

Análise de um processo farmacêutico de mistura de pós com espectroscopia NIR e ferramentas quimiométricas

Andreia Afonso Rodrigues

Dissertação para obtenção do Grau de Mestre em

Engenharia Farmacêutica

Orientador: Prof. João Pedro Martins de Almeida Lopes
Prof. José Monteiro Cardoso de Menezes

Júri:

Presidente: Prof. João Carlos Moura Bordado

Orientador: Prof. João Pedro Martins de Almeida Lopes

Vogal: Prof. Marco Paulo Seabra dos Reis

Dezembro de 2015

AGRADECIMENTOS

A realização desta dissertação de mestrado contou com importantes apoios e incentivos sem os quais não se teria tornado uma realidade e aos quais estarei eternamente grata.

Gostaria de começar por agradecer ao Prof. João Almeida Lopes, por ter sido um orientador excepcional que me acompanhou desde o início até ao fim, pelo saber que transmitiu, pelas opiniões e críticas, total colaboração na solução de dúvidas e problemas que foram surgindo ao longo da realização deste trabalho e por todas as palavras de incentivo.

À Sandra Lucas e Margarida Casa Nova por me terem proporcionado um estágio na Generis® e em simultâneo terem permitido a elaboração da minha dissertação na empresa.

À Margarida Casa Nova por ter sido mais do que uma colega de trabalho, mas sim uma amiga e todo o apoio demonstrado ao longo destes 9 meses de estágio. Agradeço ainda todos os conselhos, incentivo, força e me ter proporcionado alguma orientação que foi crucial para o término deste trabalho.

Ao Dr. Miguel Freudenthal da Generis® por confiar em mim, pela disponibilidade, apoio, conselhos, profissionalismo, ensinamentos, encorajamento e ter sido um superior hierárquico excepcional que me permitiu crescer tanto como pessoa como profissionalmente.

À equipa Financeira e Sr. João Menem pela receção e bons momentos proporcionados.

À Dra. Rosário Fonseca da Generis® por me ter proporcionado o estágio e por ser compreensível quando necessário.

Ao Luís Lopes da Generis® pela prontidão com que aceitou orientar a minha dissertação e por ao longo desta fase me ter facultado sempre que necessário toda a informação assim como matéria-prima e amostras necessárias. Agradeço ainda por todos os conselhos, saberes, opiniões e críticas que demonstraram ser um homem bastante experiente nesta área e pela preocupação demonstrada ao longo do trabalho.

À equipa do Fabrico pelo apoio prestado sempre que necessário, à Patrocínia Reis que me facultou algumas amostras quando necessário e em especial ao Carlos Gonçalves por ser um bom profissional e me colocou à vontade durante o acompanhamento do fabrico do produto, bem como pela ajuda prestada que permitiu solucionar alguns problemas que apareceram ao longo do tempo.

À Dra. Cláudia Bernardino por me ter facultado os boletins analíticos e ter criado a ligação entre mim e o laboratório de Controlo de Qualidade.

À Dra. Sofia Queirós por me ter proporcionado realizar as análises ao produto no laboratório de Controlo de Qualidade e à Prazeres Rato que me ensinou todas as técnicas utilizadas para a análise do produto e pela paciência demonstrada.

À minha família, pais e irmão pelo apoio incondicional, força, incentivo e amizade sem igual, sem eles nada disto seria possível.

Ao meu namorado pela compreensão, sobretudo nos momentos mais críticos e, a sua dedicada atenção.

Aos meus amigos, pela paciência, atenção e força que sempre transmitiram e à Marisa Rodrigues pelos conhecimentos partilhados.

Por fim, a todos os que, de uma forma ou de outra, contribuíram para a concretização da tese, e que não estão aqui referidos, muito obrigado.

RESUMO

Este trabalho teve como objetivo a análise de um processo farmacêutico de mistura de pós, com recurso à espectroscopia de infravermelho próximo (NIR) que é uma técnica que tem crescente aceitação em termos de aplicação industrial neste tipo de processos. A abordagem mais correntemente utilizada, consiste no acompanhamento do processo de mistura espectroscopicamente avaliando a variabilidade espectral em função do tempo, permite identificar o ponto final da mistura mas não garante que o produto está dentro de especificação. Propõem-se aqui duas estratégias alternativas, designadas estratégias qualitativa e quantitativa. Pretende-se com estas abordagens, avaliar duas metodologias alternativas para o acompanhamento e monitorização de processos de mistura, capazes de serem aplicadas em situações de produção, naturalmente nos casos em que os compostos relevantes são ativos em NIR. Este trabalho baseou-se num produto farmacêutico existente no mercado mas por questões de confidencialidade, não é referido o seu nome nem a natureza e identidade dos compostos envolvidos na formulação.

Ambas as estratégias propostas são baseadas em amostras produzidas à escala laboratorial de acordo com dois delineamentos experimentais. No caso qualitativo, os espetros foram avaliados recorrendo à análise de componentes principais (PCA). A abordagem quantitativa por seu lado consistiu no desenvolvimento de calibrações com base em mínimos quadrados parciais (PLS) entre os espetros e os teores dos vários componentes da mistura.

Ambas as estratégias demonstraram viabilidade para acompanhar um processo de mistura de pós, sendo alternativas viáveis e vantajosas à metodologia correntemente adotada

Palavras chave: Mistura de pós; Quimiometria; Espectroscopia de infravermelho próximo; Análise por componentes principais; Regressão por mínimos quadrados parciais

ABSTRACT

This work aimed at monitoring a pharmaceutical powder mixture process, using near-infrared (NIR) spectroscopy, which is a technique that has growing acceptance in terms of industrial application in such processes. The most commonly used approach consists in monitoring the mixing process by analyzing spectra variability. Evaluating the spectral variation over time, allows the identification of the end point of the mixture but does not ensure that the product is within specification. Herein, are proposed two alternative strategies, designated as qualitative or quantitative, in order to evaluate different methods for monitoring and controlling mixing processes, capable of being applied in production situations, where of course the relevant compounds are active in the NIR. This work was based on a commercial pharmaceutical product but, due to confidentiality reasons, its name, nature and description of the compounds involved in the formulation are not mentioned.

The proposed strategies are based on samples produced at laboratory scale according to two experimental designs. While in the qualitative case, the spectra were evaluated using principal component analysis (PCA), in the quantitative approach were developed several calibrations based on partial least squares (PLS) between the spectra and the contents of the various components.

Both strategies have demonstrated viability to accompany this powder mixing process, constituting feasible and efficient alternatives to the currently adopted methodology.

Keywords: Powder mixing; chemometrics; near infrared spectroscopy; principal component analysis; partial least squares regression

ÍNDICE GERAL

Agradecimentos.....	iii
Resumo	v
Abstract.....	vii
Índice geral	ix
Índice de figuras	xi
Índice de tabelas	xiii
Lista de siglas e convenções	xv
1. Introdução.....	1
1.1. Objetivos	2
1.2. Apresentação	3
1.3. Controlo de qualidade	4
1.4. Mistura de pós.....	6
1.5. Espectroscopia de infravermelho próximo.....	7
1.5.1. História.....	7
1.5.2. Princípios teóricos	8
1.5.3. Espectrofotómetro de transformada de Fourier de infravermelho próximo (FT-NIR)	10
1.6. Delineamento experimental	13
1.7. Quimiometria.....	15
1.7.1. Pré-tratamento espectral	15
1.7.1.1. Derivadas (Savitzky-Golay)	15
1.7.1.1. Correção multiplicativa de sinal.....	16
1.7.1.2. Variação de padrão normal	16
1.7.2. Análise multivariada.....	17
1.7.2.1. Análise de componentes principais.....	17
1.7.2.2. Regressão por mínimos quadrados parciais.....	18
2. Revisão bibliográfica	21
3. Materiais e Métodos	23
4. Resultados e Discussão.....	27
4.1. Análise dos espectros dos componentes puros	27

4.2. Controlo estatístico do processo de mistura.....	31
4.2.1. Análise qualitativa.....	32
4.2.2. Análise quantitativa.....	44
4.2.2.1. Componente B.....	50
4.2.2.2. Componente C	52
4.2.2.3. Componente D	54
4.2.2.4. Componente H	56
4.3. Comparação entre as duas abordagens.....	58
5. Conclusões.....	59
6. Perspetivas futuras.....	61
7. Referências bibliográficas	63

ÍNDICE DE FIGURAS

Figura 1 Bandas e posições relativas de picos de absorção em infravermelho próximo (cm^{-1}) (<i>Mantanus, 2012</i>).	8
Figura 2 A) Vibrações de alongamento; B) Vibrações de deformação para uma molécula triatómica (adaptado de <i>Silva et al., 2007</i>).	9
Figura 3 Esquema de espectrofotômetro FT-NIR com interferômetro de Michelson (<i>Almeida, 2009</i>).	11
Figura 4 Fluxograma ilustrativo do processo de fabrico do produto P.	24
Figura 5 Espectro NIR obtidos em reflectância difusa do componente A.	27
Figura 6 Espectros NIR obtidos em reflectância difusa dos componentes B (A) e C (B).	28
Figura 7 Espectro NIR obtidos em reflectância difusa do componente D.	29
Figura 8 Sobreposição dos espectros NIR dos componentes puros (componentes A, B, C e D).	30
Figura 9 Espectros NIR não processados obtidos em modo de reflectância difusa das 24 amostras do DoE utilizado na abordagem qualitativa.	34
Figura 10 Mapa de scores obtidos do modelo PCA com base nos espectros NIR das amostras do DoE para abordagem qualitativa (DOE-QLT).	35
Figura 11 Estatísticas de Hotelling T^2 (A) e resíduos (B) relativas ao modelo PCA construído com base no DoE para a abordagem qualitativa.	36
Figura 12 Mapa de scores relativo à validação do modelo PCA, com projeção as amostras de validação (●: amostras de calibração (DoE); ▼: amostras de validação).	37
Figura 13 Mapa de scores relativo ao modelo PCA com projeção das amostras obtidas industrialmente (●: amostras de calibração (DoE); ▼: amostras de teste).	38
Figura 14 Estatísticas de Hotelling T^2 (A) e resíduos (B) relativas ao modelo PCA construído com base no DoE para a abordagem qualitativa. com projeção das amostras obtidas industrialmente (●: amostras de calibração (DoE); ▼: amostras de teste).	39
Figura 15 Mapa de scores relativo modelo PCA com projeção das amostras obtidas durante a monitorização das misturas efetuadas em condições laboratoriais (A-Lote 1; B-Lote 2) (●: amostras de calibração (DoE); ▼: amostras de teste).	40
Figura 16 Estatísticas Hotelling T^2 relativas ao modelo PCA construído com base no DoE-QLT com projeção das amostras obtidas do processo de mistura à escala laboratorial (A-Lote 1; B-Lote 2) (●: amostras de calibração (DoE); ▼: amostras de teste).	41
Figura 17 Estatísticas dos resíduos relativas ao modelo PCA construído com base no DoE-QLT com projeção das amostras obtidas do processo de mistura à escala laboratorial (A-Lote 1; B-Lote 2) (●: amostras de calibração (DoE); ▼: amostras de teste).	42
Figura 18 Para a construção dos modelos PLS utilizou-se um conjunto de 24 amostras para calibração (A) e 6 amostras para validação (B).	47
Figura 19 Proporções experimentais versus previstas pelo modelo PLS do componente A com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).	49

Figura 20 Proporções experimentais versus previstas pelo modelo PLS do componente B com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).	51
Figura 21 Proporções experimentais versus previstas pelo modelo PLS do componente C com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).	53
Figura 22 Proporções experimentais versus previstas pelo modelo PLS do componente D com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).	55
Figura 23 Proporções experimentais versus previstas pelo modelo PLS do componente H com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).	57

ÍNDICE DE TABELAS

Tabela 1 Delineamento experimentais com maior utilização (<i>Singh et al., 2011</i>).....	14
Tabela 2 Composição do medicamento P.....	23
Tabela 3 Especificações no produto comercial P dos componentes A, B, C, D, H, tendo por base os métodos de referência.....	25
Tabela 4 Identificação das bandas no espectro do componente A.....	27
Tabela 5 Identificação das bandas nos espectros dos componentes B e C.....	29
Tabela 6 Identificação das bandas no espectro do componente D.....	30
Tabela 7 Resultados obtidos através dos métodos de referência para as amostras industriais.....	31
Tabela 8 Resultados obtidos através dos métodos de referência para as amostras laboratoriais.....	31
Tabela 9 Proporções dos componentes nas misturas delineadas pelo DoE-QLT construído para a abordagem qualitativa proposta de acordo com análise pelos métodos analíticos implementados no Departamento de CQ da Generis®.....	33
Tabela 10 Resultados obtidos pelos métodos analíticos certificados para o conjunto de amostras a utilizar na validação do modelo PCA.....	34
Tabela 11 Resumo do modelo PCA desenvolvido para a abordagem qualitativa.....	35
Tabela 12 Proporções determinadas pelo método de referência para o conjunto de 24 amostras do DoE-QNT, usadas na abordagem quantitativa.....	45
Tabela 13 Proporções determinadas pelo método de referência para as amostras de validação (V#) e teste (T#) usadas na abordagem quantitativa.....	46
Tabela 14 Resumo do modelo PLS relativo ao componente A.....	48
Tabela 15 Resumo dos resultados obtidos para o modelo PLS na previsão do componente A, para as amostras industriais e laboratoriais.....	49
Tabela 16 Proporções esperadas e previstas do componente A para as amostras industriais e laboratoriais.....	50
Tabela 17 Resumo do modelo PLS relativo ao componente B.....	50
Tabela 18 Resumo dos parâmetros obtidos para o modelo PLS na previsão do componente B, para as amostras industriais e laboratoriais.....	51
Tabela 19 Proporções esperadas e previstas do componente B para as amostras industriais e laboratoriais.....	52
Tabela 20 Resumo do modelo PLS relativo ao componente C.....	52
Tabela 21 Resumo dos parâmetros obtidos para o modelo PLS na previsão do componente C, para as amostras industriais e laboratoriais.....	53
Tabela 22 Proporções esperadas e previstas do componente C para as amostras industriais e laboratoriais.....	54
Tabela 23 Resumo do modelo PLS relativo ao componente D.....	54
Tabela 24 Resumo dos parâmetros obtidos para o modelo PLS na previsão do componente D, para as amostras industriais e laboratoriais.....	55

Tabela 25 Proporções esperadas e previstas do componente D para as amostras industriais e laboratoriais.	56
Tabela 26 Resumo do modelo PLS relativo ao componente H.....	56
Tabela 27 Resumo dos parâmetros obtidos para o modelo PLS na previsão do componente H, para as amostras industriais e laboratoriais.....	57
Tabela 28 Proporções esperadas e previstas do componente H para as amostras industriais e laboratoriais.	58

LISTA DE SIGLAS E CONVENÇÕES

AIM	Autorização de Introdução no mercado
cGMP	Boas práticas de fabrico (current good manufacturing practices)
CP	Componentes principais
CQ	Controlo de Qualidade
CQA	Analytical Quality Certificate
DoE	Delineamento experimental
EMA	Agência Europeia do Medicamento
EP	Farmacopeia Europeia
FP	Farmacopeia Portuguesa
FT-NIR	Infravermelho próximo com transformada de Fourier
ICH	International Conference on Harmonisation
INFARMED I.P.	Autoridade Nacional do Medicamento e Produtos de Saúde
IV	Infravermelho
MSC	Correção multiplicativa de sinal
NIR	Espectroscopia de Infravermelho próximo
PAT	Tecnologia analítica de processo
PCA	Análise de componentes principais
PLS	Mínimos quadrados parciais
RMSEC	Raiz quadrada do erro quadrado médio (calibração)
RMSECV	Raiz quadrada do erro quadrado médio (validação cruzada)
RMSEP	Raiz quadrada do erro quadrado médio (previsão)
SNV	Variação de padrão normal

1. INTRODUÇÃO

Com o aumento da competitividade existente nos últimos anos entre empresas do sector farmacêutico, as diferentes empresas e grupos farmacêuticos têm cada vez mais necessidade de se diferenciar dos seus principais competidores, dando crescente importância à satisfação e confiança do consumidor. Frequentemente, este esforço traduz-se também no reforço da capacidade de monitorização em processo requer a adoção de novas metodologias e diferenciado equipamento analítico. Estas novas técnicas dizem respeito a todas as operações unitárias envolvidas no fabrico de medicamentos, nomeadamente na monitorização de processos de mistura, que quando aplicadas permitem um controlo rápido e eficaz sobre os atributos de qualidade dos produtos.

A espectroscopia de infravermelho próximo (NIR), é a técnica analítica que mais tem sido implementada quando se pretende dotar um processo industrial de fabrico de medicamentos de sistemas de monitorização em tempo real das etapas de mistura. A espectroscopia NIR estuda a interação da radiação infravermelha com a matéria e apresenta como vantagens o facto de ser uma técnica não destrutiva, rápida, simples mecanicamente (eventualmente sem recurso a componentes móveis), precisa e com boa relação sinal/ruído. Neste contexto do acompanhamento de um processo de mistura pode inclusivamente não haver necessidade de qualquer tipo de calibração ou desenvolvimento elaborado de um método.

Esta técnica instrumental requer, em geral, a adoção de estratégias de análise multivariadas, podendo estas ser de natureza qualitativa ou quantitativa. A análise qualitativa tem por base o processamento de todo o espectro, permitindo o agrupamento de amostras semelhantes, enquanto que a análise quantitativa permite relacionar os espectros com os valores de referência de cada parâmetro analisado (p.e. um teor de princípio ativo), de modo a que os possa estimar em amostras de produção. Para desenvolver cada uma destas estratégias torna-se necessário a conceção de um conjunto de experiências que permitam obter o máximo de informação possível a partir de um número mínimo de experiências. Naturalmente, que produzir estas amostras em ambiente industrial se revela de natureza difícil dadas as condicionantes naturais de uma linha de produção, razão pela qual aqui se adota o método de produção destas amostras necessárias a uma escala reduzida.

Este trabalho propõe o desenvolvimento de modelos com análise de componentes principais (metodologia qualitativa) e regressão por mínimos quadrados parciais (metodologia quantitativa). Para a construção dos diversos modelos, foi necessária a análise da composição das amostras pelos métodos certificados e adotados pelo controlo de qualidade. Foram igualmente obtidos espectros de absorção utilizando radiação NIR. A abordagem mais correntemente utilizada para monitorização de processos de mistura por NIR, consiste no acompanhamento espectroscópico avaliando a variabilidade espectral em função do tempo. Esta, permite identificar o ponto final da mistura mas não garante que o produto está dentro de especificação. Propõem-se aqui duas estratégias alternativas, designadas estratégias qualitativa e quantitativa. Pretende-se com estas abordagens, avaliar duas metodologias alternativas para o acompanhamento e monitorização de processos de mistura, capazes de serem aplicadas em situações de produção, naturalmente nos casos em que os compostos relevantes são

ativos em NIR. Este trabalho baseou-se num produto farmacêutico existente no mercado mas por questões de confidencialidade, não é referido o seu nome nem a natureza e identidade dos compostos envolvidos na formulação.

1.1. Objetivos

Este trabalho teve como objetivo geral a análise de um processo farmacêutico de mistura de pós, com recurso à espectroscopia de infravermelho próximo (NIR).

Apresenta como objetivos específicos:

- avaliar duas metodologias alternativas (qualitativa e quantitativa) para o acompanhamento e monitorização de processos de mistura, tendo por base dois métodos de quimiometria;
- desenvolvimento de delineamentos experimentais que permitissem aplicar as metodologias e validar os modelos desenvolvidos;
- recolha de espectros NIR e análise através dos métodos analíticos certificados de todas as amostras produzidas, de forma a desenvolver e testar os modelos desenvolvidos;
- utilização da análise multivariada para tratamento de dados: análise de componentes principais (metodologia qualitativa) e análise pelos mínimos quadrados parciais (metodologia quantitativa) e
- avaliar comparativamente as duas metodologias.

1.2. Apresentação

A Generis® Farmacêutica, S.A. foi criada em Junho de 2002, iniciando a sua atividade de promoção e comercialização de medicamentos genéricos e tendo como compromisso disponibilizar o maior portefólio de genéricos de Portugal. É participada a 100% pelo Grupo Generis que opera há mais de 25 anos no mercado hospitalar português, sendo nº 1 no ranking das empresas nacionais em vendas hospitalares.

Encontra-se sediada na Rua João de Deus, 19, 2700-487- Amadora e possui 2 unidades fabris em Portugal, uma em Loures e outra em Venda Nova, com capacidade instalada de produção de 50 milhões de unidades.

É considerada a primeira empresa farmacêutica nacional exclusivamente vocacionada para o desenvolvimento, fabrico e comercialização de medicamentos genéricos, apresentando como principal objetivo o lançamento do 1º medicamento genérico de cada substância, promovendo a investigação própria no seu laboratório de desenvolvimento galénico. A Generis® pretende ainda, disponibilizar soluções terapêuticas de elevada qualidade e a um custo acessível.

Atualmente, a Generis® possui um valioso e invejável portefólio que compreende cerca de 250 moléculas registadas e aprovadas pelo INFARMED, permitindo assumir a liderança do mercado de genéricos desde 2011.

A Generis® é uma empresa do ramo farmacêutico que se preocupa em assegurar a qualidade, segurança, eficácia e credibilidade dos seus medicamentos junto do consumidor, procurando desenvolver técnicas que permitam simplificar aumentando em simultâneo a qualidade do controlo que faz, nomeadamente em termos de processo de produção.



1.3. Controlo de qualidade

A indústria farmacêutica sofreu grandes alterações nos últimos anos, com a globalização intensiva e com o aumento da competitividade. Os vários grupos e empresas do setor têm uma crescente necessidade de se diferenciar. A satisfação e confiança do consumidor são fatores importantes que têm recebido cada vez mais atenção durante a produção do medicamento. Na União Europeia, a indústria farmacêutica mantém elevados padrões de qualidade no que toca ao desenvolvimento, fabrico e controlo de medicamentos. Um sistema de autorizações de introdução no mercado assegura que todos os medicamentos são analisados por uma autoridade competente, a fim de garantir a conformidade com requisitos de segurança, qualidade e eficácia. Um sistema de autorização de fabrico garante que todos os produtos aos quais foi concedida autorização de introdução no mercado europeu são fabricados apenas por fabricantes autorizados, cujas atividades são regularmente inspecionadas pelas autoridades competentes. Todos os fabricantes de produtos farmacêuticos localizados na Comunidade Europeia necessitam de autorizações de fabrico, quer os produtos sejam vendidos dentro ou fora desta (*Comissão Europeia, 2008*). Para que um medicamento possa ser comercializado a nível mundial, as empresas farmacêuticas devem ter em conta as normas da Conferência Internacional de Harmonização dos Requisitos Técnicos de Medicamentos para Uso Humano (ICH), e as normas definidas na entidade reguladora de cada país para todo o ciclo de vida do medicamento (*ICH, 2015*).

A Autoridade Nacional do Medicamento e Produtos de Saúde (INFARMED I.P.) é a entidade reguladora dos produtos farmacêuticos em Portugal tutelada pelo Ministério da Saúde. Tem atribuições nos domínios da avaliação, autorização, disciplina, inspeção e controlo da produção, distribuição, comercialização e utilização de medicamentos de uso humano. Este, apresenta um manual de Boas Práticas de Fabrico de medicamentos, cujos princípios e diretrizes se aplicam a todas as operações, garantindo que os medicamentos são fabricados e controlados de acordo com as normas de qualidade apropriadas à utilização prevista e conforme mencionada na Autorização de Introdução no Mercado (AIM) (*Infarmed, 2015*).

A European Medicines Agency (EMA) é a entidade responsável pela avaliação dos pedidos de AIM a nível Europeu. Esta entidade desempenha igualmente um papel na promoção da inovação e da investigação na indústria farmacêutica. A EMA presta às empresas aconselhamento científico e assistência na elaboração de protocolos para o desenvolvimento de novos medicamentos e publica normas orientadoras sobre requisitos técnicos dos ensaios de qualidade, segurança e eficácia. Finalmente, a segurança dos medicamentos é monitorizada de forma constante pela EMA através de uma rede de farmacovigilância (*Comissão europeia, 2011*).

Face a isto, o controlo e garantia de qualidade na indústria farmacêutica assume uma importância notória, uma vez que a produção de um medicamento deve prioritariamente garantir a segurança e eficácia do mesmo. Estas garantias são apenas possíveis através da implementação de normas estritas relativas ao fabrico de medicamentos (cGMP). O controlo de qualidade é definido como sendo o conjunto de técnicas e atividades operacionais usadas para cumprir os requisitos da qualidade. Este conceito tem sofrido diversas alterações ao longo do tempo, sendo hoje muito mais abrangente. O

controle de qualidade está presente em várias fases do processo, desde a receção das matérias-primas, controle de processo, produto semi-acabado, produto acabado e estabilidade. Este controle inicia-se na fase de receção das matérias-primas pelo armazém onde as mesmas ficam de quarentena até confirmação dos atributos de qualidade. Se se confirmarem que as matérias-primas estão dentro de especificação seguem para outra zona do armazém onde estão prontas a serem utilizadas pelo Departamento de Produção. O Departamento de Produção inicia o processo de fabrico e ao longo do mesmo são retiradas e analisadas amostras de acordo com o estipulado na ordem de fabrico e requisitos de qualidade. No final do fabrico são analisadas as amostras que vão permitir a libertação do produto, ou seja, caso esteja dentro de especificações, o mesmo pode seguir para a embalagem. A embalagem é a última etapa do processo, embora o produto acabado também volte a ser analisado. A partir desta fase, são retiradas amostras que ficam armazenadas (estabilidade) e ao longo do tempo são realizados testes de controle de qualidade. Os testes que se fazem são em geral paramétricos, isto é, avalia-se determinado parâmetro e verifica-se se essa propriedade da forma farmacêutica/matéria-prima está dentro das especificações, ou seja, dentro do intervalo de valores admitido para essa propriedade de acordo com a farmacopeia. As farmacopeias, quer a Europeia (EP) quer a Portuguesa (FP9), apresentam monografias para os produtos estabelecendo quais os requisitos de qualidade assim com os testes admitidos.

1.4. Mistura de pós

A mistura de pós é um passo importante no fabrico de muitos produtos farmacêuticos, comprimidos, cápsulas, granulados e misturas de pós (*Berntsson, et al., 2002*). A mistura e homogeneização de pós é uma operação unitária que consiste na dispersão de pós de diferentes identidades de modo a que a probabilidade de encontrar uma partícula de um constituinte seja a mesma em todos os pontos da mistura, e igual à proporção em que esse constituinte se encontra na mistura. São vários os fatores que interferem com a homogeneidade da mistura, entre eles, a densidade do pó, tamanho e forma das partículas, a proporção dos diferentes componentes, tendência dos materiais a formar agregados, humidade e eficiência do misturador (*Gad, et al., 2008*). Existem diversos métodos para avaliar o desempenho dos misturadores de pó, sendo que a metodologia tradicional consiste na introdução de uma sonda ladrão dentro do misturador e é retirada uma amostra, enquanto que uma metodologia mais automatizada consiste na aplicação de um equipamento no misturador que permita a monitorização da mistura ao longo do tempo, p.e. sonda NIR (*Berntsson, et al., 2002*). Nos últimos anos, as indústrias farmacêuticas têm investido na melhoria dos métodos utilizados, optando pela automatização dos mesmos, sendo que a aplicação de uma sonda NIR no misturador para controlo da homogeneidade da mistura é dos mais comuns (*El-Hagrasy et al., 2001*).

Por vezes a mistura final não cumpre com os requisitos necessários, e o principal fator é a segregação de pós, que pode ocorrer de várias formas, desde a variação do tamanho de partícula e tempos de misturas superiores ao que eram esperados (*Gad, et al., 2008*).

1.5. Espectroscopia de infravermelho próximo

1.5.1. História

Foi William Herschel, cientista inglês, que em 1800 descobriu a região NIR, enquanto registava a quantidade de calor associada a cada cor que constitui a luz solar. Com recurso a um prisma de vidro transparente à radiação, que decompõe a luz solar, verificou com auxílio de um termómetro que existia um aumento acentuado de temperatura na zona acima da região vermelha do espectro, descobrindo assim, o que hoje se chama de radiação de infravermelho próximo (Almeida, 2009).

Karl Norris do Departamento de Agricultura dos Estados Unidos reconheceu o potencial desta técnica analítica e vulgarizou esta técnica como prática comum industrial. A partir de então, verificou-se um avanço notório como método de qualidade industrial e ferramenta de controlo de processo, através da introdução de técnicas de processamento eficientes de dados (quimiometria) e o desenvolvimento de novas configurações de equipamentos com base em sondas de fibra ótica (Reich, 2005).

Nos últimos anos, a espectroscopia NIR ganhou grande aceitação dentro da indústria farmacêutica na análise de matérias-primas, controlo de qualidade de produtos e monitorização de processo (Reich, 2005). O interesse crescente em produtos farmacêuticos é, provavelmente um resultado direto face às vantagens que apresenta relativamente a outras técnicas analíticas, isto é, a monitorização NIR é uma técnica não invasiva que minimiza os tempos de análise e a necessidade de preparação de amostras, não necessitando de pessoal qualificado para a realização das análises (El-Hagrasy, et al., 2001).

O principal interesse desta técnica é a possibilidade de obtenção de informação (espectro) que pode ser usada de forma versátil usando ferramentas matemáticas tipicamente de natureza multivariada. Um único espectro pode conter informação qualitativa e quantitativa, física e química, servindo como uma ferramenta útil na construção de um banco de dados dos atributos do produto e processo, levando a uma melhor compreensão do mesmo (Beer, et al., 2010; El-Hagrasy, et al., 2005). É uma técnica analítica viável para a avaliação do processo de mistura de pós, uma vez que, quase todos os compostos relevantes na indústria farmacêutica são ativos em termos de radiação NIR (Popo, et al., 2002; Wargo, et al., 1995; El-Hagrasy, et al., 2001). Atualmente na indústria farmacêutica, a espectroscopia NIR é a técnica analítica que se aplica mais como tecnologia analítica de processo (PAT).

1.5.2. Princípios teóricos

A espectroscopia no infravermelho (IV) estuda a interação da radiação infravermelha com a matéria. A radiação IV corresponde à região do espectro eletromagnético com números de onda entre 13000 e 10 cm^{-1} e situa-se entre a região da radiação visível e a região das micro-ondas.

O espectro eletromagnético IV encontra-se dividido em três zonas: infravermelho próximo 13000-4000 cm^{-1} , infravermelho médio 4000-200 cm^{-1} e infravermelho distante 200-10 cm^{-1} (Morisseau, et al., 1995). A absorção de radiação infravermelha é o resultado de transições entre estados vibracionais e rotacionais moleculares (torsão e flexão). Após a interação com a radiação IV, frações de radiação incidente são absorvidos em comprimentos de onda específicos (Morisseau, et al., 1995).

Uma das características do espectro de IV resulta do facto das ligações dos grupos funcionais absorverem radiação a frequências específicas (Pasquini, 2003). Para moléculas com vários grupos funcionais, observa-se que a frequência de absorção de um determinado grupo funcional específico é constante. Dado que a frequência de vibração dos grupos funcionais numa determinada molécula ocorrem a uma frequência característica, é possível construir tabelas de correlação entre os grupos funcionais e as frequências características de absorção (Figura 1) (Leitão, 2012).

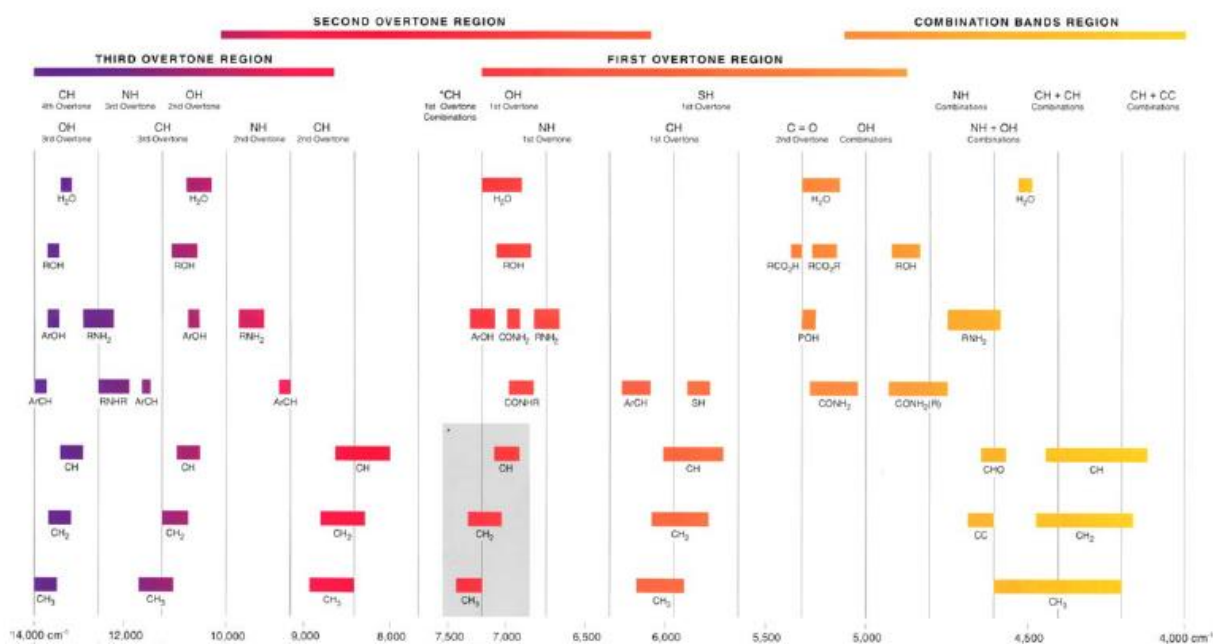


Figura 1 Bandas e posições relativas de picos de absorção em infravermelho próximo (cm^{-1}) (Mantanus, 2012).

Para um determinado composto apresentar um espectro de absorção na região do IV é necessário que a radiação incidente na molécula provoque a excitação dos modos vibracionais de forma a que a sua energia corresponda à diferença de dois níveis energéticos vibracionais (excitado e fundamental). Quando estas condições ocorrem, as substâncias dizem-se ativas no IV (Silva et al., 2007). As vibrações de uma molécula podem classificar-se em vibrações de alongamento (*stretching*), onde os átomos se movem sobre o mesmo eixo da ligação com alteração do comprimento da ligação, podendo ser simétricas ou assimétricas e em vibrações de deformação (*bending*) onde o movimento altera o

ângulo de ligação entre átomos. Os termos tesoura (*scissoring*), baloiço (*rocking*), meneio (*wagging*) e torção (*twisting*) são utilizados para descrever esse tipo de movimentos (**Figura 2**) (*Silva et al., 2007*).

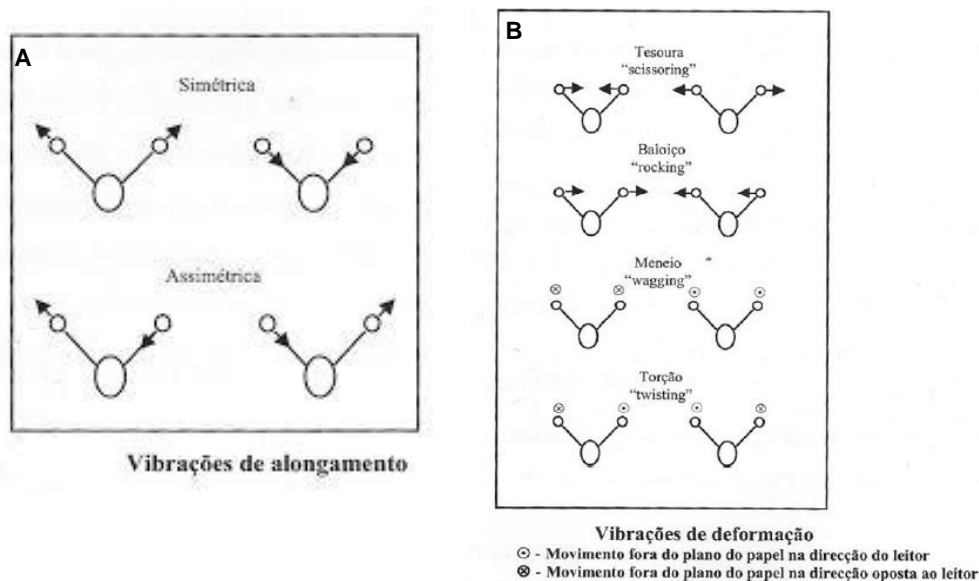


Figura 2 A) Vibrações de alongamento; B) Vibrações de deformação para uma molécula triatômica (adaptado de *Silva et al., 2007*).

Entre os vários modelos que podem explicar o comportamento vibracional das moléculas, destacam-se o modelo oscilador harmónico linear onde a distância entre cada nível se mantém, e o modelo oscilador anarmónico onde a distância vai diminuindo, sendo considerado o mais realista, uma vez que a ligação entre os átomos não é um oscilador perfeitamente harmónico e prevê a quebra de ligação da molécula (dissociação) ao se estender demasiado (*Almeida, 2009*).

O modelo do oscilador harmónico considera que a ligação entre átomos é feita por uma mola, em que a frequência de vibração da mola (ν) é descrita pela lei de Hooke e representada pela **Equação 1**.

$$\nu = \frac{1}{2\pi} \sqrt{\frac{\kappa}{\mu}}$$

Equação 1

Na **Equação 1**, κ representa a constante de força e ligação e μ a massa reduzida.

$$\mu = \frac{Mm}{m + M}$$

Equação 2

Na **Equação 2**, M e m representam a massa dos átomos presentes na ligação.

O modelo do oscilador anarmónico é descrito pela lei de Morse, no qual, caso aumente o número quântico vibracional podem-se verificar sobreposições e transições diferentes (designadas combinações), permitindo transições entre níveis não adjacentes (*Almeida, 2009*). Através da **Equação 3** podemos calcular a energia potencial do sistema para o modelo oscilador harmónico linear e para o modelo anarmónico, sendo que κ representa a constante de força, r a distância internuclear, r_e a distância internuclear de equilíbrio e x o deslocamento dos átomos.

$$E = \frac{1}{2}\kappa(r - r_e)^2 = \frac{1}{2}\kappa x^2$$

Equação 3

A importância da espectroscopia no IV resulta do facto das ligações dos grupos funcionais absorverem radiação a frequências características. Para moléculas com vários grupos funcionais observa-se que a frequência de absorção de um determinado grupo funcional específico é constante. Dado que a frequência de vibração dos grupos funcionais numa determinada molécula ocorrem a uma frequência característica, é possível construir tabelas de correlação entre os grupos funcionais e as frequências características de absorção (*Leitão, 2012*).

1.5.3. Espectrofotómetro de transformada de Fourier de infravermelho próximo (FT-NIR)

Quando pretendemos obter um espectro NIR, é imprescindível a utilização de um espectrofotómetro constituído por uma fonte luminosa, um sistema ótico, um porta amostras (ou acessório de amostragem como uma sonda) e um detetor de sinal. A fonte de luz divide-se em lâmpadas de quartzo, LED ou halogénio sendo estas últimas as mais utilizadas, pois são pequenas e robustas (*Leitão, 2012; Reich, 2005*). Os tipos de detetores incluem silício, sulfureto de chumbo (PbS) e índio-gálio-arsénio (InGaAs). Os detetores de silício são rápidos, baixo ruído, pequenos e altamente sensíveis a comprimentos de onda baixos. Já os detetores PbS são mais lentos, mas muito mais comuns pois são operacionais na região 9091-4000 cm^{-1} (comprimentos de onda altos) e fornecem boas propriedades de relação sinal-ruído. O detetor mais caro é o InGaAs e combina a velocidade e as características de dimensão do detetor de silício com a gama de comprimentos de onda do detetor de PbS (*Reich, 2005*). Dependendo da fonte de energia e do tipo de detetor, os instrumentos IR convencionais de laboratório podem operar tanto a regiões IR curtas, médias ou distantes (*Morisseau, et al., 1995*). Os instrumentos IV originais eram do tipo dispersivo, que tinham por finalidade a separação de frequências individuais da energia emitida da fonte de IV, com recurso a um prisma.

Em geral a espectroscopia NIR tem as seguintes características que a diferenciam de outras técnicas analíticas concorrentes:

- é uma técnica não-destrutiva;
- rápida;
- não requer em geral processamento da amostra;
- baixo custo por análise;
- precisa;
- boa relação sinal-ruído e
- simples mecanicamente podendo não ter partes móveis.

O espectrofotômetro de Transformada de Fourier foi desenvolvido com o intuito de ultrapassar as limitações verificadas nos instrumentos de dispersão, sendo considerada a principal dificuldade o processo de digitalização lento. Através da aplicação de um interferômetro conseguiu-se desenvolver um método para a medição de todas as frequências de IV simultaneamente, em vez de individualmente. O interferômetro produz um único tipo de sinal que tem todas as frequências de infravermelhos e permite medir o sinal muito rapidamente na ordem de um segundo ou mais. Assim, o elemento de tempo por amostra é reduzido para uma questão de segundos, em vez de vários minutos (TN, 2001). O interferômetro de Michelson é composto por dois espelhos e um divisor de feixe que transmite 50% da radiação incidente da fonte para o espelho móvel e os restantes 50% para o espelho fixo, ou seja, o feixe de luz que é transmitido pela fonte é dividido em 2 feixes de luz e cada feixe dirigido para um espelho diferente (Oliveira, 2011). Neste acessório divide-se e combina-se a radiação, para assim produzir uma interferência dependente do comprimento de onda. A radiação que sai do interferômetro é direcionada para o detector que mede a quantidade de energia em cada frequência que passou através da amostra, e é originado o interferograma (gráfico intensidade vs frequência) que contém informações relativas às frequências e características do espectro (Leitão, 2012; TN, 2001)).

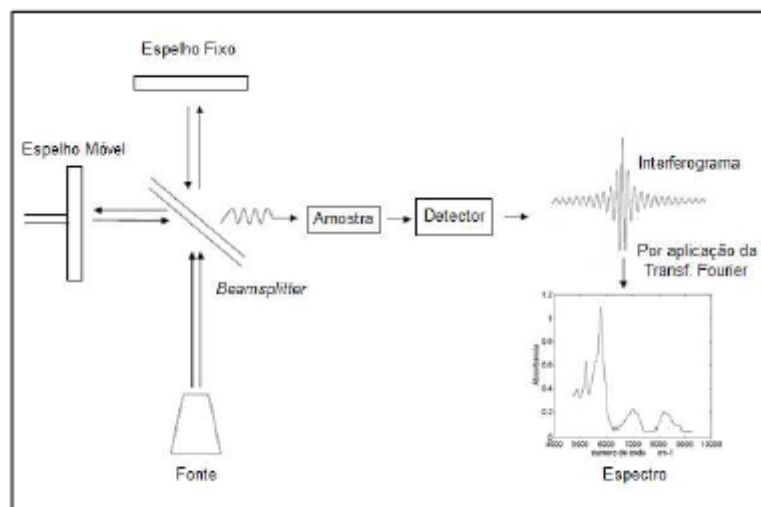


Figura 3 Esquema de espectrofotômetro FT-NIR com interferômetro de Michelson (Almeida, 2009).

A operação matemática da transformada de Fourier é depois aplicada, quer por ação de um microprocessador acoplado ao espectrofotômetro, quer por um computador associado externamente ao aparelho, distribuindo a energia absorvida pela amostra por cada comprimento de onda, apresentando os valores de cada frequência, resultando numa compreensão e leitura dos resultados simplificada (Leitão, 2012).

A espectroscopia de IV com Transformada de Fourier é preferida em relação a métodos de dispersão ou filtros (p.e. AOTF) devido a vantagens que tem nomeadamente: a vantagem de Jacquinot que consiste num maior débito de energia devido a uma abertura maior e a vantagem Fellgett (multiplexação) que deriva do processamento simultâneo ao longo de toda a gama espectral numa única medição. O equipamento mais difundido é o que tem um interferómetro de Michelson. Equipamentos com interferómetro de Michelson possuem também maior precisão e resolução que equipamentos dispersivos (vantagem de Connes).

As opções de leitura disponibilizadas pelos espectrofotômetros NIR dividem-se essencialmente nos modos de absorvância, transmitância e reflectância (Leitão, 2012). O método de análise utilizado neste trabalho foi a reflectância difusa sendo considerado um dos métodos mais comuns na análise de sólidos e líquidos viscosos (como geles), pois penetra na amostra e devolve informação sobre características físicas e químicas da mesma (Mantanus, 2012). A reflectância difusa, conhecida por DRIFTS (*Diffuse Reflectance Infrared Fourier Transform Spectrometry*), baseia-se na incidência da radiação na amostra sendo parte da radiação absorvida (alteração de estados vibracionais), e a fração que não é absorvida é refletida de volta (Leitão, 2012). Através do detetor, que recolhe a radiação refletida (radiação devolvida em todos os ângulos - e por isso designada por difusa), é possível o cálculo da intensidade da radiação que foi absorvida e assim poder identificar os componentes da amostra através do espectro obtido, uma vez que cada molécula tem um espectro característico na região do IV, tornando o método viável para identificar diferentes tipos de amostras (análise qualitativa). Os picos (ou mais frequentemente sobreposição de picos) presentes no espectro correspondem a frequências de vibrações entre os átomos que compõem cada amostra. A altura desses picos pode ser correlacionada com a quantidade de determinados compostos (análise quantitativa) (Real, 2014; Oliveira, 2011). A reflectância (R) é traduzida pelo quociente entre a intensidade da luz refletida (I_R), e a intensidade da luz incidente (I_0) (**Equação 4**)(Almeida, 2009).

$$R = \frac{I_R}{I_0}$$

Equação 4

Os espectros NIR podem ser apresentados em termos de absorvância, sendo que no caso de se efetuar uma leitura por reflectância difusa, deve ter-se em conta o seguinte cálculo: $A = \log\left(\frac{1}{R}\right)$ (Almeida, 2009). Em aplicações em processo, é usual efetuar-se as leituras recorrendo-se a sondas de fibra ótica que possibilitam o transporte da radiação NIR até ao local da medição e de volta para o detetor, com um mínimo de perda de qualidade do sinal (Mantanus, 2012).

1.6. Delineamento experimental

O delineamento experimental (DOE), é uma metodologia usada frequentemente em estatística para a análise e conceção sistemática e eficiente de experiências de forma a responder a perguntas específicas, sendo utilizado para analisar sistematicamente os diferentes tipos de problemas que surgem, por exemplo, na pesquisa, desenvolvimento e produção, fornecendo informações qualitativas e quantitativas. Para se poderem planear as experiências de forma racional, o problema tem de ser concreto e é necessário definir tanto as variáveis experimentais como as respostas, para que seja obtido o máximo de informação possível a partir de um número mínimo de experiências. No caso de uma mistura, as variáveis a ter em consideração são a proporção dos componentes em vez da sua quantidade absoluta (*Booker, 1998*).

Os três princípios do delineamento experimental são:

- replicação (para fornecer uma estimativa do erro experimental);
- aleatorização (para garantir que a estimativa é estatisticamente válida) e
- controlo local (para reduzir o erro experimental, tornando a experiência mais eficiente).

São inúmeros os exemplos da aplicação de delineamento experimental no contexto deste trabalho, sendo que os mais utilizados se encontram resumidos na **Tabela 1**.

Dentro dos delineamentos acima descritos o mais utilizado em situações similares à deste trabalho é o desenho de mistura, onde a resposta muda consoante as proporções dos componentes. Todos os componentes devem apresentar as mesmas unidades e o total deverá ser sempre o mesmo. Este tipo de delineamento pode ser de vários tipos (*Fisher et al., 1937*).

- **Simple**s (aplicável a todos os componentes que variam entre 0 e 100% (sem restrições). É necessário que a diferença entre a proporção máxima e mínima seja a mesma para todos os componentes).
- **Vértice** (especificar o grau m do polinómio (1- linear, 2- quadrático ou 3 cúbico). O desenho é então construído de $m+1$ com valores igualmente espaçados entre 1 e 0 (códigos de níveis do componente da mistura individual) sendo o número resultante de misturas dependente tanto do número de componentes (Q) como do grau de polinómio (m)).
- **Centróide** (por vezes inclui um desenho composto por 2-1 misturas distintas geradas por permutações de componentes puros (1, 0, ..., 0), misturas binárias (1/2, 1/2, 0, ..., 0), misturas ternárias (1/3, 1/3, 0, ..., 0) e assim sucessivamente até ao centróide global).
- **Ótimo** (usado quando a percentagem de componentes não é a mesma e consistem no tipo de modelos mais utilizado).

Tabela 1 Delineamento experimentais com maior utilização (Singh et al., 2011).

Tipo de delineamento	Descrição
Fatorial	Todos os níveis (X) são combinados com todos os fatores (K) dando origem a um número total de experiências de X^k .
D-ótimo	Desenhos personalizados gerados através de um algoritmo de maximização e um critério que tem em conta a quantidade de informação (determinante) na matriz de delineamento experimental.
Mistura	Utilizado para a produção de misturas com diversos componentes.
Composto central	Aplicado a problemas que requerem modelação não linear (p.e. que requerem modelos de segunda ordem). O delineamento contém embutido um desenho fatorial (2^k) ou desenho fatorial fracionado (2^{k-r}), adicionado de um grupo de ponto em estrela ($2K$) e um ponto central. O número total de combinações de fatores é dada por 2^k+2K+1 .
Box-Behnken	Requer apenas três níveis de cada um dos fatores, sendo uma alternativa ao desenho composto central

1.7. Quimiometria

Nos últimos 20 anos, houve um avanço notório da informática em termos de *hardware* e *software* que proporcionaram uma grande evolução na forma como os dados são adquiridos e processados, tornando possíveis análises complexas. A quimiometria surgiu devido à necessidade de tratar estas grandes quantidades de dados e poder extrair informações relevantes (Mark *et al.*, 2007). A quimiometria, segundo Kowalski é "... uma disciplina química que emprega métodos matemáticos e estatísticos para planejar ou selecionar experiências de forma otimizada e fornecer o máximo de informação química através da análise dos dados obtidos (Frank, *et al.*, 1982).

Os dados obtidos têm de ser tratados, analisados e correlacionados com a propriedade pretendida. Neste contexto há necessidade de desenvolver técnicas de calibração capazes de correlacionar de modo robusto os dados obtidos com as propriedades de interesse. São várias as técnicas disponíveis para o tratamento de dados espectrais, sendo impossível definir uma regra para qualquer estudo envolvendo dados de espectroscopia NIR. Dentro das várias técnicas possíveis, as mais conhecidas são a análise por componentes principais (PCA), mínimos quadrados clássicos (CLS), regressão linear múltipla (MLR), regressão por componentes principais (PCR) e regressão por variáveis latentes ou mínimos quadrados parciais (PLS), sendo esta última a técnica de regressão mais utilizada (Mark *et al.*, 2007).

1.7.1. Pré-tratamento espectral

O pré-tratamento é definido como qualquer manipulação matemática antes de qualquer análise preliminar de espectros, sendo que a principal finalidade é reduzir fontes de variação indesejáveis, que não estão relacionadas com os parâmetros de interesse, por exemplo o ruído instrumental. O tratamento dos dados é aplicado antes das fase de modelação (p.ex. calibração) para melhorar a eficiência do modelo, originando modelos com um melhor desempenho. Contudo, esta etapa de pré-tratamento deve ser realizada com algum critério, uma vez que, informações espectrais relevantes podem ser inadvertidamente atenuadas ou mesmo removidas. A escolha do tipo de pré-tratamento a aplicar aos espectros é essencial para o sucesso de qualquer análise (Maleki *et al.*, 2007). Dentro da gama de pré-tratamentos mais frequentemente utilizados, os mais relevantes são a aplicação de derivadas (p. ex. filtro de Savitzky-Golay), variação de padrão normal (SNV) e correção multiplicativa de sinal (MSC). Existem muitos outros mas estes são os que têm tido mais aceitação no tratamento de espectros NIR.

1.7.1.1. Derivadas (Savitzky-Golay)

Este tratamento consiste na derivação do sinal. Pode por isso utilizar derivadas de ordem várias. O método permite um aumento de resolução espectral e eliminação dos efeitos aditivos, dando maior ênfase aos picos do espectro (Savitzky *et al.*, 1964). A primeira derivada evidencia características

pequenas face a características com maior intensidade através da remoção dos efeitos aditivos (linha de base do espectro), facilitando a análise do espectro. No caso da segunda derivada, os mesmo sinais podem ser estudados, mas tem de ser aplicada a regiões espectrais muito restritas devido ao ruído espectral poder ser amplificado neste processo (Naes et al., 2002).

1.7.1.1. Correção multiplicativa de sinal

O método de correção multiplicativa de sinal (MSC) foi desenvolvido para reduzir o efeito do fenómeno de dispersão de luz em espectros NIR obtidos essencialmente por reflectância difusa, transmissão e transfectância (um modo híbrido) (Burns et al., 2008). Este método simplifica nomeadamente modelos de calibração (por variáveis latentes) através da redução do número de componentes necessários e aumenta a sua linearidade, sendo que os espectros transformados são semelhantes aos espectros originais (Naes et al., 2002; Maleki et al., 2007). A MSC baseia-se no cálculo de um espectro médio referente a todos os espectros presentes num conjunto. Posteriormente, cada espectro da amostra i (X_i) é transformado de acordo com a **Equação 5**, onde X'_i representa o espectro transformado para a amostra i e os coeficientes u e v são escolhidos de modo a que a diferença entre o espectro transformado e o espectro médio seja mínima (Conzen, 2006).

$$X'_i = u + vX_i$$

Equação 5

Este tipo de tratamento é considerado uma transformação dependente, pois se o conjunto de dados é alterado, o espectro médio também vai sofrer alterações e por sua vez os espectros corrigidos por MSC terão de ser recalculados (Jorgensen, 2000).

1.7.1.2. Variação de padrão normal

A variação de padrão normal (SNV) é um método que remove a variação da inclinação dos espectros causada pela dispersão de luz normalmente associada à presença de partículas em suspensão ou amostras sólidas com partículas de tamanho diverso. A transformação é aplicada a cada espectro individualmente através da subtração do valor médio do espectro e divisão pelo seu desvio-padrão, tendo como objetivo centrar cada espectro em torno de zero e dar variância unitária (Jorgensen, 2000; Wise et al., 2006). O cálculo é feito com base na **Equação 6**, onde X_i representa o espectro da amostra i , m_i é a média do espectro para a amostra i e S_i o desvio padrão para a mesma amostra (Conzen, 2006).

$$X'_i = \frac{(X_i - m_i)}{S_i}$$

Equação 6

Este pré-tratamento tem um efeito semelhante ao MSC, sendo que a principal diferença é a normalização de cada espectro utilizando apenas dados do próprio espectro (Naes *et al.*, 2002).

1.7.2. Análise multivariada

Na presença de sinais com múltiplas variáveis (como no caso dos espectros NIR) pode ser necessária a aplicação de análise multivariada, para se obter uma resposta eficaz e um eficaz aproveitamento da informação presente no sinal. São vários os algoritmos que permitem fazer análise multivariada, entre eles os já referidos métodos PCA e PLS, que serão descritos mais à frente em maior detalhe. De notar que adicionalmente à necessidade de transformar (ou pré-processar) os espectros com os métodos referidos anteriormente, a aplicação de ferramentas como o PCA ou regressão PLS requer que o conjunto de espectros seja normalizado. Esta normalização pode simplesmente consistir em subtrair a cada espectro a usar na análise o espectro médio obtido desse conjunto (centro na média ou *mean centring*) ou adicionalmente dividir cada elemento do espectro pelo desvio padrão calculado sobre todos os espectros do conjunto (auto escalonamento ou *auto-scaling*). Neste trabalho foi sempre usado o método de centro na média.

1.7.2.1. Análise de componentes principais

Esta técnica constitui a base para o entendimento de muitas das técnicas de análise multivariável. A análise de componentes principais (PCA) consiste num método estatístico que procura calcular, a partir de um conjunto de dados que contém muitas variáveis, um subconjunto que contenha o máximo de informação presente dos dados originais, permitindo desta forma minimizar a dimensão desse conjunto. A informação “residual” é aquela que não está presente no subconjunto gerado, mas está presente no conjunto original, comprometendo o modelo PCA originado a perder informação relevante.

Em termos de espectros, este método consiste na transformação da matriz de dados espectrais (matriz X), onde n linhas correspondem às observações (ou amostras) e as colunas k são as variáveis (ou comprimentos de onda). A matriz X é decomposta no produto dos scores (T) e loadings (P). O remanescente fica na matriz que contém os resíduos (E) (**Equação 7**).

$$X = TP^t + E$$

Equação 7

A principal função do método PCA é alterar o espaço original dos dados num espaço de dimensão reduzida tentando em simultâneo preservar a informação (variabilidade) presente originalmente. Com esta transformação pretende-se não apenas simplificar a análise mas também eliminar informação redundante (p.e. presença de variáveis correlacionadas) ou não sistemática (p.e. ruído ou presença de outliers). Às novas componentes dá-se o nome de componentes principais (CP). (Esbensen *et al.*,

1987). Depois de aplicado o modelo de PCA, estas novas variáveis são combinações lineares das variáveis originais. São vários os critérios que permitem determinar o número ideal de componentes principais, entre eles a percentagem de variância explicada, o critério do valor próprio, *scree-test* e a validação cruzada. Com o aumento do número de componentes principais, a informação retida também vai aumentar e por sua vez os resíduos vão diminuir (Naes et al., 2002). Os scores de cada amostra podem ser representados graficamente e permitem visualizar a distribuição das amostras eventualmente identificando características semelhantes assim como amostras atípicas (*outliers*). Os *outliers* devem ser confirmados através da análise das estatísticas dos resíduos e Hotelling T² (Naes et al., 2002).

1.7.2.2. Regressão por mínimos quadrados parciais

A regressão por mínimos quadrados parciais (PLS) foi desenvolvido na década de 1960 por Herman Wold como uma técnica econométrica para monitorizar e calibrar processos industriais. É um método utilizado para relacionar duas matrizes de dados, X e Y, correlacionando-as de modo a obter uma relação linear, baseada no uso de variáveis latentes (Wold et al., 2001). Neste modelo as informações espectrais e as concentrações (proporções dos componentes) são usadas ao mesmo tempo, na fase de calibração (Morgano et al., 2002). É um método recomendado para desenvolvimento de regressões em casos em que o número de variáveis é alto (como no caso dos espectros). As matrizes contendo os espectros dos padrões (X) são relacionadas com os vetores que armazenam dados sobre a sua concentração (Y), originando as **Equação 8** e **Equação 10**. As matrizes X e Y são decompostas nos scores (T e U), nos loadings (P e Q) e nos resíduos (E e F), tal como representado na **Equação 8** para os dados espectrais e na **Equação 9** para os dados das concentrações.

$$X = TP^t + E \quad \text{Equação 8}$$

$$Y = UQ^t + F \quad \text{Equação 9}$$

O resultado é uma equação linear de uma curva de calibração semelhante à **Equação 10**, onde \tilde{X} é a matriz contendo os dados relativos a uma série de amostras (p.e. novos espectros), \tilde{Y} é uma matriz com a concentração dessas amostras, B_{PLS} é uma matriz contendo os parâmetros do modelo (coeficientes de regressão) e \tilde{E} é a matriz dos resíduos (Morgano et al., 2002). No caso de existir apenas um analito (ou variável dependente) então \tilde{Y} , B_{PLS} e \tilde{E} são vetores.

$$\tilde{Y} = \tilde{X}B_{PLS} + \tilde{E} \quad \text{Equação 10}$$

Quando se pretende prever apenas uma variável dependente tem-se um modelo PLS1 e y é um vetor. Caso se pretendam prever várias variáveis dependentes o algoritmo a considerar é o PLS2 e Y é uma matriz (Naes et al., 2002).

No caso de modelos com base em espectros, a escolha da região espectral mais favorável deve ser efetuada com recurso a um dos muitos algoritmos disponíveis para seleção de variáveis (*Gauchi, et al., 2001*).

A construção do modelo de regressão PLS pode ser dividido em três etapas (*Conzen, 2006*):

- **Calibração:** na qual são utilizados os espectros das misturas do conjunto de calibração para a construção de um modelo matemático que melhor se ajuste aos dados espectrais e aos valores de referência do parâmetro desejado.
- **Validação:** permite verificar e otimizar o modelo. Nesta etapa é avaliado o erro de previsão relativo a amostras não usadas na calibração. O método pode usar validação interna (as amostras do conjunto de calibração são usadas para calibrar e validar, designado como validação cruzada) ou validação externa (amostras externas ao conjunto de calibração).

A validação cruzada é um dos métodos mais utilizados para ajudar a definir o número ideal de variáveis latentes (em PCR ou PLS) ou componentes principais (em PCA). Um dos métodos de validação cruzada mais utilizados, o *leave-one-out*, deixa uma amostra de fora, de uma série de n amostras, de seguida $n-1$ amostras são usadas na calibração e realiza-se a previsão da concentração da amostra que não foi utilizada na calibração. Este processo repete-se até que todas as amostras tenham sido excluídas. No final comparam-se os valores previstos com os valores de referência e verifica-se qual o modelo que originou um menor erro. Este erro pode ser calculado através da raiz quadrada do erro quadrado médio de validação cruzada (RMSECV) que é calculado de acordo com a **Equação 11**. Na **Equação 11**, n é o número de amostras de calibração, $y_{i(\text{referência})}$ o valor de referência da amostra i e $y_{i(\text{previsto},cv)}$ o valor previsto para a amostra i determinado pelo modelo de acordo com o método de validação cruzada escolhido (*Naes et al., 2002*).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_{i(\text{referência})} - y_{i(\text{previsto},cv)})^2}{n - 1}} \quad \text{Equação 11}$$

No processo de validação existem dois fatores a ter em conta: o número de variáveis latentes ou fatores a serem utilizados pelo modelo e a existência de *outliers*. O número de variáveis latentes será o correspondente ao menor valor de RMSECV. Mas nem sempre o método que apresenta um menor erro é o mais vantajoso, uma vez que pode utilizar um número de fatores em excesso. Relativamente aos *outliers* é importante eliminá-los pois são amostras que não se enquadram no restante conjunto e não devem ser consideradas (contudo, eliminar amostras apenas com justificação plausível).

Previsão: fase em que se avalia o poder preditivo do modelo construído utilizando amostras do conjunto que não participaram na etapa de calibração. É necessário verificar o modelo com um conjunto externo, onde se utilizam amostras de concentrações conhecidas, que pertencem a um conjunto distinto do conjunto da calibração. Nesta etapa também pode ser calculado um erro de forma similar ao RMSECV (neste caso RMSEP, sendo o P de previsão) (Naes *et al.*, 2002).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_{i(\text{referência})} - y_{i(\text{previsto})})^2}{n}}$$

Equação 12

O RMSEP é uma medida da incerteza sobre futuras previsões. Este valor (um para cada resposta) é uma medida da exatidão que pode ser esperada quando o modelo é usado em novas amostras (Conzen, 2006).

2. REVISÃO BIBLIOGRÁFICA

A espectroscopia NIR tem sido aplicada nas mais diversas áreas, sendo que na indústria farmacêutica tem havido um grande avanço na aplicação desta tecnologia em aplicações de controle de processo nomeadamente nas enquadradas em PAT. Cada vez mais a espectroscopia NIR é utilizada para o controle de processo de mistura e assim criar modelos que permitam definir a homogeneidade da mistura, o conteúdo da mesma e por fim definir tempos de mistura ótimos (ponto final da mistura). Na literatura são vários os trabalhos publicados neste contexto, mas em produtos/condições diferentes. Exemplos são os trabalhos de *El- Hagrany et al., 2001*; *Wargo et al., 1995*; *Berntsson et al., 2001*; *Wu et al., 2009*; *Shi et al., 2008* e *Sekulic et al., 1996*. Neste sentido a espectroscopia do infravermelho próximo apresenta-se com grande potencial como ferramenta analítica para a análise de processos de mistura de pós. *D'Amico et al., 2001* abordaram o efeito da humidade, concentração da substância ativa e velocidade do misturador como fatores críticos que afetam a mistura num misturador em V. Recorreram a um delineamento experimental, onde concluíram que tanto a humidade como a concentração de substância ativa vão influenciar no tamanho de partícula. *Plugge et al., 1992* destacam a importância da espectroscopia NIR como uma ferramenta poderosa em laboratório de controle de qualidade na indústria farmacêutica.

3. MATERIAIS E MÉTODOS

4.1 Produto comercial e processo de fabrico industrial

O produto comercial em que este trabalho se baseia é produzido na Generis Farmacêutica, S.A. É composto por sete componentes, que por questões de confidencialidade vão ser referidos por componentes A, B, C, D, E, F e G, sendo as respetivas proporções m/m (%) representadas na tabela abaixo. O produto será doravante referenciado como P.

Tabela 2 Composição do medicamento P.

Componente	Proporção m/m (%)
A	72,285
B	9,543
C	6,091
D	10,761
E	0,183
F	0,122
G	1,015

O processo de fabrico industrial do produto P inicia-se com a calibração prévia de cada componente num granulador oscilante com rede de 0,8 mm abertura luz e 0,5 mm diâmetro de fio. O componente F é triturado a pó fino num almofariz antes de seguir para o granulador. Segue-se o processo de mistura num misturador em V com capacidade para 1000 litros, sendo que este processo se divide em três misturas distintas. A mistura 1 com duração de 20 minutos e 30 minutos para as restantes duas misturas. Como mostra a **Figura 4**, os componentes são adicionados em momentos diferentes às misturas assim como as quantidades do componente A são repartidas nas três misturas, pois este é um componente maioritário e espera-se que a mistura final se encontre homogénea. No final a mistura segue para fase de embalagem, onde é acondicionada corretamente. Durante a fase da embalagem primária, são retidas amostras para análise no controlo de qualidade, designadas neste trabalho por amostras industriais.

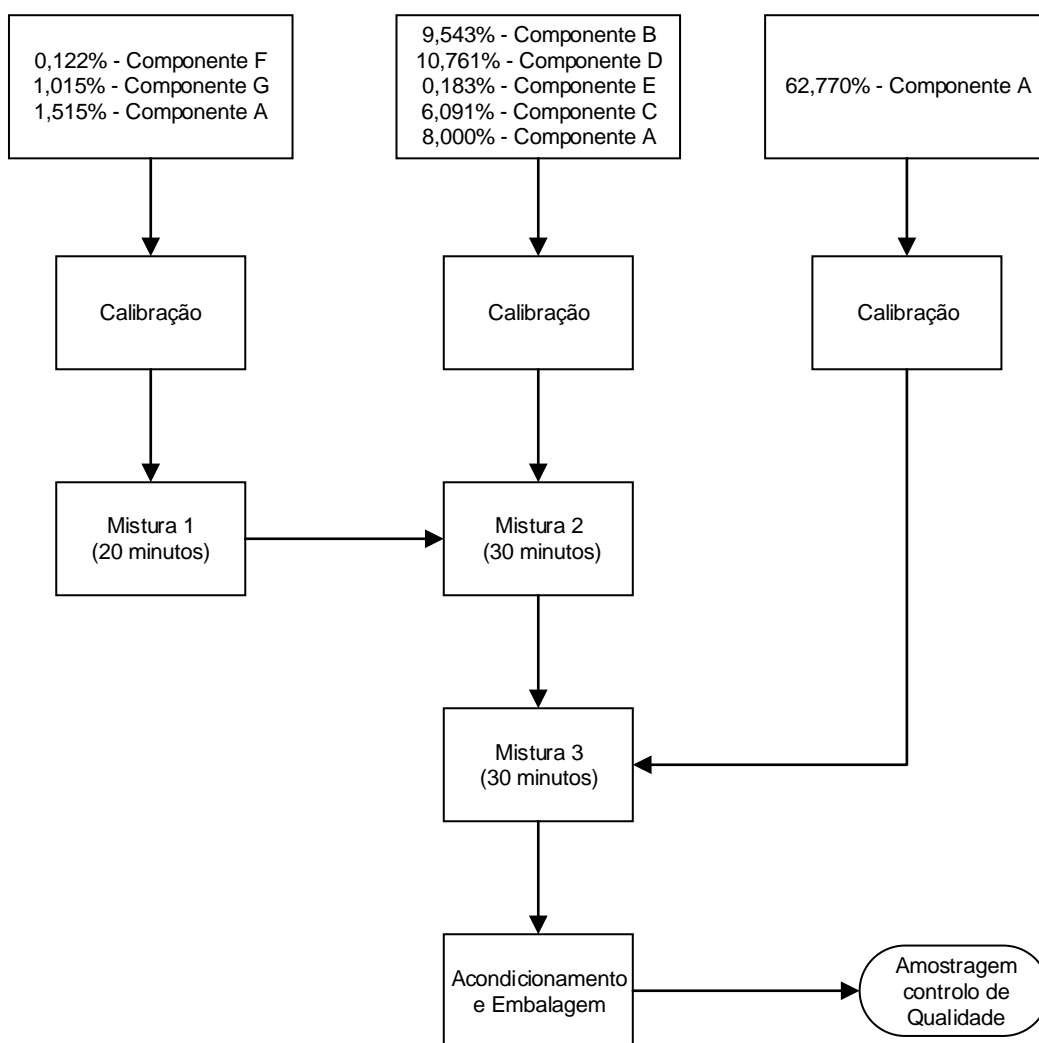


Figura 4 Fluxograma ilustrativo do processo de fabrico do produto P.

4.2 Doseamento e amostragem do processo industrial

O doseamento dos componentes da mistura foi efetuado através dos métodos certificados e feito no laboratório de Controlo de qualidade da Generis SA. São doseados apenas os componentes A, B, C e D, uma vez que estes componentes representam mais de 98% da mistura final. Não é expectável que o NIR detete componentes vestigiais, não se justificando o doseamento dos restantes componentes. Os componentes B e C são bastante semelhantes em termos químicos, sendo que durante o doseamento é doseado um componente H que pertence aos componentes B e C. Relativamente ao componente D, é apenas doseado o ião correspondente. Os métodos de referência por questões de confidencialidade não serão apresentados detalhadamente.

Abaixo está representada uma tabela com as especificações em razão massa/massa(%) dos respectivos excipientes doseados.

Tabela 3 Especificações no produto comercial P dos componentes A, B, C, D, H, tendo por base os métodos de referência.

Componente	Especificação (%m/m)
A	68,63 – 75,94
B	5,36 – 5,93
C	3,03 – 3,35
D	7,35 – 8,12
H	8,26 – 9,12

4.3 Amostras

Foi reunido um total de nove amostras industriais. Destas, oito estavam dentro de especificação e uma correspondia a um lote rejeitado. As oito amostras dentro de especificação são identificadas de 1-8 e a amostra fora de especificação como R1.

Fez-se ainda a reprodução de dois lotes à escala laboratorial com massa total de 600 gramas, num misturador em V com capacidade de 2 litros, onde se monitorizou a mistura 3 ao longo do tempo, pois é a única mistura que tem a totalidade dos componentes. Esta análise foi feita através de leituras NIR onde a cada três minutos foi inserida a sonda no braço direito do misturador. No final da mistura 3, foram realizadas leituras NIR em ambos os braços do misturador, designadas por 30 min DRT (braço direito) e 30 min ESQ (braço esquerdo) e foi retirada uma amostra do misturador (30 min Final) para análise no controlo de qualidade (doseamentos) e espectrofotómetro NIR (amostras referenciadas como Lote 1 e Lote 2).

Realizaram-se dois DoEs diferentes tendo por base o software MODDE versão 10.1 (Umetrics, Suécia). Pesaram-se as respetivas quantidades para um copo de amostragem perfazendo uma massa total de 50g. Essa mistura de pós contida dentro do copo de amostragem foi agitada durante 5 minutos. Em ambos os DoEs fez-se variar apenas os componentes maioritários (excipientes A, B, C e D), sendo que num primeiro DoE, designado como DoE-QLT (DoE utilizado para a abordagem qualitativa) a gama de variação foi de $\pm 1,5\%$ em relação ao valor teórico das proporções no produto comercial. No segundo DoE, utilizado para a abordagem quantitativa (DoE- QNT), a variação foi de $\pm 10\%$ em relação ao valor teórico das proporções no produto comercial.

4.4 Espectroscopia NIR

Os espectros NIR foram obtidos em modo de reflectância difusa, num espectrofotómetro Thermo Nicolet FT-NIR modelo ANTARIS, com detetor InGaAs, intervalo espectral entre os 10000 e os 4000 cm^{-1} (1000 - 2500 nm), resolução de 4cm^{-1} e considerada a média de 36 leituras e controlado pelo software RESULT (Thermo Nicolet, EUA). As leituras foram efetuadas por intermédio de uma sonda de reflectância difusa modelo SabIR (Solvias, Suíça) com área de iluminação de ca. de $0,2\text{ cm}^2$. Todas as amostras foram medidas em duplicado e considerado a média desses espectros. Antes de iniciar as aquisições espectrais e com a periodicidade de 1 hora foi feito um branco com Teflon (spectralon®).

Foram analisadas todas as amostras referidas no ponto anterior assim como cada componente presente na mistura (espectros puros).

4.5 Tratamento quimiométrico

O tratamento quimiométrico foi feito com recurso ao software MATLAB, versão 8.3 (MathWorks, Natick, MA) e PLS Toolbox, versão 7.5 (Eigenvector Research Inc., Wenatchee, WA).

RESULTADOS E DISCUSSÃO

3.1. Análise dos espectros dos componentes puros

Numa primeira fase procedeu-se à análise dos espectros dos componentes puros A, B, C e D da mistura. Fez-se uma análise preliminar dos espectros (identificação dos principais picos e bandas) e suas principais características por forma a poder inferir sobre a atividade de cada componente na zona NIR assim como a maior ou menor seletividade de cada componente na mistura.

Nas figuras seguintes estão representados os espectros dos componentes A, B, C e D, identificação dos picos/bandas principais e interpretação.

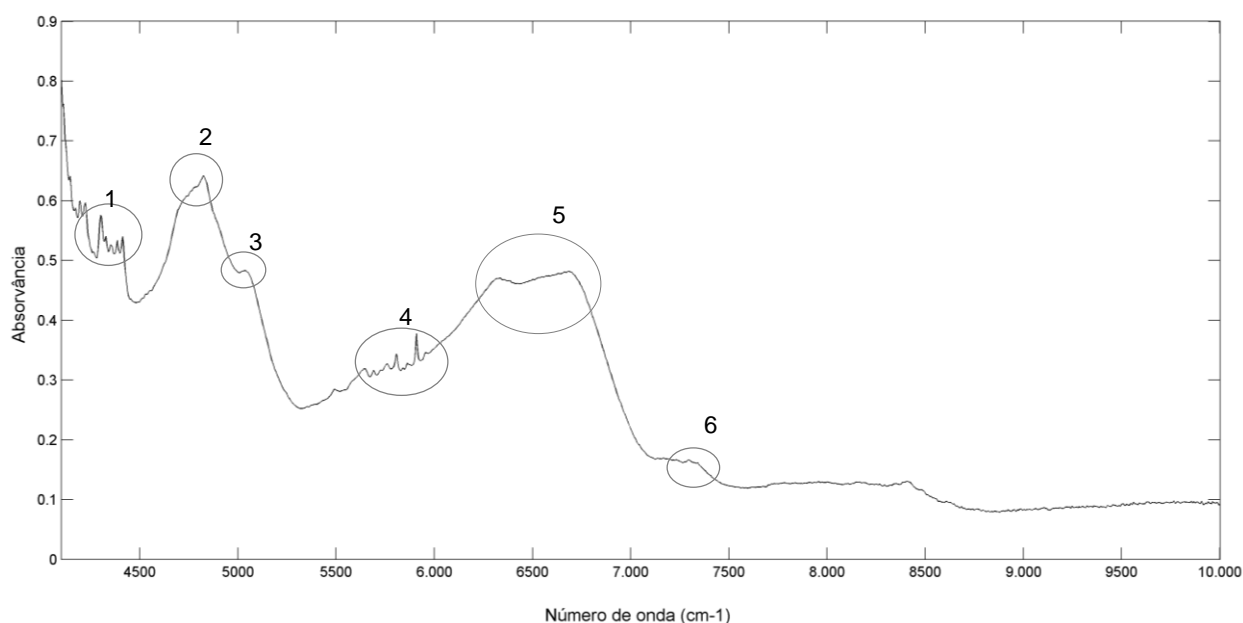


Figura 5 Espectro NIR obtidos em reflectância difusa do componente A.

Tabela 4 Identificação das bandas no espectro do componente A.

Número	Zona espectral (cm ⁻¹)	Tipo de vibração
1	4223	CH; CH ₂
2	4825	ROH
3	5033	OH
4	5909	CH; CH ₂
5	6680	CH
6	7293	H ₂ O

O espectro de absorção correspondente ao componente A é composto por vários picos sendo que os principais são os picos de absorção da água (H₂O), carbonos primários (CH), carbonos secundários (CH₂) e álcool (ROH).

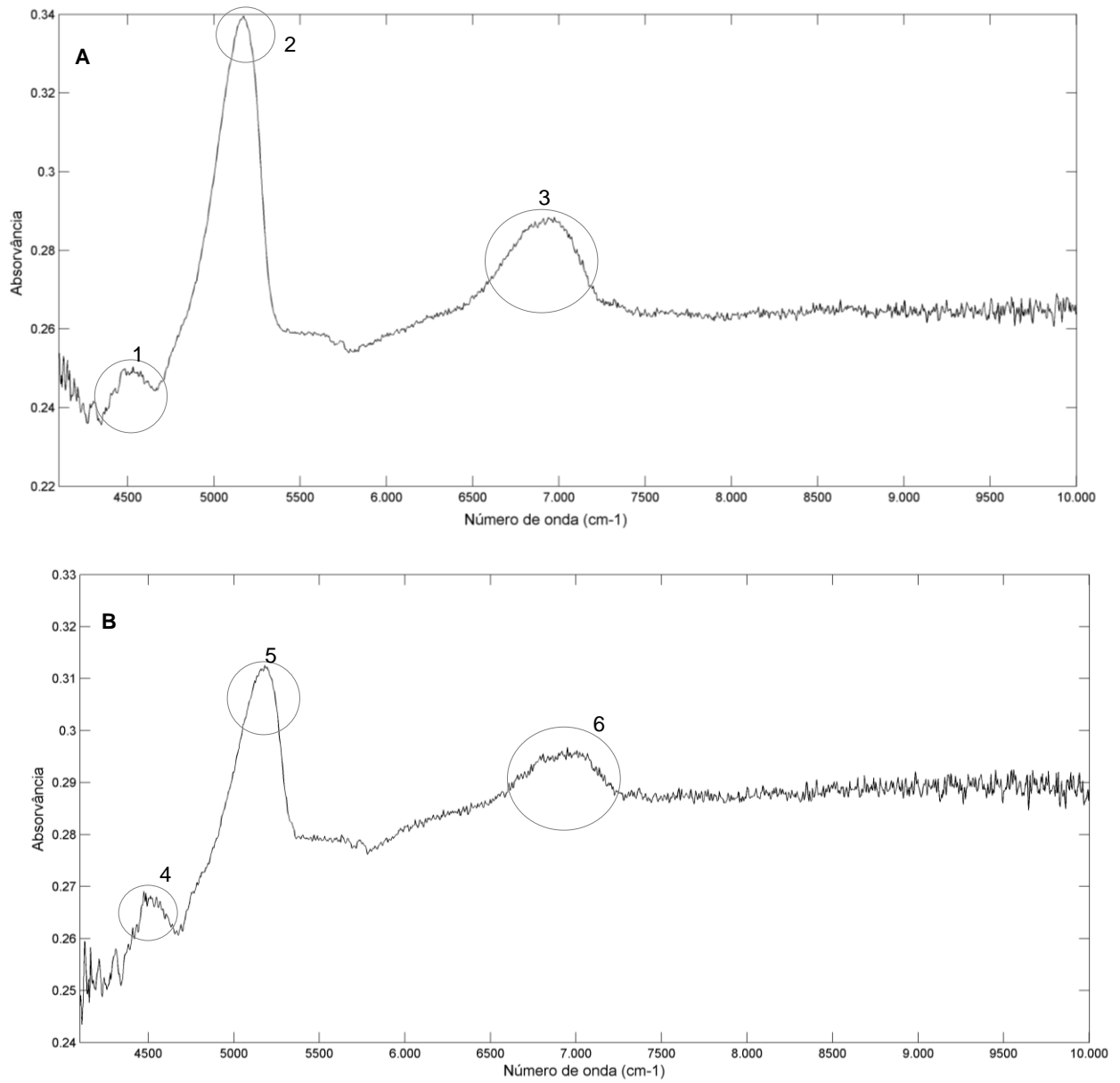


Figura 6 Espectros NIR obtidos em reflectância difusa dos componentes B (A) e C (B).

Tabela 5 Identificação das bandas nos espectros dos componentes B e C.

Número	Zona espectral (cm ⁻¹)	Tipo de vibração
1	4532	H ₂ O
2	5172	
3	6973	
4	4474	
5	5180	
6	6950	

Ambos os espectros da **Figura 6** apresentam algum ruído no início e fim. Estão representados apenas três picos que correspondem a zonas de absorção da água (H₂O). Os espectros dos componentes B e C são bastante semelhantes e variam essencialmente na intensidade de absorção, sendo que picos maiores significam que o composto apresenta maior quantidade de água (ou maior grau de hidratação). Desta forma, o componente B apresenta uma maior quantidade de água quando comparado com o componente C.

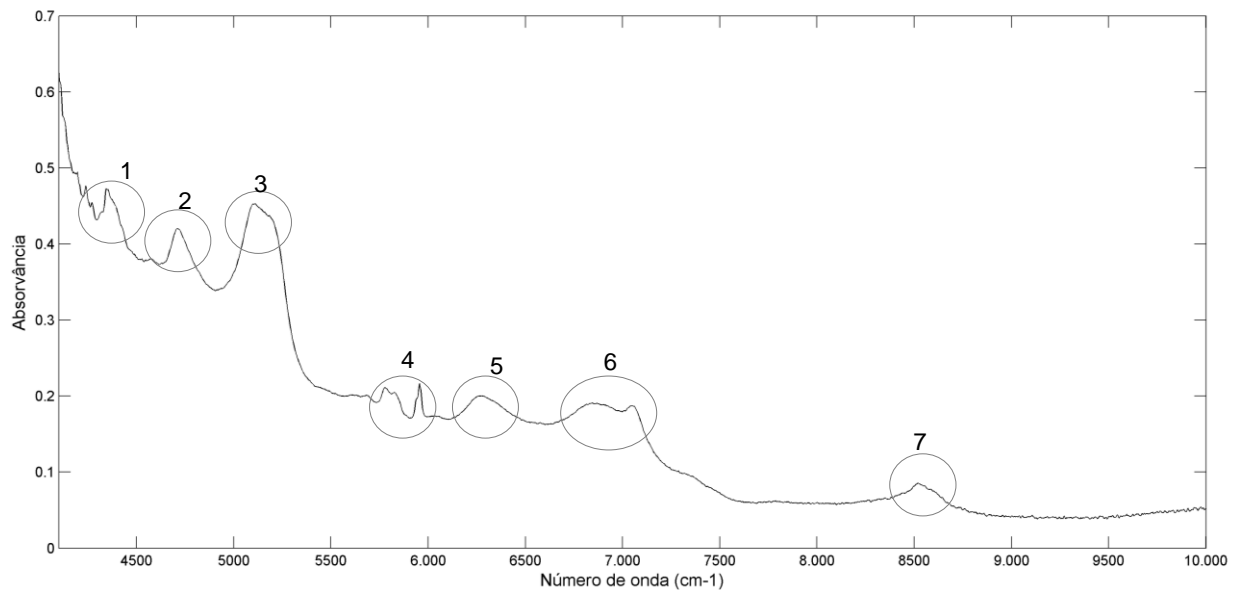


Figura 7 Espectro NIR obtidos em reflectância difusa do componente D.

Tabela 6 Identificação das bandas no espectro do componente D.

Número	Zona espectral (cm ⁻¹)	Tipo de vibração
1	4343	CH ₂
2	4709	ROH
3	5110	H ₂ O
4	5955	CH ₂
5	6275	CH ₂
6	6842	ROH
7	8520	CH ₂

Relativamente ao componente D, o espectro de absorção apresenta mais picos que os restantes, sendo que os mesmos correspondem a picos de absorção da água (H₂O), de carbonos secundários (CH₂) e de álcoois (ROH).

Na **Figura 8** estão representados todos os espectros dos componentes puros, onde se pode verificar que os componentes A e D se aproximam um do outro, enquanto que os componentes B e C são muito semelhantes, apresentando os mesmos picos, tal como descrito anteriormente.

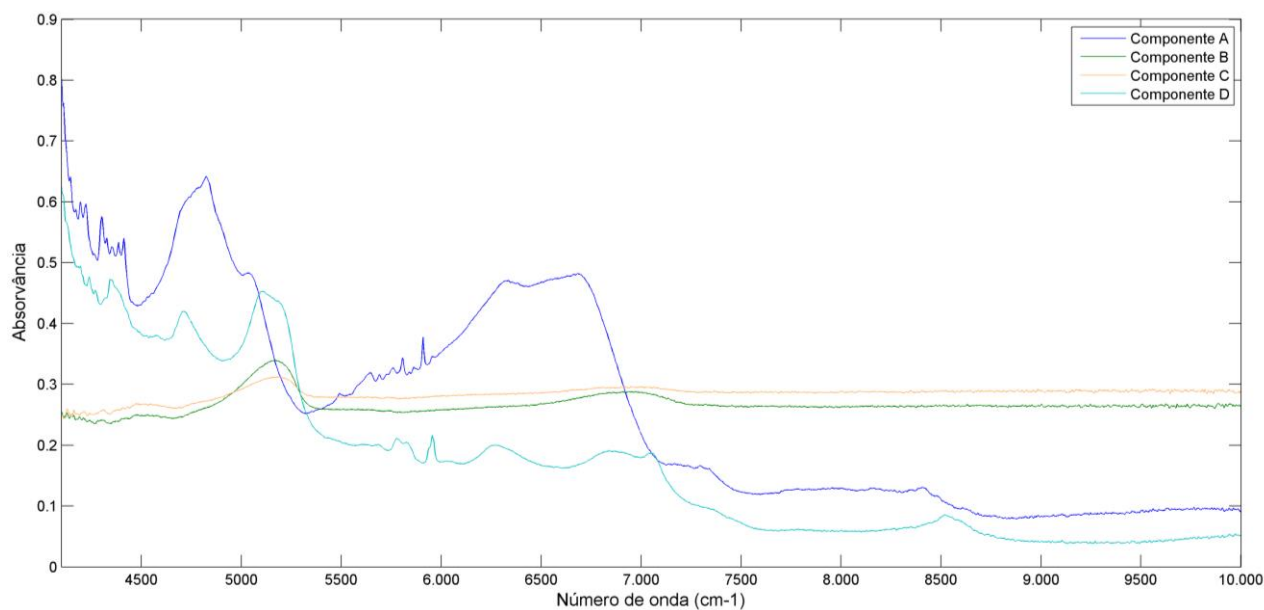


Figura 8 Sobreposição dos espectros NIR dos componentes puros (componentes A, B, C e D).

3.2. Controlo estatístico do processo de mistura

Para a análise dos diversos dados obtidos foram aplicadas dois métodos de análise distintos. A análise qualitativa através da construção de modelos PCA e a análise quantitativa através de modelos PLS. Para ambos os modelos, foram gerados DoEs diferentes. Na primeira análise gerou-se um DoE-QLT, onde a gama de variação foi menor, quando comparado com DoE-QNT utilizado para a construção de modelos PLS (análise quantitativa). Nas duas análises foram projetadas os espectros NIR obtidos das amostras industriais e laboratoriais, sendo os resultados comparados com os valores obtidos pelos métodos de referência (**Tabela 7**).

Tabela 7 Resultados obtidos através dos métodos de referência para as amostras industriais.

Número de ensaio	Proporção de cada componente (%m/m)				
	A	B	C	D	H
1	72,15	5,56	3,20	7,45	8,69
2	71,27	5,44	3,11	7,37	9,00
3	71,47	5,62	3,23	7,47	8,99
4	74,12	5,69	3,25	8,11	8,83
5	71,50	5,69	3,29	7,69	8,30
6	69,44	5,60	3,23	8,03	8,71
7	69,24	5,50	3,17	8,10	8,41
8	74,60	5,62	3,29	7,30	8,27
R1	74,52	5,40	3,20	7,20	8,60

Tabela 8 Resultados obtidos através dos métodos de referência para as amostras laboratoriais.

Amostra	Proporção de cada componente (%m/m)				
	A	B	C	D	H
Lote 1	74,52	5,66	3,19	7,36	9,11
Lote 2	68,70	5,93	3,35	7,90	9,10

3.2.1. Análise qualitativa

Para a análise qualitativa procedeu-se à construção de um modelo PCA com recurso a amostras obtidas por um DoE-QLT, onde se fez variar apenas os quatro componentes A,B C e D, numa gama bastante restrita ($\pm 1,5\%$), sendo que os restantes componentes E, F e G foram mantidos constantes. Para a elaboração deste DoE, teve-se em consideração a especificação do produto considerado em termos da proporção de cada componente como o valor central. Foram produzidas 24 amostras neste DoE utilizando a metodologia de delineamento de mistura e o algoritmo D-ótimo (maximização do determinante da matriz de delineamento experimental). Todas as amostras produzidas foram analisadas no Departamento de CQ da Generis® usando os métodos certificados usados em rotina para este produto (**Tabela 9**). Realizou-se a análise NIR das 24 amostras que permitiram a calibração do modelo PCA, e através de um conjunto de 6 amostras adicionais validou-se o modelo PCA (**Tabela 10**). Estas amostras foram produzidas adicionalmente usando novo DoE do mesmo tipo mas restringindo a variação de modo a estarem dentro dos limites de proporções das amostras do DoE.

Projetaram-se as amostras recolhidas industrialmente e acompanhou-se a evolução das misturas monitorizadas laboratorialmente. Todo este processo encontra-se descrito detalhadamente a seguir.

Tabela 9 Proporções dos componentes nas misturas delineadas pelo DoE-QLT construído para a abordagem qualitativa proposta de acordo com análise pelos métodos analíticos implementados no Departamento de CQ da Generis®.

Amostra	Proporção na mistura (%m/m)				
	A	B	C	D	H
C1	71,88	5,64	3,94	7,05	9,59
C2	71,52	5,11	3,18	9,03	7,09
C3	72,15	5,62	2,76	8,02	8,07
C4	68,66	6,09	3,64	8,82	9,56
C5	70,43	6,08	2,92	8,19	9,04
C6	75,13	5,07	2,96	6,64	7,40
C7	73,62	4,97	3,25	7,41	7,38
C8	71,56	5,56	2,70	9,29	7,42
C9	73,62	5,07	3,35	7,75	7,57
C10	71,61	5,76	3,06	7,31	9,04
C11	71,56	5,58	3,25	8,16	8,34
C12	76,51	4,85	2,80	6,53	7,04
C13	73,68	5,76	3,53	6,23	9,45
C14	71,88	5,60	3,69	7,28	8,89
C15	73,32	5,52	2,62	7,32	7,78
C16	75,12	5,38	3,17	5,77	8,49
C17	70,04	6,09	3,69	7,33	9,69
C18	71,87	5,95	3,65	6,12	10,32
C19	74,73	4,99	3,33	6,70	7,60
C20	71,61	5,81	3,02	7,44	9,08
C21	73,65	5,38	2,54	6,53	8,03
C22	73,29	5,22	2,56	8,80	6,84
C23	73,68	5,38	3,25	6,44	8,29
C24	70,01	5,34	3,08	8,82	7,83

Tabela 10 Resultados obtidos pelos métodos analíticos certificados para o conjunto de amostras a utilizar na validação do modelo PCA.

Amostra	Proporção na mistura (%m/m)				
	A	B	C	D	H
V1	73,01	5,37	3,33	7,37	9,00
V2	72,05	5,48	3,82	7,65	9,11
V3	74,20	5,39	3,03	7,36	8,30
V4	72,50	5,70	3,30	8,11	9,08
V5	73,00	5,80	3,11	8,01	8,30

Na **Figura 9** encontram-se representados os espectros NIR obtidos para as 24 amostras, sendo visível que a região espectral entre os 4115-8670 cm^{-1} e a região em torno de 4073 cm^{-1} são as regiões que apresentam absorções mais significativas, sendo consideradas estas as regiões comportando as vibrações de interesse. Teve-se em consideração a zona em torno de 4073 cm^{-1} pois os componentes A e D apresentam absorções significativas nesta zona e ao longo da construção do modelo PCA verificou-se que era uma zona muito importante para a elaboração dos modelos.

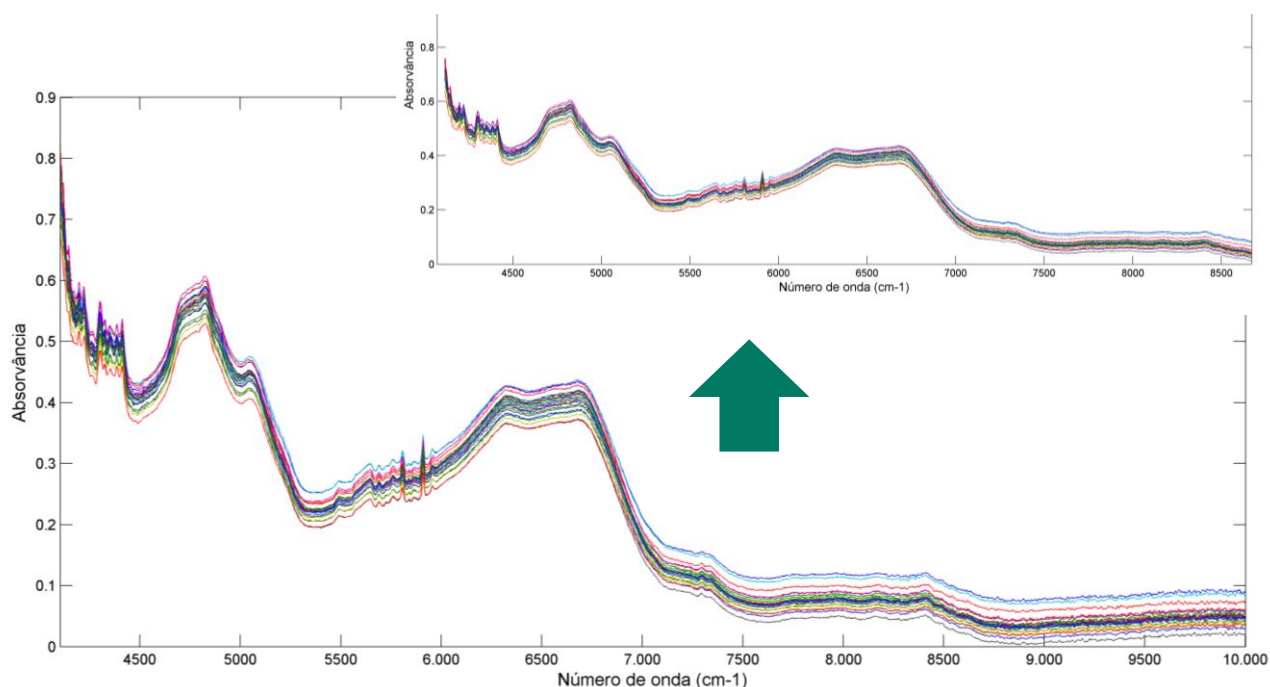


Figura 9 Espectros NIR não processados obtidos em modo de reflectância difusa das 24 amostras do DoE utilizado na abordagem qualitativa.

Previamente à construção do modelo PCA com base nas 24 amostras do DoE-QLT foi definido o pré-processamento a aplicar aos espectros. Uma vez que se trata de espectros obtidos de um mistura de

pós é normalmente aconselhado a utilização de um filtro com derivada por forma a cancelar variações devidas à dispersão de luz resultante de fenómenos físicos de dispersão devidos a variações no tamanho das partículas assim como do percurso ótico equivalente (resultando por exemplo de densidades aparentes distintas). Foi por isso selecionado o método de Savitzky-Golay com primeira derivada. Adicionalmente foi usada a centralização na média antes da aplicação do modelo PCA. Na **Tabela 11** encontram-se resumidos os parâmetros relativos ao modelo PCA desenvolvido.

Tabela 11 Resumo do modelo PCA desenvolvido para a abordagem qualitativa.

Característica	Valor
Região espectral (cm ⁻¹)	8670 - 4115 + zona em torno de 4073
Pré-tratamento espectral	Savitzky-Golay (1ª derivada, 21 pontos de tamanho de filtro e polinómio de 2ª ordem)
Número de componentes principais	2
Variância acumulada (%)	99,62

De seguida, realizou-se a análise ao modelo, com recurso à análise dos scores, estatísticas dos resíduos e Hotelling T², permitindo verificar se existem *outliers* no conjunto de dados e se é necessário alguma alteração ao modelo.

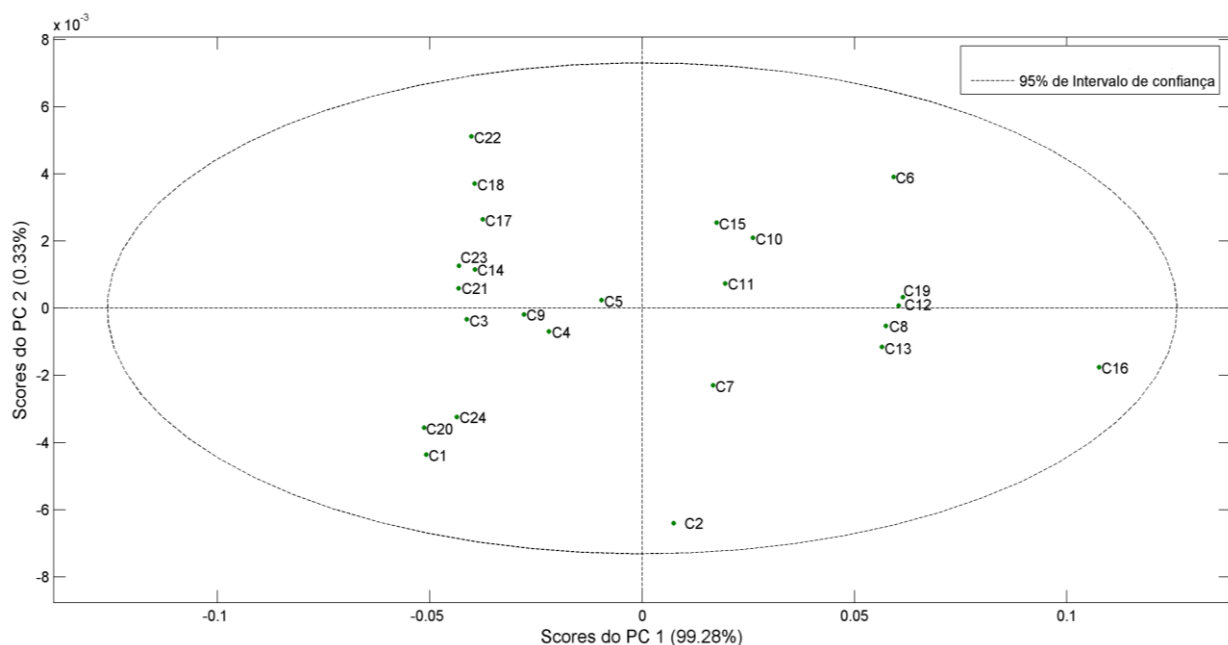


Figura 10 Mapa de scores obtidos do modelo PCA com base nos espetros NIR das amostras do DoE para abordagem qualitativa (DOE-QLT).

Através do gráfico dos scores pode-se verificar que existe alguma dispersão das amostras sem que no entanto exista um padrão ou aglomerados de amostras o que não seria expectável dado tratar-se de amostras produzidas através de um delineamento experimental. Percebe-se também que o DoE-QLT permitiu produzir amostras representando vários cenários possíveis em torno do que seria uma mistura contendo as proporções teóricas para o produto P.

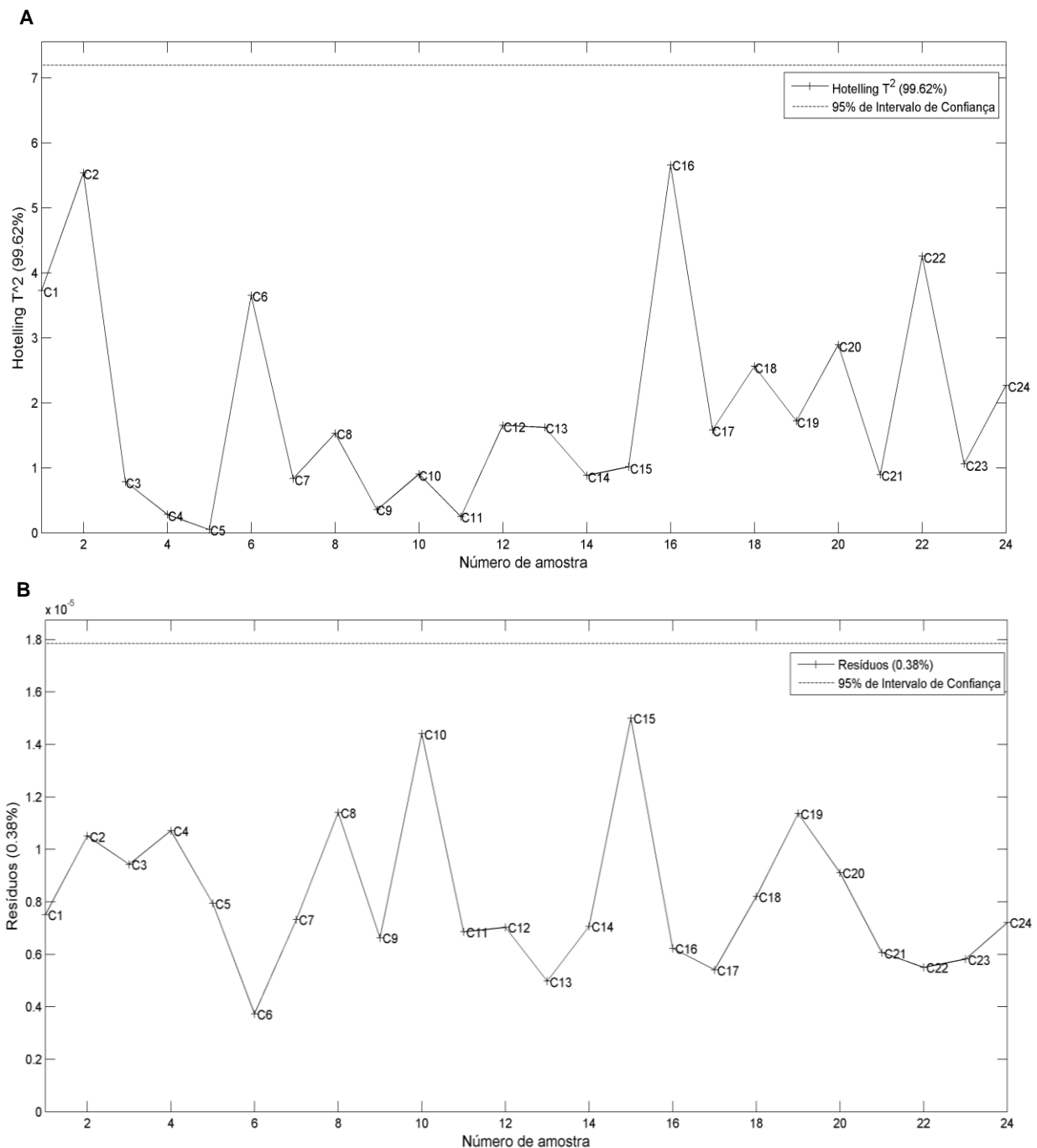


Figura 11 Estatísticas de Hotelling T² (A) e resíduos (B) relativas ao modelo PCA construído com base no DoE para a abordagem qualitativa.

Através da análise das estatísticas de Hotelling T^2 e dos resíduos pode-se verificar que o modelo não assinala *outliers*, sendo portanto um modelo consistente apto para a análise qualitativa proposta.

Procedeu-se à validação do modelo PCA, onde se verificou que todas as amostras utilizadas para validação do modelo se encontravam dentro da elipse, tal como era previsto.

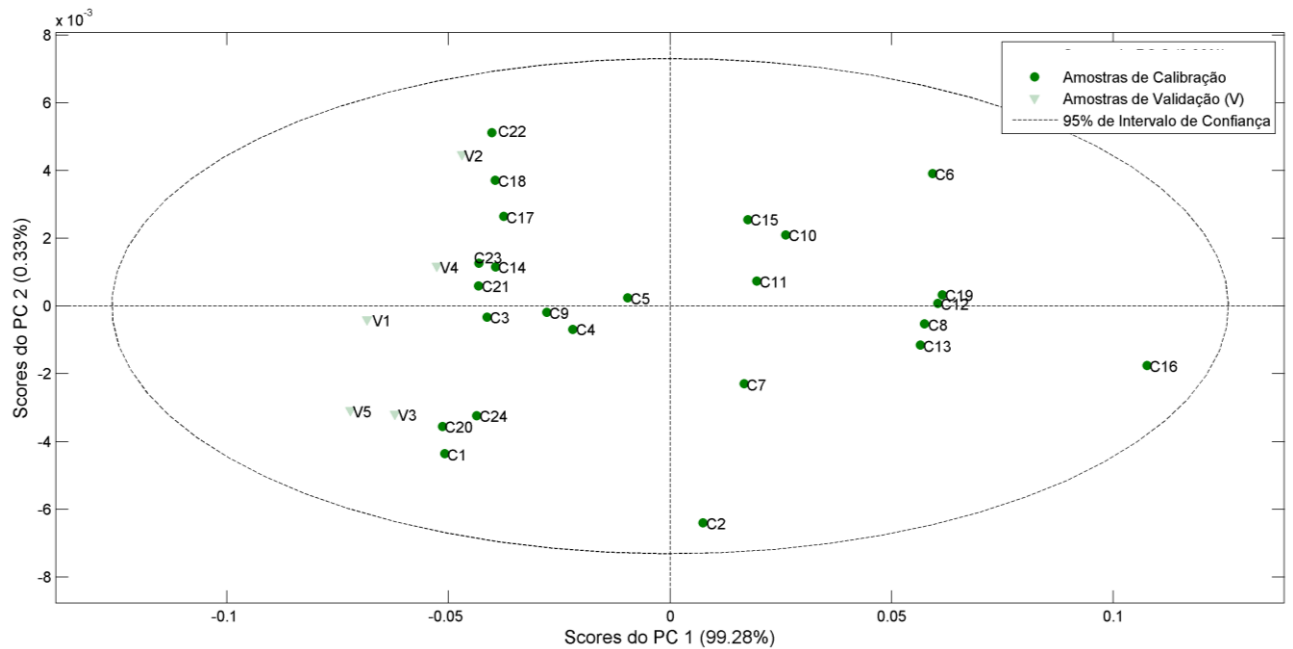


Figura 12 Mapa de scores relativo à validação do modelo PCA, com projeção as amostras de validação (●: amostras de calibração (DoE); ▼: amostras de validação).

Após validação do modelo, projetaram-se as amostras de teste (industriais e laboratoriais). Os resultados encontram-se representados nas figuras 13 e 15.

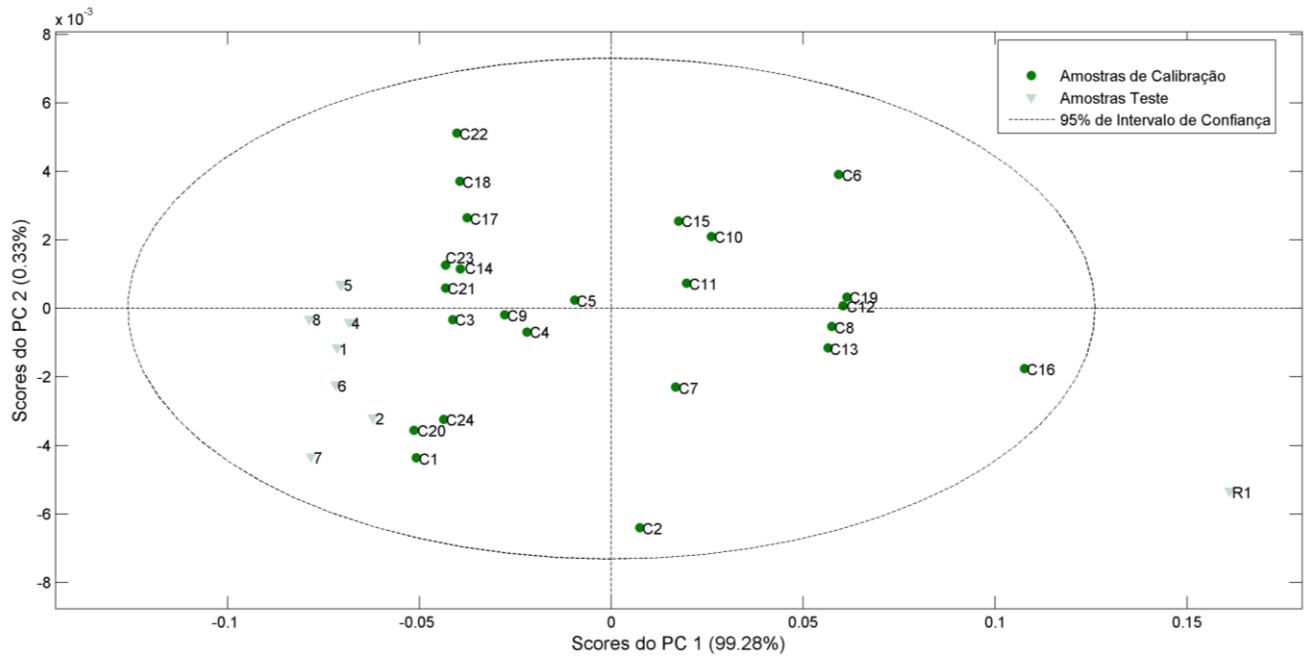


Figura 13 Mapa de scores relativo ao modelo PCA com projeção das amostras obtidas industrialmente (●: amostras de calibração (DoE); ▼: amostras de teste).

Através da **Figura 13** pode-se verificar que o modelo representa as amostras como seria previsível. Todas as amostras apresentam-se dentro dos limites de confiança, o que indica que as mesmas estão dentro de especificação sendo que a amostra R1 aparece fora da elipse e do lado contrário das restantes amostras de teste. Este resultado está em consonância com o fato de ser um lote rejeitado (fora de especificação). Conclusões idênticas podem ser tomadas quando se representam as estatísticas de Hotelling T^2 e resíduos para estas amostras projetadas (**Figura 14**), sendo que na estatística de Hotelling T^2 verifica-se que apenas o lote rejeitado se distancia do modelo, tal como era de esperado e no gráfico dos resíduos verifica-se que todas as amostras se encontram fora do intervalo de confiança indicando que toda a informação de interesse está contida no modelo.

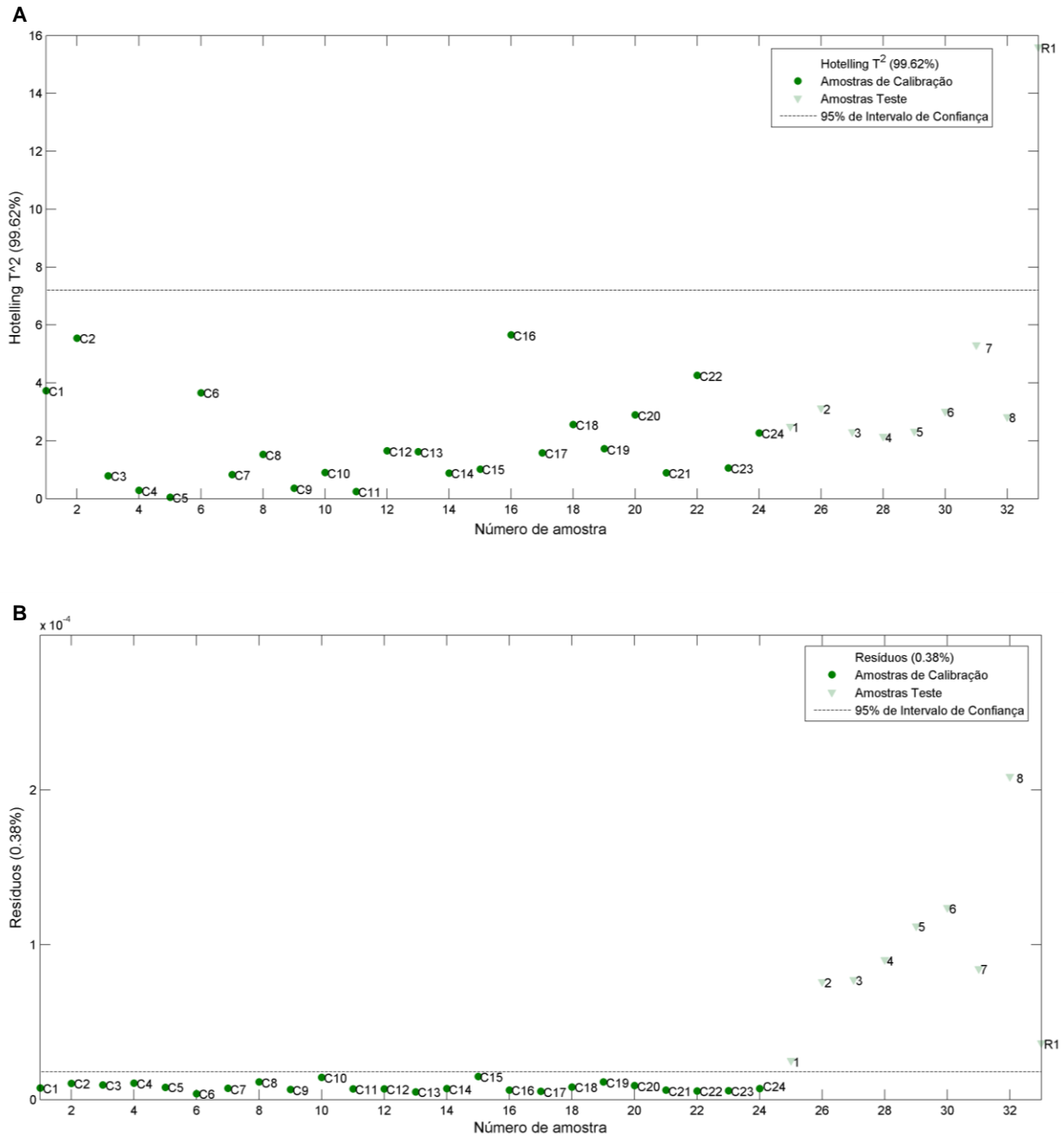


Figura 14 Estatísticas de Hotelling T^2 (A) e resíduos (B) relativas ao modelo PCA construído com base no DoE para a abordagem qualitativa, com projeção das amostras obtidas industrialmente (●: amostras de calibração (DoE); ▼: amostras de teste).

Relativamente à projeção das amostras obtidas da monitorização do processo de mistura produzido em condições laboratoriais (Lote 1 e Lote 2) verifica-se que é possível acompanhar a evolução da mistura ao longo do tempo. Esta evolução inicia-se com as amostras fora da elipse o que indica que a mistura ainda não se encontra homogénea. Aproximadamente após 27 minutos a mistura entra dentro da elipse sendo por isso expectável que esteja já homogénea. Os últimos três minutos são cruciais para o término da mistura, sendo que a mistura final se encontra dentro de especificação, informação

esta comprovada tanto pelo modelo PCA como pelos métodos de referência. Na **Figura 15** está representado todo o descritivo feito anteriormente, onde se aumentou a zona de projeção das amostras e a seta que aponta para o interior da elipse indica a tendência da mistura ao longo do tempo.

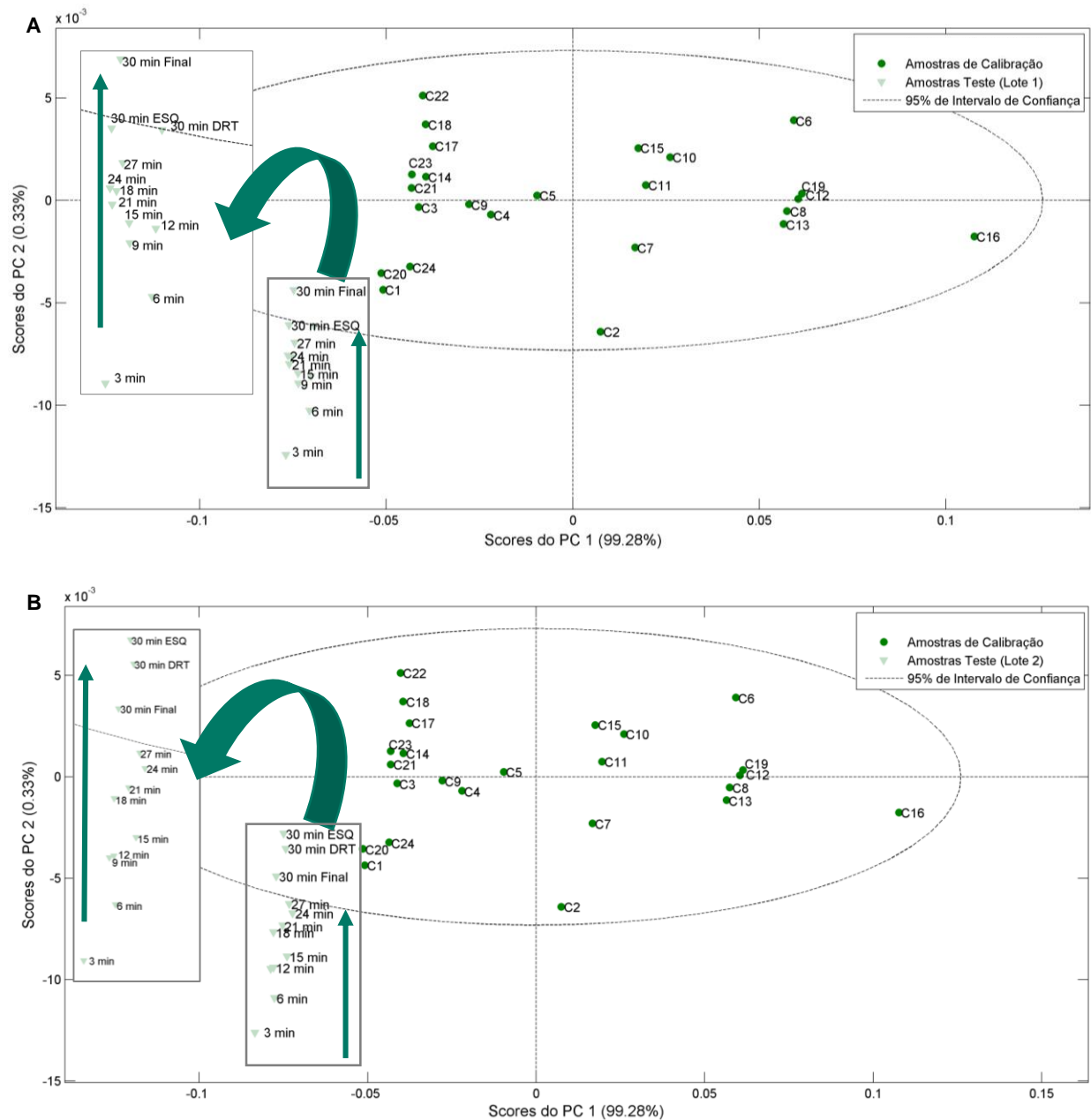


Figura 15 Mapa de scores relativo modelo PCA com projeção das amostras obtidas durante a monitorização das misturas efetuadas em condições laboratoriais (A-Lote 1; B-Lote 2) (●: amostras de calibração (DoE); ▼: amostras de teste).

Nas **Figura 16** estão representadas estatísticas de Hotelling T^2 e resíduos, para as amostras recolhidas ao longo da reprodução dos lotes laboratoriais (Lote 1 e 2). Através das estatísticas de Hotelling T^2 pode-se verificar que as amostras recolhidas ao longo do tempo tendem a aproximar-se do intervalo de confiança, ou seja, tendem a aproximar-se do modelo. Esta tendência verifica-se também no mapa

dos scores. Já na **Figura 17** verifica-se que todas as amostras se encontram fora do intervalo de confiança, indicando que não se está a perder informação relevante para o modelo.

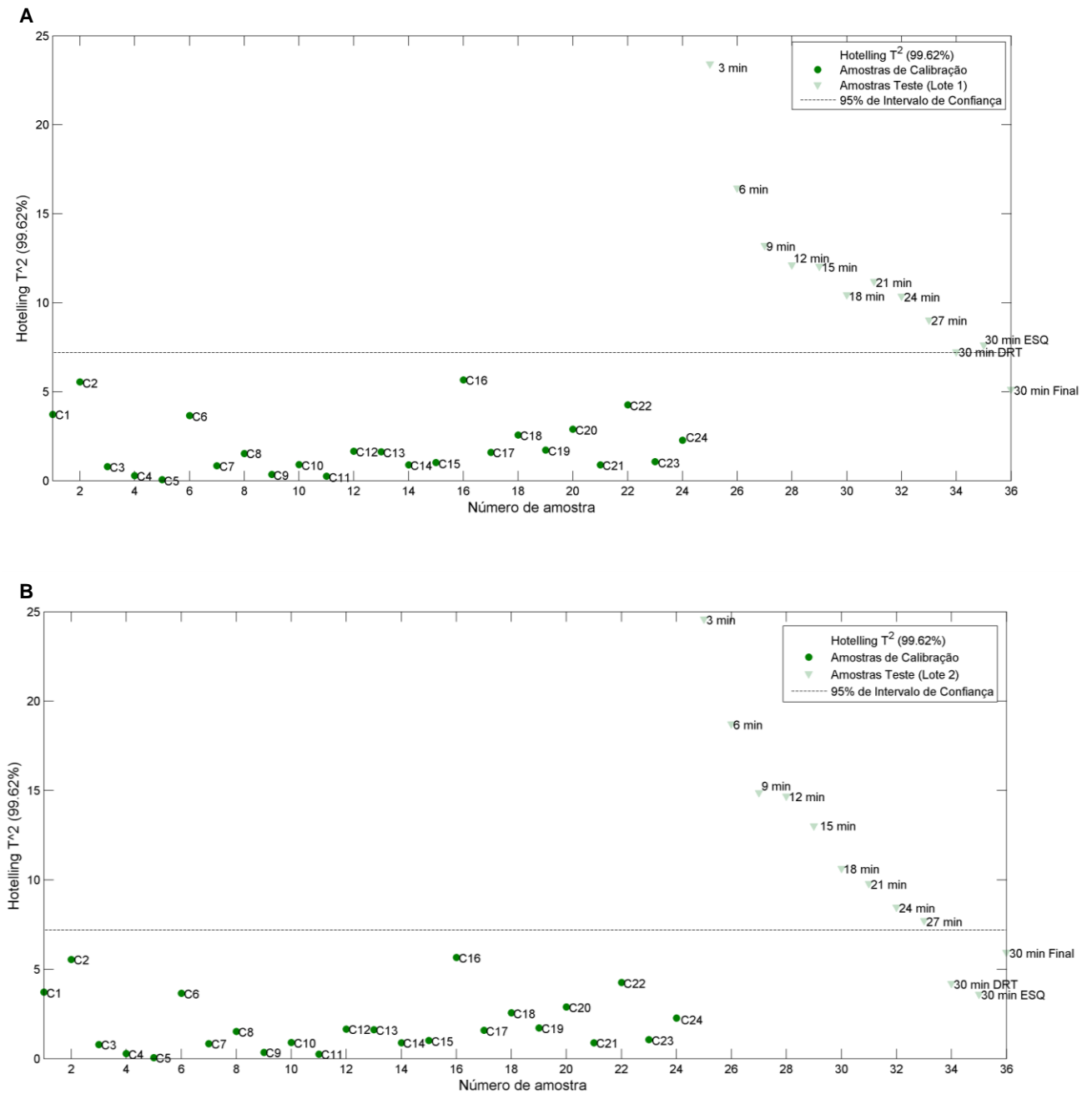


Figura 16 Estatísticas Hotelling T^2 relativas ao modelo PCA construído com base no DoE-QLT com projeção das amostras obtidas do processo de mistura à escala laboratorial (A-Lote 1; B-Lote 2) (●: amostras de calibração (DoE); ▼: amostras de teste).

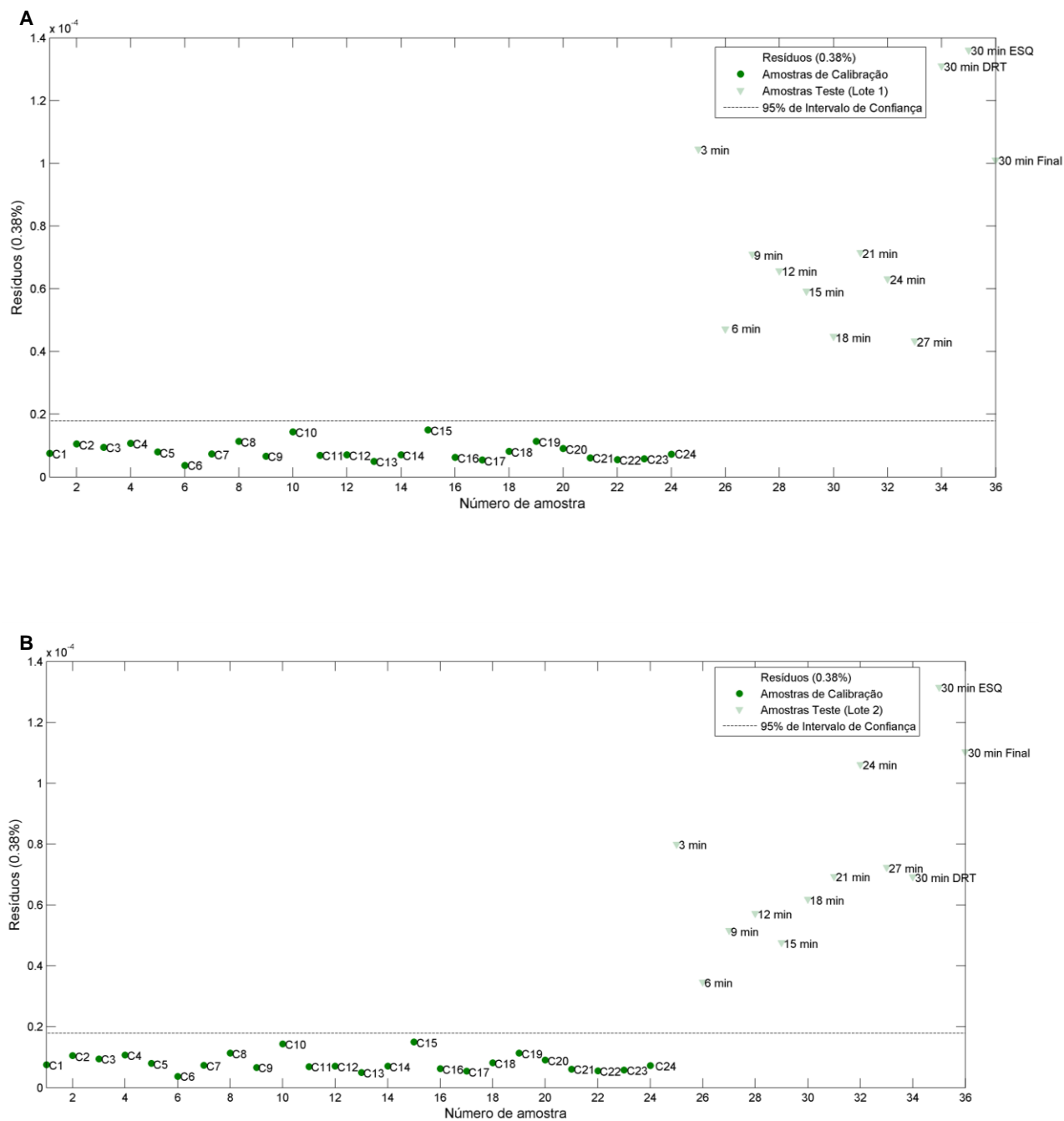


Figura 17 Estatísticas dos resíduos relativos ao modelo PCA construído com base no DoE-QLT com projeção das amostras obtidas do processo de mistura à escala laboratorial (A-Lote 1; B-Lote 2) (●: amostras de calibração (DoE); ▼: amostras de teste).

Esta análise de monitorização da mistura permite verificar que o tempo de mistura pode não ser o suficiente para que a mesma se encontre homogénea. Naturalmente que as condições usadas à escala laboratorial não são idênticas às usadas à escala industrial e por isso dever-se-á ter em conta que os tempos ideais de mistura poderão não ser os mesmos. Seria necessário aplicar procedimentos de transposição de escala (ou neste caso de redução de escala) mas que estão fora do âmbito deste

trabalho. Este resultado mostra no entanto que o tempo necessário para que a mistura esteja homogénea é de fato relativamente longo e está próximo daquele que é utilizado em condições industriais sendo por isso provavelmente uma variável que deve ser considerada crítica em termos de qualidade (CQA).

3.2.2. Análise quantitativa

Para a análise quantitativa apenas se tiveram em consideração 4 componentes (A, B, C e D) por se apresentarem em concentrações mais elevadas na mistura, permitindo verificar a sensibilidade que o modelo PLS tem quando comparado com a sua proporção na mistura.

Com recurso ao delineamento experimental, fez-se variar cerca de 10% relativamente ao valor padrão de cada componente, obtendo-se um conjunto de 24 amostras com variação nos componentes A, B, C e D. Todos os componentes foram variados em simultâneo. Esta variação é maior, quando comparada com o DoE anterior (DoE-QLT), pois nesta análise pretende-se criar modelos que traduzam dados quantitativos, referentes aos doseamentos dos diferentes componentes. De acordo com as 24 amostras obtidas, gerou-se um conjunto adicional de 6 amostras para otimização dos modelos (designadas por amostras de validação) e 6 amostras para a validação externa (designadas por amostras de teste). Estas amostras foram geradas dentro do intervalo utilizado para a calibração dos diferentes modelos. As proporções destas 12 amostras foram geradas aleatoriamente tendo apenas como requisito que todos os componentes estivessem dentro do intervalo de 10% em torno do valor médio para as proporções dos componentes. Partiu-se assim de um total de 36 amostras para desenvolver esta metodologia. Todas as amostras foram analisadas através dos métodos de referência assim como por NIR. As tabelas 12 e 13 mostram as proporções medidas de cada componentes para o DoE da abordagem quantitativa (DOE-QNT) de acordo com o doseamento dado pelos métodos de referência para o conjunto de 36 amostras.

Tabela 12 Proporções determinadas pelo método de referência para o conjunto de 24 amostras do DoE-QNT, usadas na abordagem quantitativa.

Amostra	Proporções (%m/m)				
	A	B	C	D	H
C1	77,94	2,88	1,80	8,96	4,91
C2	77,94	2,88	5,01	4,91	4,91
C3	77,89	6,53	1,80	4,29	11,14
C4	75,87	7,70	1,80	4,27	13,14
C5	77,94	3,38	5,01	4,27	5,76
C6	70,24	7,70	5,02	4,28	13,12
C7	74,96	2,88	1,80	11,11	4,91
C8	64,73	7,70	2,76	11,09	13,12
C9	66,39	7,70	1,80	11,10	13,12
C10	64,68	5,50	5,01	11,09	9,37
C11	64,69	7,70	4,26	9,21	13,13
C12	64,71	6,96	5,02	9,20	11,87
C13	77,94	2,88	2,87	7,63	4,91
C14	77,94	4,10	1,80	7,40	6,99
C15	77,93	4,43	3,94	4,28	7,55
C16	72,20	2,88	5,01	9,04	4,91
C17	71,20	2,88	3,94	11,11	4,91
C18	72,70	7,70	1,80	6,55	13,12
C19	68,38	7,69	5,01	5,61	13,12
C20	72,75	7,57	2,83	4,20	12,91
C21	68,05	5,99	1,77	10,92	10,21
C22	72,79	6,26	5,01	4,27	10,67
C23	67,79	3,75	5,01	11,10	6,39
C24	70,61	5,36	3,45	7,75	9,13

Tabela 13 Proporções determinadas pelo método de referência para as amostras de validação (V#) e teste (T#) usadas na abordagem quantitativa.

Amostra	Proporções (%m/m)				
	A	B	C	D	H
V1	72,89	4,13	3,91	8,29	7,04
V2	67,67	6,67	4,22	8,42	11,36
V3	71,15	6,37	3,27	7,50	10,86
V4	70,81	6,38	3,28	7,70	10,89
V5	72,51	4,69	2,74	7,88	7,99
V6	75,91	5,29	2,77	6,09	9,02
T1	72,28	5,56	3,72	7,18	9,48
T2	70,58	6,14	3,36	8,08	10,47
T3	72,89	4,93	3,99	7,30	8,41
T4	72,0	6,21	4,29	5,77	10,58
T5	72,78	4,94	2,66	6,06	8,42
T6	71,95	5,26	4,01	7,39	8,97

Nesta etapa, o procedimento para construção de modelos PLS para os vários componentes seguiu sempre os mesmos procedimentos, explicitando-se de uma forma geral os passos envolvidos. No desenvolvimento dos modelos PLS entre os espectros NIR e as proporções uma decisão crítica é a seleção da região espectral mais adequada em cada caso. Essa seleção pode ser feita de múltiplas formas (automáticas ou não). Uma forma possível consiste em avaliar o espectro do composto puro e privilegiar as zonas de máxima absorção. De notar no entanto, que como todos os componentes foram variados em simultâneo e a técnica NIR não é seletiva, é necessário ter também atenção a zonas espectrais mais dadas a interferências pelos outros componentes. Neste trabalho a região espectral a ter em conta em cada componente, teve por base a identificação dos picos dos espectros dos excipientes puros.

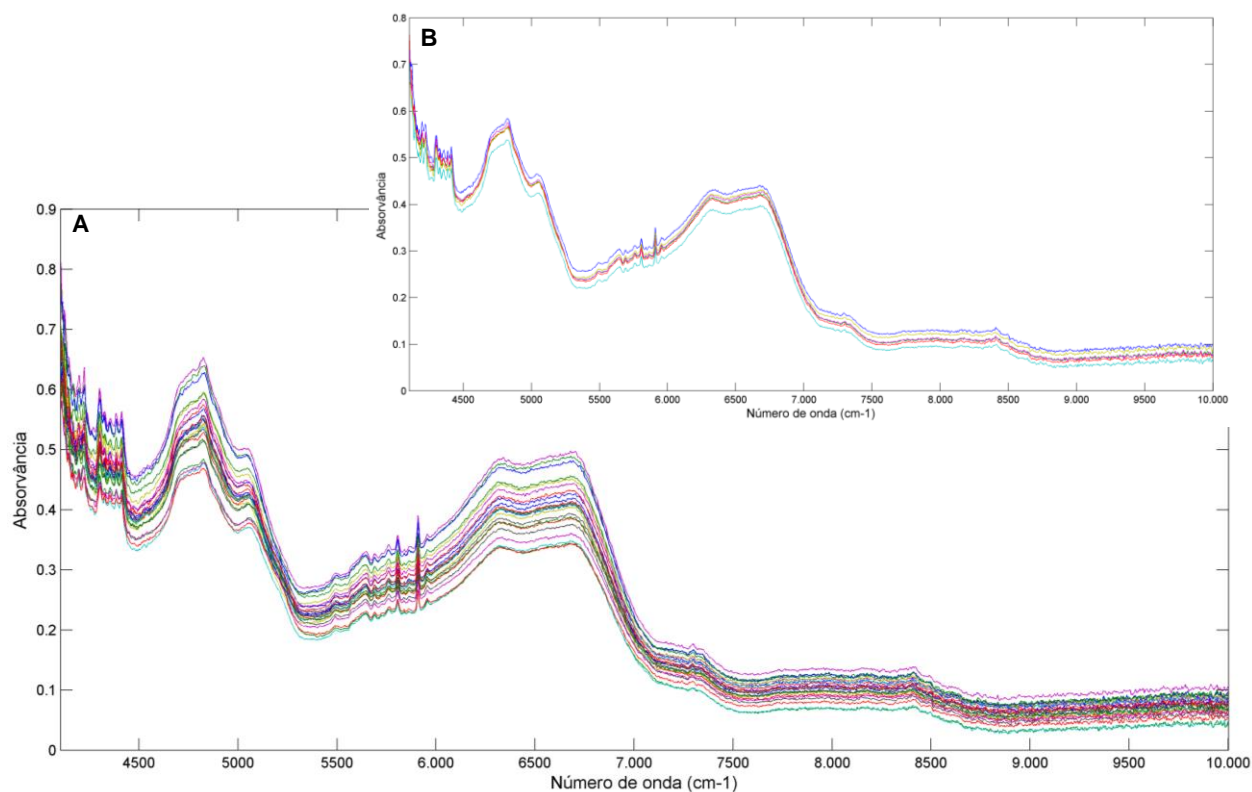


Figura 18 Para a construção dos modelos PLS utilizou-se um conjunto de 24 amostras para calibração (A) e 6 amostras para validação (B).

Também é necessário definir qual o pré-processamento espectral ideal em cada modelo. Nem sempre o método que apresenta um menor erro é o mais vantajoso, uma vez que pode utilizar um número de fatores em excesso. O melhor método de pré-processamento foi identificado tendo em conta o menor erro possível para o conjunto de validação (conjunto usado para definir o melhor número de variáveis latentes) ou RMSECV assim como o menor número possível de variáveis latentes. Note-se que se designa o erro do conjunto de validação por CV embora não se trate de um procedimento de validação cruzada, mas para se diferenciar do erro de previsão (RMSEP) que é obtido projetando outros espectros no modelo. A centralização na média é um pré-tratamento aplicado a todos os espectros NIR após o(s) pré-processamento(s) selecionado(s). Os pré-processamentos referidos neste trabalho anteriormente foram todos testados isolados e em combinação. A otimização do modelo PLS (prevenindo os fenómenos de *overfitting* ou *underfitting*) para cada componente foi feita da seguinte forma:

- 1) escolha de uma região espectral;
- 2) escolha de um pré-processamento (ou combinação) dos mencionados;
- 3) calibração de modelos considerando as amostras do DoE e diferentes números de variáveis latentes;
- 4) projeção das amostras de validação (V) em cada modelo e obtenção do RMSECV respetivo e

- 5) decisão pelo modelo que produz um menor RMSECV para um menor número de variáveis latentes.

A região espectral obtida para a otimização do modelo pode ainda ser ajustada no caso de se considerar que é necessário adicionar mais informação espectral ou retirar informação no caso de se estar a adicionar ruído ao modelo. Por fim é importante observar a reta de calibração e o respetivo R^2 para o conjunto de validação (V) e verificar se é necessário alguma afinação adicional do modelo.

Após otimização dos modelos para os vários componentes, são projetadas as amostras de teste (T) para o teste do modelo com recurso a um conjunto externo e confirmar o desempenho do modelo em termos de exatidão (obtendo-se o RMSEP). Por fim são projetadas as amostras industriais (conjunto de 9 amostras) e amostras laboratoriais calculando-se o erro (RMSEP).

5.2.2.1 Modelo para o componente A

Na **Tabela 14** encontram-se resumidos as características do modelo PLS desenvolvido para previsão do componente A.

Tabela 14 Resumo do modelo PLS relativo ao componente A.

Parâmetro	Valor
Região espectral NIR (cm ⁻¹)	9735-5307
Pré-tratamento espectral	Savitzky-Golay (1 ^a derivada, 19 pontos de tamanho de filtro, Polinómio de 2 ^a ordem)
Número de variáveis latentes	4
RMSEC	0,93
RMSECV	1,70
R^2_{cv}	0,920
Variância acumulada nos espectros (%)	90,3
Variância acumulada na proporção (%)	96,4

A região espectral selecionada abrange quase toda a gama espectral disponível. Uma vez que o espectro do componente puro apresenta picos ao longo de toda a gama parece ser importante não restringir demasiado a região espectral, pois está-se potencialmente a retirar informação importante ao modelo.

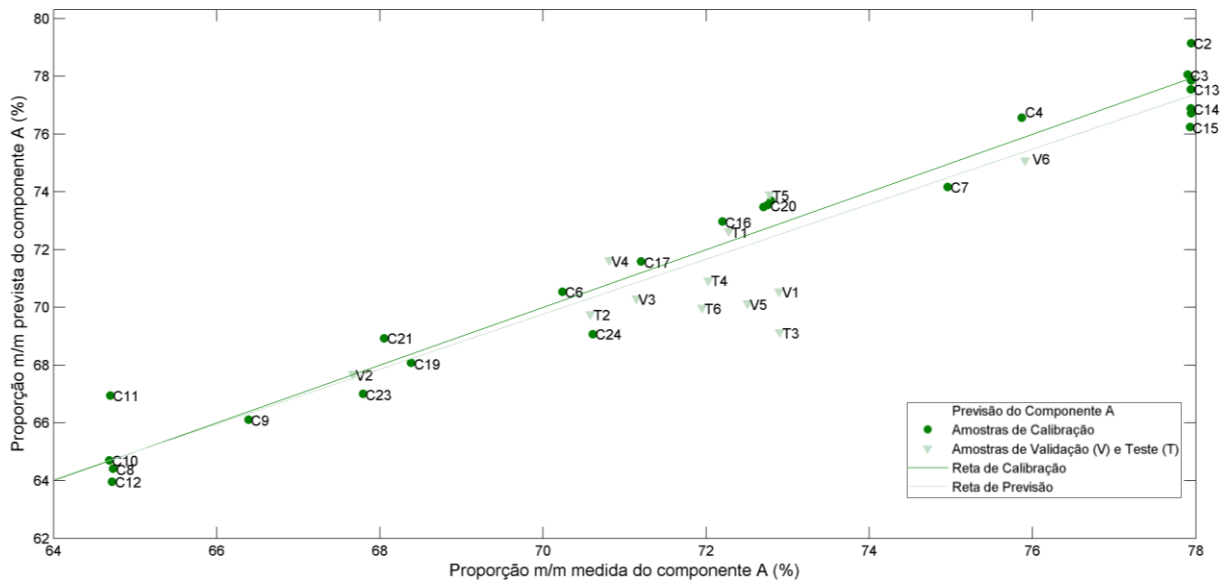


Figura 19 Proporções experimentais versus previstas pelo modelo PLS do componente A com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).

Pode-se comprovar através da **Figura 19** a capacidade preditiva do modelo, pois apresenta um R^2 alto, e um erro de previsão (RMSEP) baixo, atendendo à gama de variação do componente A. Este resultado era o esperado pois o componente A é muito ativo em NIR e é o componente maioritário da mistura.

Projetaram-se as amostras industriais e laboratoriais neste modelo PLS, obtendo-se os resultados representados na **Tabela 15**. Pode-se verificar que o modelo apresenta um bom R^2 e um erro de previsão sensivelmente mais baixo que o obtido durante a validação.

Tabela 15 Resumo dos resultados obtidos para o modelo PLS na previsão do componente A, para as amostras industriais e laboratoriais.

Parâmetro	Valor
RMSEC	0,930
RMSEP	1,58
R^2_P	0,956

Na **Tabela 16** estão representados os valores previstos para as amostras industriais e laboratoriais e o respetivo erro associado, sendo que as previsões dadas pelo modelo são muito semelhantes às esperadas.

Tabela 16 Proporções esperadas e previstas do componente A para as amostras industriais e laboratoriais.

Designação da amostra	Experimental (%)	Previsão NIR (%)	Erro (%)
1	72,2	74,0	-0,026
2	71,3	71,0	0,004
3	71,5	69,1	0,034
4	74,1	73,0	0,016
5	71,5	73,0	-0,021
6	69,4	67,8	0,023
7	69,2	65,9	0,049
8	74,6	73,1	0,021
R1	74,5	74,1	0,006
Lote 1	74,5	74,1	0,005
Lote 2	68,7	68,5	0,002

3.2.2.1. Componente B

Para o componente B, a região espectral selecionada foi mais restrita, uma vez que o espectro obtido do componente B puro, apenas contém três picos. Neste modelo o número de variáveis latentes foi superior, quando comparado com o modelo anterior e o erro de previsão também é considerável, pois este componente representa aproximadamente 9% da mistura final. Na **Tabela 17** estão representados os parâmetros obtidos para o modelo PLS na previsão do componente B.

Tabela 17 Resumo do modelo PLS relativo ao componente B.

Parâmetro	Valor
Região espectral NIR (cm ⁻¹)	7671 - 4389
Pré-tratamento espectral	Savitzky-Golay (1ª derivada, -15 pontos de tamanho de filtro, Polinómio de ordem 2)
Número de variáveis latentes	5
RMSEC	0,198
RMSECV	1,32
R ² _{cv}	91,3
Variância acumulada nos espectros (%)	99,0
Variância acumulada na proporção (%)	0,798

Na **Figura 20** está representado o conjunto de amostras de calibração, validação e teste. Pode-se verificar que o modelo representa bem as amostras de calibração, mas as amostras de validação e teste encontram-se dispersas, traduzindo-se no R² obtido para o modelo, sendo considerado baixo.

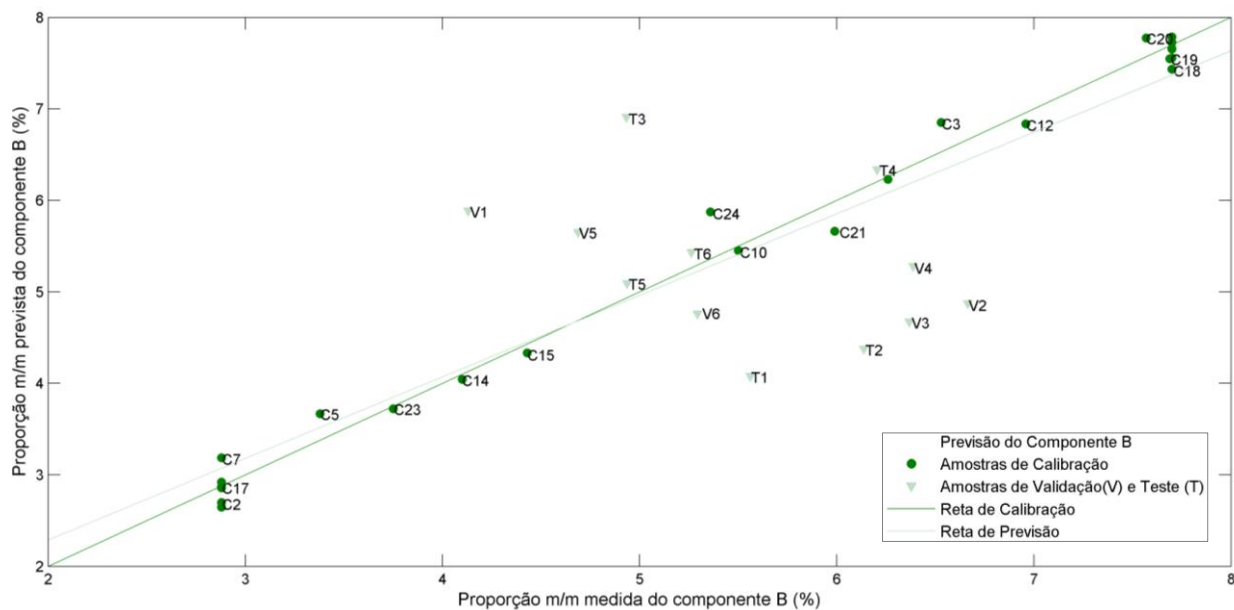


Figura 20 Proporções experimentais versus previstas pelo modelo PLS do componente B com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).

Abaixo encontram-se representados os resultados obtidos para a projeção das amostras industriais e laboratoriais. Pode-se verificar que o erro de previsão (RMSEP) aumentou relativamente ao erro obtido para a validação do modelo, assim como o R^2 diminuiu. Na **Tabela 19**, quando comparadas as amostras industriais e laboratoriais, pode-se verificar que o erro obtido é considerável. Este facto pode ser explicado devido à baixa concentração do componente B presente na mistura e à baixa absorção em NIR.

Tabela 18 Resumo dos parâmetros obtidos para o modelo PLS na previsão do componente B, para as amostras industriais e laboratoriais.

Parâmetro	Valor
RMSEC	0,198
RMSEP	2,04
R^2	0,680

Tabela 19 Proporções esperadas e previstas do componente B para as amostras industriais e laboratoriais.

Designação da amostra	Valor Medido (%)	Valor Previsto (%)	Erro
1	5,56	5,23	0,059
2	5,44	5,43	0,002
3	5,62	6,11	-0,087
4	5,69	5,51	0,032
5	5,69	3,80	0,332
6	5,60	1,06	0,811
7	5,50	1,28	0,767
8	5,62	4,18	0,256
R1	5,40	6,31	-0,169
Lote 1	5,66	5,10	0,099
Lote 2	5,93	5,60	0,056

3.2.2.2. Componente C

Procedeu-se à construção de modelo PLS para a previsão do componente C, onde se seleccionou a mesma região espectral o componente B, pois os espectros de absorção em NIR são bastantes semelhantes. Relativamente ao erro de previsão obtido é baixo quando comparado com o modelo anterior.

Tabela 20 Resumo do modelo PLS relativo ao componente C.

Parâmetro	Valor
Região espectral NIR (cm ⁻¹)	7671 - 4389
Pré-tratamento espectral	Savitzky-Golay (1ª derivada, -15 pontos de tamanho de filtro, Polinómio de ordem 2)
Número de variáveis latentes	3
RMSEC	0,584
RMSECV	0,791
R ² _{cv}	0,694
Variância acumulada nos espectros (%)	88,9
Variância acumulada na proporção (%)	82,2

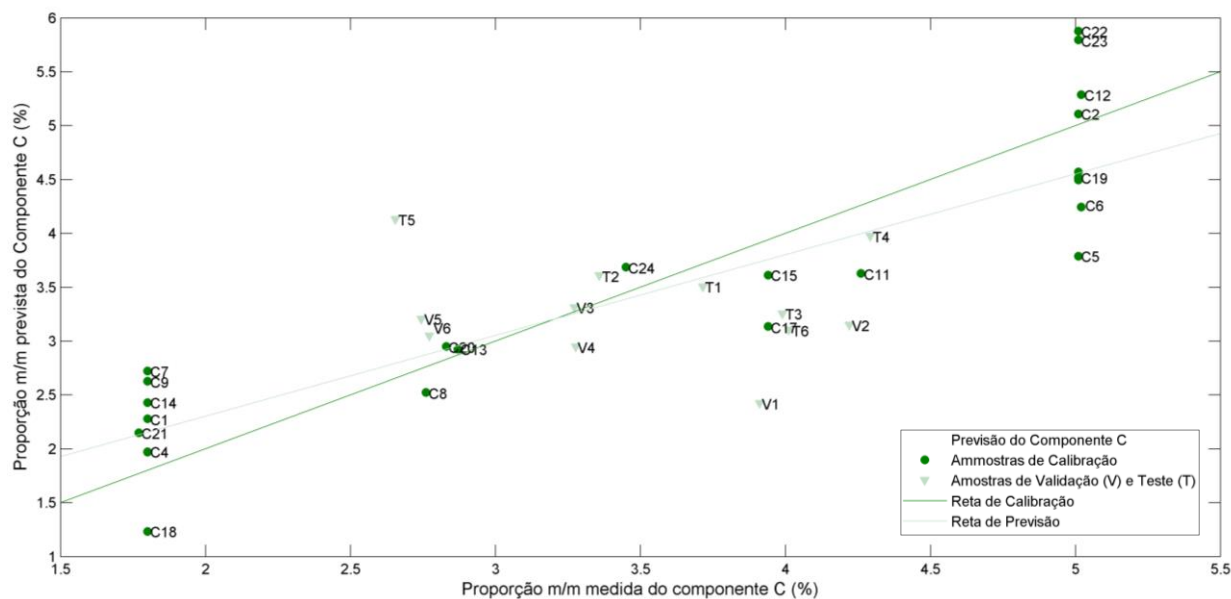


Figura 21 Proporções experimentais versus previstas pelo modelo PLS do componente C com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).

Na **Figura 21** pode-se verificar o modelo PLS com o conjunto de amostras de calibração e a projeção das amostras de validação e teste. Verifica-se que o R^2 é baixo, traduzindo-se na forma como as amostras se distribuem no modelo.

Por fim projetaram-se as amostras industriais e laboratoriais no modelo obtido e verificou-se que o erro de revisão aumentou quase o dobro e o R^2 diminuiu, o que indica que o modelo desenvolvido não é o melhor para a previsão deste composto. Todos estes resultados refletem-se no erro obtido, quando comparado os valores referência com os previstos. Todos estes resultados eram expectáveis, pois o componente C apenas apresenta bandas de absorção em NIR correspondentes à água e a sua proporção m/m na mistura também é baixa.

Tabela 21 Resumo dos parâmetros obtidos para o modelo PLS na previsão do componente C, para as amostras industriais e laboratoriais.

Parâmetro	Valor
RMSEC	0,584
RMSEP	1,44
R^2	0,532

Tabela 22 Proporções esperadas e previstas do componente C para as amostras industriais e laboratoriais.

Designação da amostra	Valor Medido (%)	Valor Previsto (%)	Erro
1	3,20	3,24	-0,013
2	3,11	3,87	-0,244
3	3,23	3,67	-0,136
4	3,25	4,33	-0,332
5	3,29	4,86	-0,477
6	3,23	6,49	-1,01
7	3,17	4,75	-0,498
8	3,29	5,35	-0,626
R1	3,10	2,45	0,210
Lote 1	3,19	2,97	0,069
Lote 2	3,35	4,09	-0,221

3.2.2.3. Componente D

Na **Tabela 23** estão representados os parâmetros obtidos na construção de modelos PLS para a previsão do componente D. Verifica-se que se considerou praticamente toda a região espectral, pois como se pode verificar no espectro puro do componente, este é aquele que apresenta mais picos e absorve em quase todo o espectro. O erro obtido de previsão é razoável e obteve-se um R^2 elevado. Através da **Figura 22** verifica-se que as amostras de validação e teste se encontram bem distribuídas no modelo.

Tabela 23 Resumo do modelo PLS relativo ao componente D.

Parâmetro	Valor
Região espectral NIR (cm^{-1})	9735 - 4385
Pré-tratamento espectral	Savitzky-Golay (1ª derivada, 15 pontos de tamanho de filtro, Polinómio de 2ª ordem) + SNV
Número de variáveis latentes	3
RMSEC	0,247
RMSECV	0,869
R^2_{cv}	0,954
Variância acumulada nos espectros (%)	80,4
Variância acumulada na proporção (%)	92,8

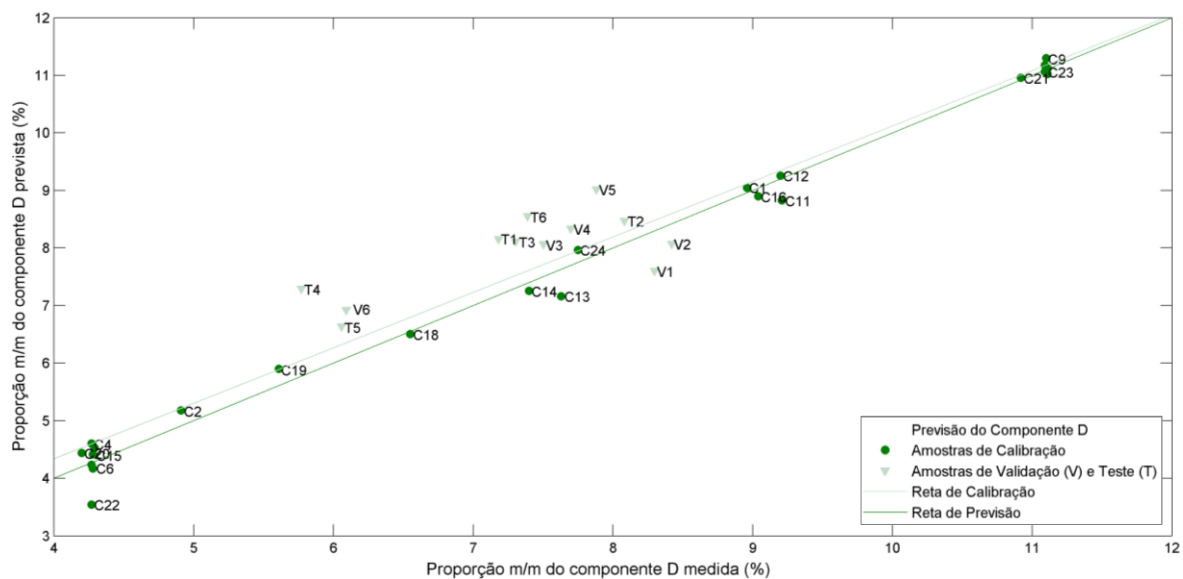


Figura 22 Proporções experimentais versus previstas pelo modelo PLS do componente D com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).

Procedeu-se à projeção das amostras industriais e laboratoriais e verificou-se que o erro de previsão não aumentou muito, assim como o R^2 não diminuiu muito (**Tabela 24**). Estas observações verificam-se nos resultados de previsão, sendo que quando comparados com os resultados dos métodos de referência não se distanciam muito uns dos outros(exceto a amostra 7 e 8)

Tabela 24 Resumo dos parâmetros obtidos para o modelo PLS na previsão do componente D, para as amostras industriais e laboratoriais.

Parâmetro	Valor
RMSEC	0,247
RMSEP	1,08
R^2	0,930

Tabela 25 Proporções esperadas e previstas do componente D para as amostras industriais e laboratoriais.

Designação da amostra	Valor Medido (%)	Valor Previsto (%)	Erro
1	7,45	7,06	0,052
2	7,37	6,60	0,104
3	7,47	6,97	0,067
4	8,11	8,53	-0,052
5	7,69	7,61	0,010
6	8,03	9,02	-0,123
7	8,10	10,5	-0,293
8	7,30	5,65	0,226
R1	7,20	6,28	0,128
Lote 1	7,36	6,23	0,154
Lote 2	7,90	7,91	-0,001

3.2.2.4. Componente H

Por último, realizou-se um modelo PLS para previsão do componente H, sendo que a região espectral seleccionada não foi a mesma que o componente B e C, pois obtinha-se um erro de previsão muito alto, assim como não se conseguia calibrar o modelo da melhor forma.

Na projeção das amostras de validação e teste, verifica-se que as mesmas se encontram bem distribuídas ao da reta de calibração, embora algumas se apresentem distanciadas.

Tabela 26 Resumo do modelo PLS relativo ao componente H.

Parâmetro	Valor
Região espectral NIR (cm ⁻¹)	9735 - 4385
Pré-tratamento espectral	Savitzky-Golay (1ª derivada, 15 pontos de tamanho de filtro, Polinómio de 2ª ordem)
Número de variáveis latentes	3
RMSEC	0,906
RMSECV	1,90
R ² _{cv}	0,836
Variância acumulada nos espectros (%)	80,4
Variância acumulada na proporção (%)	92,8

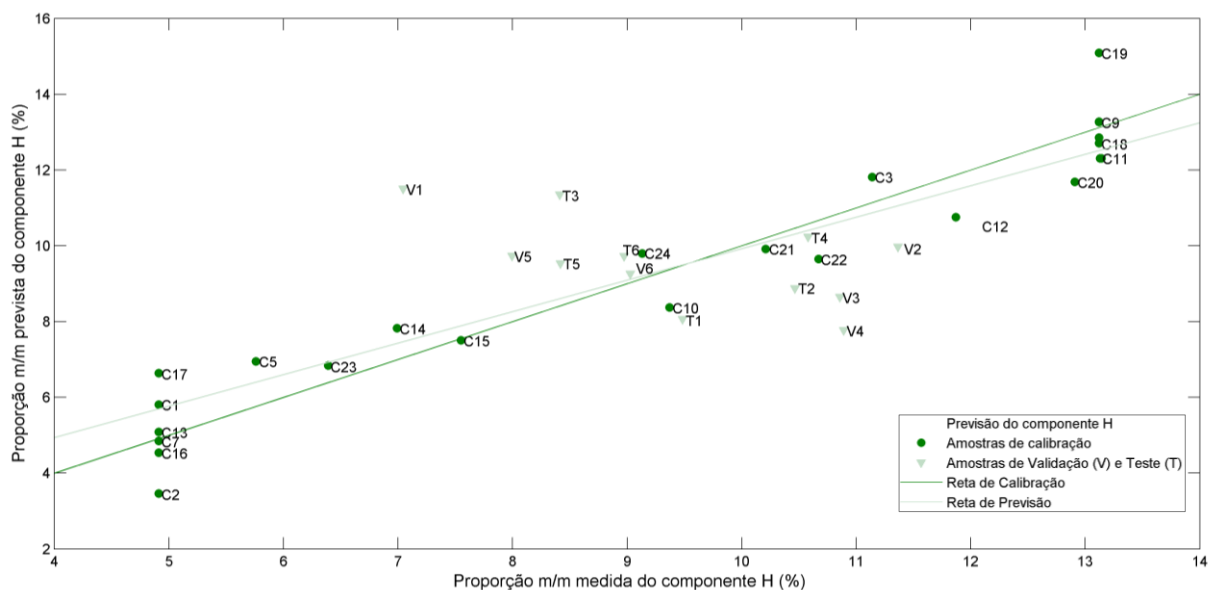


Figura 23 Proporções experimentais versus previstas pelo modelo PLS do componente H com indicação das amostras usadas na calibração (●,C#) validação (▼,V#) e teste (▼,T#).

Projetaram-se as amostras industriais e laboratoriais no modelo e verificou-se que o erro de previsão diminuiu assim como o R^2 aumentou. Os valores previstos são consideravelmente diferentes dos valores obtidos pelos métodos de referência, pois uma vez que o componente H absorve muito pouco em NIR, o modelo irá ser pouco sensível na previsão deste componente, refletindo-se em erros altos.

Tabela 27 Resumo dos parâmetros obtidos para o modelo PLS na previsão do componente H, para as amostras industriais e laboratoriais

Parâmetro	Valor
RMSEC	0,906
RMSEP	1,53
R^2	0,840

Tabela 28 Proporções esperadas e previstas do componente H para as amostras industriais e laboratoriais.

Designação da amostra	Valor Medido (%)	Valor Previsto (%)	Erro
1	8,69	8,84	-0,017
2	9,00	10,5	-0,166
3	8,99	10,2	-0,129
4	8,83	8,98	-0,017
5	8,30	7,28	0,123
6	8,71	5,44	0,375
7	8,41	5,40	0,358
8	8,27	7,92	0,042
R1	8,60	8,48	0,014
Lote 1	9,11	9,65	-0,059
Lote 2	9,10	9,40	-0,033

3.3. Comparação entre as duas abordagens

Ambas as abordagens aplicadas neste trabalho, demonstraram ser aplicáveis na monitorização da mistura. Embora as análises aplicadas permitissem obter conclusões diferentes, verificou-se que os modelos PCA obtidos na análise qualitativa eram mais simples e robustos quando comparados com os modelos PLS obtidos na análise quantitativa. São vários os fatores relacionados com a fraca reprodutibilidade de modelos PLS, entre eles a fraca absorção em NIR de alguns componentes, respetiva proporção presente na mistura, sendo que os modelos PLS pioram com a diminuição da proporção do componente na mistura. O facto de se variar os quatro componentes ao mesmo tempo, também vai intervir na construção dos modelos PLS, nomeadamente em termos de introdução de interferências que devem ser compensadas pelo modelo. Numa tentativa de melhoria dos modelos PLS, poder-se-ia construir um novo DoE, onde se fizesse variar apenas um componente de cada vez, possibilitando maior eficiência preditiva, nomeadamente nos casos dos componentes minoritários na mistura.

4. CONCLUSÕES

Conclui-se que a espectroscopia NIR, associada às técnicas de quimiometria, permite a construção de modelos capazes de acompanhar o processo de mistura. As duas abordagens revelaram-se possíveis de implementação.

Os modelos criados possibilitaram duas estratégias diferentes. A análise qualitativa permitiu verificar quais as amostras que se encontravam dentro ou fora de especificação. De acordo com os resultados obtidos, verificou-se que as amostras industriais (produto comercial) usadas se encontravam dentro da elipse que define o intervalo de confiança exceto o ponto correspondente ao lote rejeitado. Relativamente à monitorização da mistura dos dois lotes verificou-se que os últimos três minutos são cruciais para a mistura ficar homogénea, podendo indicar que o tempo de mistura não é o suficiente.

Relativamente à análise quantitativa pode-se considerar que se obtiveram modelos com boa capacidade preditiva para a previsão dos componentes A e D e modelos menos bons para os restantes. Este resultado pode ser explicado devido ao facto dos componentes A e D apresentarem bastante atividade em termos da radiação NIR (visto nos espectros puros obtidos) e daí ter-se selecionado uma região espectral considerável, enquanto que os restantes componentes (B, C e H) apenas apresentam três picos de absorção em NIR que correspondem a bandas da água (e por isso o sinal espectral é muito pouco característico da sua estrutura sendo essencialmente apto para verificar o seu grau de hidratação). Estes três componentes também se encontram em proporções mais baixas na mistura e o NIR é também pouco sensível a componentes “vestigiais” (abaixo de 1%/m deve ser usado com precaução). Uma estratégia para melhorar os modelos seria a de variar apenas um dos componentes de cada vez, tornando o modelo mais sensível e robusto para cada um dos compostos.

Comparando ambas as abordagens, verifica-se que na primeira análise conseguimos obter resultados mais fiáveis, enquanto que na segunda análise devido à fraca absorção dos componentes B, C e D não nos permite ter uma boa previsão dos mesmos. Também é importante salientar que as duas análises são importantes, pois permitem-nos tirar conclusões diferentes, embora a análise quantitativa é mais completa nos resultados obtidos. Evidentemente que a tecnologia NIR a ser usada na prática se revelaria menos sensível à variação nos componente B, C e D. Por outro lado, a espectroscopia NIR é bastante sensível à humidade que poderá ser detetada quando a mistura não estiver dentro dos valores adequados da mesma. Essa deteção poderá ser feita essencialmente através da abordagem qualitativa.

De realçar que no ambiente industrial existem muitas vezes vibrações que afetam as leituras realizadas, bem como as variações de temperatura. Sendo este tipo de equipamento sensível a estas variações, a sala onde se encontra deve ter controlo de temperatura, e as leituras devem ser realizadas sem se verificarem vibrações intensas.

Em geral, os modelos obtidos permitem a monitorização da mistura e eventuais ajustes que sejam necessários, e analisar as amostras finais de forma rápida e pouco dispendiosa, antes de seguirem para o controlo de qualidade.

PERSPETIVAS FUTURAS

Nesta secção apresentam-se algumas possibilidades de continuação e consolidação deste trabalho:

- Comparação dos resultados obtidos com as abordagens propostas com a metodologia correntemente adotada que se baseia na análise da variação espectral em função do tempo.
- Reprodução de lotes laboratoriais com um tempo de mistura final superior aos trinta minutos, de forma a verificar se o tempo de mistura definido na instrução de fabrico é o mais adequado.
- Realização de um novo DoE para um modelo PCA, onde se fizesse variar apenas um dos componentes de cada vez, possibilitando uma análise qualitativa mais explícita para cada componente.
- Definição de DoEs distintos por componente para estimativa dos modelos PLS, tornando-os desta forma mais robustos e precisos.
- Desenvolvimento de modelos para outros produtos Generis Farmacêutica, S.A, permitindo o estudo do controlo de processo e diminuindo o tempo de análise do produto final.
- Avaliar o desempenho dos dois métodos com análises espectrais obtidas diretamente do processo industrial de fabrico.
- Os métodos desenvolvidos requerem manutenção e atualização, para que continuem a responder eficazmente.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Abdi, H.. Partial Least Square Regression PLS- Regression, <https://www.utdallas.edu/~herve/Abdi-PLSR2007-pretty.pdf>, Acedido em: 20-08-2015.
- Almeida, F. M. N.. Espectroscopia de Infravermelho Próximo com Transformada de Fourier (FT-NIR) na Caracterização de Farinhas para Alimentação Pueril. Tese de Mestrado, Instituto Superior Técnico, Lisboa. 2009.
- Beer, T., Burggraeve, A., Fonteyne, M., Saerens, L., Remon, J.P., Vervaet, C.(2011). Near Infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *Int. J. Pharm.*,417(1-2), 32-47.
- Berntsson, O., Danielsson, L.G., Lagerholm, B., Folestad, S. (2002). Quantitative in-line monitoring of powder blending by near infrared reflection spectroscopy. *PowderTechnol.*, 123, 185-193.
- Booker, A.J.(1998). Design and analysis of computer experiments. American Institute of Aeronautics and Astronautics. A98-39713.
- Burns, D.A., Ciurczak, E.W.. Handbook of Near- Infrared Analysis. Third Edition, CRC Press (2007).
- Comissão Europeia (2008). EudraLex - Normas que Regulam os Medicamentos na União Europeia. Normas orientadoras sobre Boas Práticas de Fabrico Medicamentos para uso Humano e Veterinário. Vol. 4-Parte I.
- Conzen, J. P.. Multivariate Calibration. A Practical Guide for the Method Development in the Analytical Chemistry. Bruker Optick GmbH. Second Edition. 2006.
- El-Hagrasy, A., D'Amico, F.,S., Drennen, J.K.(2005). A Process Analytical Technology Approach to Near-infrared Process Control of Pharmaceutical Powder Blending. PartI: D-Optimal Design for Characterization of Powder Mixing and Preliminary Spectral Data Evaluation. *J. Pharm. Sci*, 95 (2), 392-406.
- El-Hagrasy, A.S., Morris, H.R., D'Amico, F., Lodder, R.A., Drennen, J. K..(2001). Near – Infrared Spectroscopy and imaging for the Monitoring of Powder Blend Homogeneity. *J. Pharm. Sci*, 90 (9), 1298- 1307.
- Esbensen, K., Geladi, P.(1987). Principal Component Analysis. *Chemom. Intell. Lab. Syst.*, 2 (1-3), 37-52.
- European Medicines Agency. Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requirements for new submissions and variations, www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/06/WC500167967.pdf, Acedido em: 26-10-2015.
- Fisher, R.A., D.,Sc., S., F.R.. The Design Experiments.Second Edition, London (1937).

Frank, I.E., Kowalski, B.R.(1982). Chemometrics. Am. Chem. Soc., 54, 232R-243R.

Gad, S.C., PH.D., D.A.B.T.. Pharmaceutical Manufacturing Handbook. John Wiley & Sons, New Jersey (2008).

Gauchi, A.P., Chagnon, P. (2001). Comparison of selection methods of explanatory variables in PLS regression with application to manufacturing process data. Chemom. Intell. Lab. Syst, 58 (2), 171-193.

ICH- Harmonisation for better health. www.ich.org/home.html, Acedido em 09-10-2015.

Infarmed - Autoridade Nacional do Medicamento e dos Produtos de Saúde I.P, Perguntas frequentes sobre o Infarmed, www.infarmed.pt/portal/page/portal/INFARMED/PERGUNTAS_FREQUENTES/SOBRE_INFARMED, Acedido em 26-10-2015.

Isidro, E.M.. Regulatory Intelligence: Desafios e oportunidades sob a perspectiva de uma filiada nacional. Obtenção do título de especialista do colégio de assuntos regulamentares da ordem dos farmacêuticos, Lisboa, 2011.

Jorgensen, A.. Clustering Excipient Near Infrared Spectra Using Different Chemometric Methods. Seminar, Department of Pharmacy University, Helsinki. 2000.

Leitão, T.M.D.. Aplicações da espectroscopia de infravermelho próximo em Ciências Farmacêuticas. Tese de Mestrado, Faculdade de Ciências da Saúde da Universidade Fernando Pessoa, Porto, 2012.

Maleki, M.R., Mouazen, A.M., Ramon, H., Baerdemaeker, J.(2007). Multiplicative Scatter Correction during On-line Measurement with Near Infrared Spectroscopy. Biosyst. Eng., 96 (3), 427-433.

Mantanus, J.. New pharmaceutical applications involving Near Infrared spectroscopy as a PAT compliant process analyser. Doctoral thesis, Faculty of Medicine, Liege, 2012.

Mark, H., Workman, J.. Chemometrics in spectroscopy. First Edition. Elsevier Inc, USA (2007).

Morgano, M., A., Moriya, C.(2003). Determinação do Teor de Sacarose em Açúcar Cristal por Espectroscopia FT-IR/ATR e Regressão Multivariada. Braz. J. Food Technol, 6, 77-83.

Morisseau, K.M., Rhodes, C.(1995). Pharmaceutical uses of near-infrared spectroscopy. Drug Dev. Ind. Pharm., 21 (9), 1071-1090.

Naes, T., Isaksson, T., Feam, T., Davies, T.. A User Friendly Guide to Multivariate Calibration and Classification. Sixth Edition NIR Publications (2002).

Oliveira, M.C.P.P.. Espectrofotometria no Infravermelho- por metodologia FTIR (Fourier Transform Infrared): Validação da análise do teor de uréia e de outros parâmetros de qualidade do leite. Tese de Mestrado, Escola de Veterinária da Universidade Federal, Minas Gerais, 2011.

Pasquini, C.(2003). Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. J. Braz. Chem. Soc., 14 (2), 198-219.

Plugge, W., Vlies, V.D.(1992). The use of near infrared spectroscopy in the quality control laboratory of the pharmaceutical industry. *J. Pharm. Biomed. Anal.*, 10 (0-12), 797-803.

Popo, M., Torres, S.R., Conde, C., Romanãch, R.(2002). Blend Uniformity Analysis Using Stream Sampling and Near Infrared Spectroscopy. *AAPS PharmSciTech*, article 24.

Real, C.P.M.V.. Estudo de métodos analíticos alternativos para quantificação do teor de cálcio, ferro e vitamina C em pré-misturas enriquecidas em micronutrientes. Universidade Católica Portuguesa do Porto. 2014.

Reich, G.(2005). Near- Infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Int. J. Pharm.*, 57 (8), 1109-1143.

Savitzky, A., Golay, M.J.E.(1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures.*Anal. Chem.*, 36(8), 1627–39.

Sekulic, S.S., Ward, H.W., Brannegan, D.R., Stanley, E.D., Evans, C.L., Sciavolino, S.T., Hailey, P.A., Aldridge, P.K.(1996). On-line Monitoring of Powder Blend Homogeneity by Near-Infrared Spectroscopy. *Anal. Chem.*, 68 (3), 509- 513.

Shi, Z., Cogdill, R.P., Short, S.M., Anderson, C.A.(2008). Process characterization of powder blending by near-infrared spectroscopy: Blend end-points and beyond. *J. Pharm. Biomed. Anal.*, 47, 738-745.

Silva, H.M.A.R., Rodrigues, M.P.M.C., Eusébio, M.I.. Espectroscopia no infravermelho por transformada de Fourier (FTIR). Primeira Edição, LNEC, Lisboa (2007).

Singh, B., Bhatowa, R., Bhushan, T., Kapil, R.(2011).Developing micro-/nanoparticulate drug delivery systems using “design of experiments. *Int. J. Pharm. Investig.*, 1 (2), 75-87.

Stat-Ease. Handbook for Experiments- Version 09.01.03. 2014.

TN- Thermo Nicolet. Introduction to Fourier Transform Infrared Spectrometry. 2001.

Tobias, R.D.. An Introduction to Partial Least Squares Regression. citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.460.1258&rep=rep1&type=pdf, Acedido em: 22-08-2015.

Wargo, D.J., Drennen, J.K.(1996). Near- infrared spectroscopic characterization of pharmaceutical poder blends. *J. Pharm. Biomed. Anal.*, 14, 1415-1423.

Wise, B., M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., Koch, R.S.. *PLS_Toolbox 4.0*. Elgenvektor Research. 2006.

Wold, S., Sjöström, M., Eriksson, L.(2001). PLS- regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.*, 58, 109-130.

Wu, H., Tawakkul, M., White, M., Khan, M.A.(2009). Quality-by-design (QbD): An integrated multivariate approach for the component quantification in powder blends. *J. of Pharm. Sci.*, 98 (8), 2784- 2798.