponíveis muitas vezes pode não ser o mais indicado, pois acarretaria, por exemplo, em um tempo computacional longo, em tendência a superajuste (*overfitting*), colinearidades, na alta complexidade nos modelos, etc., Por causa disso, dois critérios de seleção de variáveis foram avaliados: (i) utilização de um subconjunto de variáveis diretamente relacionadas com a propriedade a ser predita, selecionado de acordo com conhecimentos químicos; (ii) utilização de um subconjunto de variáveis selecionadas de acordo com dois gráficos fornecidos pelo programa estatístico utilizado: coeficientes de regressão e cargas das variáveis X (*X-loadings weights*).

Os coeficientes de regressão são usados principalmente para verificar a importância das diferentes variáveis X em prever Y. Valores absolutos altos de coeficientes de regressão indicam variáveis de grande importância (significância) e valores pequenos (perto de 0) indicam variáveis sem importância. O valor do coeficiente indica o aumento médio em Y quando a variável X correspondente é aumentada de uma unidade, mantidas todas as outras variáveis constantes [110].

O gráfico das cargas das variáveis X contra o número da variável, para um determinado componente principal, também é usado para descobrir quais variáveis X são as importantes para prever Y. Esse gráfico mostra a relação entre o componente especificado e as diferentes variáveis X. Se uma variável tiver um grande valor de carga, positivo ou negativo, isto significa que a variável é importante para o componente. Variáveis com valores de carga altos nos primeiros componentes são aquelas que explicam a maior variação em Y [110].

Para ilustrar como são utilizados tais gráficos, na Figura 88 são apresentados os coeficientes de regressão e na Figura 89 as cargas das variáveis X do modelo PLS construído a partir dos espectros inteiros do conjunto calibração ($\Delta\lambda = 30$ nm), para predição da concentração de 9ATHA, com os dados centrados em torno da média. Nesse caso, observa-se que os valores mais altos de coeficientes de regressão e as maiores cargas nos componentes principais 1 e 2 correspondem às variáveis na região de 300 a 350 nm. No caso das cargas das variáveis X, Figura 89, vale ressaltar que o componente principal 2, para o qual observamos valores positivos altos entre 300 e 350 nm, explica 99% da variação dos dados em Y, ou seja, da variação da concentração da 9ATHA. Além disso, a Figura 85 mostra que o espectro de varredura sincronizada da 9ATHA, usando $\Delta\lambda$ de 30 nm, praticamente não sofre interferência dos outros analitos na região de 300 a 350 nm.

Para aquele modelo construído com os espectros inteiros, obteve-se RSEP% igual a 5,6% para o conjunto validação, utilizando-se quatro componentes principais. Com a seleção da região do espectro de 300 a 350 nm, obteve-se RSEP% igual a 3,6% para as mesmas amostras, utilizando-se apenas um componente principal. Portanto, a seleção de variáveis proporcionou uma diminuição no erro da predição, além de diminuir o número de componentes principais utilizados.

Para todos os modelos PLS construídos foi necessário algum tipo de seleção de variáveis. Essa seleção foi feita, na maioria dos casos, combinando-se subconjuntos de variáveis selecionados pelos dois critérios mencionados acima. As variáveis utilizadas nos modelos finais estão apresentadas na Tabela 17, apresentada no item 5.5.5.



Figura 88 - Coeficientes de regressão do modelo PLS construído com os espectros inteiros (250 a 450 nm) de varredura sincronizada, $\Delta\lambda = 30\text{nm}$, das misturas do conjunto calibração para a predição de 9ATHA.

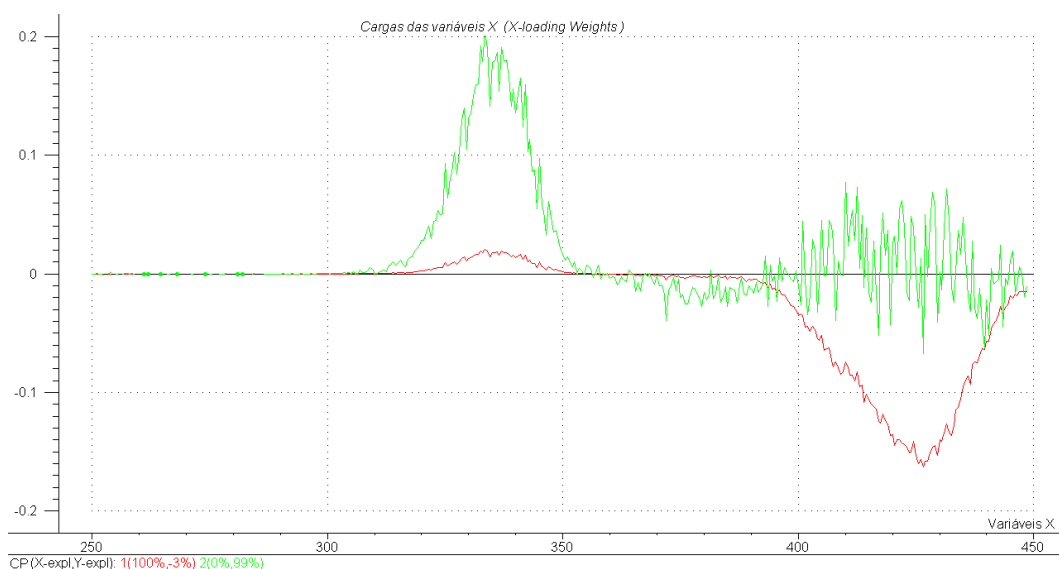


Figura 89 - Cargas das variáveis X nos componentes principais 1 e 2 do modelo PLS construído com os espectros inteiros (250 a 350 nm) de varredura sincronizada, $\Delta\lambda = 30\text{nm}$, das misturas do conjunto calibração para a predição de 9ATHA. PC1 (—) e PC2 (—)

5.5.4. Detecção de amostras anômalas

Primeiramente, observando os espectros obtidos para os conjuntos calibração e validação com $\Delta\lambda = 30$ nm, apresentados na Figura 90, pode-se verificar que, aparentemente, não há nenhuma amostra suspeita de ser anômala. No entanto, ao se acrescentar os espectros obtidos para os extratos de QAV, fortificado e não-fortificado, Figura 91, verifica-se que, em algumas regiões, os comportamentos espectrais são bem distintos. Nessa mesma Figura 91, para facilitar a visualização, foram incluídos espectros de apenas algumas soluções-padrão (10, 50 e 100 $\mu\text{g L}^{-1}$ para todos os analitos), o espectro do extrato de QAV sem fortificação e os espectros dos extratos de QAV fortificado com 6,0 $\mu\text{g L}^{-1}$ de cada analito (30 $\mu\text{g L}^{-1}$ no extrato para 100% de recuperação).

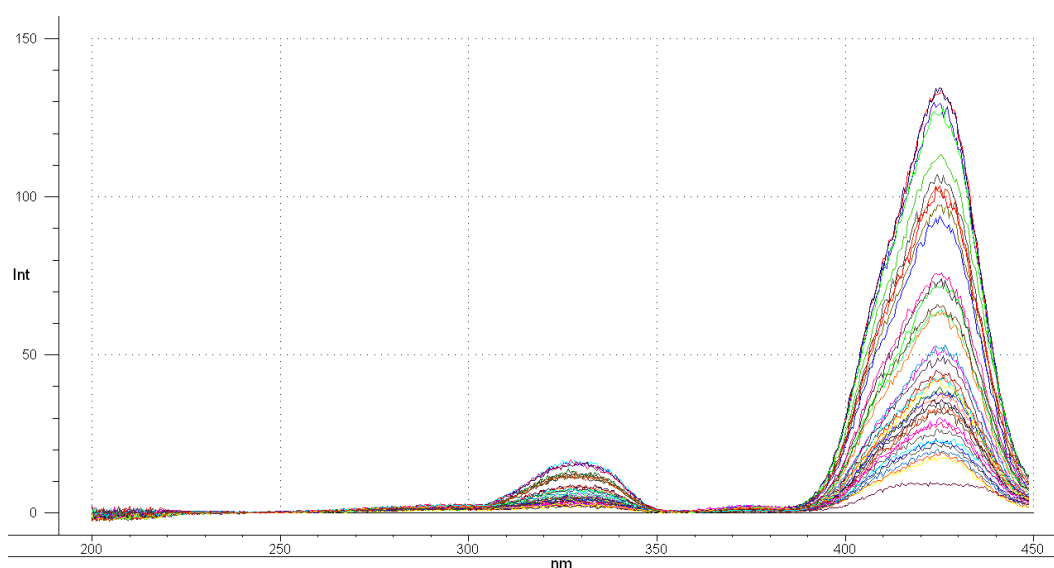


Figura 90 – Espectros de varredura sincronizada, $\Delta\lambda = 30$ nm, das soluções dos conjuntos calibração e validação.

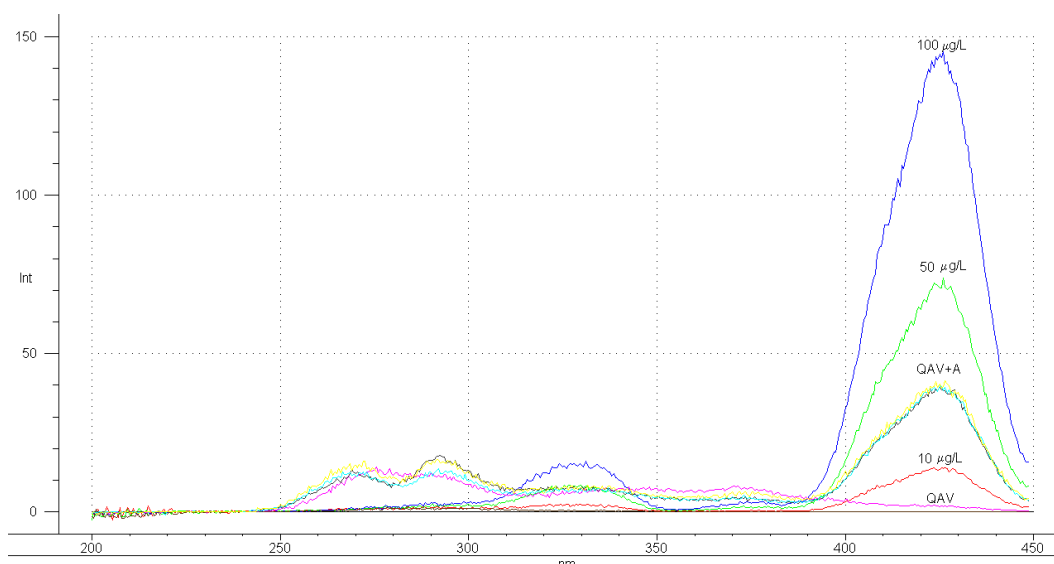


Figura 91 – Espectros de varredura sincronizada, $\Delta\lambda = 30$ nm, de soluções padrão (10, 50 e 100 $\mu\text{g L}^{-1}$ para cada um dos analitos), do extrato de QAV sem fortificação e dos extratos de QAV fortificado com 6,0 $\mu\text{g L}^{-1}$ de cada analito, QAV+A, (30 $\mu\text{g L}^{-1}$ no extrato para 100% de recuperação).

A análise dos componentes principais, utilizando $\Delta\lambda = 30$ nm e os espectros em toda a região estudada (200 – 450 nm), com os dados centrados em torno da média, confirmou que as amostras de QAV e QAV fortificado poderiam ser classificadas como anômalas quando comparadas com as misturas sintéticas dos conjuntos calibração e validação. Nessa análise, foram necessários quatro componentes principais para explicar 100% da variação de X. Os gráficos dos “scores” e de influência estão apresentados, respectivamente, nas Figura 92 e Figura 93. Nesses gráficos, os pontos envolvidos com um círculo referem-se às amostras de QAV.

Por outro lado, ao se fazer a análise dos componentes principais apenas com as variáveis selecionadas para cada analito, as amostras de QAV podem não ser consideradas anômalas, já que se estaria trabalhando em uma região de menor interferência espectral. Como exemplo, podemos citar o caso da DBA, para a qual foram utilizados o $\Delta\lambda$ de 30 nm e a região espectral de 380 a 450 nm para a construção do modelo PLS. Nessa análise, foi necessário apenas um componente principal para explicar 100% da variação de X. Os gráficos dos “scores” e de influência estão apresentados, respectivamente, na Figura 94 e na Figura 95. Nesse caso, a análise dos componentes principais com os dados

centrados em torno da média não aponta as amostras de QAV (destacadas com um círculo) como anômalas.

No gráfico de influência mostrado na Figura 95 pode-se observar que há uma amostra com alto resíduo (extrato de QAV não fortificado), o que indica que ela não é bem descrita pelo modelo. No entanto, o seu valor de “leverage” é comparável aos valores apresentados pelas soluções-padrão utilizadas como os conjuntos calibração e validação.

As análises dos componentes principais foram realizadas com os espectros de varredura sincronizada obtidos apenas com as amostras dos conjuntos calibração e validação, utilizando as regiões espectrais e os $\Delta\lambda$ selecionados para cada analito (Tabela 17), e nenhuma foi considerada anômala. No entanto, quando os espectros dos extratos de QAV foram incluídos na análise, estas amostras foram apontadas como anômalas para a 9ATHA, conforme mostrado nos gráficos dos *scores* (Figura 96) e de influência (Figura 97). Isso pode ser explicado pela interferência espectral proveniente da matriz do QAV sobre o sinal da 9ATHA, já mencionada anteriormente.

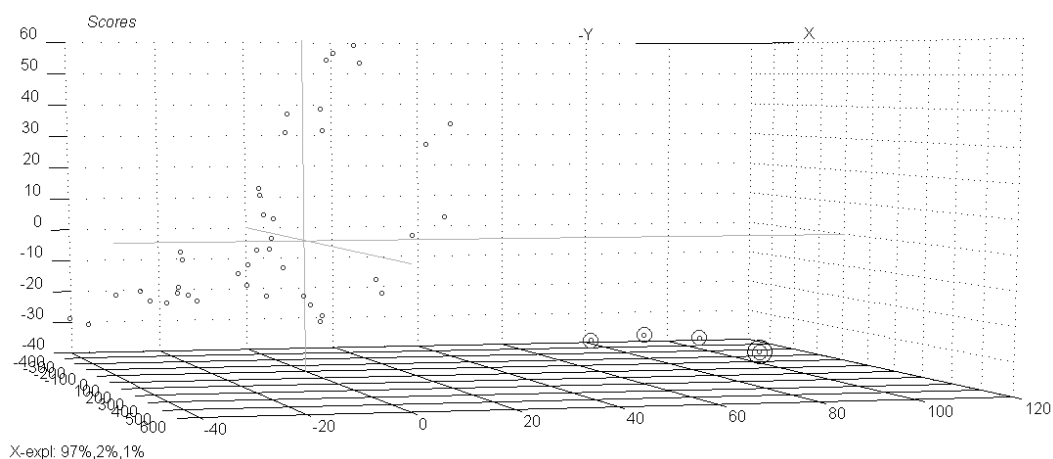


Figura 92 - Gráfico dos “scores” nos três componentes principais dos espectros de varredura sincronizada ($\Delta\lambda = 30$ nm, espectros inteiros - 200 a 450 nm) das amostras dos conjuntos calibração e validação e dos extratos de QAV. Os pontos marcados com um círculo representam os extratos dos QAV fortificados e o ponto marcado com dois círculos representa o extrato do QAV sem fortificação.

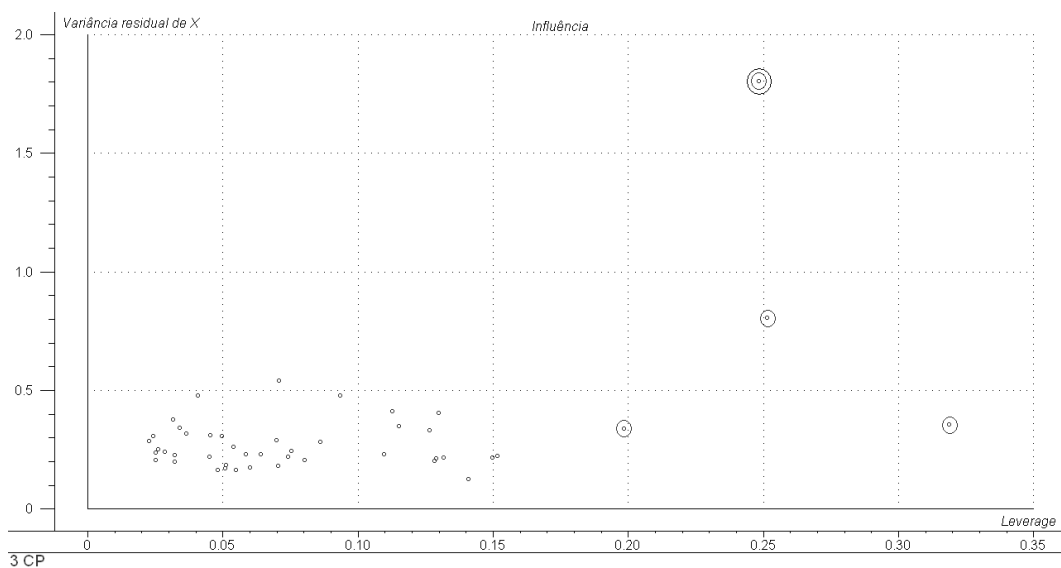


Figura 93 - Gráfico de influência nos três componentes principais dos espectros de varredura sincronizada ($\Delta\lambda = 30$ nm, espectros inteiros - 200 a 450 nm) das amostras dos conjuntos calibração e validação e dos extratos de QAV. Os pontos marcados com um círculo representam os extratos dos QAV fortificados e o ponto marcado com dois círculos representa o extrato do QAV sem fortificação.

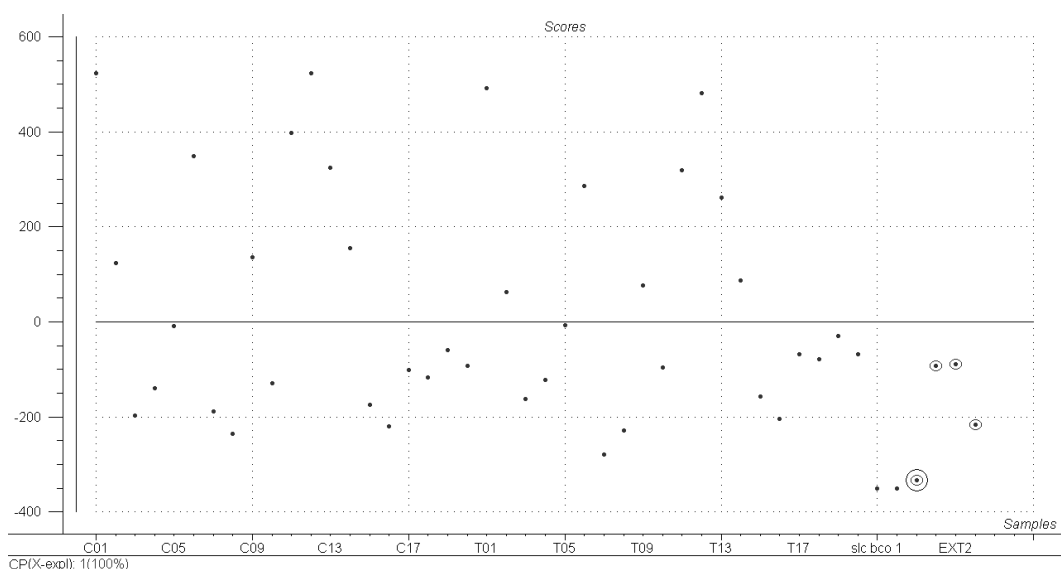


Figura 94 - Gráfico dos “scores” no primeiro componente principal dos espectros de varredura sincronizada, utilizando as variáveis selecionadas para a DBA ($\Delta\lambda = 30$ nm, 380 a 450 nm), das amostras dos conjuntos calibração e validação e dos extratos de QAV. Os pontos marcados com um círculo representam os extratos dos QAV fortificados e o ponto marcado com dois círculos representa o extrato do QAV sem fortificação.

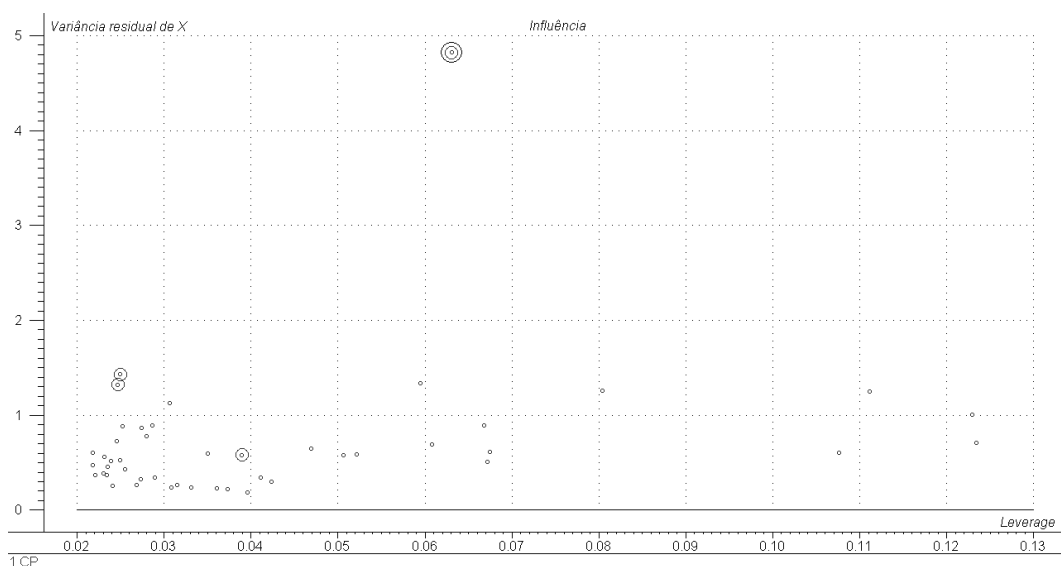


Figura 95 - Gráfico de influência no primeiro componente principal dos espectros de varredura sincronizada, utilizando as variáveis selecionadas para a DBA ($\Delta\lambda = 30$ nm, 380 a 450 nm) das amostras dos conjuntos calibração e validação e dos extratos de QAV. Os pontos marcados com um círculo representam os extratos dos QAV fortificados e o ponto marcado com dois círculos representa o extrato do QAV sem fortificação.

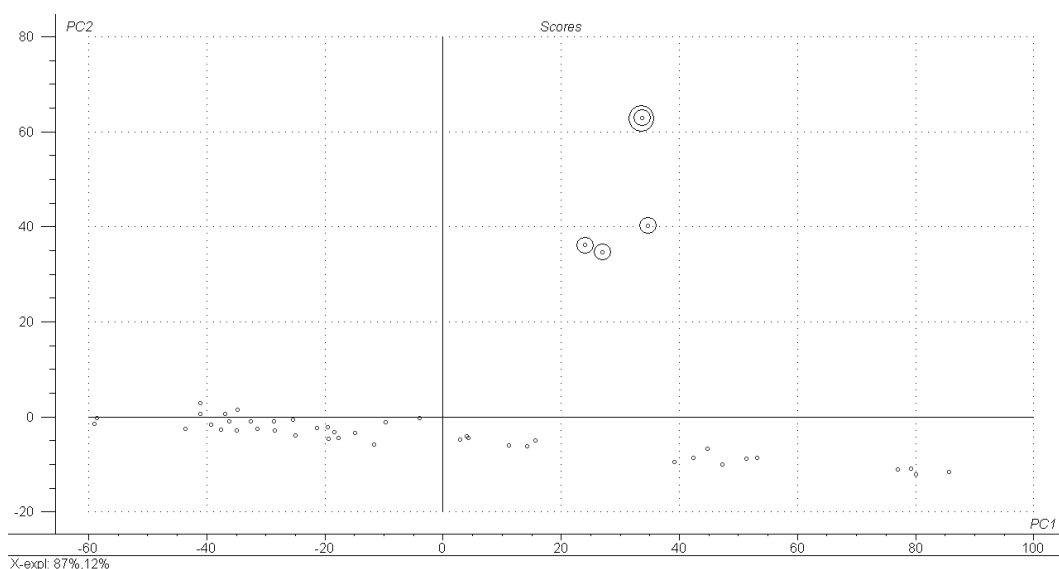


Figura 96 - Gráfico dos "scores" nos dois componentes principais dos espectros de varredura sincronizada, utilizando as variáveis selecionadas para a 9ATHA ($\Delta\lambda = 30$ nm, 300 a 350 nm) das amostras dos conjuntos calibração e validação e dos extratos de QAV. Os pontos marcados com um círculo representam os extratos dos QAV fortificados e o ponto marcado com dois círculos representa o extrato do QAV sem fortificação.

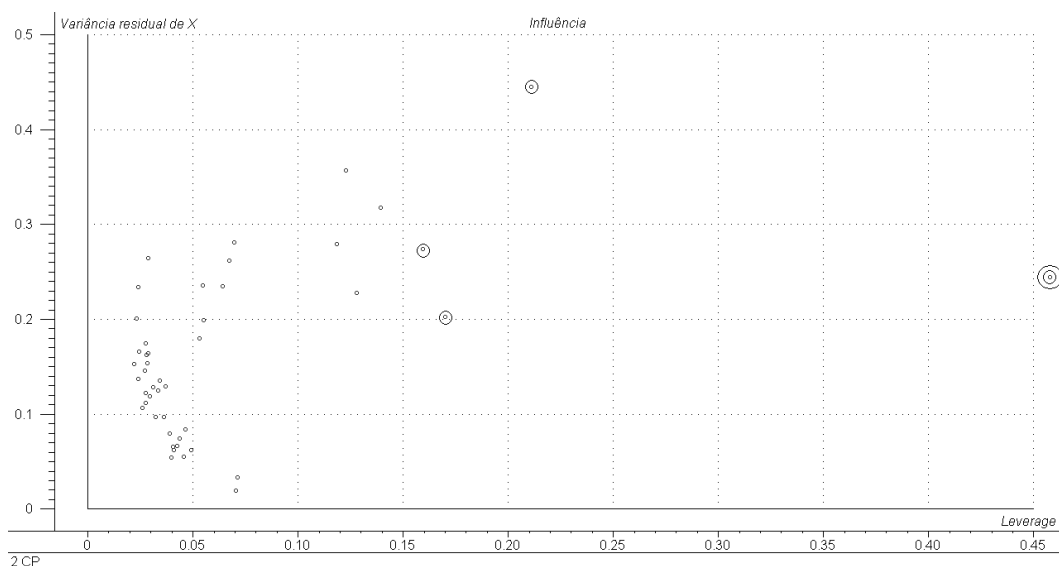


Figura 97 - Gráfico de influência nos dois componentes principais dos espectros de varredura sincronizada, utilizando as variáveis selecionadas para a 9ATHA ($\Delta\lambda = 30$ nm, 300 a 350 nm) das amostras dos conjuntos calibração e validação e dos extratos de QAV. Os pontos marcados com um círculo representam os extratos dos QAV fortificados e o ponto marcado com dois círculos representa o extrato do QAV sem fortificação.

5.5.5.

Os modelos de regressão PLS

Para todos os analitos, foram obtidos diferentes modelos PLS, utilizando-se: (i) espectros inteiros, sem pré-tratamentos; (ii) espectros inteiros, com cada um dos pré-tratamentos já mencionados; (iii) subconjuntos de variáveis, sem pré-tratamentos; (iv) subconjuntos de variáveis, com cada um dos pré-tratamentos.

A habilidade preditiva de cada modelo foi avaliada utilizando-se o RMSEP, que representa uma medida da diferença média entre os valores fornecidos pelo modelo e os valores de referência, nos estágios de validação e predição. O RMSEP pode ser interpretado como o erro médio da predição, expresso na mesma unidade da propriedade que está sendo medida, e é bastante sensível à presença de amostras com erros elevados [118].

Como já foi dito, durante o processo de validação, seja com um conjunto de amostras independentes da calibração ou por validação cruzada, um passo crítico é estabelecer o número ótimo de componentes principais que serão utilizados no modelo. Para exemplificar, na Figura 98 é apresentado o gráfico do RMSEP *versus* o número de componentes principais utilizado, do modelo construído para a predição de 9MA. Verifica-se que a utilização de quatro componentes principais fornece o menor valor de RMSEP. O número de componentes principais utilizados nos modelos finais para todos os analitos estão apresentados na Tabela 17.

Com relação aos pré-tratamentos testados, os modelos que forneceram resultados com menores valores de RMSEP foram aqueles construídos a partir dos espectros centrados em torno da média, sem nenhum outro tipo de pré-tratamento. A Tabela 17 apresenta também os parâmetros dos modelos de regressão selecionados para a predição dos azaarenos, mostrando as regiões do espectro (variáveis X), o número de componentes principais utilizados, e os valores de erro padrão relativo (RSEP%) obtidos nas etapas de calibração, validação cruzada e teste, além da predição das amostras de QAV fortificado.



Figura 98 - Gráfico RMSEP versus o número de componentes principais, do modelo PLS construído para a determinação de 9MA ($\Delta\lambda = 120 \text{ nm}$, 340 – 400nm).

Como pode ser observado, nenhum modelo de regressão PLS foi apresentado para a determinação da acridina, pois nenhum dos modelos testados forneceu resultados satisfatórios, já que os erros de predição foram muito altos. Isso deve-se ao fato de a acridina ter um espectro de fluorescência praticamente idêntico ao da 9MA, mas com intensidade consideravelmente menor, como pode ser visto na Figura 85 e na Figura 86. Optou-se, então, pela construção de um modelo para a predição da combinação de A+9MA que apresentou resultados de qualidade um pouco inferior ao dos outros analitos, mas ainda satisfatórios, inclusive para os erros da predição nas amostras de QAV.

Outro ponto importante a ser observado é que não foi possível determinar 9ATHA e 78BQ nas amostras de QAV, visto que os erros de predição foram muito altos. Isso ocorreu devido à interferência espectral proveniente de outras substâncias presentes no extrato de QAV sobre o sinal da 9ATHA e porque a 78BQ apresenta sinal fluorescente muito baixo quando comparado com os dos outros analitos (Figura 85 à Figura 87).

Tabela 17 - Modelos PLS selecionados para predição dos azaarenos.

	Região (nm)	$\Delta\lambda$ (nm)	CP ^a	RMSE ($\mu\text{g L}^{-1}$) ^b			
				CAL	FCV	PRED	QAV+A ^c
78BQ	240 - 400	120	5	3,3	4,7	3,2	104
79DMBA	270 - 450	120	3	1,5	2,4	1,3	6,1
9ATHA	300 - 350	30	1	1,7	1,9	1,7	54
9MA	340 - 400	120	4	3,3	5,1	5,0	2,4
A+9MA	320 - 400	150	4	5,8	13	8,9	5,9
DBA	380 - 450	30	1	2,7	3,1	2,2	2,5

^a CP = número de componentes principais.

^b Erro médio quadrático para as amostras: no conjunto calibração (CAL); na validação cruzada total (FCV); no conjunto validação (PRED) e nas amostras de QAV fortificado (QAVP+A).

^c Três amostras de QAV fortificadas com $6 \mu\text{g L}^{-1}$ de cada analito ($30 \mu\text{g L}^{-1}$ no extrato, para 100% de recuperação), sendo as concentrações de referência determinadas pelo método HPLC-FD [79].

5.5.6.

Parâmetros analíticos de mérito

O resultado do cálculo das figuras analíticas de mérito está apresentado na Tabela 18. Os dados de regressão parecem ajustar-se satisfatoriamente bem ao modelo linear. O conjunto de figuras que inclui da Figura 99 à Figura 104 mostra os gráficos dos valores de concentração dos analitos de referência contra os mesmos valores estimados, que demonstram o bom ajuste do modelo linear e a apresentação do modelo PLS em sua forma pseudounivariada. Os valores dos coeficientes de correlação estão apresentados na Tabela 18.

Tabela 18 – Figuras analíticas de mérito estimadas para os modelos Fluorescência-PLS para a predição dos azaarenos.

Figuras de mérito		78BQ	79DMBA	9ATHA	9MA	A+9MA	DBA
Exatidão (RSEP%)	Calibração	6,7	3,1	3,4	6,7	6,4	5,4
	FCV ^a	9,5	4,7	3,8	10	15	6,3
	Validação	6,6	3,0	3,6	10	10	4,7
Precisão (µg L ⁻¹) ^b	Repetitividade	1,5	0,78	0,67	4,1	4,4	0,40
	Intermediária entre dias	2,1	2,2	2,4	1,2	3,5	2,0
	Média	2,5	1,9	2,4	5,3	5,9	2,0
Sensibilidade ^c		3,8	6,3	1,4	2,0	1,3	8,6
Inverso da sensibilidade analítica (γ ⁻¹) ^d		0,40	0,22	0,59	0,38	0,66	0,015
Razão Sinal-Ruído (conjunto calibração)	Máxima	248	449	171	266	152	611
	Mínima	25	45	17	27	15	61
Ajuste (referência x estimado)	Inclinação	0,996 ± 0,019	0,994 ± 0,009	0,997 ± 0,010	0,962 ± 0,024	0,929 ± 0,031	0,990 ± 0,014
	Interseção	0,446 ± 0,942	0,112 ± 0,420	0,350 ± 0,488	1,145 ± 1,190	4,352 ± 2,715	-0,222 ± 0,700
r ²		0,9860	0,9970	0,9962	0,9763	0,9605	0,9925
LOD (µg L ⁻¹)		1,2	0,67	1,8	1,1	2,0	0,045
LOQ (µg L ⁻¹)		4,0	2,2	5,9	3,8	6,6	0,15

^a FCV = validação cruzada total.

^b Para o nível de concentração de 50 µg L⁻¹.

^c Valores expressos como a razão entre a intensidade fluorescente e a unidade de concentração (µg L⁻¹).

^d Valores expressos em µg L⁻¹.

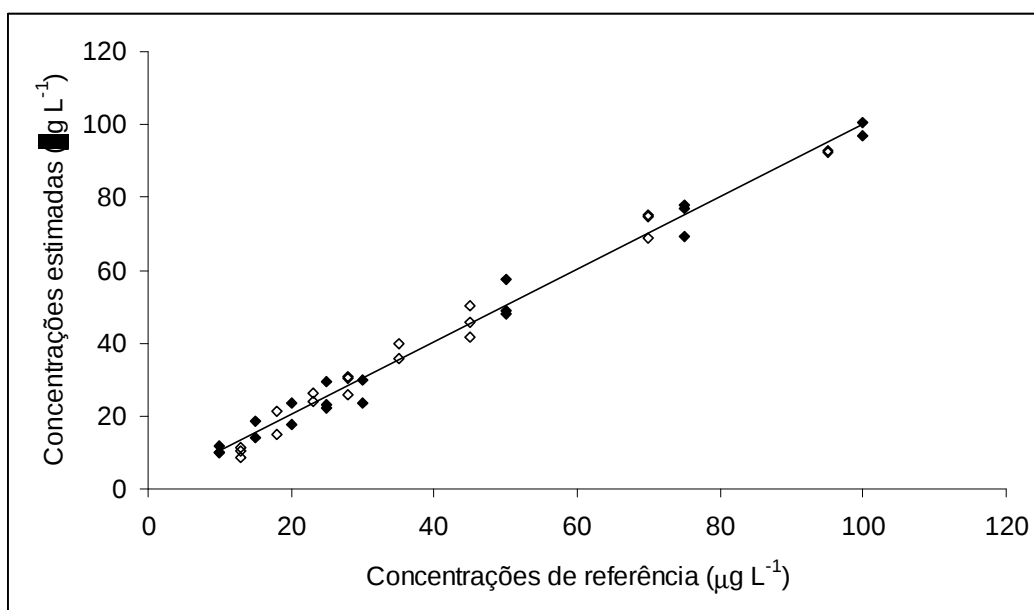


Figura 99 - Valores de referência *versus* estimados pelo modelo Fluorescência-PLS para a 78BQ. ($\blacklozenge$) conjunto calibração, ($\diamond$) conjunto validação e (—) reta ajustada.

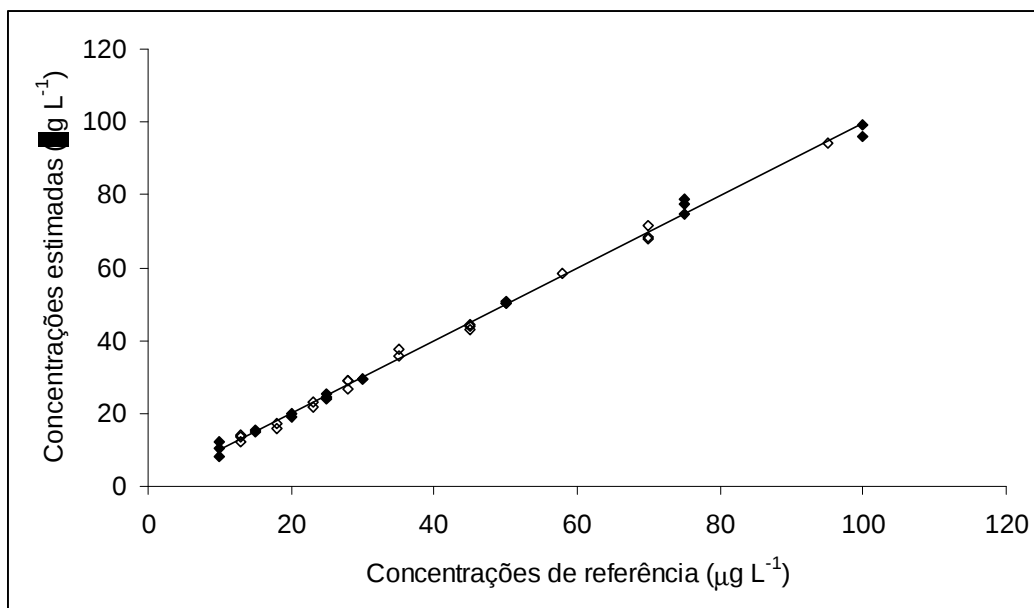


Figura 100 - Valores de referência *versus* estimados pelo modelo Fluorescência-PLS para a 79DMBA. ($\blacklozenge$) conjunto calibração, ($\diamond$) conjunto validação e (—) reta ajustada.

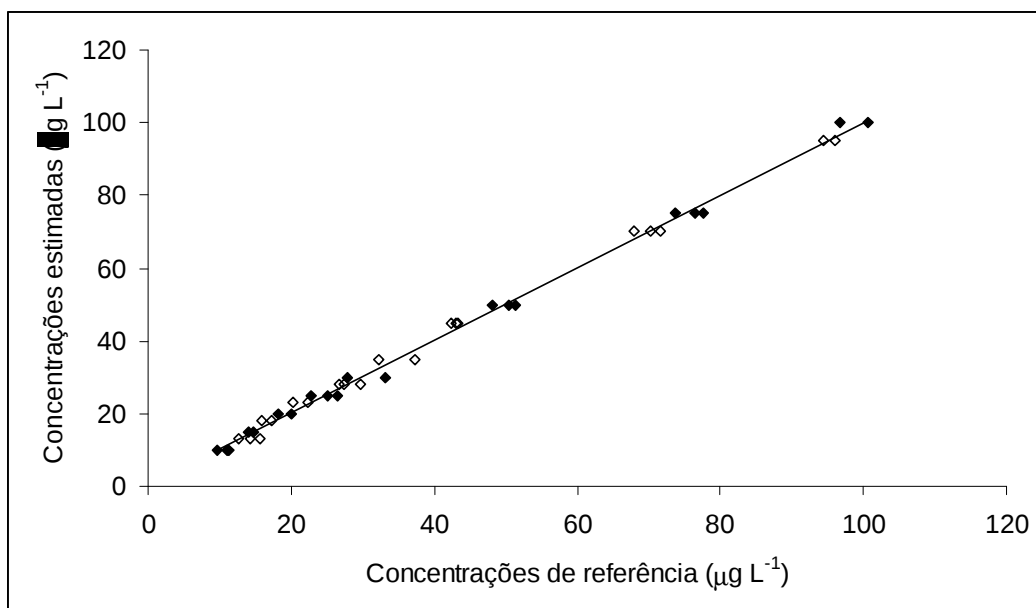


Figura 101 - Valores de referência *versus* estimados pelo modelo Fluorescência-PLS para a 9ATHA. ($\blacklozenge$) conjunto calibração, ($\diamond$) conjunto validação e (—) reta ajustada.

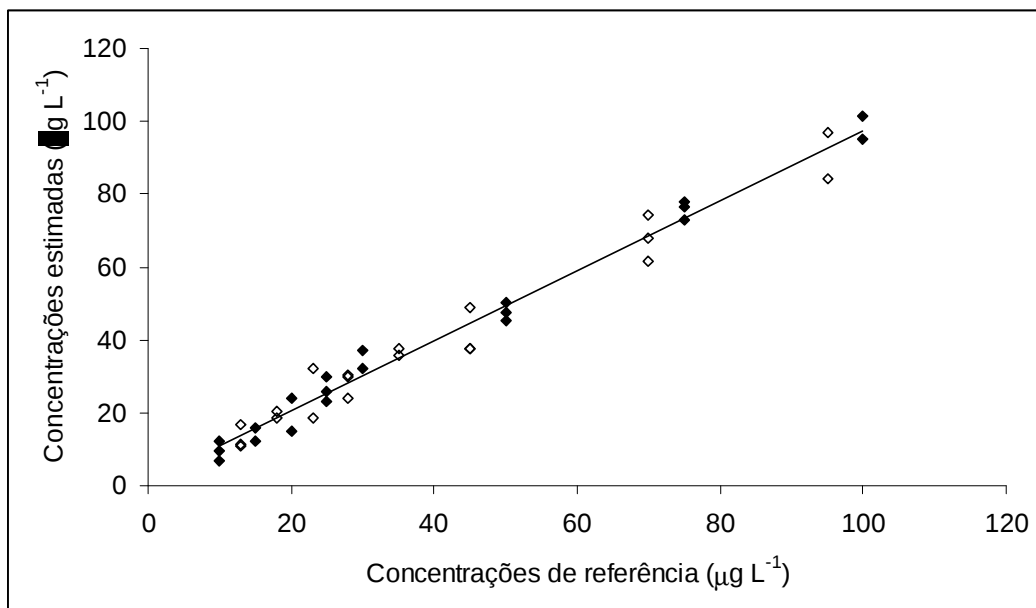


Figura 102 - Valores de referência *versus* estimados pelo modelo Fluorescência-PLS para a 9MA. ($\blacklozenge$) conjunto calibração, ($\diamond$) conjunto validação e (—) reta ajustada.

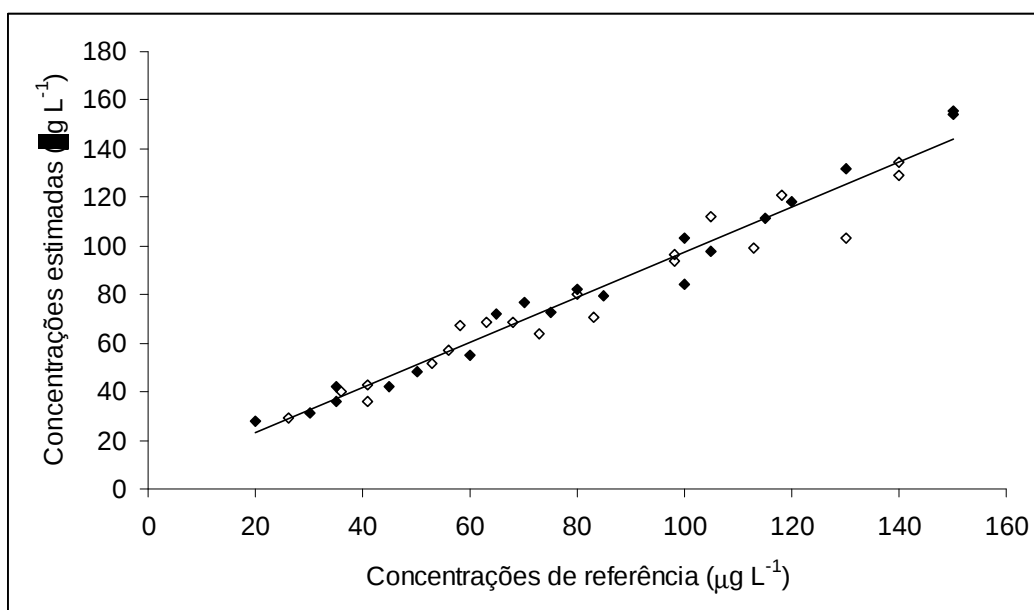


Figura 103 - Valores de referência *versus* estimados pelo modelo Fluorescência-PLS para a A+9MA. ($\blacklozenge$) conjunto calibração, ($\diamond$) conjunto validação e (—) reta ajustada.

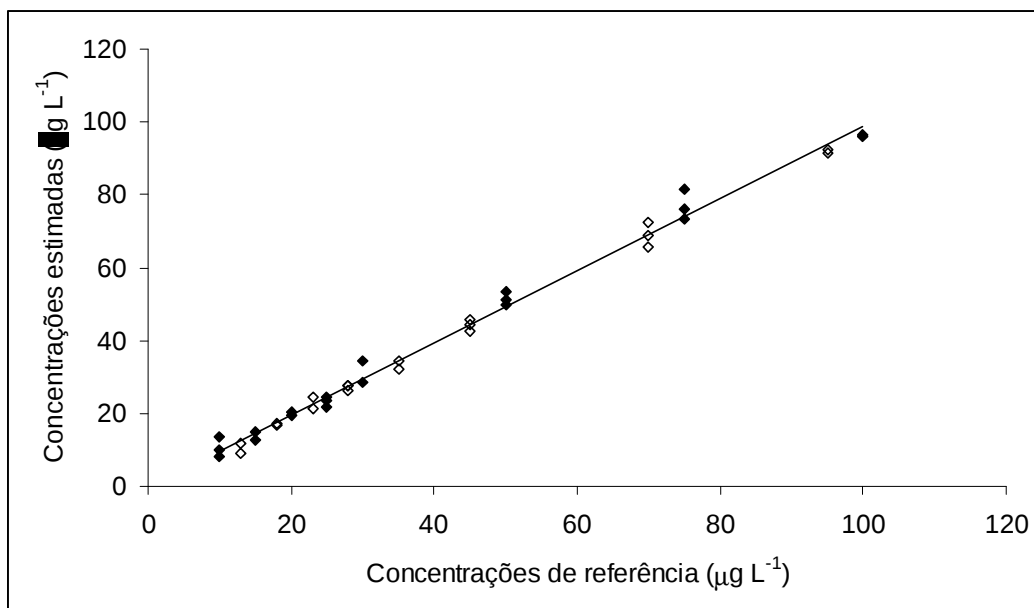


Figura 104 - Valores de referência *versus* estimados pelo modelo Fluorescência-PLS para a DBA. ($\blacklozenge$) conjunto calibração, ($\diamond$) conjunto validação e (—) reta ajustada.

No conjunto de figuras que vai da Figura 105A à Figura 110A são mostrados os gráficos dos *scores* do primeiro componente principal contra a concentração, com exceção da 78BQ (Figura 105A) para a qual o terceiro componente principal foi utilizado por ser o que explicou a maior parte da variação de Y. Nas mesmas figuras, nas partes denominadas pela letra B, são apresentados os gráficos dos resíduos. Para a 9ATHA (Figura 107) e DBA (Figura 110), claramente observa-se um comportamento linear nos gráficos dos *scores* e erros com comportamento aleatório nos gráficos dos resíduos. Para as outros azaarenos, os pontos que apresentam maior desvio do comportamento linear nos gráficos dos *scores* não devem ser encarados, necessariamente, como indício de falta de linearidade dos dados, uma vez que os modelos PLS utilizam mais componentes principais do que os utilizados na confecção destes gráficos.

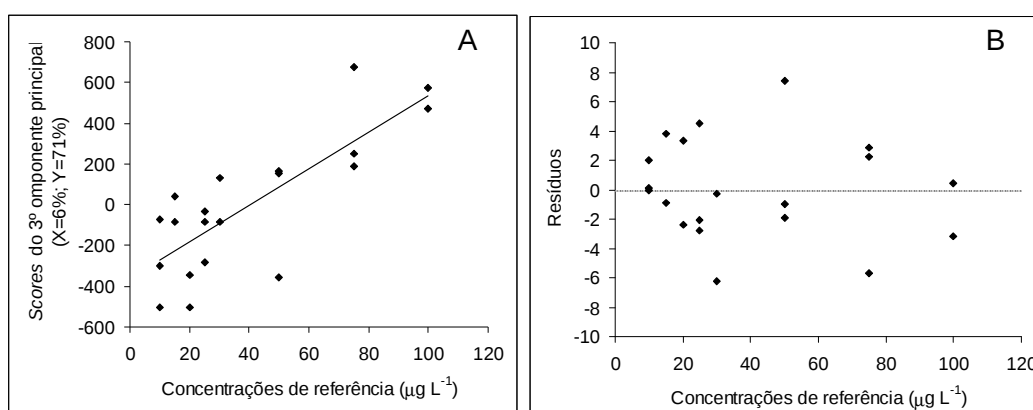


Figura 105 – A: (♦) *scores* do terceiro componente principal (6% de variância em X e 71% de variância em Y) contra as concentrações de referência da 78BQ para as amostras do conjunto calibração, (—) reta ajustada. B: gráfico dos resíduos contra as concentrações de referência da 78BQ para os conjuntos calibração (♦) e validação (□).

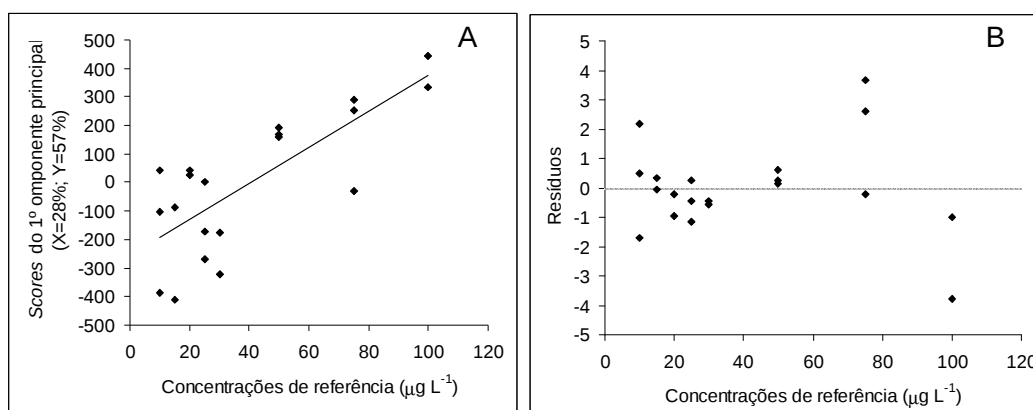


Figura 106 – A: (♦) scores do primeiro componente principal (28% de variância em X e 57% de variância em Y) contra as concentrações de referência da 79DMBA para as amostras do conjunto calibração, (—) reta ajustada. B: gráfico dos resíduos contra as concentrações de referência da 79DMBA para os conjuntos calibração (♦) e validação (□).

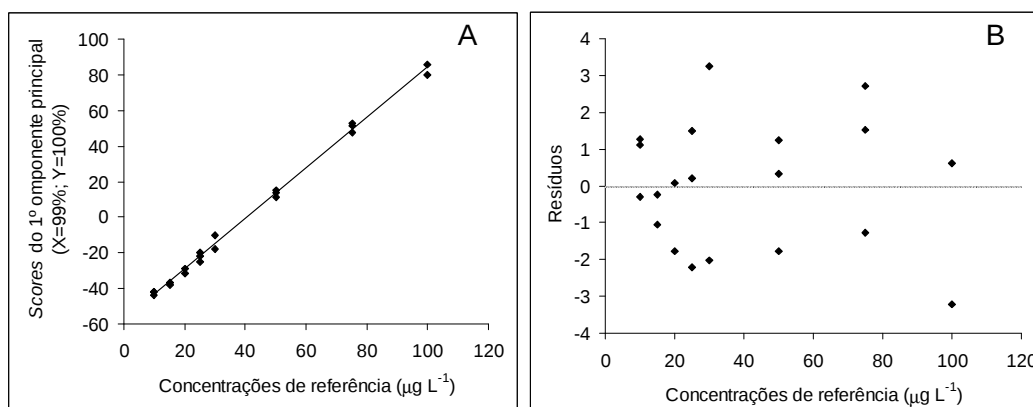


Figura 107 – A: (♦) scores do primeiro componente principal (99% de variância em X e 100% de variância em Y) contra as concentrações de referência da 9ATHA para as amostras do conjunto calibração, (—) reta ajustada. B: gráfico dos resíduos contra as concentrações de referência da 9ATHA para os conjuntos calibração (♦) e validação (□).

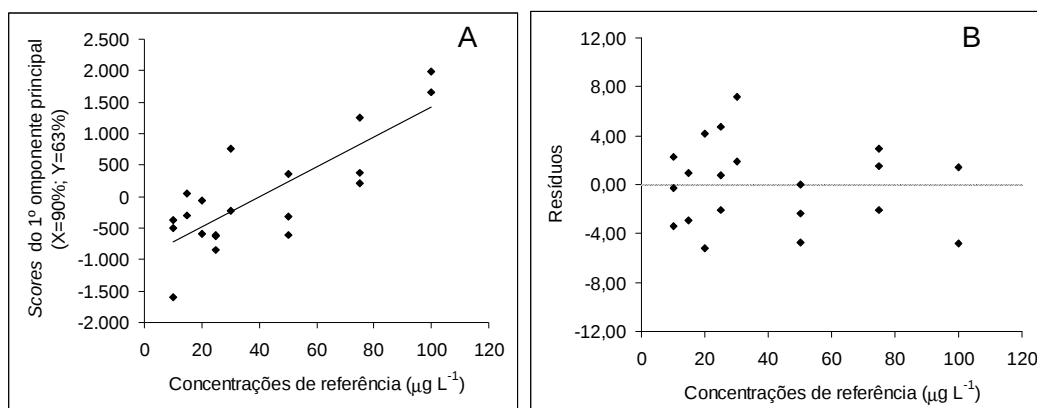


Figura 108 – A: (♦) scores do primeiro componente principal (90% de variância em X e 63% de variância em Y) contra as concentrações de referência da 9MA para as amostras do conjunto calibração, (—) reta ajustada. B: gráfico dos resíduos contra as concentrações de referência da 9MA para os conjuntos calibração (♦) e validação (□).

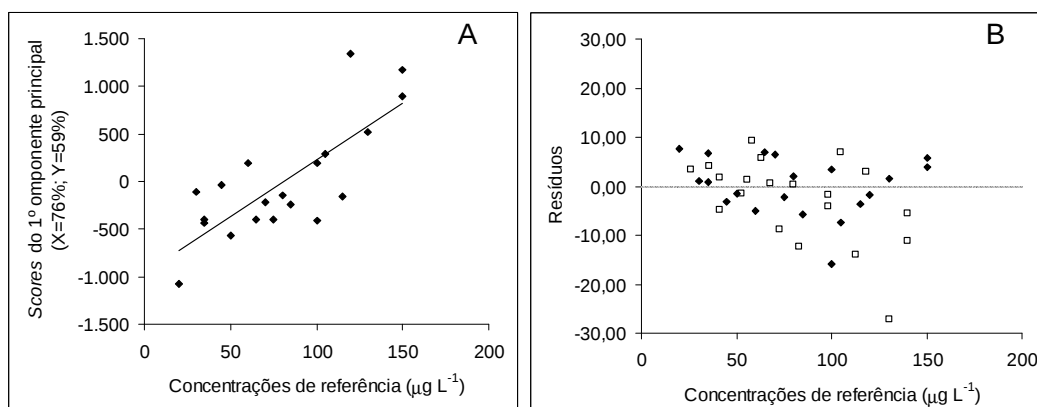


Figura 109 – A: (♦) scores do primeiro componente principal (76% de variância em X e 59% de variância em Y) contra as concentrações de referência da A+9MA para as amostras do conjunto calibração, (—) reta ajustada. B: gráfico dos resíduos contra as concentrações de referência da A+9MA para os conjuntos calibração (♦) e validação (□).

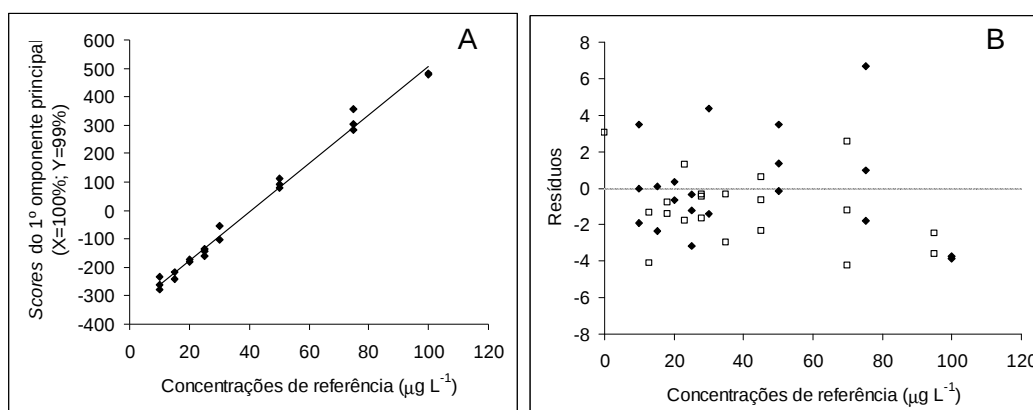


Figura 110 - A: (♦) scores do primeiro componente principal (100% de variância em X e 99% de variância em Y) contra as concentrações de referência da DBA para as amostras do conjunto calibração, (—) reta ajustada. B: gráfico dos resíduos contra as concentrações de referência da DBA para os conjuntos calibração (♦) e validação (□).

Os indicadores de exatidão, mostrados na Tabela 18, apresentam um nível aceitável de dispersão, indicando que o valor do RMSECV é uma boa estimativa dos erros de previsão observados no conjunto validação. Todos os valores de RSEP% foram menores que 10%, com exceção do valor obtido na etapa de validação cruzada para A+9MA que foi igual a 15%.

Para avaliar a sensibilidade dos modelos o melhor parâmetro é o inverso da sensibilidade analítica (γ^{-1}), pois pode ser interpretado de forma mais clara por sua relação direta com a concentração. Segundo esses valores, os modelos seriam capazes de distinguir entre amostras com diferença de concentração da ordem de 0,015 a 0,66 $\mu\text{g L}^{-1}$, dependendo do analito.

Os valores de precisão apresentaram um nível de dispersão aceitável. Os resultados para repetitividade e precisão intermediária foram obtidos para três níveis de concentração (correspondentes aos extremos e ao ponto central da calibração: 10, 50 e 100 $\mu\text{g L}^{-1}$) e calculados a partir de seis medidas de uma mesma amostra realizadas no mesmo dia, e a partir de seis medidas de uma mesma amostra por dia, realizadas em seis dias consecutivos, respectivamente. Os resultados para os três níveis de concentração, em valores percentuais, estão apresentados na Tabela 19. Para precisão média, foi calculado o desvio padrão médio das mesmas três amostras que foram medidas seis vezes em seis dias consecutivos. Os resultados de repetitividade e de precisão intermediária para o nível de concentração de 50 $\mu\text{g L}^{-1}$, na unidade de concentração, e os resultados de precisão média estão apresentados na Tabela 18.

Tabela 19 – Resultados de precisão, em valores percentuais, para os modelos fluorescência-PLS.

		78BQ	79DMBA	9ATHA	9MA	A+9MA	DBA
10 $\mu\text{g L}^{-1}$	Repetitividade	12	3,2	3,5	16	8,1	1,1
	Intermediária ^a	19	1,1	3,9	12	1,8	4,8
50 $\mu\text{g L}^{-1}$	Repetitividade	2,8	1,2	1,2	6,5	4,4	0,75
	Intermediária ^a	4,0	3,5	4,3	2,0	3,6	3,7
100 $\mu\text{g L}^{-1}$	Repetitividade	2,0	1,3	1,4	6,0	4,6	0,44
	Intermediária ^a	3,6	2,0	3,7	4,9	3,7	3,2

^a Precisão intermediária entre dias.

Os resultados apresentados na Tabela 19 mostram que, nos casos da 9MA e A+9MA, os valores obtidos para precisão intermediária entre dias são menores que os obtidos para a repetitividade, apresentando uma ordem inversa da que é apresentada teoricamente. Essa diferença pode ser atribuída a possíveis variações instrumentais que ocorreram ao longo do dia em que foram realizados os estudos de repetitividade. No entanto, se um teste F for aplicado, a partir dos quadrados dos respectivos valores de precisão intermediária e repetitividade (variâncias), observa-se que os valores calculados, 0,086 para 9MA e 0,633 para A+9MA, são menores que o valor $F_{95\%}$ tabelado (com 5 e 30 graus de liberdade) de 2,534. Uma vez que os valores calculados não excedem o valor crítico, pode-se concluir que não há diferença estatisticamente significativa entre as duas estimativas de precisão, para os dois casos.

Os valores para razão sinal ruído apresentados mostram o quanto o escalar NAS está acima do desvio padrão da flutuação do sinal instrumental. De acordo com os resultados, a menor razão sinal ruído observada para as amostras do conjunto calibração foi 15 (A+9MA), que representa um valor cinco vezes maior que o limite mínimo aceitável para determinações quantitativas [97].

5.5.7.

Determinação dos azaarenos básicos em amostras comerciais de QAV

Para avaliar a qualidade dos resultados, além do RMSEP, foram também calculados os erros relativos (Equação 26) para as amostras de QAV e QAV fortificado, apresentados na Tabela 20, onde $y_{\text{referência}}$ representa o valor de

concentração determinado pelo método HPLC-FD apresentado neste trabalho [79].

$$\text{Erro\%} = [(y_{\text{estimado}} - y_{\text{referência}}) / y_{\text{referência}}] \times 100 \quad \text{Equação 26}$$

Tabela 20 – Erros relativos para a determinação dos analitos nas amostras de QAV.

	78BQ			79DMBA			9ATHA		
	$\mu\text{g L}^{-1}$			$\mu\text{g L}^{-1}$			$\mu\text{g L}^{-1}$		
	estimado	referência	erro%	estimado	referência	erro%	estimado	referência	erro%
Branco-extração	177	1,0	17303	5,9	0	--	59	0	--
Extrato 1	123	26	365	26	27	-1,0	57	9,5	493
Extrato 2	121	26	362	26	26	-1,1	54	0	--
Extrato 3	147	28	420	28	30	-7,6	62	0	--
	9MA			DBA			A+9MA		
	$\mu\text{g L}^{-1}$			$\mu\text{g L}^{-1}$			$\mu\text{g L}^{-1}$		
	estimado	referência	erro%	estimado	referência	erro%	estimado	referência	erro%
Branco-extração	0	0	--	0	0	--	8,4	0	--
Extrato 1	30	27	12	25	25	-2,3	58	54	7,7
Extrato 2	22	25	-10	25	29	-12	54	50	6,8
Extrato 3	21	22	-4,6	26	29	-8,2	58	50	17
A = A+9MA - 9MA									
	$\mu\text{g L}^{-1}$								
	estimado	referência	erro%						
Branco-extração	8,4	0	--						
Extrato 1	28	28	3,4	RMSEP ($\mu\text{g L}^{-1}$)		6,6			
Extrato 2	31	25	23						
Extrato 3	37	28	35	RSEP%		24			

Na Tabela 20, os extratos das amostras de QAV fortificado estão denominados como Extratos 1, 2 e 3, e o extrato de QAV não fortificado como Branco-extração. Os valores em negrito indicam os casos em que o método de referência (HPLC-FD) não detectou a presença do analito e o método Fluorescência-PLS estimou um valor acima do limite de quantificação, previamente determinado. Observa-se que os erros relativos encontrados para a determinação de 79DMBA, 9MA e DBA foram todos menores que 15%, o que é bastante razoável para o método em estudo, e os erros para a determinação de A+9MA foram menores que 20%. Os valores não mencionados referem-se a amostras que não contém o analito, não sendo possível, portanto, o cálculo do erro relativo.

Como já mencionado, não foi possível determinar 9ATHA e 78BQ nas amostras de QAV, visto que os erros de predição foram muito altos. Isso ocorreu devido à interferência espectral proveniente de outras substâncias presentes no

extrato de QAV sobre o sinal da 9ATHA e porque a 78BQ apresenta sinal fluorescente muito baixo quando comparado com os dos outros analitos.

Além do cálculo dos erros relativos para a composição A+9MA, o mesmo foi feito para a acridina, por diferença entre os teores de A+9MA e 9MA. Os valores de RMSEP e RSEP% também foram calculados e estão apresentados na Tabela 20. Vale mencionar, em termos de comparação, que o RMSEP encontrado para a 9MA foi de $2,4 \mu\text{g L}^{-1}$ e para a composição A+9MA foi de $5,9 \mu\text{g L}^{-1}$.