

Análise Multivariada no diagnóstico da Produção de Biodiesel

Susana Margarida Coelho Lima Melo

Mestrado em Química

Departamento de Química

FCTUC

Setembro de 2011



UNIVERSIDADE DE COIMBRA

Análise Multivariada no diagnóstico da Produção de Biodiesel

Susana Margarida Coelho Lima Melo

**Dissertação apresentada para provas de Mestrado em Química,
área de especialização em Química Controlo da Qualidade e
Ambiente**

Orientador – Professor Doutor Jorge L.G.F.S. Costa Pereira, Professor Auxiliar, Departamento de Química, FCTUC

Co-Orientador - Doutor Renato H. Carvalho, Responsável pelo Departamento de Desenvolvimento Tecnológico da Iberol, Alhandra

Setembro de 2011

Universidade de Coimbra

Índice

Abreviaturas.....	5
Agradecimentos.....	6
Objectivos.....	7
Resumo.....	8
1. INTRODUÇÃO.....	10
.....	10
1.1. Enquadramento do trabalho.....	10
1.2. A Iberol.....	11
1.3. O biodiesel como combustível alternativo.....	12
1.3.1. Óleos Vegetais e Seus Derivados.....	12
1.3.2. Biodiesel e aspectos gerais.....	13
.....	15
1.4. Química do Biodiesel e Propriedades.....	16
1.5. Química da reacção de transesterificação.....	18
2. DESENVOLVIMENTO.....	23
2.1. Processo de Produção de Biodiesel.....	23
2.1.1. Refinação Alcalina dos óleos.....	24
2.1.1.1. Desgomagem Química.....	24
2.1.1.2. Neutralização.....	25
2.1.1.3. Lavagem do óleo.....	25
2.1.1.4. Secagem do óleo.....	25
2.1.2. Transesterificação.....	26
2.1.3. Lavagem e Secagem do Biodiesel.....	26
2.1.4. Recuperação do Metanol.....	27
.....	27
2.2. Propriedades e especificações do biodiesel.....	28
2.2.1. Ponto de inflamação.....	28
2.2.2. Viscosidade cinemática	28
2.2.3. Densidade	29
2.2.4. Índice de cetano.....	29
2.2.5. Resíduo carbonoso.....	29
2.2.6. Teor de enxofre	30
2.2.7. Teor de cinzas sulfatadas.....	30
2.2.8. Teste de corrosão à lâmina de cobre.....	30
2.2.9. CFPP.....	31
2.2.10. CP.....	31
2.2.11. Estabilidade oxidativa	31
2.2.12. Índice de acidez.....	31
2.2.13. Índice de iodo.....	32
2.2.14. Teor em ésteres alquílicos ou pureza	32
2.2.15. Teor em ésteres metílicos de ácido linolénico.....	32
2.2.16. Teores em mono-, di- e tri-glicéridos.....	32
2.2.17. Teor em glicerol livre e total.....	33
2.2.18. Teor em água	33
2.2.19. Teor em metanol.....	33

2.2.20 Teor em metais do grupo I (e grupo II).....	33
2.2.21 Contaminação total	34
2.3. A quimiometria.....	36
2.3.1. Análise por componentes principais.....	36
2.3.2. Regressão pelos mínimos quadrados.....	37
3. Parte Experimental.....	40
3.1. Material e métodos	40
3.1.1. Descrição geral da metodologia utilizada na recolha e obtenção dos dados.....	40
3.1.2. Análise de componentes principais.....	40
3.1.3. Análise de PLS.....	41
4. Resultados e discussão.....	43
4.1 Verificação das variáveis.....	43
4.1.1 Remoção de variáveis qualitativas.....	43
4.1.2 Remoção de linhas com falhas de dados.....	44
4.1.3 Substituição de valores e arredondamento dos números à mesma precisão de cada variável	44
4.2 Análise de variáveis.....	44
4.2.1 Matéria-prima.....	45
4.2.2 Processo.....	52
4.2.3 Predictores.....	57
4.2.4 Resultados.....	60
4.2.5 Análise conjunta.....	64
4.3 Interdependência.....	66
4.3.1 Modelação implícita.....	66
4.3.2 Correlação.....	73
4.4 Discussão.....	77
4.4.1 PCA.....	77
4.4.1.1 Matéria-prima.....	77
4.4.1.2. Processo.....	78
4.4.1.3. Predictores.....	79
.....	79
.....	80
4.4.1.3 Respostas.....	81
4.4.1.4 Processo global.....	81
4.4.2 PLS.....	83
4.4.3 Correlação.....	86
4.4.4. Análise do processo.....	90
4.4.4.1. Análise da interdependência da resposta y53 com 15 variáveis do processo de transesterificação.....	90
4.4.4.2. Relação entre x071 e y11, y24 e y37.....	93
4.4.4.3. Relação entre x002 e y09, y10, y23 e y35.....	93
4.4.4.4. Relação entre x007 e y09, y22 e y35.....	94
5 CONCLUSÕES.....	96
Bibliografia.....	100
Anexos.....	103
A1. Codificação das variáveis associadas às matérias primas.....	103

A2. Codificação das variáveis associadas ao processo de desgomagem e neutralização dos óleos (matérias primas) usados na reacção de transesterificação (produção de Biodiesel).....	103
A3. Codificação das variáveis do processo associadas à produção de Biodiesel (Etapa da Transesterificação).....	104
A4. Codificação das variáveis do processo associadas à produção de Biodiesel (Etapa da Recuperação do Metanol).....	104
A5. Codificação das variáveis do processo associadas à produção de Biodiesel (Etapa da secagem da Glicerina).....	105
A6. Codificação das respostas obtidas para a análise do produto final do processo produtivo (Biodiesel).....	105

Abreviaturas

AA	Absorção Atómica
c.d.o.	Comprimento de onda
CE	Comunidade Europeia
CFPP	Cold Filtering Plugging Point
CP	Cloud Point
CO ₂	Dióxido de Carbono
DQ	Departamento de Química
EUA	Estados Unidos da América
FAME	Fatty Acid Methyl Esthers
FCTUC	Faculdade de Ciências e Tecnologia da Universidade de Coimbra
FFA	Free Fatty Acid
ICP	Plasma Acoplado Indutivamente
ISP	Imposto sob produtos petrolíferos
MeOH	Metanol
Na-Methylat	Metilato de Sódio
NP EN ISO/IEC	Norma Portuguesa; European Normalization ; International Organization for Standardization/ International Electrotechnical Commission
PC	Componente principal
PCA	Análise de Componentes Principais
PLS	Regressão parcial dos mínimos quadrados
ppm	partes por milhão
S.A.	Sociedade Anónima
S _x	Desvio Padrão
TriG.	triglicéridos
UPB	Unidade de Produção de Biodiesel
x _m	média
2D	duas dimensões (Bidimensional)
3D	três dimensões (Tridimensional)

Agradecimentos

Ao meu orientador de estágio, Professor Doutor Jorge Luís Gabriel F. S. Costa Pereira, pela oportunidade de participar neste projecto e pela disponibilidade sempre que solicitada, atenção, dedicação e empenho neste projecto. Obrigada pela paciência, compreensão e incentivo.

Ao meu co-orientador na Iberol, Doutor Renato Carvalho, pela disponibilidade e atenção sempre disponibilizadas, dados fornecidos e explicações sobre o processo de produção de biodiesel, que permitiram a realização deste trabalho.

À Iberol – Sociedade Ibérica de Biocombustíveis e Oleaginosas, minha entidade empregadora, o meu muito obrigada pela autorização e disponibilidade em realizar este trabalho nas suas instalações fabris.

À minha chefe, Eng.^a Maria do Carmo Lobato, Chefe do Laboratório Iberol, por tudo o que me tem ensinado durante estes quatro anos ao nível do Laboratório e Produção, que foram de extrema importância para a realização e entendimento deste trabalho.

Aos meus colegas da Unidade de Produção de Biodiesel (UPB), que sempre que eu tinha uma dúvida me esclareciam com enorme entusiasmo e entreatajuda.

Obrigada às minhas colegas do Laboratório que sempre me deram apoio e incentivo para este trabalho de projecto.

À minha família pela dedicação e apoio que sempre tiveram durante todo este tempo e que é sempre incondicional.

Obrigada aos meus amigos e colegas que passando dias comigo, sejam de trabalho ou de lazer, me incentivaram de alguma forma.

Obrigada a todos.

Objectivos

Este trabalho pretendeu trabalhar dados recolhidos do processo produtivo relativos à Unidade de Produção de Biodiesel (UPB) bem como relacioná-los com os dados recolhidos no Laboratório Iberol, que se encontra em processo de acreditação de acordo com a NP EN ISO/IEC 17025:2005 e rege as suas análises de acordo com as especificações da Norma EN 14214 e analisando matérias-primas, produto final e sub-produtos envolvidos e resultantes da produção do Biodiesel.

Com este projecto pretendeu-se, numa primeira fase, analisar o impacto da matéria-prima utilizada, bem como as condições processuais sobre as características do produto obtido ao nível da transesterificação e subsequente processamento.

Após esta fase inicial procurou-se, com base na variabilidade natural do processo, realizar uma análise factorial impondo a modelação do sistema por mínimos quadrados parciais (PLS). Esta perspectiva pretende encontrar quais os factores mais relevantes do sistema e com base nestes procurar obter uma melhoria nos parâmetros desejáveis para o processo.

O objectivo deste trabalho passou pela recolha de dados relativos à transformação de matéria-primas, análise do processo de fabrico e sua transformação em Biodiesel e seus subprodutos. Pretende-se fazer uma análise multivariada do processo de produção do Biodiesel com vista à obtenção de correlações e estabelecer relações que sejam de alguma forma esclarecedoras para alguma conjecturas que se fazem relativamente ao processo produtivo e respectivos resultados obtidos, esperando-se assim que os resultados que se irão obter no final deste trabalho espelhem de alguma forma o que são as melhorias processuais que se pretendem.

Resumo

No contexto actual em que se insere a produção e consumo dos Biocombustíveis em Portugal, foram desenvolvidos combustíveis alternativos como recorrência a uma forma de energia renovável de duração ilimitada e que cause menor impacto ambiental do que os combustíveis de origem fóssil.

A Iberol, possui nas suas instalações fabris uma unidade de produção de Biodiesel (UPB) cujo objectivo foi aumentar a competitividade da empresa bem como promover a sua preocupação com o meio ambiente.

As condições de fabrico de biocombustíveis são um exemplo excelente de um sistema multivariado em que é por demais relevante realizar um diagnóstico do processo de fabrico com vista a procurar otimizar e rentabilizar a produção, mantendo os parâmetros do produto dentro dos intervalos de confiança impostos pela norma EN 14214.

Este trabalho começou com a recolha de dados quer ao nível da produção, quer ao nível do laboratório da Iberol.

Após a recolha de dados e constituição da matriz original de dados estes foram tratados de modo a iniciar os cálculos que iriam conduzir aos resultados obtidos.

Sendo este sistema constituído por cerca de 200 variáveis o que constituiu demasiada informação para ser processada, realizou-se uma abordagem preliminar a cada um dos três sectores de dados – as matérias-primas, as condições de produção e as características finais do produto na primeira fase de tratamento.

Considerou-se que dos predictores (x) fazem parte toda a informação relativa à matéria-prima, bem como todos os dados relativos ao processo produtivo. As respostas (y) dizem respeito a todos os parâmetros de análise de produto final (análise completa de lote de expedição).

Tentou-se verificar alguma eventual relação entre as matérias-primas e as condições de produção explorando o sub-espaco dos predictores.

Quando o sub-espaco dos predictores (matérias-primas e condições de produção) estavam estudados seguiu-se para a exploração das inter-relações entre predictores e respostas.

Como um dos objectivos deste trabalho era encontrar as inter-relações entre variáveis, a normalização das variáveis foi realizada em todos os casos no sentido de minimizar os efeitos de escala e, deste modo, permitir a maximização das respectivas inter-relações.

Efectuou-se um cálculo e análise das componentes principais que antecedeu a realização de uma análise factorial através da modelação do sistema por mínimos quadrados parciais (PLS).

Foram obtidas correlações importantes entre variáveis e que são o espelho de todos os factores e problemáticas do processo produtivo do biodiesel.

No sentido de compatibilizar a transferência de resultados entre aplicações de software (octave/excel/word), na representação de números reais, optou-se por utilizar o símbolo “.” (ponto) como separador decimal em vez do símbolo “,” (vírgula) adoptado pela escrita portuguesa.

1. INTRODUÇÃO

1. INTRODUÇÃO

A maioria das necessidades energéticas mundiais é suprida através do petróleo e seus derivados, que constituem recursos não renováveis. Mantendo-se o consumo actual, a diminuição das reservas será um facto inevitável que aumentará a dificuldade de extração e respectivo custo. É neste contexto que surgem os Biocombustíveis como propulsores dos combustíveis renováveis para o setor dos transportes.

A Iberol – Sociedade Ibérica de Biocombustíveis e Oleaginosas S.A. foi pioneira em Portugal no desenvolvimento de Biodiesel produzido a partir de óleos vegetais com características adequadas à sua incorporação no diesel convencional.

1.1. Enquadramento do trabalho

Nos últimos anos tem-se verificado uma escalada dos preços do petróleo, devida a factores de produção e instabilidade nos países produtores, bem como as especulações financeiras a nível internacional. Este facto tem-se denotado bastante prejudicial para o desenvolvimento económico de muitos países desenvolvidos e em vias de desenvolvimento, principalmente para os mais dependentes desta fonte de energia como Portugal, aliadas aos problemas económicos e políticos, além dos ambientais causados pelo uso generalizado e desmedido de combustíveis fósseis, tornam necessário que se recorra a uma energia renovável de duração ilimitada e que cause menor impacto ambiental que as anteriores. Tal alternativa terá que ser competitiva economicamente, viável em termos técnicos, aceitável ambientalmente e com elevada disponibilidade.

Assim sendo, torna-se imperativo reduzir rapidamente a dependência do petróleo, pelo que, a Directiva 2009/28/CE, mais conhecida como RED (Renewable Energy Directive), para atingir os seus objectivos, definiu as seguintes metas para a UE:

1. Obter no consumo final bruto de energia até 2020 um valor de 20% de energia proveniente de energias renováveis.
2. Obter no sector dos transportes um valor de 10% de energia proveniente de energias renováveis.

A Directiva sobre Energias Renováveis (RED) definiu que todos os seus Estados-Membros devem elaborar e aprovar um Plano Nacional para as Energias Renováveis (PNAER). O PNAER serve como explicação das medidas tomadas por cada Estado-Membro para atingir as metas estabelecidas, tendo sempre como objectivo a maior eficiência energética e a menor poluição.

Assim, o PNAER português tem como metas:

- Obter no consumo final bruto de energia até 2020 um valor de 31% de energia proveniente de

energias renováveis

· Obter no sector dos transportes um valor de 10% de energia proveniente de energias renováveis.

O PNAER segue os princípios definidos pela Estratégia Nacional para a Energia (ENE2020). A ENE2020 é um conjunto de medidas tomada pelo conselho de ministros, cujo principal objectivo é tornar Portugal num líder europeu na indústria das energias renováveis. Para se atingir este objectivo o ENE2020 definiu os seguintes princípios:

1. Agenda para a competitividade, o crescimento e a independência energética e financeira;
2. Aposta nas energias renováveis;
3. Promoção da eficiência energética;
4. Garantia da segurança de abastecimento;
5. Sustentabilidade da estratégia energética.

Para tal será necessário incentivar a indústria para a inovação tecnológica, de forma a otimizar os processos produtivos para consumirem menos energia, generalizar a utilização de energias renováveis (como a produzida pelas ondas, pela biomassa, pelo vento, pelo sol e os biocombustíveis) e incentivar e investir na energia hídrica [1].

Neste contexto, a Iberol decidiu investir na construção de uma fábrica de produção de biodiesel.

O laboratório da Iberol foi reequipado, de modo a poder efectuar a maioria das análises de acordo as especificações recomendadas na norma EN 14214.

O presente trabalho pretendeu verificar a possibilidade da utilização de técnicas quimiométricas, na caracterização e análise do processo produtivo do biodiesel. Além disso, pretendeu-se relacionar as matérias-primas, suas características, processo produtivo do Biodiesel e caracterização e análise de produto final, como forma de estabelecer relações entre todos estes factores.

1.2. A Iberol

A Iberol é uma empresa com sede e instalações fabris em Alhandra, no concelho de Vila Franca de Xira, a 20 km de Lisboa. O facto de, a dada altura, existir no mercado um excesso de óleo, principal produto da Iberol, conjugado com as directivas europeias de incorporação de biocombustíveis nos combustíveis fósseis das fontes móveis de combustão (conferência de Estocolmo (1972), Protocolo de Kyoto (1997), Directiva 2001/77/EC, Directiva 2003/30/EC e por fim Directiva 2009/28/EC (esta já depois da implementação da UPB)) levou ao avanço com um projecto de implementação de uma unidade destinada à produção de biodiesel, que entrou em pleno funcionamento em Março de 2006.

A produção de biodiesel na Iberol resulta da utilização de óleo neutralizado aí produzido. Além disso, de forma a cumprir algumas das especificações na EN 14214, a Iberol efectua a mistura de diferentes óleos.

Em termos da comercialização do biodiesel, sendo os seus custos de produção ainda elevados

relativamente ao diesel convencional, o Governo Português aprovou, em Março de 2006, a isenção parcial e total do imposto sobre os produtos petrolíferos e energéticos (ISP) aos biocombustíveis, quando incorporados na gasolina e no gasóleo, utilizados nos transportes, isenção essa que deixou de estar em vigor no presente ano de 2011.

A produção, no caso da Iberol, está a ser comprada maioritariamente pela GalpEnergia, que iniciou a incorporação de biodiesel no início de 2006, sendo a Iberol o seu maior fornecedor.

No entanto, a Iberol possui outros clientes que compram o seu Biodiesel tais como a CEPSA, REPSOL e BP.

Além disso, no decorrer dos últimos anos, a Iberol conseguiu a certificação com base na ISO 9001 e 14001, que garantem a qualidade dos seus serviços e produtos.

O Laboratório da Iberol encontra-se na fase final do processo de acreditação pela NP EN ISO/IEC 17025:2005 [2].

1.3. O biodiesel como combustível alternativo

A hipótese de utilização de óleos vegetais como alternativa aos combustíveis fósseis surgiu há mais de 100 anos, com Rudolph Diesel em 1885, o inventor do motor a diesel. Depois de testar óleo de amendoim num dos seus motores, Diesel afirmou que: *“The use of vegetable oils for engine fuels may seem insignificant today, but such oils may in course of time be as important as petroleum and the coal tar products of the present time”* [3,4].

1.3.1. Óleos Vegetais e Seus Derivados

Devido à abundância de petróleo no século XX, aliada aos baixos custos dos seus derivados, estes combustíveis não convencionais nunca acompanharam o mercado dos combustíveis fósseis.

Actualmente, os óleos vegetais podem ser utilizados em motores diesel desde que as suas características específicas sejam devidamente consideradas. Comparativamente ao gasóleo, os óleos vegetais possuem maior densidade, maior viscosidade (cerca de 11-17 vezes maior) e menor poder calorífico. As suas características físico-químicas colocam assim limitações à sua utilização:

- × Elevadas perdas de pressão no sistema de alimentação de combustível
- × Entupimento frequente dos filtros de combustível, provocado por depósitos de carbono
- × Más condições de lubrificação dos órgãos de injeção de combustível, particularmente a baixas temperaturas
- × Modificações geométricas dos jactos emergentes dos bicos dos injectores

- × Alterações significativas das propriedades do óleo de lubrificação

Esta série de limitações impõe sérios condicionamentos à sua utilização em motores diesel, particularmente do tipo de injeção directa, os mais utilizados hoje em dia.

No entanto, e apesar das dificuldades aqui referidas, a Elsbett (Alemanha e Malásia) e a Diesel Morten und Geraetebau GmGH (Alemanha e Estados Unidos) conseguiram construir motores que demonstram boas performances quando abastecidos com óleos vegetais de diferentes composições [5,6].

Como alternativa à utilização directa dos óleos vegetais como combustíveis existe a possibilidade de utilizar os produtos resultantes de transformações destes óleos através, por exemplo, das reacções de transesterificação e de pirólise [6]. A transesterificação é, sem dúvida, a transformação mais importante e permite produzir um combustível limpo e ambientalmente aceite: o biodiesel. Como será referido num item posterior, a transesterificação consiste na reacção entre um óleo vegetal e um álcool, na presença de um catalisador. Esta reacção promove a quebra das ligações entre as cadeias de triglicéridos e leva a produção de ésteres metílicos dos ácidos gordos constituintes e de glicerina como subproduto.

1.3.2. Biodiesel e aspectos gerais

O biodiesel é um combustível limpo, alternativo aos combustíveis convencionais, produzido através de fontes biológicas renováveis como óleos vegetais ou gorduras animais. O biodiesel é não inflamável, não tóxico (10 vezes menos tóxico que o sal de mesa), biodegradável e livre de enxofre ou de compostos aromáticos.

Embora o biodiesel possa ser obtido a partir de uma gama muito distinta de óleos vegetais, os óleos de colza e de girassol são os mais utilizados a nível mundial (ver **Fig.1.1.**) e europeu.

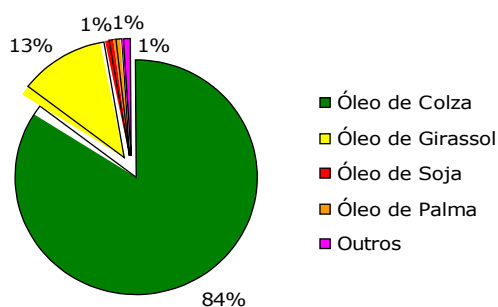


Figura 1.1 – Distribuição mundial de óleos vegetais utilizados para a produção de biodiesel. (Adaptado de [6]).

A Iberol utiliza na sua produção de Biodiesel **óleo de Colza, óleo de Soja e Oleína de Palma**.

A utilização de determinado óleo vegetal para produção de biodiesel encontra-se intimamente ligada à distribuição geográfica do país produtor.

Um aspecto interessante que se tem vindo a observar é a aplicação da Engenharia Genética neste campo. Assim, começam a surgir, por exemplo, novas variedades de colza, geneticamente modificadas, que apresentam maiores conteúdos em ácidos gordos. De facto, enquanto que a colza comum apresenta um conteúdo de ácidos gordos na ordem dos 60%, a variedade LZ 7632 (geneticamente modificada) apresenta um conteúdo de ácidos gordos até 87% [6,7].

As consequências decorrentes da produção e utilização de biodiesel têm merecido uma particular atenção, uma vez que a valorização da sua utilização é indubitavelmente marcada não só por benefícios ambientais dela decorrentes, como por benefícios a nível da saúde humana. Assim, as emissões de diversos produtos resultantes da queima do diesel convencional considerados cancerígenos são significativamente diminuídas com a substituição por biodiesel.

Por outro lado, um dos principais objectivos do protocolo de Quioto, para evitar o aquecimento global é a redução das emissões de dióxido de carbono.

A origem dos óleos vegetais utilizados para a produção de biodiesel está muito associada a questões geográficas dos países onde são produzidos [8]. Com efeito, na Europa utilizam-se mais os óleos de girassol e de colza; nos EUA a opção é pelo óleo de soja; e nos países asiáticos a opção é pelo óleo de palma.

Neste contexto, evidencia-se na **Tabela 1.1** quais os ácidos gordos de diversos óleos vegetais e gorduras animais, possíveis matérias-primas para a produção de biodiesel. A nomenclatura associada aos ácidos gordos tem como base o número de átomos de carbono e número de ligações duplas. A nomenclatura referente aos ésteres alquílicos constituintes do biodiesel é similar.

Tabela 1.1 – Ácidos gordos presentes nos óleos utilizados na transesterificação

Composição em Ácidos Gordos			
	Óleo de Palma (%)	Óleo de Soja (%)	Óleo de Colza (%)
Ácido Láurico (C12:0)	0-0,4	<0,1	-
Ácido Mirístico (C14:0)	0,5-2,0	≤0,2	0-2
Ácido Palmítico (C16:0)	32,0-45,0	9,9-12,2	1,0-5,0
Ácido Palmitoleico (C16:1)	0-0,6	0-0,2	-
Ácido Heptadecanóico (C17:0)	-	-	-
Ácido Heptadecenóico (C17:1)	-	-	-
Ácido Esteárico (C18:0)	2,0-7,0	3,0-5,4	0-3,0
Ácido Oleico (C18:1)	38,0-52,0	17,7-26,0	59,7-63,0
Ácido Linoleico (C18:2)	6,5-12,0	49,7-58,0	20,0-21,3
Ácido Linolénico (C18:3)	0-0,3	5,0-10,0	8,5-9,3
Ácido Araquídico (C20:0)	-	0,1-0,3	-
Ácido Gadoleico (C20:1)	-	0,2-0,5	0-1,0
Ácido Beénico (C22:0)	-	0,3-0,7	0-1,0
Ácido Erúcico (C22:1)	-	-	-
Ácido Linhocérico (C24:0)		≤0,4	-
Ácido Nervónico (C24:1)	-	-	-
Isómeros transoleicos (T-C18:1)	-	-	-
Isómeros trans (linoleico+linolénico) (T-C18(2+3))	-	-	-

1.4. Química do Biodiesel e Propriedades

No que diz respeito à composição química dos óleos vegetais e das gorduras animais, os seus constituintes principais são os triacilgliceróis, normalmente designados de triglicéridos.

Em termos químicos, os triglicéridos são três ésteres de vários ácidos gordos associados a uma molécula de Glicerol (1,2,3-propanotriol).

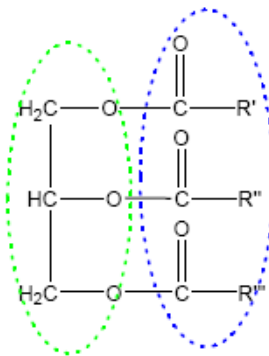


Figura 1.2 – Fórmula estrutural genérica de triglicéridos, onde se evidencia a estrutura química do glicerol (a verde) e as moléculas de ácidos gordos (a azul, onde normalmente $R' \neq R'' \neq R'''$).

Como diversos ácidos gordos podem estar ligados à estrutura química do glicerol e, considerando que os diferentes ácidos gordos têm propriedades físicas e químicas diferentes, então, conforme a composição em ácidos gordos nos triglicéridos, os óleos adquirem propriedades diferentes (viscosidade, ponto de fusão, índice de cetano e outros). Deste modo, através da reacção de transesterificação, as moléculas de triglicéridos de óleos vegetais ou gorduras animais são convertidas em ésteres monoalquílicos de cadeia longa, que normalmente se designam de biodiesel.

A origem dos óleos vegetais utilizados, provêm de Colza, Soja e Palma e que são as matérias-primas de origem vegetal utilizadas para a produção de Biodiesel na Iberol, como referido anteriormente.

O óleo pode ser comprado ou extratado nas instalações das Iberol a partir de grão importado.

O processo de extração do óleo vegetal a partir do grão é um processo bastante complexo, no entanto, apresenta-se um esquema simplificado do mesmo na Figura 1.3 para uma melhor compreensão.

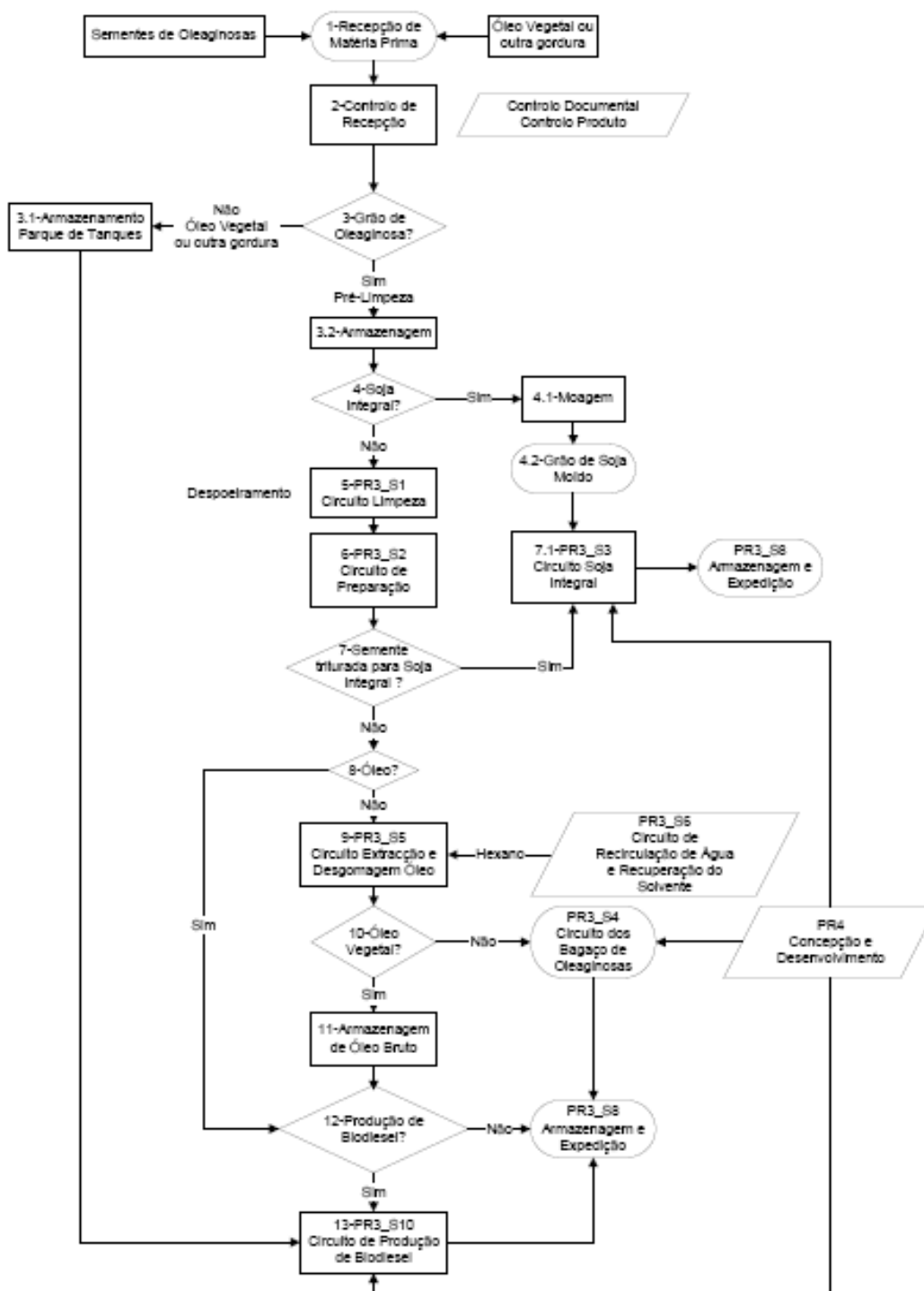


Figura 1.3 – Diagrama esquemático representativo do processo de extração do óleo a partir do grão

A nomenclatura associada aos ácidos gordos tem como base o número de átomos de carbono e números de ligações duplas, ou seja, um ácido gordo C18:1 terá uma cadeia com 18 átomos de carbonos e uma ligação dupla. A nomenclatura referente aos ésteres alquílicos constituintes do Biodiesel é similar.

A reacção de transesterificação ocorre recorrendo ao uso de um catalisador que aumenta a velocidade, selectividade e rendimento da reacção sem afetar o seu equilíbrio (seja a reacção a principal ou a secundária). Dependendo do tipo de catalisador e das suas propriedades, tem-se diferentes mecanismos de reacção.

1.5. Química da reacção de transesterificação

Como foi referido, a reacção mais comum para a transformação dos óleos é a reacção de transesterificação, que é efectuada recorrendo ao auxílio de um catalisador que aumenta a velocidade, selectividade e o rendimento da reacção. Assim, o tipo de catálise que se utiliza determina o tipo de transesterificação utilizada, como se descreve:

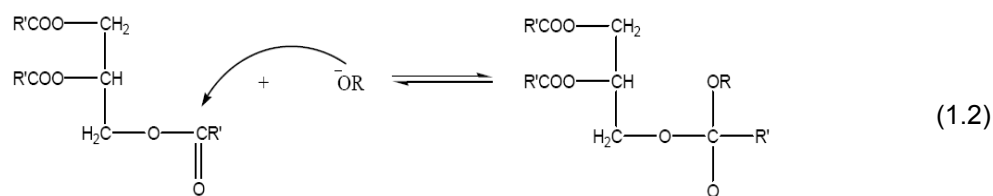
- Transesterificação catalisada por uma base;
- Transesterificação catalisada por um ácido (utilizada principalmente quando o óleo tem um elevado teor de ácidos gordos livres e um elevado teor de humidade. O rendimento é bastante elevado, mas a reacção é lenta, sendo necessárias temperaturas elevadas e tempos de reacção elevados. Temperaturas elevadas e tempos de reacção elevados é uma razão molar normalmente superior em metanol:óleo) [9];
- Transesterificação catalisada por enzimas [10] (trata-se de um processo ainda em estudo, em que enzimas, mais particularmente lipases, catalisam a reacção de formação do biodiesel) e
- Catálise heterogénea – em alternativa a catálise homogénea, tem-se igualmente procurado um catalisador sólido ácido ou alcalino. Sabe-se que alguns óxidos metálicos (como os de chumbo, magnésio e zinco) catalisam a reacção, mas actuam seguindo o mecanismo da catálise homogénea, formando sabões dos metais [11]. Existem também alternativas de óxidos mistos de zinco e alumínio, fixos em reactores em determinadas proporções, que promovem a catálise sem perda do catalisador [12].

Presentemente, a escolha recai sobre a catálise homogénea, mais particularmente sobre a catálise alcalina, principalmente por conseguir melhores rendimentos e ser muito mais rápida.

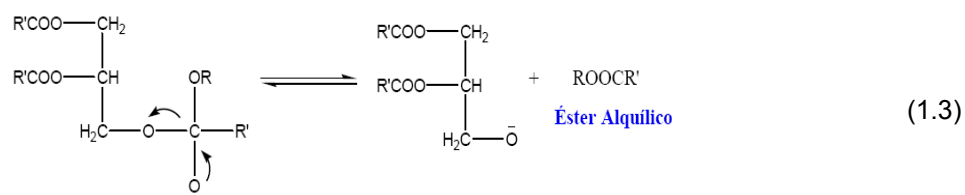
O processo começa com a reacção do catalisador, uma base forte (B) com o álcool, originando um alcóxido e o catalisador protonado como mostra a seguinte equação,



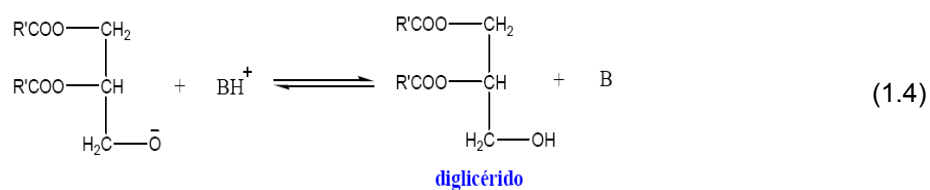
Seguidamente, o alcóxido efectua um ataque nucleófilo ao grupo carbonilo do triglicérido,



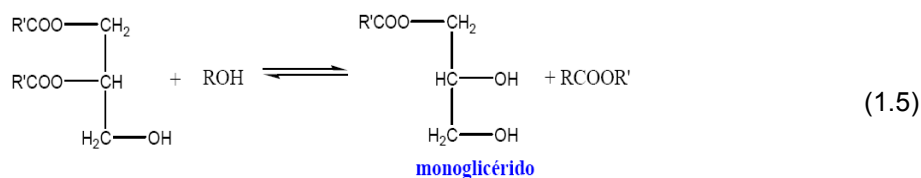
originando um produto intermédio que, por deslocalização electrónica, leva à formação do éster monoalquílico (equação 1.3)



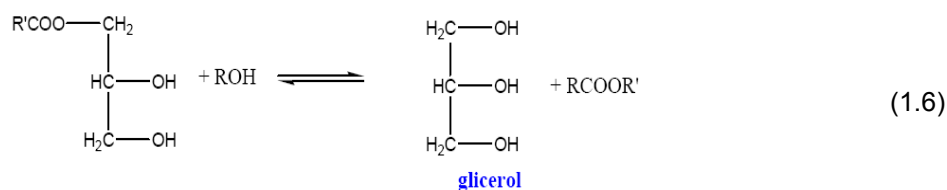
e de uma espécie intermédia (que é o anião do diglicérido), tal como se pode verificar pela equação 1.4 que é estabilizada pelo próton do catalisador, o que regenera a espécie activa.



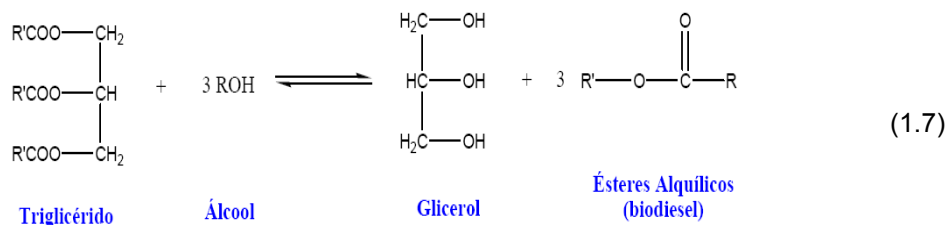
Isso possibilita que se realize um novo ciclo catalítico, com um mecanismo reaccional semelhante, efectuando-se a passagem dos diglicéridos a monoglicéridos (equação 1.5)



e destes últimos a glicerol (equação 1.6).



Isto corresponde à formação de três moles de ésteres monoalquílicos por cada mole de triglicéridos utilizada e cuja reacção global é dada pela equação 1.7 onde R e o grupo alquílico do álcool e R' e uma mistura de várias cadeias de ácidos gordos

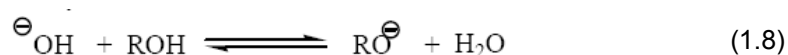


Todas estas reacções são reversíveis, daí que seja utilizado um excesso do álcool ou a remoção de um dos produtos da reacção (por exemplo, o glicerol) para deslocar os equilíbrios para a formação dos ésteres. Na presença de um excesso de álcool, a reacção de formação dos ésteres é de pseudo primeira ordem, enquanto que a reacção inversa é de segunda ordem, o que justifica um deslocamento do equilíbrio no sentido dos ésteres, uma vez que a velocidade de reacção neste sentido é muito superior à do sentido inverso [13].

Assim, em conformidade com o mecanismo reaccional apresentado, enquanto que os triglicéridos diminuem muito rapidamente, vão-se formando diglicéridos e a partir destes monoglicéridos com consequente formação do biodiesel. Normalmente, o álcool de cadeia curta utilizado (ROH) é o metanol (designando-se reacção de metanólise), apesar de também se poderem utilizar outros alcoóis.

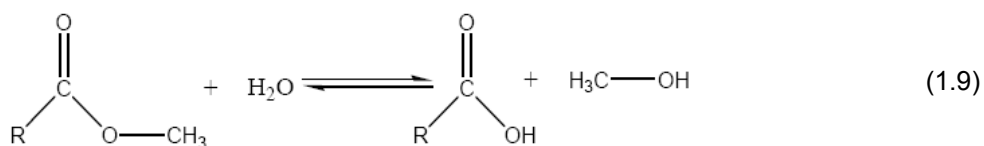
A utilização do metanol justifica-se pelas seguintes razões, o seu consumo é inferior ao do etanol em condições idênticas (uma vez que a separação é mais fácil; as constantes de solubilidade são menores que o etanol); o tempo de reacção necessário é menor do que quando se utiliza o etanol; a dimensão dos equipamentos utilizados é inferior aos utilizados para o etanol; e o consumo de vapor para aquecimento é aproximadamente 20% do consumo quando se utiliza etanol, o que leva a menor consumo de electricidade [13].

Quanto aos catalisadores, os mais utilizados são normalmente o hidróxido de sódio ou potássio ainda, os respectivos alcóxidos dissolvidos em metanol. Uma das vantagens de utilizar os alcóxidos é a não formação de água, que ocorre quando se utilizam os hidróxidos:



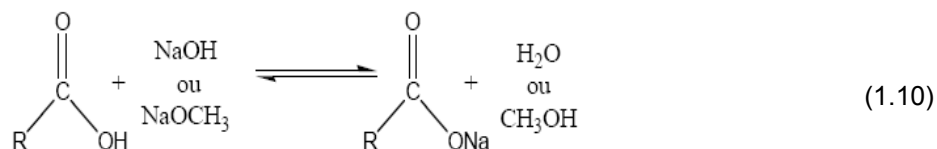
Equação 1.8 - Reacção de formação de água entre o álcool e o hidróxido.

A existência de água é problemática, pois pode provocar a hidrólise dos ésteres alquílicos formados em ácidos gordos livres:



Equação 1.9 – Reacção de hidrólise de um éster metílico.

Como consequência desta hidrólise, formam-se sabões, que, além de reduzirem o rendimento da reacção, dificultam a separação entre a fase de ésteres e a de glicerol, uma vez que provocam a formação de emulsões.



Equação 1.10 – Reacção de formação de sabões.

2. DESENVOLVIMENTO

2. DESENVOLVIMENTO

O biodiesel é produzido quando ocorre uma reacção de transesterificação entre um éster e um dado álcool em excesso, com catálise ácida ou básica, que dá origem a um novo éster e um álcool.

O biodiesel produzido na Iberol é obtido por uma reacção de transesterificação entre os triglicéridos existentes na mistura de óleos neutros e o metanol (álcool), na presença do metilato de sódio (catalisador – acelera a reacção). Uma vantagem do uso de metilato de sódio como catalisador em detrimento de outro (NaOH, por ex.) é o facto de não ocorrer a formação de água durante a catálise, problema que já foi referido e explicado anteriormente. Assim, por cada molécula de triglicérido e três de metanol, formam-se três moléculas de metiléster (Biodiesel) e uma de glicerol (Glicerina).

No entanto, a transformação do metanol é igual quer seja com o hidróxido ou com o alcóxido. O que acontece é que os alcóxidos já estão um passo à frente na reacção que os hidróxidos, pois estes últimos têm que se transformar em alcóxidos com o metanol.

2.1. Processo de Produção de Biodiesel

O óleo vegetal que não é produzido na Iberol é recepcionado e armazenado (por exemplo a oleína de palma).

O óleo vegetal bruto (extraído e desgomado fisicamente na unidade fabril de Preparação/Extracção da Iberol) é armazenado em tanques de onde é bombeado via *pipeline* para os tanques de mistura.

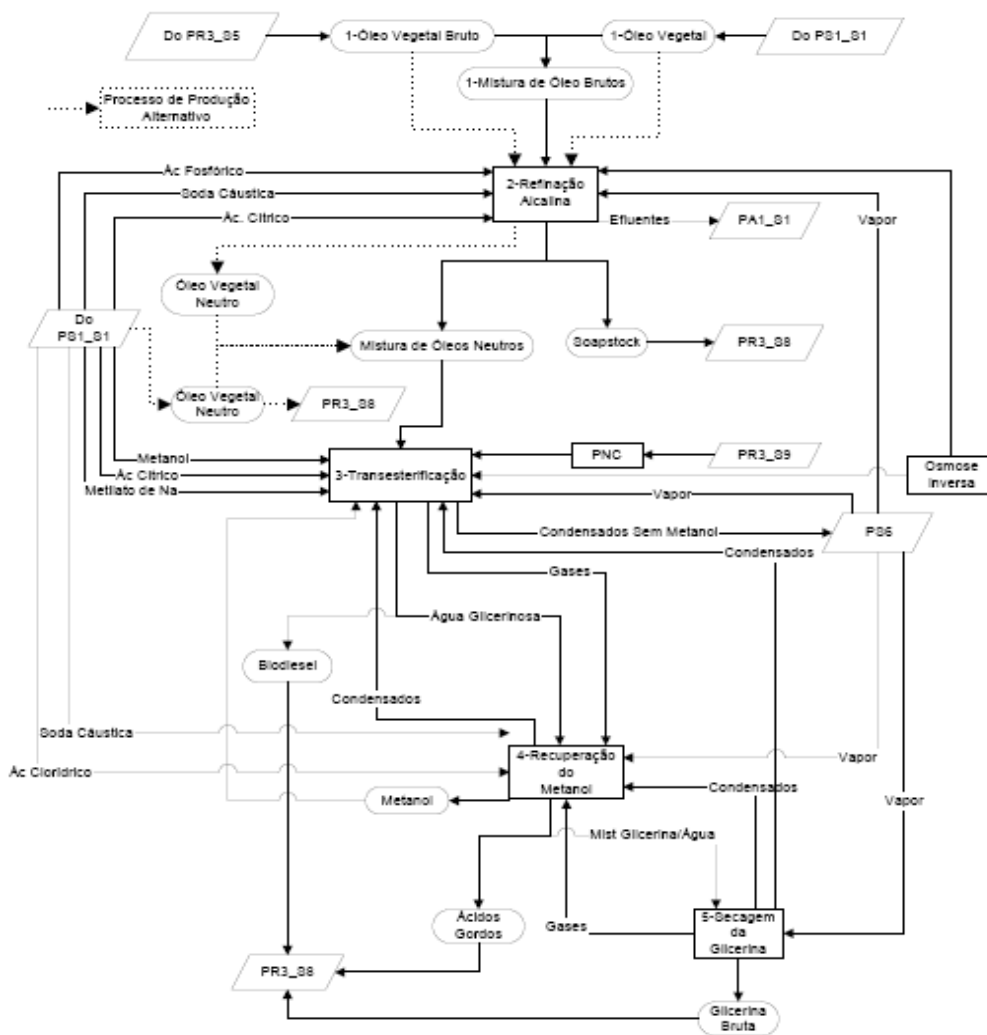


Figura 2.1 – Diagrama abreviado do processo de produção de biodiesel.

2.1.1. Refinação Alcalina dos óleos

A refinação alcalina da mistura de óleos ocorre em quatro fases distintas:

2.1.1.1. Desgomagem Química

Consiste na conversão de fosfatídeos não-hidratáveis (e que por isso não foram removidos do óleo na Desgomagem Física) em ácido fosfatídico (gomas – insolúveis no óleo) para que sejam eliminadas juntamente com as Pastas de Neutralização. De um tanque intermédio a mistura de óleo/ou óleo de um só tipo é bombeado ao permutador de placas, que funciona como um economizador que aproveita a temperatura do óleo neutro seco para aquecer o óleo bruto até 60 °C, à saída é colhida uma amostra do óleo bruto para determinação do valor de acidez. Após determinação do valor de acidez este valor é introduzido no software de supervisão da UPB desencadeando o doseamento e controlo

automático dos diferentes fluídos. O aquecimento do óleo prossegue num permutador, onde atinge os 75 °C. À saída do permutador é doseado no óleo bruto aproximadamente 0.06 % (dependendo da acidez do óleo bruto) de ácido fosfórico a partir de um tanque, ocorrendo mistura hidrodinâmica na tubagem e dinâmica num misturador. A mistura do óleo vegetal com o ácido fosfórico reage ao longo do percurso num tanque com entrada e saída contínua de produto [15].

2.1.1.2. Neutralização

A acidez natural do óleo aumenta com a desgomagem química pelo que se torna necessário neutralizar os ácidos gordos livres, ocorrendo uma reacção de saponificação. A neutralização do óleo antes da transesterificação é uma etapa importante porque os ácidos gordos livres vão diminuir o rendimento da reacção, formando sabões na presença de água.

A neutralização do óleo é realizada com soda cáustica. De um tanque a soda é direccionada por bomba doseadora a um misturador, variando o doseamento de acordo com a determinação da acidez do óleo bruto, utilizando-se um excesso de ± 20 %. Antes da entrada no misturador, ocorre na tubagem a diluição da soda em água osmotizada quente (90 °C), proveniente de um tanque, de modo a que a concentração da soda atinja os 12-14 %.

A solução de soda cáustica prossegue para um misturador, onde se junta ao óleo/gomas proveniente do tanque de retenção e a água de aglomeração a 0.45 %.

A emulsão segue para um tanque onde fica em agitação e após a neutralização dos ácidos gordos livres resultam duas fases distintas de diferentes densidades. Por um processo que passa pela adição de água quente dá-se a separação das duas fases (óleo neutro/mistura de gomas e sabões) [15].

2.1.1.3. Lavagem do óleo

O óleo neutro que se obtém da descrição anterior ainda apresenta vestígios de sabão pelo que é lavado por água osmotizada quente previamente acidulada com ácido cítrico e cujo doseamento depende do teor de sabão do óleo. A mistura de água e óleo leva à formação de duas fases que são separadas por centrifuga [15].

2.1.1.4. Secagem do óleo

O óleo neutro após processo de lavagem encontra-se com elevado teor de humidade pelo que é conduzido a um secador.

À saída do secador o óleo neutro é amostrado e é feita a análise do teor de acidez e sabão. Se a acidez se encontrar acima dos 0.08 % e o sabão acima dos 80 ppm é feito um reajuste nos parâmetros de desgomagem ao nível da sala de controlo da UPB de modo a que a dosagem de soda cáustica seja corrigida em modo automático.

Se a acidez se mantiver acima dos 0.10 % ou o sabão acima dos 120 ppm o óleo poderá ter que ser reprocessado [15].

2.1.2. Transesterificação

Na reacção de transesterificação estão envolvidos reagentes químicos que se encontram previamente armazenados em tanques específicos.

Numa primeira fase do processo, óleo neutro seco, metanol e metilato de sódio são doseados e misturados de forma hidrodinâmica numa tubagem, sendo que em primeira instância é feita a mistura de metanol com metilato de sódio. Os três são misturados em atmosfera inertizada com azoto e aquecida de forma, a que à saída do permutador a mistura apresente uma temperatura de 60 °C que é a adequada para que ocorra a conversão de glicéridos em metil-ésteres.

A mistura quente entra numa primeira coluna de transesterificação pelo topo estando sempre cheia nos diferentes estágios da reacção passando a mistura pelo fundo acima da linha de separação da Glicerina para a segunda coluna onde continua a reacção.

O metil-éster (Biodiesel) menos denso, é sobrenadante em relação à Glicerina que decanta no fundo das colunas sendo direccionada à entrada de uma centrífuga para separação do Biodiesel que possa arrastar.

Da segunda coluna de transesterificação o biodiesel prossegue para a centrífuga para separar a Glicerina residual que contenha. O Biodiesel prossegue para a segunda etapa de transesterificação que visa converter os triglicéridos que ainda não tenham reagido em metil-ésteres para que, a reacção seja mais completa possível [15].

É adicionada uma nova mistura de metanol com metilato de sódio à entrada do segundo par de colunas com o biodiesel e cuja mistura é sujeita a uma aumento de temperatura de 60 °C seguindo-se o enchimentos das terceira e quarta colunas de transesterificação em tudo idêntico à primeira fase de transesterificação.

2.1.3. Lavagem e Secagem do Biodiesel

Após a reacção de transesterificação o Biodiesel é lavado com água acidulada com ácido cítrico.

O ácido cítrico acidifica a água para melhor remover os sabões e evitar a criação de emulsões e respectivas perdas (diminuição de rendimento). A água é separada do Biodiesel por centrifugação. O Biodiesel é lavado outra vez com água quente osmotizada e separado novamente por centrifugação e segue para secagem.

O Biodiesel é aquecido aos 100 °C antes de entrar no secador sob forma pulverizada sendo a humidade e metanol residuais arrastados pelo vácuo.

O Biodiesel à saída do secador tem uma temperatura menor sendo arrefecido por transferência de

calor num permutador. A entalpia sensível será a que fornecerá energia para a latente e por isso vaporizar os contaminantes. O Biodiesel é arrefecido aos 35 °C por passagem por um permutador de calor onde circula água de refrigeração.

Este arrefecimento tem como finalidade a insolubilização de impurezas que não foram removidas do Biodiesel nas etapas anteriores. As impurezas são removidas por centrifugação seguindo para tanques de armazenagem intermédios B01 e B02. Quando cheios estes tanques são amostados e analisados no Laboratório Iberol e de acordo com a conformidade do produto são transferidos para os tanques de expedição C01/C02 ou em alternativa os tanque C03/04/05, se o produto final não se encontrar conforme segue para reprocessamento [15].

2.1.4. Recuperação do Metanol

A Glicerina que flui por gravidade das colunas de transesterificação e das centrífugas para um tanque é constituída por uma mistura de Glicerina, metanol, água, catalisador e sabões formados.

O primeiro passo consiste na conversão dos sabões por adição de HCl uma vez que este por ser um ácido forte causa a inversão da reacção de saponificação ocorrendo a formação de ácidos gordos insolúveis em água e sal (neste caso NaCl).

A mistura de ácidos gordos e glicerina é conduzida a um decantador onde se dá a separação por decantação de fases. Os ácidos gordos que ficam retidos num compartimento central são bombeados a um permutador aquecendo aos 100 °C para que se liberte o metanol residual existe nos ácidos gordos.

Após este processo a Glicerina encontra-se com um valor de pH muito baixo devido à adição de HCl pelo que é necessário alcalinizar a mistura (até pH 5) com NaOH (soda cáustica) ocorrendo a formação de NaCl e água.

A Glicerina segue para um processo de secagem através de conjuntos Aquecedor/Coluna de Evaporação. Os gases que saem pelo topo das colunas de evaporação são condensados e recolhidos para um Tanque de condensados [15].

2.2. Propriedades e especificações do biodiesel

O controlo de qualidade do biodiesel passa não só pelo controlo do produto final, o biodiesel, mas também da própria matéria-prima. Assim, nos óleos, é importante analisar parâmetros, como o teor em água, a sua acidez e o seu teor em ácidos gordos livres, uma vez que esses parâmetros vão influenciar o processo de fabrico de biodiesel. Por exemplo, uma elevada acidez fará com que o catalisador seja consumido a neutralizar o óleo e não a catalisar a reacção.

A qualidade do produto final é dependente de diversos aspectos, como sejam:

- Inerentes à matéria-prima seleccionada: por exemplo, o tipo de óleo utilizado pode afectar parâmetros como o índice de iodo, a densidade, as suas propriedades de ignição, a viscosidade e o *cold filter plugging point* (CFPP); e

- Relativos ao processo: por exemplo, diferentes condições experimentais como a quantidade de catalisador, metanol, temperatura e outros parâmetros que influenciam a pureza do biodiesel e o teor de contaminantes, como o glicerol, os glicéridos e o metanol residual que este contém. O principal problema, associado ao não cumprimento dos parâmetros de qualidade do biodiesel, passa pela produção de efeitos nefastos nos motores aquando da sua utilização, sobretudo a corrosão de metais e degradação de borrachas, a formação de depósitos e o entupimentos de filtros.

Desta forma, é necessário normalizar todos os parâmetros que asseguram a qualidade do biodiesel para garantir a sua utilização segura nos motores diesel. De facto, a normalização é um instrumento necessário para promover a introdução e incorporação do biodiesel no mercado de combustíveis. De seguida, apresentam-se os parâmetros de qualidade do biodiesel que são analisados de acordo com norma do Biodiesel EN 14214 [15].

2.2.1. Ponto de inflamação

Refere-se à temperatura a que o combustível liberta vapor suficiente para sofrer ignição quando sujeito a uma fonte de ignição, encontrando-se o respectivo método de determinação na norma EN ISO 3679. O facto do ponto de inflamação ("flash point") do biodiesel (>150 °C) ser superior ao do diesel convencional ([62-80]°C torna o biodiesel um combustível mais seguro, em termos de transporte e de armazenamento [17,18,19].

2.2.2. Viscosidade cinemática

Ter-se um combustível ser mais ou menos viscoso interfere com a atomização no injector dos motores diesel, podendo levar à formação de depósitos. Em termos estruturais, a viscosidade aumenta com o comprimento das cadeias das moléculas do combustível e com a diminuição do grau de insaturação [13]. Considerando que a viscosidade do biodiesel pode variar substancialmente, não

só devido a diferentes matérias-primas, mas também devido à presença de contaminantes, torna-se necessário garantir que este parâmetro não se afaste muito do valor de viscosidade do diesel convencional, que é de 2,31 mm²/s a 40 °C [21]. A sua determinação pelo método de referência passa pela medição do tempo de escoamento do biodiesel num viscosímetro, conforme a norma EN 3104.

2.2.3. Densidade

A densidade absoluta ou massa específica é uma característica própria de cada material, por isso é classificada como sendo uma propriedade específica. A densidade absoluta é definida como sendo a razão entre a massa de uma amostra e o volume ocupado por esta massa.

No caso do biodiesel, esta propriedade depende, por exemplo, do óleo de origem e do grau de transesterificação da amostra. Este parâmetro é determinado usando um densímetro e um termómetro calibrados, conforme a norma ISO 3675, já que a densidade varia com a temperatura.

2.2.4. Índice de cetano

É uma propriedade que mede a facilidade com que se dá a ignição e a suavidade da combustão, que se relaciona com o comprimento e ramificações das cadeias constituintes das moléculas do combustível. Assim, cadeias compridas e lineares aumentam este parâmetro, enquanto que cadeias ramificadas diminuem-no. Isto justifica a razão das moléculas de triglicéridos e, conseqüentemente, dos ésteres alquílicos se adequarem estruturalmente como combustíveis, já que se assemelham às dos n-alcenos. A influência desta propriedade na performance dos motores passa por, se o índice de cetano for muito elevado pode ocorrer combustão antes do combustível e de o ar estarem misturados convenientemente, o que resulta numa combustão incompleta e na libertação de maior quantidade de poluentes; se o índice de cetano for muito baixo atingem-se temperaturas mais elevadas, o que provoca um arranque mais lento do motor e a combustão também pode ser incompleta [16]. A sua determinação pelo método de referência encontra-se descrita na norma EN ISO 5165. A análise deste parâmetro não é efectuada na Iberol, mas solicitada a um Laboratório Acreditado externo.

2.2.5. Resíduo carbonoso

É uma medida que reflecte a tendência do combustível para provocar depósitos de carbono, encontrando-se a descrição da sua determinação pelo método de referência na norma ISO 10370. Assim, os filtros dos motores diesel retêm partículas de dimensão superior a 10 µm (ou superior a 2 µm para modelos mais recentes) impedindo estas de entrar no sistema de injeção de combustível. Porém, quando os combustíveis são sujeitos a temperaturas elevadas e na presença de oxigénio, podem ocorrer reacções químicas que formem compostos insolúveis no combustível, os quais

formam depósitos e sedimentos que podem entupir orifícios e danificar peças do motor. Devido à elevada concentração de compostos insaturados, o biodiesel é mais susceptível à degradação oxidativa do que o diesel convencional, estando relacionado principalmente com a presença de contaminantes, como ácidos gordos livres, sabões, glicéridos, polímeros e outros [13].

2.2.6. Teor de enxofre

O biodiesel contém quantidades residuais de enxofre, uma vez que este está presente também residualmente nos óleos vegetais e ainda em menor quantidade nas gorduras animais. Ao contrário do biodiesel, no diesel fóssil o enxofre é removido artificialmente, através de um método com emissão adicional de dióxido de carbono e perda de capacidade de lubrificação. A sua determinação é efectuada por espectrometria de fluorescência de raios X de comprimento de onda (c.d.o.) dispersivo, conforme a norma EN ISO 20846. A análise deste parâmetro não é efectuada na Iberol, sendo solicitada a um Laboratório Acreditado externo.

2.2.7. Teor de cinzas sulfatadas

Diz respeito ao resíduo que permanece após a amostra ter sido carbonizada, tratada com ácido sulfúrico, aquecida e pesada até massa constante. As impurezas inorgânicas existentes nos combustíveis (no caso do biodiesel são constituídas principalmente por vestígios de catalisador) podem produzir cinzas que causam fricção entre os pistões e os cilindros dos motores. Este parâmetro determina-se através de uma carbonização e adição de ácido à amostra, conforme a norma ISO 3987.

2.2.8. Teste de corrosão à lâmina de cobre

Refere-se à tendência do combustível para corroer metais, sendo por esse motivo, um indicador importante para prolongar a vida útil do motor, determinado pelo método descrito na norma EN ISO 2160 e que consiste em submeter uma lâmina de cobre submersa em biodiesel, durante três horas à temperatura de 50°C. No final, a lâmina é retirada e comparada com um padrão de cor e é lida a escala a que corresponde a cor que se obteve na determinação. O normal obtido nas determinações é a escala 1a, dentro da classe 1, da especificação da norma.

2.2.9. CFPP

É um parâmetro importante para garantir a boa operacionalidade do combustível mesmo com temperaturas baixas. Assim, procede-se ao arrefecimento de uma amostra a um ritmo constante, fazendo-a passar através de um filtro de arame (crivo) sob vácuo e verificando-se se há retenção de partículas no filtro, conforme a norma EN 116.

Como tal, o CFPP é a temperatura mais baixa à qual 40 mL de amostra passa através do filtro durante 60 s. O biodiesel, à semelhança do diesel convencional, pode utilizar-se a temperaturas baixas (até -20 °C), no entanto, verifica-se que em algumas situações com o arrefecimento, o biodiesel forma pequenos cristais que o tornam mais viscoso, sendo que as partículas formadas podem entupir os filtros do motor, fazendo com que o combustível não flua normalmente, o que influencia o desempenho das bombas e do injector. Quando tal acontece terão que ser adicionados ao biodiesel anti-congelantes (aditivos). Este é um valor contratual com cada cliente e que depende do tipo de óleos utilizados na produção do Biodiesel.

2.2.10. CP

É outro parâmetro, determinado conforme a EN 23015, que garante a boa operacionalidade do combustível mesmo com temperaturas baixas, pois estabelece a temperatura a que começa a turvação do biodiesel.

2.2.11. Estabilidade oxidativa

Os óleos vegetais e seus derivados possuem insaturações que os tornam mais susceptíveis de deterioração através de reacções hidrolíticas ou oxidativas. Essas reacções originam produtos insolúveis, causadores de problemas, por exemplo, na bomba injectora. Esta característica do combustível tem particular relevância no armazenamento e manuseamento do combustível, daí que devam ser escolhidos locais secos, ao abrigo da luz, de preferência em tanques de armazenagem de alumínio, polietileno, polipropileno ou teflon. Um factor que influencia bastante a estabilidade oxidativa do biodiesel e a matéria-prima utilizada para a sua produção.

2.2.12. Índice de acidez

É um indicador da quantidade de ácidos gordos livres no biodiesel. Se for elevado, é favorecida a deposição de resíduos no motor, assim como a corrosão das suas peças. A norma, EN 14104, descreve a sua determinação por titulação potenciométrica.

2.2.13. Índice de iodo

É uma medida do grau de insaturação dos ésteres alquílicos, sendo um indicador da propensão do material para polimerizar e para formar depósitos que reduzem o tempo de vida dos motores. O índice de iodo do biodiesel é semelhante ao do óleo que lhe dá origem (apenas ligeiramente inferior), daí que, por vezes, se tenha que misturar diferentes tipos de óleos de forma a cumprir o valor imposto pela norma (por exemplo, o óleo de soja tem um índice de iodo elevado, pelo que poderá proceder-se à mistura com óleo de palma ou óleo de colza para baixar esse valor). O método de referência relativo ao índice de iodo é o referido na norma EN 14111.

2.2.14. Teor em ésteres alquílicos ou pureza

Diz respeito à pureza do próprio biodiesel, ou seja, à quantidade de ésteres alquílicos que deverá ser a mais elevada possível, garantindo uma menor quantidade de contaminantes que irão acarretar problemas de ordem diversa bem como a operacionalidade do biodiesel. O método analítico de referência é a cromatografia gasosa, conforme a norma EN 14103.

2.2.15. Teor em ésteres metílicos de ácido linolénico

Este parâmetro (teor em C18:3) quantifica um dos ésteres alquílicos com três insaturações, sendo um indicativo da sua reactividade, por exemplo, resistência à oxidação, que como foi anteriormente referido, tem influência na utilização do biodiesel. O teor em ésteres metílicos de ácido linolénico é igualmente determinado por cromatografia gasosa, conforme a norma EN 14103.

2.2.16. Teores em mono-, di- e tri-glicéridos

Na reacção de transesterificação há passagem progressivamente de tri- a diglicéridos e de di- a monoglicéridos, daí que a quantidade final destes glicéridos seja um dos principais indicadores da extensão da reacção de transesterificação. Elevados teores de glicéridos levam à formação de depósitos no motor e ao entupimento de filtros. Relativamente à determinação destes parâmetros, esta é efectuada por cromatografia gasosa, conforme a norma EN 14105.

2.2.17. Teor em glicerol livre e total

O glicerol livre refere-se ao glicerol do biodiesel, enquanto que o total se refere ao livre e ao que está ligado às cadeias dos ácidos gordos dos mono-, di- e triglicéridos. Esta determinação é realizada por cromatografia gasosa, conforme a norma EN 14105.

2.2.18. Teor em água

A água pode estar presente no biodiesel, quer como água dissolvida quer como água em suspensão. O biodiesel é insolúvel em água, mas o teor em água no biodiesel pode atingir os 1500 ppm contra os 50 ppm no diesel fóssil. A especificação do Biodiesel estipula que o máximo permitido seja 500 ppm. A água pode causar o desenvolvimento de microorganismos que entopem filtros e pode causar corrosão no sistema de injeção dos motores. A determinação do teor em água utiliza uma titulação Karl-Fischer, conforme a norma EN ISO 12937.

2.2.19 Teor em metanol

O principal problema resultante da presença de metanol no biodiesel é a diminuição do ponto de inflamação do mesmo, o que pode acarretar problemas de segurança. O cumprimento dos dois parâmetros está de facto correlacionado e verifica-se que, quando um biodiesel cumpre a norma no que diz respeito ao teor em metanol (inferior a 0,2%), também cumpre o limite para o ponto de inflamação (superior a 120°C). A determinação do teor em metanol efectua-se por cromatografia gasosa de *head space*, conforme a norma EN 14110.

2.2.20 Teor em metais do grupo I (e grupo II)

A fonte destes metais provém essencialmente do catalisador. Contudo, grande parte do catalisador é removida com a fase do glicerol, sendo os restantes vestígios removidos no passo de lavagem. Estes metais, em conjugação com o indicador das cinzas sulfatadas, promovem a formação de depósitos e a abrasão a nível do motor. Os métodos de determinação destes metais passam pela utilização da absorção atómica (AA) ou de espectrometria de emissão com plasma (ICP) (conforme as normas EN 14108 e EN 14538).

2.2.21 Contaminação total

Corresponde ao teor de substâncias insolúveis no biodiesel (de dimensão $>0.8 \mu\text{m}$). Estes contaminantes podem dar origem a avarias e ao entupimento dos filtros. A sua determinação tem como base uma filtração laboratorial, conforme descrito na norma EN 12662.

Por fim, apresenta-se na tabela 2.1 os parâmetros de controlo de qualidade do biodiesel em vigor, de acordo com a norma EN 14214.

Tabela 2.1 Especificações do Biodiesel de acordo com a EN 14214

Parâmetro	Método / Norma	Limites		Unidade
		Min	Máx	
Densidade a 15 °C	ISO 3675	860	900	kg/m ³
Viscosidade a 40 °C	ISO 3104	3.50	5.00	mm ² /s
Cold Filter Plugging Point (CFPP)	EN 116	contratual		°C
Cinzas Sulfatadas	ISO 3987	-	0.02	%m/m
Teor de Água	EN ISO 12937	-	500	mg/kg
Corrosão à lâmina cobre (3h/50°C)	EN ISO 2160	classe 1		classe
Número de Acidez	EN 14104		0.50	mgKOH/g
Teor de Metanol	EN 14110		0.20	%m/m
Teor de Ésteres	EN 14103	96.5		%m/m
Éster metílico ácido linolénico	EN 14103		12.1	%m/m
Cloud Point	EN 23015	contratual		°C
Estabilidade Oxidativa (110°C)	EN 14112	6.0		h
Ponto de Inflamação	EN ISO 3679	101		°C
Resíduo Carbonoso - Método micro (10% do residuo da dest. - ASTM D1160)	ISO 10370		0.30	%m/m
Contaminação total	EN 12662		24	mg/kg
Glicerol / Glicéridos	EN 14105			
Monoglicéridos			0.80	%m/m
Diglicéridos			0.20	%m/m
Triglicéridos			0.20	%m/m
Glicerol livre			0.02	%m/m
Glicerol total			0.25	%m/m
Índice Iodo	EN 14111		120	gl/100g
Índice Iodo (est)	EN 14214		120	gl/100g
Teor de enxofre (*)	EN ISO 20846		10	mg/kg
Índice de Cetano (*)	EN ISO 5165	51		-
Fósforo	EN 14107		4.0	mg/kg
Sódio	EN 14538		1.0	mg/kg
Potássio	EN 14538		1.0	mg/kg
Cálcio	EN 14538		1.0	mg/kg
Magnésio	EN 14538		1.0	mg/kg
Metais grupo I (Na+K)	EN 14538		5.0	mg/kg
Metais grupo II (Ca+Mg)	EN 14538		5.0	mg/kg

2.3. A quimiometria

O termo quimiometria foi proposto em meados de 1970 para descrever o uso de modelos matemáticos, princípios estatísticos e outros baseados em lógica, no campo da química e, em especial, a dados químicos que apresentam natureza multivariada [29].

Dados de natureza multivariada são aqueles onde duas ou mais respostas instrumentais estão relacionadas com a propriedade de interesse.

2.3.1. Análise por componentes principais

A base fundamental da maioria dos métodos para tratamento de dados multivariados é a PCA, a qual tem por objectivo reduzir a dimensão do conjunto de dados original, a partir de combinações lineares dessas variáveis, preservando a maior quantidade de informação possível (variância).

Esta análise vectorial permite descrever a variação ou dispersão de um determinado conjunto de dados [30].

A PCA é um método conhecido de decomposição de matrizes bidimensionais, onde a variância na matriz X , com m observações e n variáveis, é decomposta em novas variáveis denominadas componentes principais (p , número de componentes principais ($p < m$)) e que se definem como sendo uma nova combinação linear de variáveis que constitui um sistema de eixos ortonormados. O número de objectos (n) diz respeito ao número de ensaios, enquanto que o número de variáveis (m) definem a dimensão do sistema multivariado.

A primeira PC formada é uma função que descreve o máximo de variância das observações no espaço X . As novas coordenadas das amostras, no novo sistema de eixos das PC's são denominadas de "scores" e "loadings". Os "scores" são a representação dos objectos no sub-espaço das componentes principais e os "loadings" a contribuição de cada variável para a componente principal [31].

No caso de análises qualitativas, surgem duas abordagens quimiométricas diferentes: a classificação multivariável, que se subdivide em análise de clusters e análise discriminativa. A primeira é utilizada para distinguir grupos num determinado conjunto de dados sem existir uma estrutura predefinida. É usada nas fases iniciais de investigação para explorar sub-populações. Pode ser efectuada por técnicas visuais, utilizando, por exemplo, a PCA que será descrito mais adiante. A segunda passa por construir classificações, nas quais são depois introduzidas amostras desconhecidas. Para análises quantitativas são utilizados métodos de regressão multivariável.

Como primeira abordagem ao tratamento de dados pode-se experimentar a redução de variáveis por PCA.

Uma vez que os dados contêm um elevado número de variáveis correlacionadas, torna-se necessário a redução de variáveis, ou seja, de tentar descrever a variabilidade dos dados através de poucas

variáveis não correlacionadas e que contenham a informação relevante. Desta forma, este método de compressão dos dados, ao ser aplicado a uma matriz de amostras por variáveis, constrói novas variáveis, designadas de componentes principais de tal forma que: as novas variáveis são combinações lineares das variáveis originais (as quais são combinações unitárias e ortogonais entre si); as novas variáveis não se correlacionam; a primeira das novas variáveis capta a maior variância possível dos dados originais; e cada nova variável seguinte capta o máximo possível da variância restante.

A utilidade desta decomposição, além de facilitar a visualização da variabilidade dos dados (por exemplo, em representações gráficas do primeiro componente principal face ao segundo), permite a utilização dos primeiros componentes como input em métodos de regressão múltipla ou de análise discriminatória, que não conseguiriam receber dados tão multivariados.

2.3.2. Regressão pelos mínimos quadrados

A matemática que suporta o algoritmo da PLS, talvez o método de regressão mais utilizado, é bastante complexa, sendo necessário algum conhecimento de conceitos matemáticos para melhor utilização do *software* disponível. O *input* para o PLS são matrizes X e y, tendo ambos sido centrados na sua média.

$$\begin{cases} X = TP' + E \\ y = Tq + f \end{cases} \quad (2.1)$$

Equação 2.1 – Sistema de equações que representa a estrutura do modelo PLS

As colunas da matriz T são as variáveis latentes (contendo os *scores*) responsáveis pelas variações sistemáticas tanto de X como de y. A matriz P e o vector q são designados de *loadings* e descrevem a forma como as variáveis em T relacionam as matrizes de dados originais X e y. A matriz E e o vector f são os resíduos, isto é, representam o ruído ou a variabilidade irrelevante em X e y.

Um outlier é uma amostra cujos dados diferem da maioria das amostras no conjunto, por algum motivo desconhecido, e que por isso tem uma influência negativa no modelo que está a ser desenvolvido. A detecção dos outliers melhora a capacidade de previsão do modelo, enquanto que a detecção de amostras anormais aquando da previsão, com o modelo implementado e a ser utilizado em rotina, ajuda a validar o próprio método de referência.

A detecção de *outliers* pode ser feita através da análise de dados, tanto espectrais como nos de concentração. Para os dados espectrais (X), os parâmetros mais utilizados são o *leverage* e os resíduos-Q. O parâmetro estatístico Hotelling T² é uma medida da variação de cada amostra dentro do modelo. Normalmente, o *software* estabelece os intervalos de confiança a 95% para os dois últimos

parâmetros, servindo estes intervalos de indicação do grau de desvio das amostras.

Considera-se que uma amostra é um *outlier* se o valor do *leverage* é superior à razão dada pela equação 2.2,

$$2 \times \frac{(1 + LV)}{N} \quad (2.2)$$

em que LV é o número de variáveis latentes do modelo e N o número de amostras utilizados, ou se os resíduos-Q (ou o Hotelling T₂) dessa amostra se encontrarem fora do intervalo de confiança de 95% para o modelo. Já para os dados de concentração, o parâmetro mais comum para detecção de outliers são os resíduos de y de Student, que são resíduos normalizados.

Deste modo, uma amostra é considerada um *outlier* se o seu resíduo de y de Student for >|2|. Este tipo de estatística é usualmente representado graficamente. No que diz respeito ao pré-processamento dos dados, é normalmente efectuado para reduzir ou minimizar as variações. Os pré-processamentos efectuados contribuem para reduzir a complexidade do modelo, nomeadamente fazendo reduzir o número de variáveis latentes, e muitas vezes contribuindo para tornar o modelo mais robusto contra perturbações inesperadas de amostras futuras. A seleção de variáveis deverá ser feita depois de optimizados estes dois parâmetros.

3. PARTE

EXPERIMENTAL

3. Parte Experimental

Neste trabalho efectuou-se a recolha de dados relativos a um ano de laboração em regime contínuo. Alguns dados foram importados e outros inseridos manualmente numa folha de excel, da qual resultou uma matriz de dados, matriz essa a partir da qual foram efectuados todos os cálculos e tratamento estatístico. Usaram-se cálculos com recorrência a técnicas Quimiométricas.

3.1. Material e métodos

Para este trabalho foram recolhidos dados relativos ao processo produtivo e análise de produto final relativos ao ano de 2009, para a produção de Biodiesel da empresa IBEROL. Os dados relativos ao processo produtivo foram obtidos diariamente nos três turnos de laboração. Os dados recolhidos no Laboratório incluem a análise diária dos três turnos de laboração bem como a análise completa de todos os parâmetros que a Norma EN 14214 manda reportar relativamente ao Biodiesel analisado. À excepção dos parâmetros: índice de cetano e teor de enxofre, que são solicitados a um Laboratório externo acreditado pela NP EN ISO/IEC 17025, todos os restantes são obtidos no Laboratório Iberol. Assim, a matriz de dados obtida e que serviu de base a este trabalho, contém informação relativa à matéria-prima, controlo de processo, análise de produto acabado e produto para expedição.

3.1.1. Descrição geral da metodologia utilizada na recolha e obtenção dos dados

Efectuou-se o tratamento dos dados, sendo que os modelos que foram desenvolvidos foram obtidos utilizando o *software* Octave 3.0.5 e também um algoritmo desenvolvido inteiramente no Departamento de Química da FCTUC. Este algoritmo foi desenvolvido com base na literatura existente para cálculos PLS [48].

3.1.2. Análise de componentes principais

Para a análise das relações entre matéria-prima, dados do processo e produto final, começou por se efectuar a PCA que permite identificar tendências no conjunto de amostras, isto é, verificar se existem parâmetros analíticos a variar segundo algum dos componentes, assim como extrair informação relativamente, por exemplo, à existência de *outliers*. Um dos aspectos que pode influenciar a análise de dados é a qualidade dos modelos desenvolvidos e o tipo de pré-tratamento a que os dados são sujeitos [32]. A forma mais indicada de extrair informação qualitativa do PCA é por observação de gráficos em que se comparem vários componentes principais. Para tal efectua-se o gráfico do segundo componente principal relativamente ao primeiro, verificando se existiam amostras que se

afastavam do aglomerado das restantes amostras (*outliers*) e analisando as tendências de distribuição das amostras.

3.1.3. Análise de PLS

De forma a obter os resultados apresentados utilizou-se o algoritmo desenvolvido no DQ-FCTUC para os cálculos PLS. O processo foi iniciado com a importação para das variáveis X e y. Depois de seleccionado o tipo de pré-tratamento a aplicar aos dados, conforme mencionado para o PCA, foi necessário escolher o tipo de validação a efectuar. Neste procedimento de validação é utilizado um processo iterativo que retira sucessivamente uma amostra (*leave one out*) ou um bloco de amostras (*contiguous block* ou *random subsets*), construindo um modelo sem a(s) incluir para depois efectuar a previsão utilizando essa(s) amostra(s). Após selecção do algoritmo de regressão, pode efectuar-se a modelação. Depois de corrido o algoritmo escolhido, o *software* produz uma tabela em que se observa a percentagem de variância capturada em X e em y em função do número de factores/variáveis latentes considerados.

A identificação de eventuais *outliers* foi efectuada em simultâneo com a escolha do pré-tratamento mais adequado. Assim, efectuou-se o pré-tratamento aos dados, depois a regressão e analisou-se no gráfico dos resíduos a existência de *outliers*.

4. RESULTADOS

E DISCUSSÃO

4. Resultados e discussão

A obtenção de resultados passou por um análise multivariada do processo, tendo em consideração a matriz original de dados referentes à produção entre Janeiro e Dezembro de 2009 dados apenas referentes à produção do biodiesel.

O conjunto de dados original era composto por 349 linhas, cada uma delas representativas de turnos de produção. O conjunto de dados encontra-se sub-dividido entre condições do processo (predictores) e características do biodiesel (respostas). O espaço dos predictores contém informação sobre matérias primas (15 variáveis codificadas através dos símbolos x001 a x015) e sobre condições de produção (129 variáveis representadas por x016 a x144). Quanto às respostas, cada turno de produção foi avaliado atendendo a 66 características de qualidade do produto (codificadas através das variáveis y01 a y66).

Todas as variáveis e respostas que seguidamente iram ser analisadas nos capítulos seguintes encontram-se codificadas nas tabelas apresentadas no capítulo ANEXOS.

4.1 Verificação das variáveis

Após completa a matriz de dados original e após a numeração das variáveis, foi necessário proceder à verificação das mesmas, o que passou essencialmente pela sua eliminação por não constituírem informação relevante, isto é, tiveram de ser tratadas como *“missing data”* ou *“bad data”*.

4.1.1 Remoção de variáveis qualitativas

A remoção de variáveis qualitativas consistiu em remover aquelas variáveis que não são susceptíveis de terem tratamento estatístico.

Assim, as variáveis eliminadas foram a y04, y17, y30, y42, y45, y54, y59-y63, y65.

As variáveis y04, y17 e y30 correspondem ao teor de glicerol livre no biodiesel dos três turnos de laboração e que é um valor praticamente constante ao longo do ano, salvo situações pontuais no processo produtivo e que têm o valor de < 0.01 %m/m.

A variável y54 é exactamente a mesma situação, mas refere-se ao glicerol livre do biodiesel analisado para lote de expedição.

A variável y42 corresponde ao teor de cinzas sulfatadas no biodiesel analisado para lote de expedição, sendo também um valor constante dado por < 0.005 %m/m. Este valor corresponde ao limite de quantificação imposto pela norma e os resultados obtidos para o biodiesel da Iberol cumprem sempre a especificação.

A variável y45 corresponde ao ponto de inflamação no biodiesel analisado para lote de expedição e

que tem um valor também quase constante de $> 150\text{ }^{\circ}\text{C}$ que é a especificação do biodiesel de acordo com a norma EN 14214. Salvo certas exceções em que o valor exacto é determinado e reportado no boletim final de análise, o excesso de valor $> 150\text{ }^{\circ}\text{C}$, levou à eliminação desta variável.

A variável y65 corresponde ao resultado do teste da corrosão à lâmina de cobre e tem um valor constante de “1a”, pois trata-se de uma escala num padrão de cor utilizado na determinação e cuja especificação da norma indica que o biodiesel deve ter o seu resultado na classe 1, e que por sua vez dentro desta classe tem a cor correspondente à cor “1a”, sendo a variável foi eliminada por se tratar de um “*bad data*”.

As variáveis de y59 a y63 também foram eliminadas e correspondem à determinação de metais e fósforo no biodiesel.

Este parâmetro tem um limite de quantificação de $<1.0\text{ mg/kg}$ para o sódio (Na), cálcio (Ca), magnésio (Mg) e potássio (K), enquanto que para o fósforo (P) (y59) o limite de quantificação é de $<4.0\text{ mg/kg}$. Assim, como o biodiesel analisado cumpriu sempre a especificação, os valores obtidos foram sempre inferiores ao limite de quantificação imposto pela norma.

4.1.2 Remoção de linhas com falhas de dados

De 349 linhas que constavam na matriz original, isto é, 349 dias de laboração, apenas foram considerados 162 dias, uma vez que foram eliminadas 187 linhas. Isto constituiu uma redução de 53.6% de informação. Tal terá acontecido uma vez que essas mesmas linhas continham pouca informação.

4.1.3 Substituição de valores e arredondamento dos números à mesma precisão de cada variável

Os valores de variáveis que em alguma linha tiveram um valor $<xxx$ ou $>xxx$, isto é, que pontualmente se situavam abaixo ou acima dos respectivos limites de quantificação, apresentando apenas um valor qualitativo, foram substituídos por zero (0).

Procedeu-se também ao arredondamento dos números à mesma precisão de cada variável, tendo em consideração o respeito pelas casas decimais.

4.2 Análise de variáveis

Sendo este sistema constituído por cerca de 200 variáveis o que constitui demasiada informação para ser processada, vamos realizar uma abordagem preliminar a cada um dos três sectores de dados – as matérias-primas, as condições de produção e as características finais do produto na primeira fase de tratamento.

Numa aproximação sucessiva iremos verificar alguma eventual relação entre as matérias primas e as

condições de produção explorando o sub-espço dos predictores.

Quando o sub-espço dos predictores (matérias-primas e condições de produção) estiverem estudadas então seguiremos para a exploração das inter-relações entre predictores e respostas.

Como um dos objectivos deste trabalho é encontrar as inter-relações entre variáveis, a normalização¹ das variáveis foi realizada em todos os casos no sentido de minimizar os efeitos de escala e, deste modo, permitir a maximização das respectivas inter-relações.

4.2.1 Matéria-prima

As primeiras quinze variáveis referem-se à matéria-prima, revelam a caracterização de três turnos em termos de características de óleo desgomado (humidade e acidez) e óleo neutro (humidade, acidez e sabão).

Uma vez que a normalização utiliza estimativas paramétricas da distribuição e estas por sua vez são afectadas pela presença de *outliers*, é conveniente verificar se esta operação não poderá ter sido de algum modo influenciada por estes valores anómalos.

Na figura 4.1 apresentamos os valores sob a forma de histogramas.

¹ A normalização das variáveis consiste em centrar cada variável em relação à média e escalar de seguida com base na estimativa da dispersão (desvio padrão).

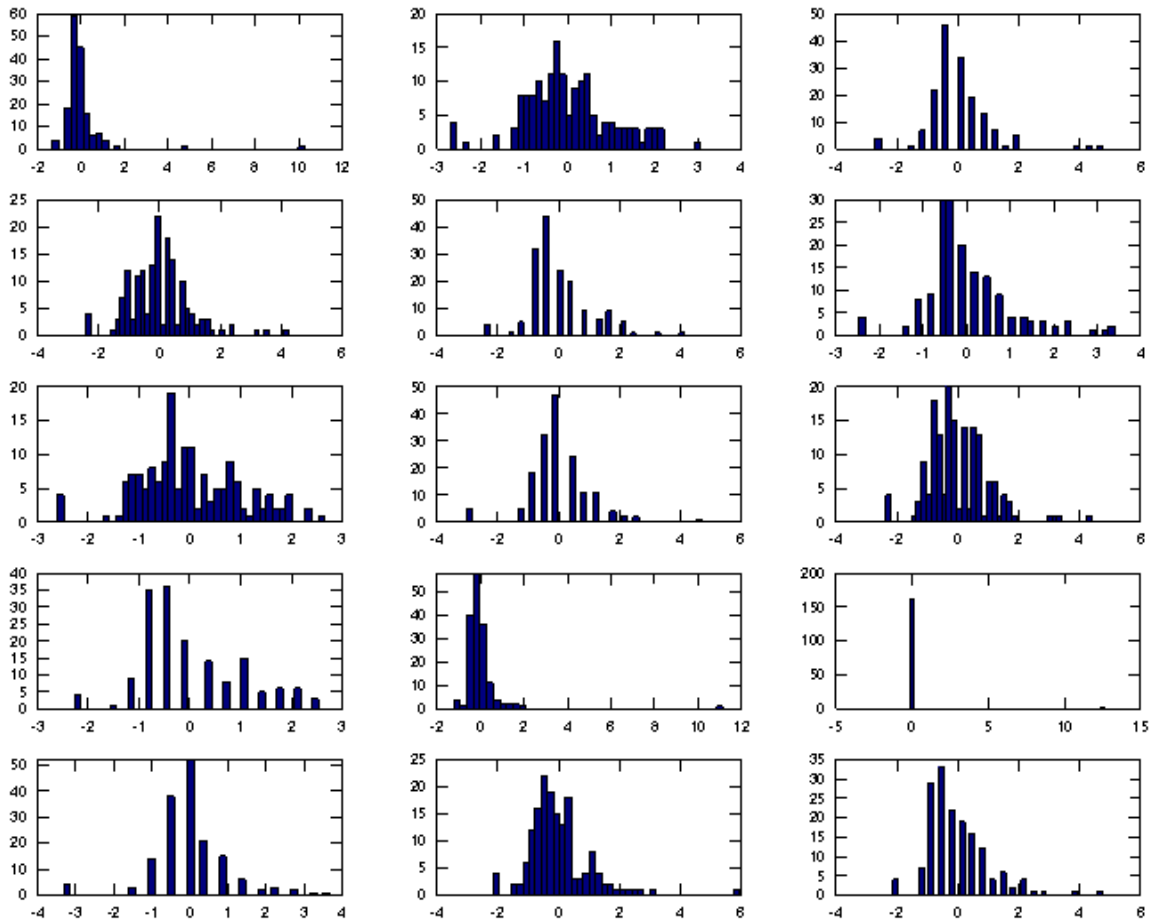


Figura 4.1 – Representação das variáveis normalizadas sob a forma de histogramas: as 15 variáveis estão ordenadas segundo a sequência normal convencional para a escrita, desde a primeira (canto superior esquerdo) até à 15ª (canto inferior direito).

Da figura 4.1 verifica-se que após a normalização as variáveis encontram-se agora em situação similar em termos de dimensão contudo, os seus perfis são bastante distintos. Na tabela 4.1 encontram-se as quatro primeiras estimativas que servem para caracterizar distribuições por via paramétrica – média (X_m), desvio padrão (S_x), simetria e curtose.

Tabela 4.1 – Estimativas paramétricas obtidas sobre as variáveis normalizadas.

Variável	Xm	Sx	Simetria	Curtose
x001	0.00	1.00	7.42	69.55
x002	0.00	1.00	0.13	0.61
x003	0.00	1.00	1.30	5.98
x004	0.00	1.00	0.83	2.64
x005	0.00	1.00	0.85	1.97
x006	0.00	1.00	0.89	1.95
x007	0.00	1.00	0.16	0.14
x008	0.00	1.00	0.24	4.06
x009	0.00	1.00	0.93	2.77
x010	0.00	1.00	0.59	0.06
x011	0.00	1.00	8.64	92.69
x012	0.00	1.00	12.46	154.51
x013	0.00	1.00	0.17	3.44
x014	0.00	1.00	1.64	7.25
x015	0.00	1.00	1,31	3.53

Como referido anteriormente aquando da análise da figura 4.1, através da tabela 4.1 verifica-se que as variáveis normalizadas foram de facto aferidas na posição e na dispersão o que é vantajoso para colocar todas as séries de valores dentro da escala das unidades. Verifica-se ainda que algumas variáveis apresentam distorções muito significativas – as variáveis x012 (12.46), x011 (8.64) e x001 (7.42) apresentam uma elevada tendência para valores acima da respectiva estimativa central.

No que diz respeito ao perfil da distribuição, as mesmas variáveis (x012, x011 e x001) são exemplos extremos de distribuições leptocúrticas enquanto que x010 (0.06), x007 (0.14) e x002 (0.64) comportam-se como platicúrticas.

Relacionando esta informação com a da figura 4.1 verifica-se que as anomalias observadas com a simetria das distribuições x012, x011 e x001 estão relacionadas com alguns valores extremos que surgem na parte superior do seu domínio.

Apesar destas constatações, nenhum esforço foi realizado no sentido de verificar se estes valores são ou não discrepantes porque, entendeu-se que estas situações correspondem a ocorrências reais que foram verificadas num certo período de laboração e que podem inclusivamente corresponder a casos muito favoráveis para conseguir extrair mais informação do processo.

Quer-se com isto referir, que durante a avaliação de uma empresa no seu normal funcionamento, as condições impostas não podem ser significativamente perturbadas com vista a obter-se mais informação sobre o comportamento desse mesmo sistema. Se uma perturbação fosse imposta poderia provocar danos na linha de produção ou conduzir a lotes cuja qualidade estaria fora de especificação. (Lote denominado não conforme). Por este motivo entendemos que as situações extremas observadas podem inclusivamente ser muito benéficas neste caso.

Realizando a análise PCA sobre este sub-espaco obteve-se a decomposição da variabilidade deste sub-sistema de variáveis.

Estes resultados estão condensados na tabela 4.2.

Tabela 4.2 – Resultados obtidos através da análise PCA sobre o conjunto de variáveis referentes à matéria-prima.

#	λ	V%	$\Sigma V\%$
1	5.041	33.61	33.61
2	1.737	11.58	45.19
3	1.324	8.83	54.01
4	1.185	7.90	61.91
5	1.034	6.89	68.80
6	0.910	6.07	74.87
7	0.786	5.24	80.11
8	0.741	4.94	85.05
9	0.533	3.56	88.60
10	0.502	3.35	91.95
11	0.350	2.34	94.29
12	0.277	1.84	96.13
13	0.267	1.78	97.91
14	0.218	1.45	99.36
15	0.096	0.64	100.0

- índice da componente principal; λ - valor próprio; V% e $\Sigma V\%$ - variabilidade e variabilidade cumulativa justificada (em percentagem).

Da tabela 4.1 verifica-se que a primeira componente principal contribui com cerca de um terço (33.6%) para descrever a informação contida neste sub-espaco de 15 variáveis sendo a que mais se evidencia em relação às outras.

No sentido de melhor se verificar o impacto das componentes principais na descrição do sistema, na figura 4.2 encontra-se representada a evolução dos valores próprios em função do número de componentes principais.

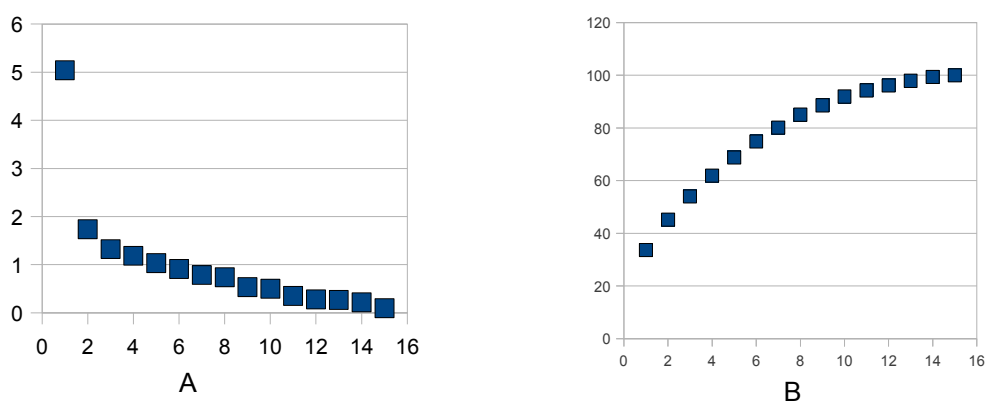


Figura 4.2 – Descrição da variabilidade do conjunto x001 a x015 em função do número de componentes principais consideradas: A) evolução dos valores próprios (λ) e B) variabilidade descrita através das componentes principais (V%) em função do número de componentes principais.

Embora as cinco primeiras componentes principais, PC1 a PC5, sejam as mais significativas, da figura 3.2 verifica-se que a primeira componente principal é a que de facto mais contribui para a descrição deste excerto do sistema. Contudo a capacidade de descrever este conjunto através das

primeiras cinco componentes ($p = 5$) é algo insatisfatório já que apenas preservam 68.8% dessa informação.

Importa agora ver quais as variáveis que mais se distinguem ao longo da componente PC1. Na tabela 4.3 encontram-se ordenadas as variáveis que mais se evidenciam em PC1 através dos seus *loads*.

Tabela 4.3 – Contribuição das variáveis para as primeiras cinco componentes principais.

Variável	PC1	PC2	PC3	PC4	PC5
x001	0.172	-0.157	0.448	0.173	-0.424
x002	0.287	0.119	-0.201	-0.495	-0.362
x003	0.279	0.143	0.049	0.263	-0.181
x004	0.292	0.223	0.429	-0.029	-0.094
x005	0.264	0.008	0.074	-0.207	-0.040
x006	0.269	-0.396	-0.037	0.137	0.030
x007	0.298	0.064	-0.456	-0.361	-0.205
x008	0.270	0.033	-0.419	0.424	0.080
x009	0.270	0.418	0.109	0.058	0.209
x010	0.287	-0.352	0.122	-0.138	0.244
x011	0.138	-0.391	0.063	-0.100	0.036
x012	0.058	0.130	0.043	-0.291	0.646
x013	0.294	-0.020	-0.301	0.394	0.097
x014	0.266	0.386	0.199	0.074	0.130
x015	0.294	-0.332	0.140	-0.077	0.219

Da tabela 4.3 verifica-se que sobre a primeira componente principal, aquela que se comporta como mais relevante, todas as variáveis apresentam um impacto positivo – a primeira componente traduz algo em que todas as variáveis estão em consonância.

Uma segunda constatação está relacionada com o facto de haver uma forte contribuição de quase todas as variáveis – apenas três destas quinze (x001, x011 e x012) não apresentam peso significativo sobre PC1.

Estas variáveis menos representadas em PC1 encontram seu máximo em componentes menos significativas indicando também a sua falha na capacidade de descrever os resultados.

Embora as componentes principais evidenciem a relevância das variáveis na caracterização dos dados, sobre a mesma componente principal é frequente estar representada informação correlacionada. Para melhor distinguir este efeito ter-se-á que observar a respectiva matriz de correlação, tabela 4.4.

Tabela 4.4 – Valores da correlação para as primeiras quinze variáveis.

	x001	x002	x003	x004	x005	x006	x007	x008	x009	x010	x011	x012	x013	x014
x002	0.167													
x003	0.273	0.381												
x004	0.439	0.406	0.451											
x005	0.268	0.374	0.237	0.313										
x006	0.334	0.241	0.290	0.172	0.258									
x007	0.039	0.818	0.284	0.269	0.384	0.334								
x008	0.114	0.218	0.422	0.209	0.227	0.384	0.478							
x009	0.022	0.330	0.394	0.540	0.340	0.164	0.300	0.341						
x010	0.192	0.303	0.289	0.343	0.379	0.553	0.308	0.237	0.198					
x011	0.162	0.141	0.062	0.095	0.139	0.347	0.151	0.110	0.008	0.296				
x012	-0.010	0.073	0.016	0.123	0.083	-0.013	0.096	0.047	0.127	0.096	-0.010			
x013	0.196	0.245	0.402	0.256	0.323	0.391	0.425	0.677	0.320	0.320	0.142	0.048		
x014	0.136	0.308	0.374	0.541	0.330	0.155	0.279	0.274	0.687	0.183	0.062	0.096	0.329	
x015	0.221	0.294	0.283	0.379	0.348	0.522	0.303	0.258	0.225	0.719	0.316	0.051	0.371	0.232

Curiosamente, esta matriz diagonal inferior contém apenas três valores negativos (todos referentes a x012, assinalados com cor vermelha) que, atendendo ao seu valor absoluto, são muito próximos de zero. Por forma a facilitar a visualização e interpretação dos 105 valores representados na tabela 4.4, de forma a ver quais os mais relevantes, foi necessário representar os valores da tabela 4.4 sob a forma de um gráfico de barras, figura 4.3.

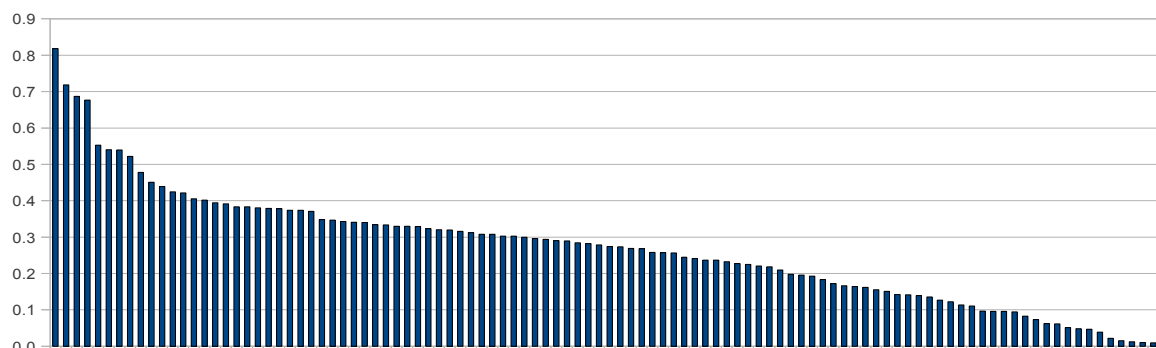


Figura 4.3 – Representação ordenada das correlações em valor absoluto.

Da figura 4.3 fica-se com a noção de que existem eventualmente 4 variáveis bem relacionadas (valores de correlação entre 0.67 e 0.82 (assinalados na tabela 4.4 com sublinhado). O conjunto seguinte, com valores entre 0.52 e 0.55 (valores assinalados a negrito) pode já não apresentar grande interação. Para facilitar esta interpretação foram representadas as relações mais significativas aqui evidenciadas, figura 4.4.

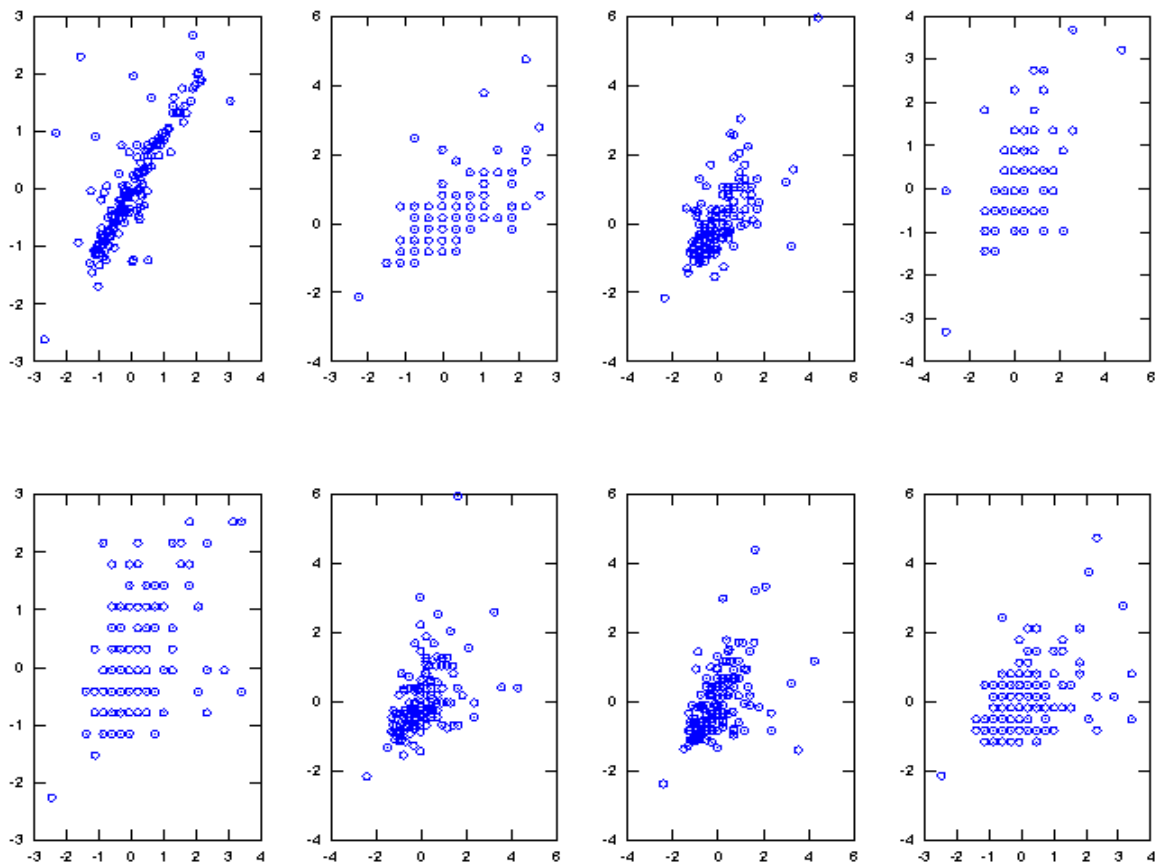


Figura 4.4 – Representação bivariada das variáveis mais correlacionadas, assinaladas na tabela 4.4, por ordem decrescente de correlação. Variável de menor índice sobre o eixo das abcissas.

Da figura 4.4 é fácil de reconhecer que o primeiro gráfico (x007 vs. x002) apresenta uma relação linear muito significativa (correlação de 0.818) enquanto que os outros casos não parecem ser tão evidentes.

Curiosamente, as variáveis x002 e x007 referem-se ao valor de acidez encontrado ao nível do óleo desgomado nos primeiro e segundo turnos, respectivamente, levando a admitir que ambos os turnos estarão a utilizar o mesmo tipo de matéria prima no fabrico.

Sendo a análise de componentes principais vocacionada para a compressão de informação relevante, vamos representar este sub-conjunto de pontos sob a forma 2D e 3D apenas para que se possa tentar evidenciar e eventualmente reconhecer situações de anomalia na produção, em termos de matéria-prima, figura 4.5.

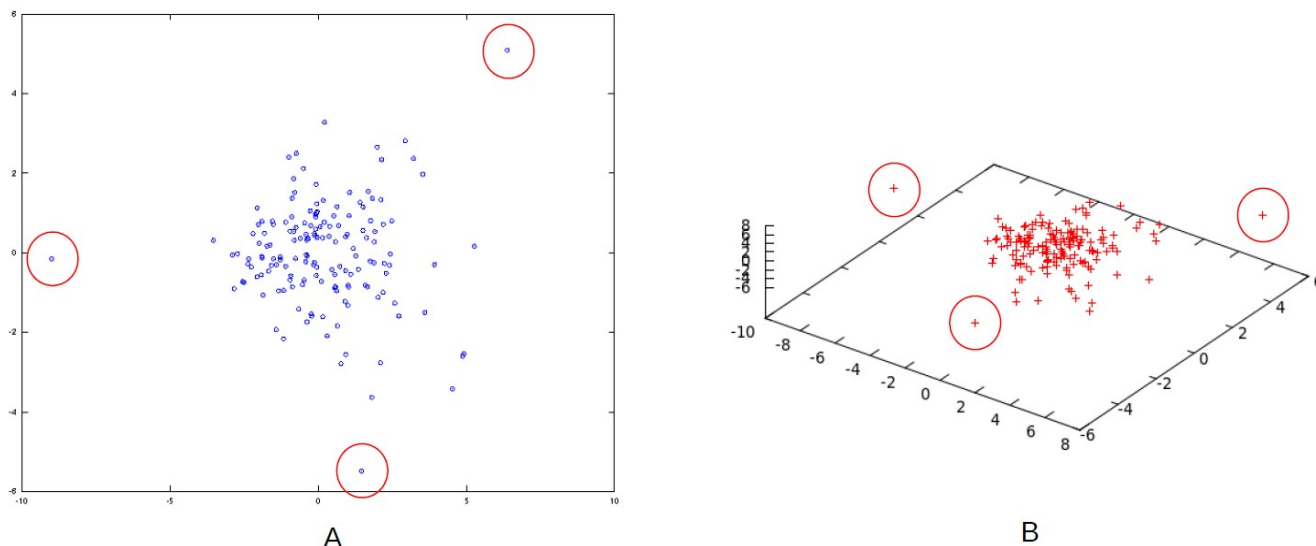


Figura 4.5 – Representação 2D (A) e 3D (B) da informação contida nas primeiras componentes.

Como já anteriormente evidenciado na análise prévia de diagnóstico das variáveis normalizadas, evidenciam-se algumas situações extremas (pelo menos três, assinaladas na figura 4.5) ao nível da matéria-prima aos quais convém estar atento nas conclusões subjacentes.

4.2.2 Processo

Analisa-se agora a secção de dados referentes ao processo de fabrico. Trata-se de um conjunto vasto composto por 129 variáveis (x016 a x144) que caracterizam o processo.

No sentido de avaliar as situações em que os dados transportam valores extremos, fora do comum, seria necessário avaliar cada uma destas variáveis. Contudo, não havendo possibilidade de fazer este diagnóstico em modo visual de todas estas variáveis, optou-se por avaliar os estimadores paramétricos já que estes são muito afectados, especialmente pela parte da dispersão, pela presença de valores discrepantes na distribuição.

Para esse efeito foram calculadas os terceiros (simetria) e quartos (curtose) momentos de cada distribuição.

Na figura 4.6 estão representadas a simetria (em valor absoluto) e a curtose destas variáveis sob a forma de barras, ordenadas por ordem decrescente quanto ao seu valor absoluto.

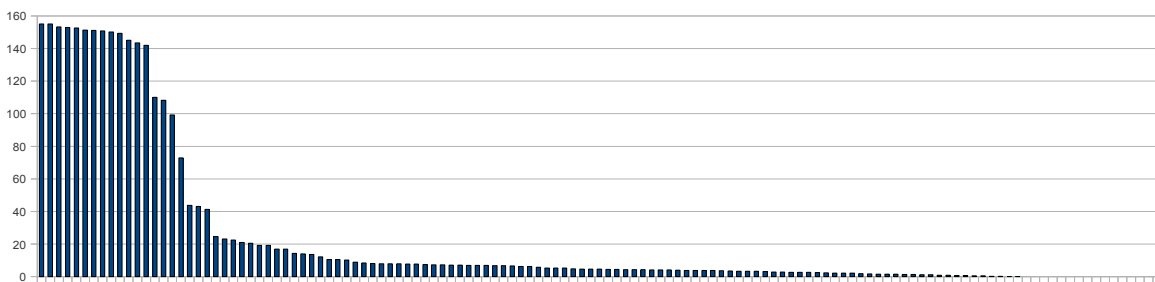
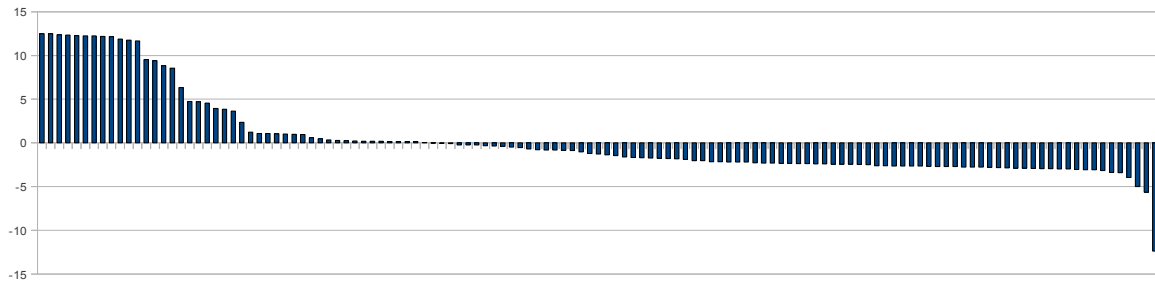


Figura 4.6 – Representação da simetria (imagem superior) e da curtose (em baixo) das 129 variáveis normalizadas.

Da figura 4.6 resulta que algumas variáveis (19 em 129) apresentam valores de assimetria muito grandes (acima de 5) e que existem 20 variáveis com curtose acima de 40.

No sentido de verificar até que ponto poderemos estar a trabalhar com variáveis eventualmente perniciosas² fez-se um diagnóstico cruzado de simetria-curtose, figura 4.7.

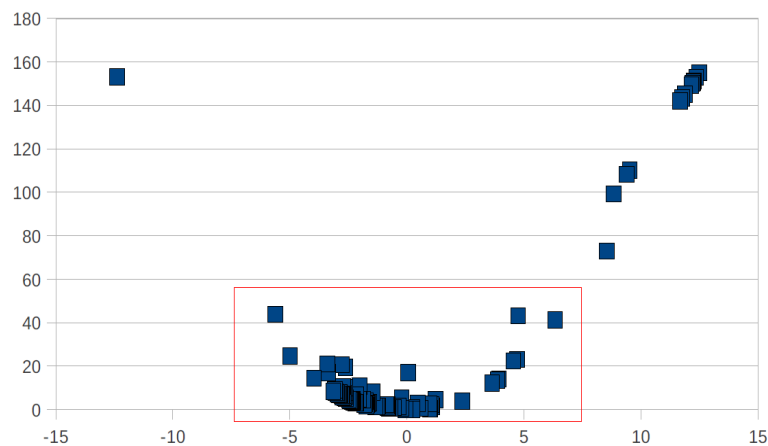


Figura 4.7 – Representação da curtose em função da simetria.

2 No tratamento estatístico de dados é frequente procurar situações de normalidade (simetria nula e curtose 3) para se estar dentro do formalismo convencional da estatística.

Da figura 4.7 reconhece-se que existem de facto algumas variáveis que podem eventualmente trazer anomalias nos dados.

Após a identificação destas 16 variáveis extremas (x045, x117, x105, x102, x042, x135, x019, x035, x119, x039, x048, x061, x066, x091, x079, x069) estas foram representadas sob a forma de histograma na figura 4.8.

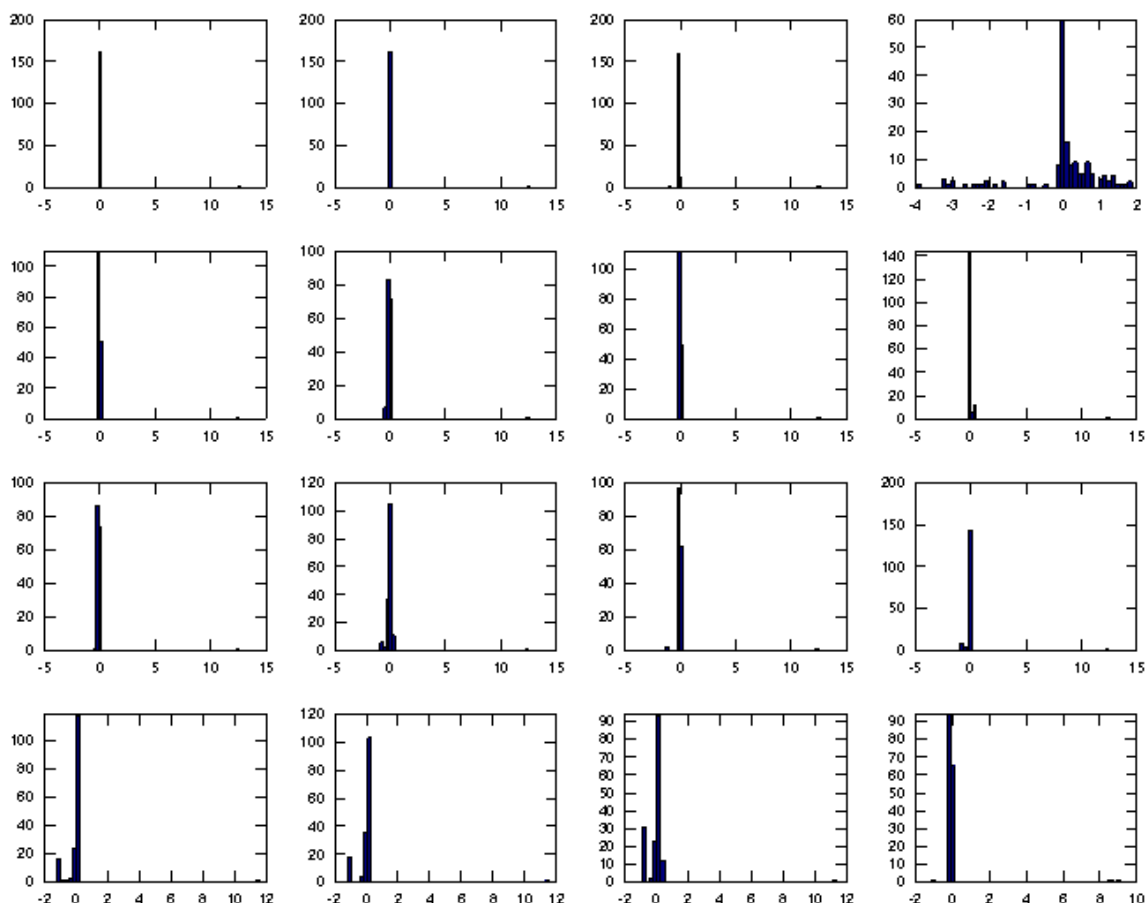


Figura 4.8 – Representação das variáveis mais extremas em termos de simetria e curtose: por ordem (linha e depois coluna, de acordo com a convenção da escrita), x045, x117, x105, x102, x042, x135, x019, x035, x119, x039, x048, x061, x066, x091, x079, x069.

Da figura 4.8 verifica-se que as variáveis identificadas através dos valores extremos em simetria e curtose têm de facto presentes valores extremos na sua distribuição.

Na tabela 4.5 apresenta-se, de uma forma ordenada em sentido decrescente de relevância, a evolução dos primeiros 10 valores próprios deste subconjunto de variáveis.

Tabela 4.5 – Listagem ordenada dos valores próprios do conjunto de variáveis referentes ao processo de fabrico.

#	λ	V%	$\Sigma V\%$
1	46.985	36.42	36.42
2	16.485	12.78	49.20
3	6.564	5.09	54.29
4	4.650	3.60	57.89
5	3.610	2.80	60.69
6	2.787	2.16	62.85
7	2.526	1.96	64.81
8	2.355	1.83	66.64
9	2.213	1.72	68.35
10	2.099	1.63	69.98

- índice da componente principal; λ - valor próprio; V% - variabilidade justificada; $\Sigma V\%$ - variabilidade justificada cumulativa (em percentagem).

A tabela 4.5 permite concluir que existem dois valores próprios extremos que são capazes de descrever cerca de metade da informação contida nos dados. Contudo surge uma série de valores próprios superiores à unidade podendo indicar eventualmente que existem outras componentes principais ainda assim relevantes.

A sequência de valores próprios acima da média termina na componente 24. Para melhor avaliar este conjunto total de informação contida nestas variáveis foi representada a figura 4.9, tendo sido suprimidos os valores acima de 30.

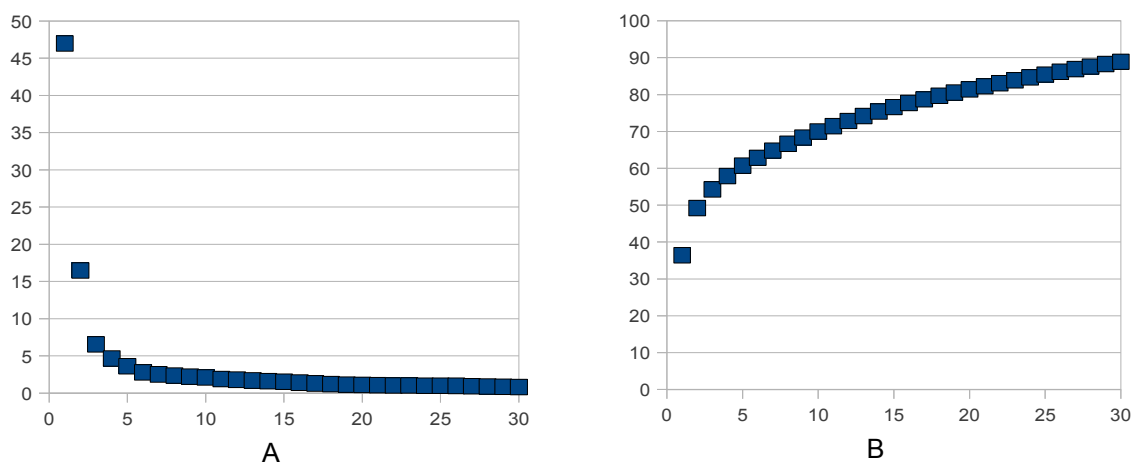


Figura 4.9 – Descrição da variabilidade referente ao processo de fabrico em função do número de componentes principais consideradas: A) evolução dos valores próprios e B) variabilidade descrita através das componentes principais em função do número das 30 primeiras componentes principais.

Da figura 4.9 é patente que de facto, apesar de haver alguns valores próprios superiores à média, os dois primeiros são de facto os mais relevantes já que se evidenciam muito acima da contribuição basal e conseguem descrever cerca de 50% da informação contida no conjunto de dados.

Na tabela 4.6 encontra-se a listagem das variáveis que mais influem nas duas primeiras componentes.

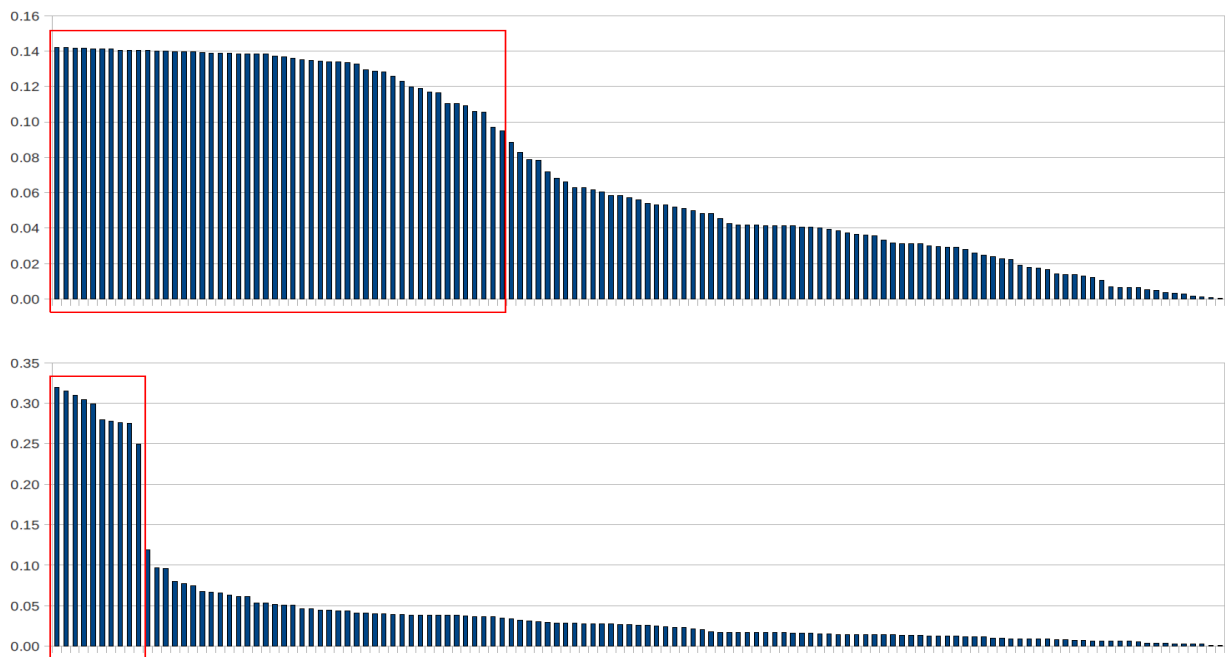


Figura 4.10 – Relevância das variáveis sobre as duas primeiras componentes principais (valores ordenados em módulo).

Da figura 4.10 é patente que na primeira componente principal estão representadas 51 das 129 variáveis originais e que somente 29 se encontram expressas de forma significativa na segunda componente. No cômputo global ($p = 2$) estão representadas 80 variáveis no total (sem repetições em PC1 e PC2) de forma que estas manifestam a informação de cerca de 62.0% das variáveis, descrevendo 49.2% do conjunto definido pelas condições de produção.

No sentido de verificar a homogeneidade de condições de produção, fez-se a representação bi e tri-dimensional, ver figura 4.11.

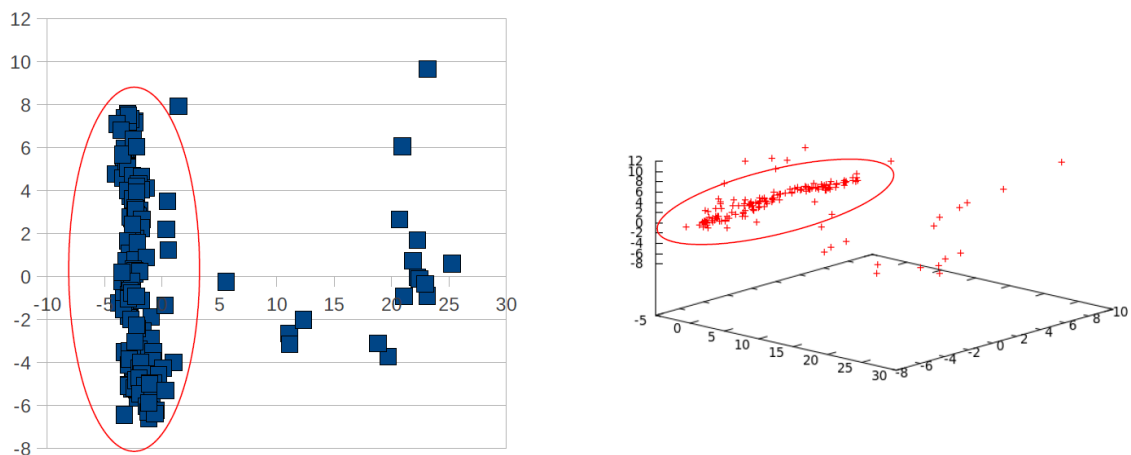


Figura 4.11 – Representação dos ciclos de produção sobre duas componentes ($p = 2$; 36.4% informação recuperada) e três componentes principais ($p = 3$; 49.2% informação preservada).

Da figura fica-se com uma noção de que neste período de laboração as condições mantiveram-se aproximadamente constantes havendo evidência de 25 situações muito distintas das outras restantes 136 (162 linhas de produção em análise).

4.2.3 Predictores

Uma vez que já foram previamente caracterizados os sub-espacos referentes à matéria-prima e ao processo de laboração, analisa-se agora como se comporta o sub-espaco predictor em conjunto, que irá servir de base à descrição da resposta.

Este sub-espaco envolve um total de 144 variáveis.

A análise PCA conduz à indicação de que as primeiras 29 componentes são relevantes, pois os respectivos valores próprios situam-se acima da média. Na tabela 4.6 apresenta-se uma listagem abreviada apenas das primeiras 10 componentes principais mas, na figura 4.12 representa-se a informação relevante referente às primeiras 30 componentes principais.

Tabela 4.6 – Listagem ordenada dos valores próprios do conjunto de variáveis referentes ao conjunto dos predictores.

#	λ	V%	$\Sigma V\%$
1	47.082	32.7	32.7
2	17.061	11.8	44.5
3	6.853	4.8	49.3
4	6.058	4.2	53.5
5	4.536	3.2	56.7
6	3.717	2.6	59.2
7	2.868	2.0	61.2
8	2.533	1.8	63.0
9	2.436	1.7	64.7
10	2.318	1.6	66.3

- índice da componente principal; λ - valor próprio; V% - variabilidade justificada; $\Sigma V\%$ - variabilidade justificada cumulativa (em percentagem).

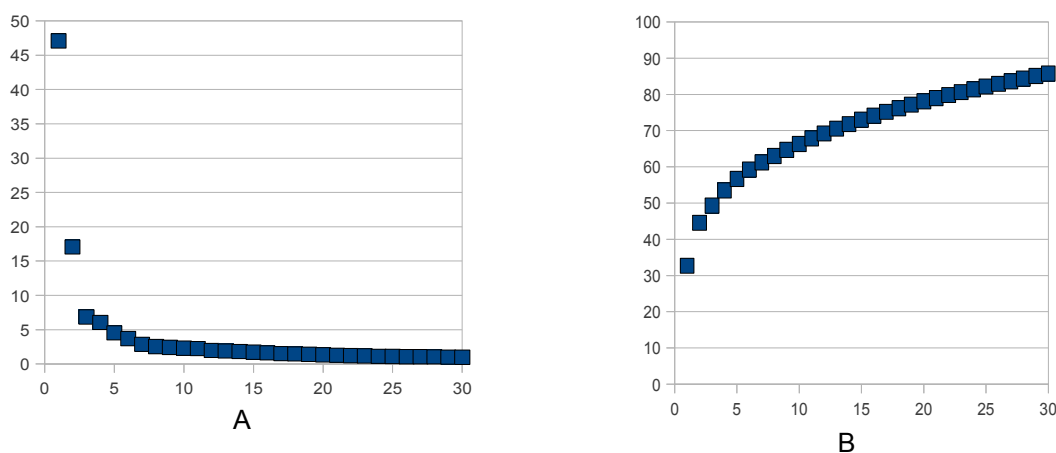


Figura 4.12 – Descrição da variabilidade referente ao sub-espaco predictor em função do número de componentes principais consideradas: A) evolução dos valores próprios e B) variabilidade descrita através das componentes principais em função do número das 30 primeiras componentes principais.

Da análise da figura 4.12 conclui-se que as duas primeiras componentes principais são as mais relevantes em relação às restantes – distinguem-se muito em relação à contribuição basal (não específica).

Contudo, considerando apenas duas componentes principais ($p = 2$) a variabilidade descrita situa-se abaixo dos 45% da informação contida nos dados.

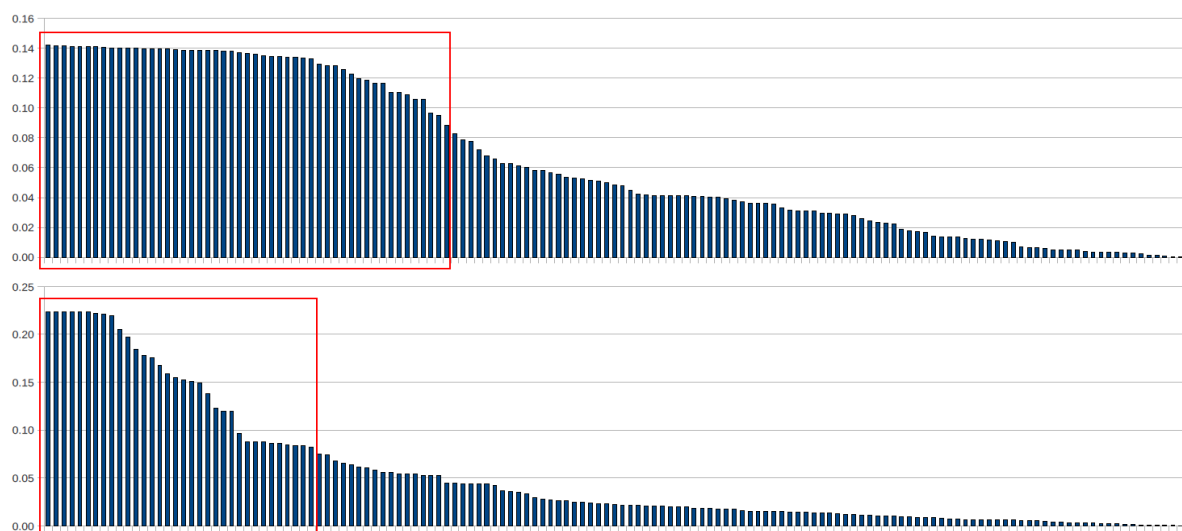


Figura 4.13 – Relevância das variáveis do sub-espço predictor sobre as duas primeiras componentes principais (valores ordenados em módulo).

Na figura 4.13 fica patente que sobre a primeira componente estão essencialmente representadas 52 variáveis enquanto que sobre a segunda componente se encontram outras 33 representadas. Verifica-se ainda que não há repetição de qualquer destas variáveis cruciais sobre estas duas componentes. Sendo assim, as duas primeiras componentes repercutem a informação contida em 85 variáveis. Embora este número de componentes principais ($p = 2$) seja insuficiente na descrição dos dados (é capaz de justificar 44.5% da informação nos predictores) trata-se de uma taxa de compressão de informação muito elevada.

Em relação às situações anteriores, verifica-se que quando se analisa o espaço predictor em conjunto, as variáveis do processo são aquelas que contêm mais informação sobre a variabilidade total – sobre PC1 não existe nenhuma variável referente à matéria-prima e só em PC2 é que surgem 2 variáveis (x007 e x002) na parte menos significativa das variáveis mais relevantes.

Na figura 4.14 encontra-se a representação 2D e 3D referentes à informação contida no sub-espço dos predictores obtida nos turnos de produção durante o período em estudo.

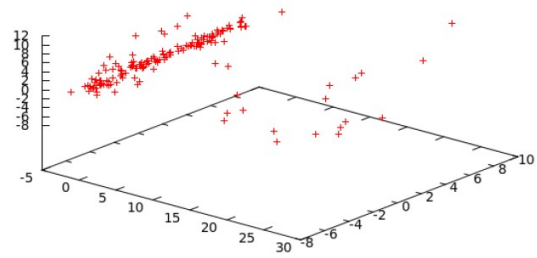
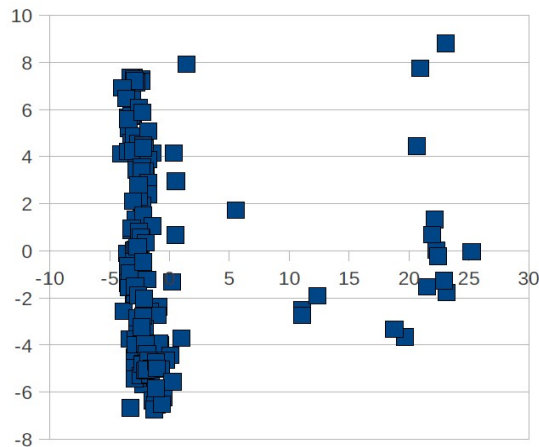


Figura 4.14 – Representação da informação contida no sub-espaco predictor em duas ($p = 2$) e três ($p = 3$) dimensões.

Notavelmente as figuras 4.11 e 4.14 são demasiado parecidas – a razão de ser é que o sub-espaco predictor depende essencialmente do sub-espaco das condições de produção, como já foi visto através da avaliação dos *loads* sobre as duas componentes principais.

Concentrando agora a atenção sobre o conjunto de variáveis mais relevantes que se encontram representadas sobre a primeira componente principal (PC1, 32.7%) verifica-se que nela estão representadas 52 variáveis (x099, x098, x122, x106, x107, x092, x129, x052, x100, x124, x097, x118, x041, x080, x120, x047, x054, x125, x095, x093, x062, x094, x096, x056, x074, x072, x128, x101, x114, x141, x110, x075, x121, x111, x068, x112, x123, x127, x126, x139, x133, x103, x142, x065, x131, x081, x137, x040, x051, x116, x058 e x077), nenhuma delas referente ao sub-conjunto das matérias primas.

No que diz respeito à segunda componente principal (PC2, 11.8%) esta contém informação referente a 33 variáveis (x029, x113, x140, x024, x046, x067, x031, x057, x022, x020, x063, x037, x082, x055, x083, x076, x016, x032, x134, x087, x108, x085, x043, x086, x007, x078, x071, x089, x073, x053, x132, x136 e x002)

4.2.4 Resultados

Analisando agora o sub-espço dos resultados onde estão avaliadas 54 propriedades características do óleo transformado nesta fase do fabrico.

Na figura 4.15 encontra-se representada os valores da simetria e da curtose das variáveis ordenadas e na figura 4.16 a representação simultânea de ambas.

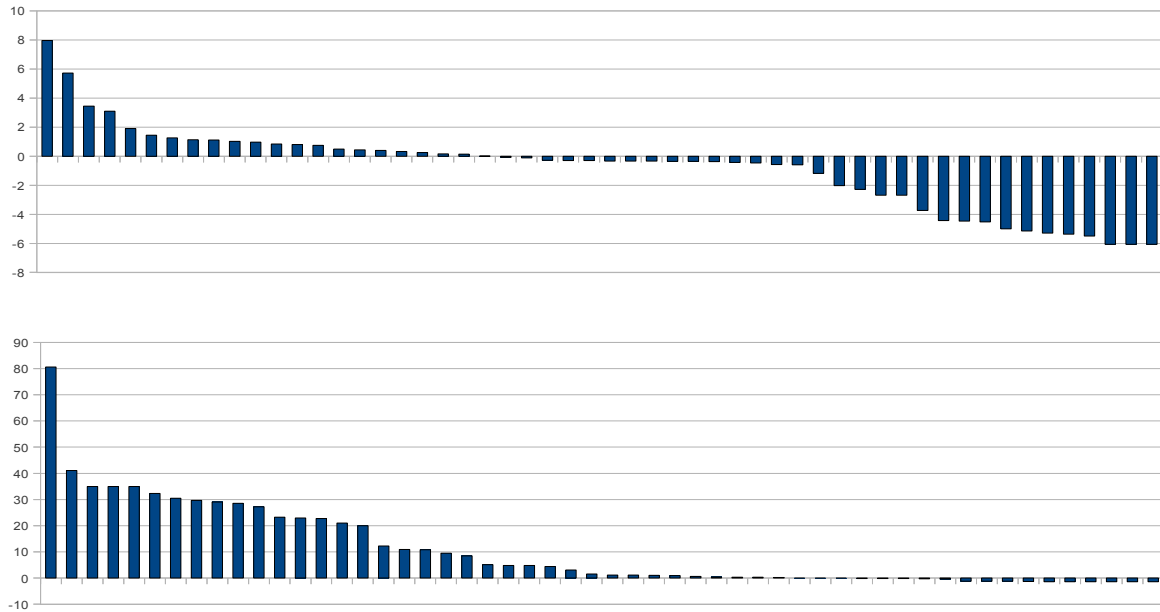


Figura 4.15 – Avaliação das variáveis do sub-espço resposta em termos de simetria (A) e de curtose (B).

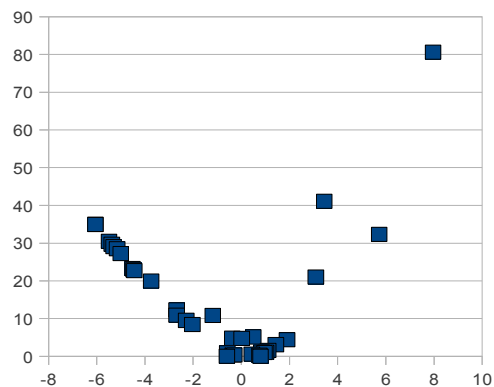


Figura 4.16 – Avaliação de anomalias nas variáveis do sub-espço resposta através da representação simultânea de simetria-curtose.

Das figuras 4.15 e 4.16 fica-se com a noção de que estas variáveis não são tão anómalas quanto as que foram anteriormente avaliadas, referentes ao sub-espço do processo. Em todo o caso existem algumas variáveis eventualmente desfavoráveis que estão apresentadas como histograma na figura 4.17.

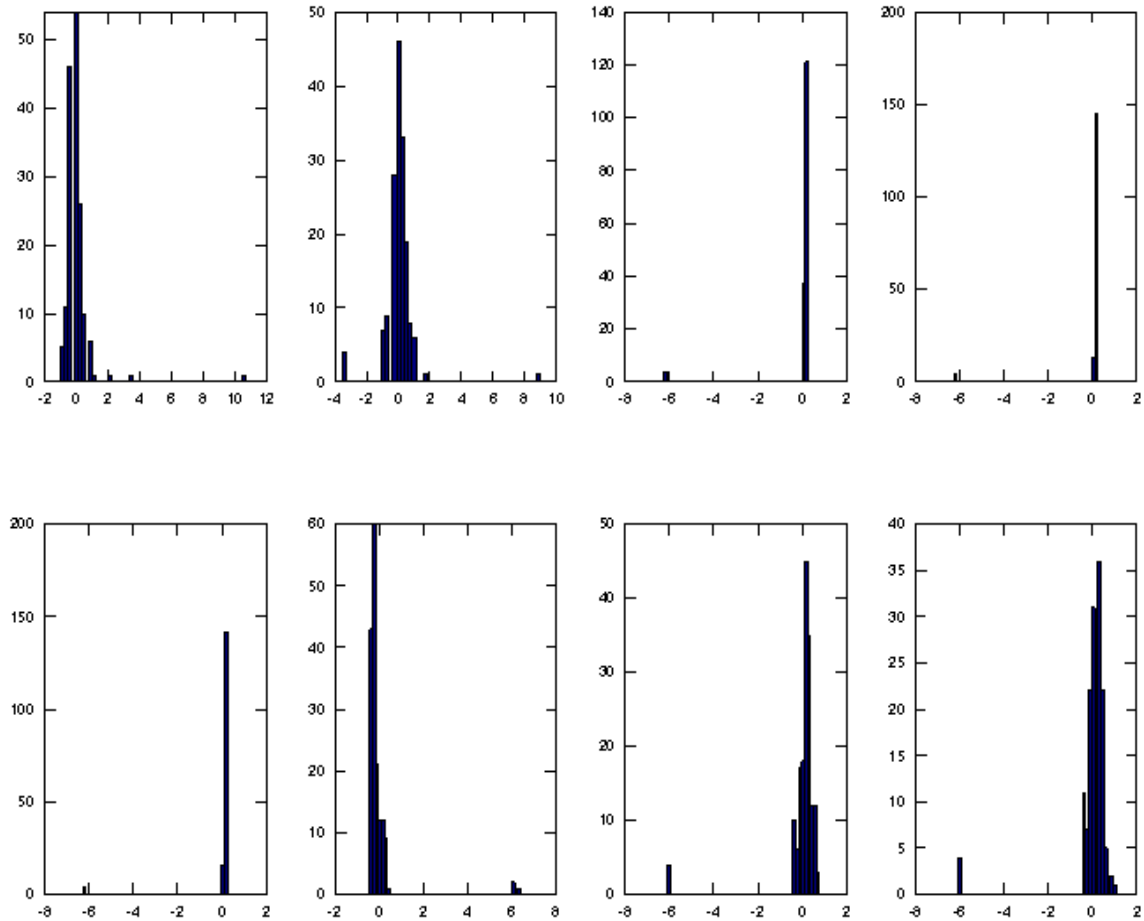


Figura 4.17 – Representação das oito variáveis com curtose mais elevada. Da esquerda para a direita e depois para baixo temos: y16, y15, y32, y19, y06, y66, y27 e y01.

A avaliar pela figura 4.17 verifica-se mais uma vez distribuições com casos de valores extremos – condições descontínuas eventualmente desfavoráveis na normalização, no tratamento estatístico de populações e na avaliação de correlações.

Realizando agora a análise PCA, verifica-se de novo ser possível comprimir esta informação do sub-espço resposta num número pequeno de componentes principais, valores na tabela 4.7.

Tabela 4.7 – Listagem ordenada dos primeiros dez valores próprios do sub-espço resposta.

#	λ	V%	$\Sigma V\%$
1	17.526	32.46	32.46
2	13.229	24.50	56.95
3	3.769	6.98	63.93
4	3.647	6.75	70.69
5	1.897	3.51	74.20
6	1.551	2.87	77.07
7	1.361	2.52	79.59
8	1.217	2.25	81.85
9	1.132	2.10	83.94
10	1.049	1.94	85.89

- índice da componente principal; λ - valor próprio; V% - variabilidade justificada; $\Sigma V\%$ - variabilidade justificada cumulativa (em percentagem).

A tabela 4.7 leva a assumir que existem duas componentes principais mais relevantes, capazes de descrever cerca de 57% da informação contida neste sub-espço das respostas. Na figura 4.18 encontra-se a evolução das primeiras vinte componentes principais.

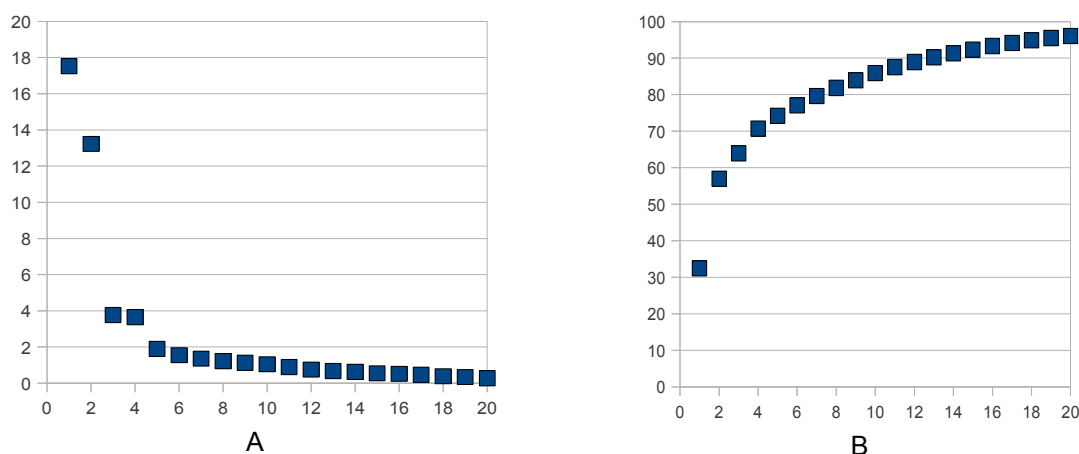


Figura 4.18 – Evolução dos valores próprios (A) e da variância justificada (B) em função do número de componentes principais consideradas (representação apenas das primeiras 20).

Da figura 4.18 verifica-se que existem duas componentes principais muito relevantes seguidas de outras duas com valores próprios mais altos que o normal – quase o dobro do valor seguinte. Contudo, atendendo ainda à parte da variância justificada, apenas as primeiras duas componentes ($p = 2$) devem ser consideradas como significativas, sedo estas capazes de descrever 57% da informação contida neste sub-espço resposta.

No sentido de avaliar a quantidade de respostas que contribuem para estas duas primeiras componentes foi representada a figura 4.19.

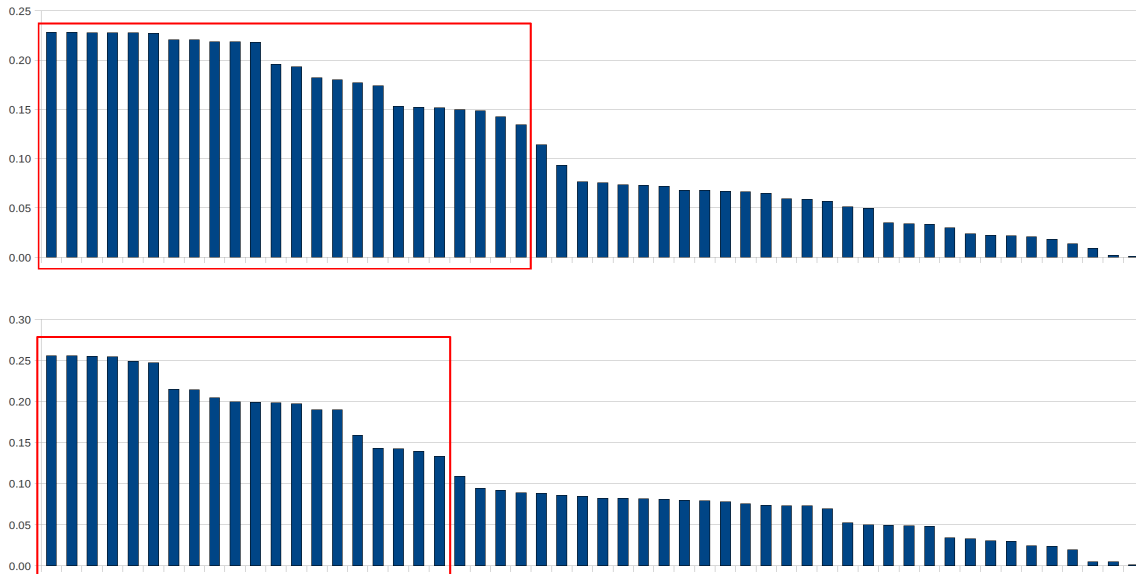
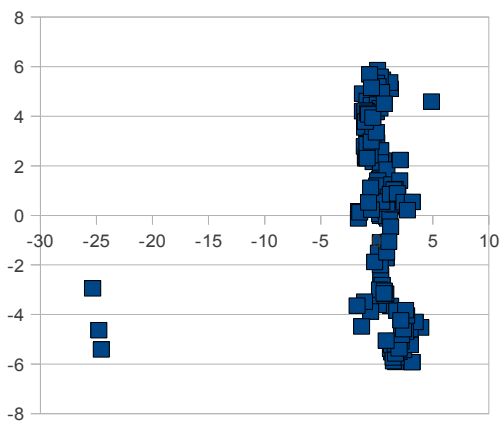


Figura 4.19 – Avaliação do número de variáveis que mais contribuem para a componente PC1 (A) e PC2 (B).

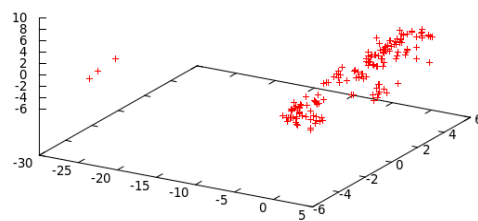
Da figura 4.19 verifica-se que sobre a componente PC1 se encontram representadas 24 respostas enquanto que sobre PC2 estão 20 respostas. Uma inspeção mais detalhada permitiu verificar que há três variáveis que se repetem sobre estas duas componentes (y13, y38 e y66).

Daqui resulta que, nestas condições, a compressão da informação ($p = 2$) processa-se à custa de 41 das 54 respostas (cerca de 76% das variáveis), sendo estas componentes capazes de descrever cerca de 57% da informação do sub-espaço resposta.

Na figura 4.20 encontra-se representada a informação contida no sub-espaço resposta sobre as primeiras duas e três componentes principais.



A



B

Figura 4.20 – Representação do sub-espaço resposta sob as duas (A) e três (B) primeiras componentes principais.

A disposição dos objectos na figura 4.20 faz lembrar o que já foi anteriormente visto no sub-espaco predictor, havendo aqui evidência de pelo menos três anomalias em que as características do produto devem ter saído das normais especificações de fabrico.

4.2.5 Análise conjunta

Na tentativa de verificar se PCA consegue detectar relações entre os sub-espacos dos predictores e respostas foi realizada a análise global do sistema.

Agora o número de variáveis é de 198, avaliadas ao longo de 162 lotes de produção.

Na tabela 4.8 encontra-se os valores indicativos da abordagem PCA (10 primeiras componentes) e na figura 4.21 estão representados os primeiros 20 valores mais relevantes.

Tabela 4.8 – Listagem ordenada dos primeiros dez valores próprios do conjunto global de dados.

#	λ	V%	$\Sigma V\%$
1	49.605	25.05	25.05
2	22.378	11.30	36.36
3	19.013	9.60	45.96
4	11.721	5.92	51.88
5	6.935	3.50	55.38
6	5.329	2.69	58.07
7	4.416	2.23	60.30
8	3.902	1.97	62.27
9	3.433	1.73	64.01
10	3.140	1.59	65.59

- índice da componente principal; λ - valor próprio; V% - variabilidade justificada; $\Sigma V\%$ - variabilidade justificada cumulativa (em percentagem).

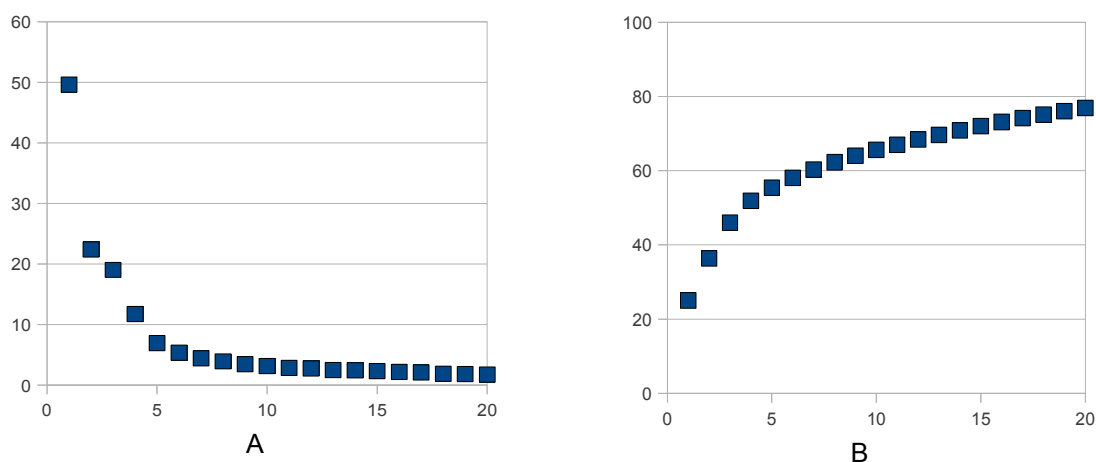


Figura 4.21 – Evolução da abordagem PCA sobre o conjunto global de dados sob a forma (A) de valores próprios e (B) variabilidade cumulativa justificada em função do número de componentes consideradas.

Pela tabela 4.8 verifica-se que o sistema é difícil de ser descrito com um número restrito de componentes principais já que até à décima componente ($p = 10$) apenas se consegue descrever 65.6% (cerca de 2/3) da informação total. Contudo, da figura 4.21 (A) tem-se a noção de que existe também uma primeira componente mais relevante (PC1, 25.1%) seguida de 3 componentes com contribuições significativas (PC2-PC4).

Se forem consideradas as primeiras 4 componentes ($p = 4$) o sistema global encontra-se descrito em 51.9%.

Sobre a primeira componente principal (PC1) estão representadas 54 variáveis apenas referentes ao processo (x040, x041, x047, x049, x050, x051, x052, x054, x056, x058, x062, x065, x068, x072, x074, x075, x077, x080, x081, x092, x093, x094, x095, x096, x097, x098, x099, x100, x101, x103, x106, x107, x110, x111, x112, x114, x116, x118, x120, x121, x122, x123, x124, x125, x126, x127, x128, x129, x131, x133, x137, x139, x141 e x142).

Na segunda componente principal (PC2) estão representadas 59 variáveis que relacionam matérias-primas (6), processo (26) e resultados (27) (x002, x003, x005, x007, x008, x013, x015, x016, x020, x022, x024, x029, x031, x032, x037, x043, x046, x055, x057, x063, x067, x071, x076, x082, x083, x085, x086, x087, x108, x113, x134, x140, y01, y02, y05, y06, y08, y09, y10, y12, y13, y14, y18, y19, y21, y22, y23, y25, y27, y28, y31, y32, y34, y35, y36, y38, y48, y49 e y53).

Curiosamente, as componentes PC1 e PC2 não têm nenhuma variável em comum e ainda existem 85 variáveis das 198 para serem utilizadas.

Em PC3 estão representadas 55 variáveis que envolvem todos os sub-espacos, 6 sobre matérias-primas, 13 do processo e 36 respostas, (x003, x004, x005, x006, x008, x010, x013, x015, x016, x020, x032, x040, x053, x071, x073, x082, x108, x134, x135, y01, y02, y05, y06, y07, y08, y11, y12, y14, y18, y19, y20, y21, y24, y25, y27, y28, y29, y31, y32, y33, y34, y37, y38, y40, y41, y43, y47, y48, y49, y51, y53, y55, y57, y58 e y64).

Sobre PC4 encontram-se representadas as contribuições de 52 variáveis, 18 do processo e 34 respostas (x022, x024, x026, x029, x030, x031, x037, x046, x055, x057, x059, x063, x067, x077, x083, x113, x135, x140, y01, y02, y05, y06, y07, y10, y11, y12, y14, y15, y18, y19, y20, y23, y24, y25, y27, y28, y31, y32, y33, y36, y37, y38, y43, y44, y47, y48, y51, y52, y55, y57, y58 e y64).

Atendendo a estes resultados, se se pretender utilizar a análise PCA para avaliar a eventual inter-relação das variáveis entre predictores e respostas tem-se que recorrer à informação contida em nas componentes PC2, PC3 e PC4 já que a PC1 apenas está a descrever a variabilidade no espaço predictor (processo).

4.3 Interdependência

Uma vez que os sub-espços de predictores e de respostas já estão suficientemente bem caracterizados, importa agora descrever a inter-relação entre ambos os espaços.

Pelo menos dois procedimentos são passíveis de poderem a vir ser utilizados – as abordagens de modelação implícita (“soft modelling”³) e de modelação explícita (“hard modelling”⁴).

4.3.1 Modelação implícita

A modelação implícita é preferencial no caso de sistemas multivariados em que pouco se sabe sobre interdependência de variáveis. Este tipo de abordagem é algo “fluida” e genérica não permitindo contudo retirar grandes conclusões sobre o sistema em termos de comportamento intrínseco já que se assume como generalidade dependências lineares entre variáveis dos dois sub-espços, preditor e resposta.

Um pouco à semelhança do que foi anteriormente descrito para o caso do estudo de sub-espços através de PCA, na tabela 4.9 encontram-se resultados sobre a evolução dos primeiros 20 factores latentes do algoritmo PLS na justificação da variabilidade entre espaços dos predictores (X) e o das respostas (Y).

3 Na modelação implícita não se necessita de um modelo previamente descrito. Esta abordagem desenvolve-se iterativamente, assumindo dependências lineares.

4 Na modelação explícita os dados são directamente ajustados a um modelo de parâmetros bem definidos que procura descrever a resposta em função dos descritores definidos.

Tabela 4.9 – Evolução dos primeiros 20 factores latentes na descrição da informação.

#	SSx%	dVx%	$\Sigma Vx\%$	SSy%	dVy%	$\Sigma Vy\%$	Param.
1	69.8	30.2	30.2	92.1	7.9	7.9	26.248
2	59.5	10.3	40.5	81.3	10.8	18.7	30.630
3	51.5	8.0	48.5	69.8	11.5	30.2	31.594
4	48.8	2.7	51.2	59.7	10.1	40.3	29.678
5	46.4	2.4	53.6	55.6	4.1	44.4	18.841
6	44.4	2.0	55.6	52.0	3.6	48.0	17.672
7	42.3	2.1	57.7	49.3	2.7	50.7	15.398
8	38.7	3.5	61.3	47.9	1.4	52.1	10.888
9	37.2	1.5	62.8	46.2	1.7	53.8	12.321
10	35.8	1.4	64.2	44.8	1.3	55.2	10.701
11	34.8	1.0	65.2	42.8	2.0	57.2	13.328
12	33.5	1.3	66.5	41.4	1.4	58.6	11.209
13	32.3	1.2	67.7	40.1	1.3	59.9	10.428
14	31.3	1.1	68.7	38.8	1.3	61.2	10.463
15	29.9	1.4	70.1	38.1	0.8	61.9	8.171
16	28.8	1.1	71.2	37.2	0.9	62.8	8.719
17	27.9	0.9	72.1	36.2	1.0	63.8	9.189
18	27.1	0.8	72.9	35.2	1.0	64.8	9.542
19	26.1	1.0	73.9	34.5	0.7	65.5	7.901
20	25.1	0.9	74.9	33.8	0.7	66.2	7.782

- Índice do factor latente; SSx%, SSy% - variabilidade residual nos sub-espacos X e Y; dVx%, dVy% - variabilidade descrita; $\Sigma Vx\%$, $\Sigma Vy\%$ - variabilidade cumulativa descrita; Param - parâmetro do modelo.

Da tabela 4.9 resulta uma primeira rápida constatação – os predictores utilizados não conseguem descrever muito bem as respostas do sistema, por exemplo, para justificar 50.7% da informação contida na resposta são necessários 7 factores latentes que utilizam 57.7% da informação contida nos predictores, revelando uma eficiência de 87.9%, revelando esta dificuldade na descrição da resposta.

Verifica-se ainda que é necessário um elevado conjunto de factores latentes para descrever a resposta – 20 factores latentes descrevem apenas 66.2% (cerca de 2/3) com base em 74.9% (cerca de 3/4) da informação contida nos predictores, revelando de novo esta baixa eficiência na modelação.

Na figura 4.22 avalia-se um pouco melhor as questões de eficiência na modelação.

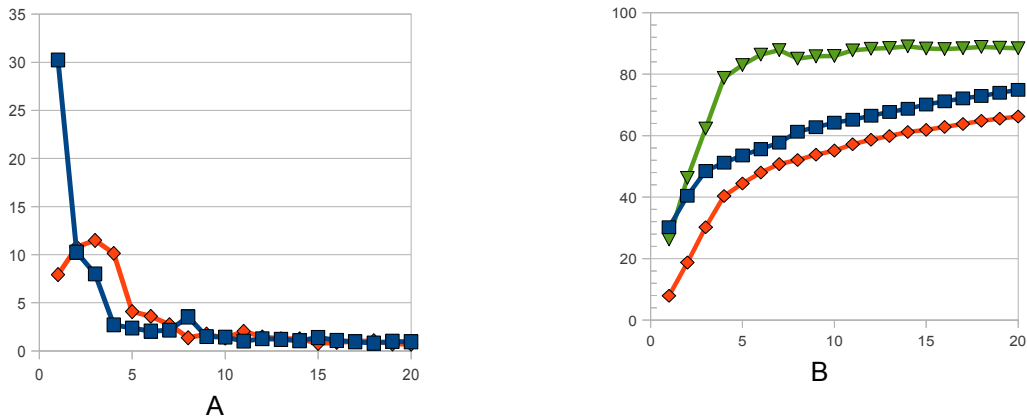


Figura 4.22 – Evolução dos primeiros 20 factores latentes na justificação da informação contida nos sub-espacos predictor (X) e resposta (Y): A) variabilidade utilizada no sub-espaco X (a azul) e Y (a vermelho) em função do número de factores considerados na descrição; B) evolução da variabilidade justificada cumulativa em cada sub-espaco (azul para X, vermelho para Y e verde para Y/X) em função do número de factores.

Da figura 4.22 pode-se inferir um pouco melhor o comportamento dos factores latentes em termos de eficiência na modelação. Em A) verifica-se que os 3 primeiros factores latentes são os que consomem mais informação do espaco predictor e que os 4 primeiros factores parecem ser relevantes na descrição da resposta. Já em B), obtém-se evidência de que com cinco factores latentes ($f = 5$) é conseguida uma boa descrição do sistema com uma eficiência global da ordem de 82.9%.

A interacção das variáveis entre os dois sub espacos pode ser obtida através dos respectivos *loads* (predictores e respostas).

As duas seguintes figuras esboçam a relevância das variáveis sobre os primeiros cinco factores latentes em relação aos sub-espacos dos predictores (fig. 4.23) e das respostas (fig. 4.24).

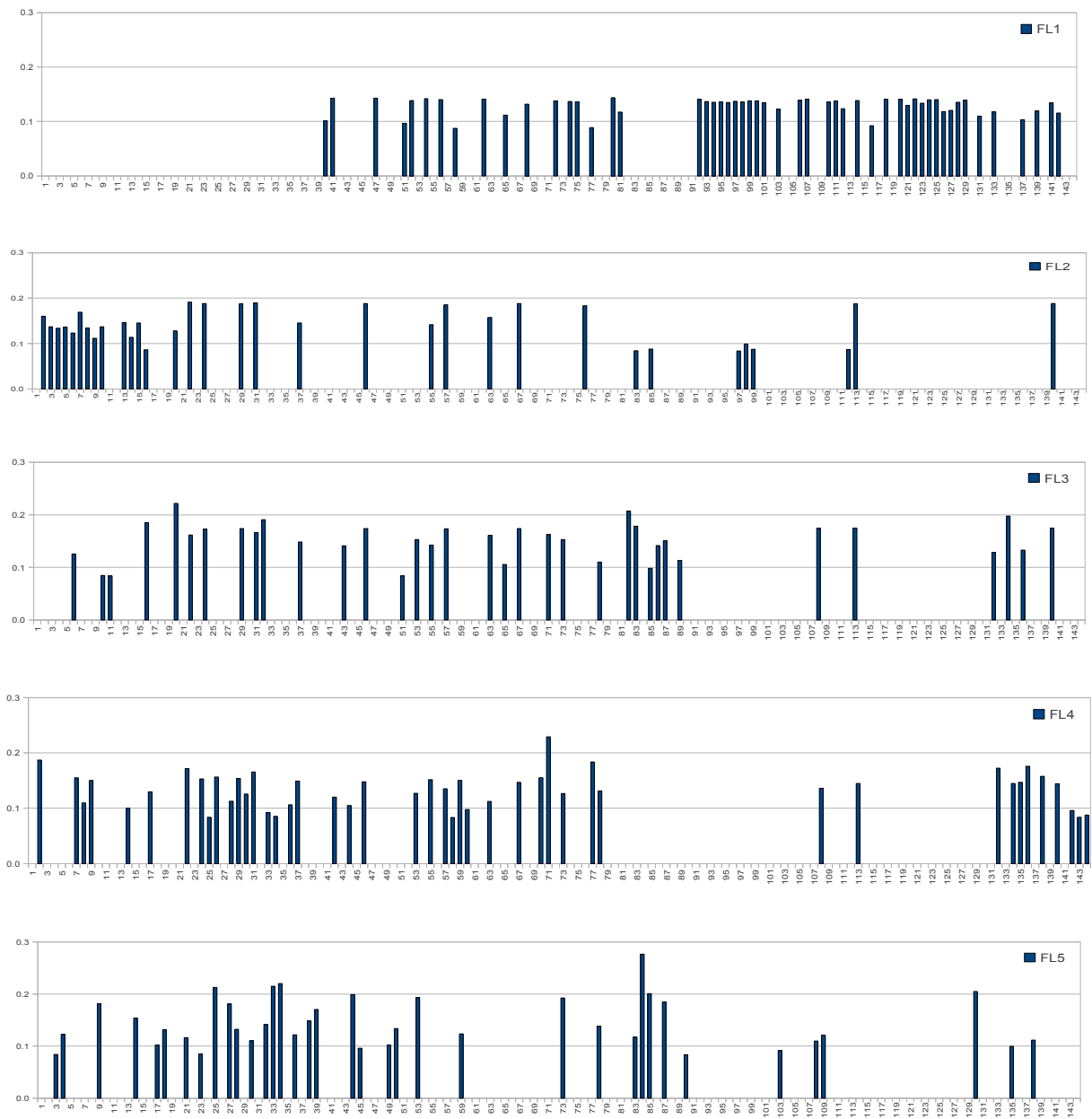


Figura 4.23 – Representação dos *loads* mais significativos (em valor absoluto) dos primeiros 5 factores latentes referentes ao sub-espaco dos 144 predictores.

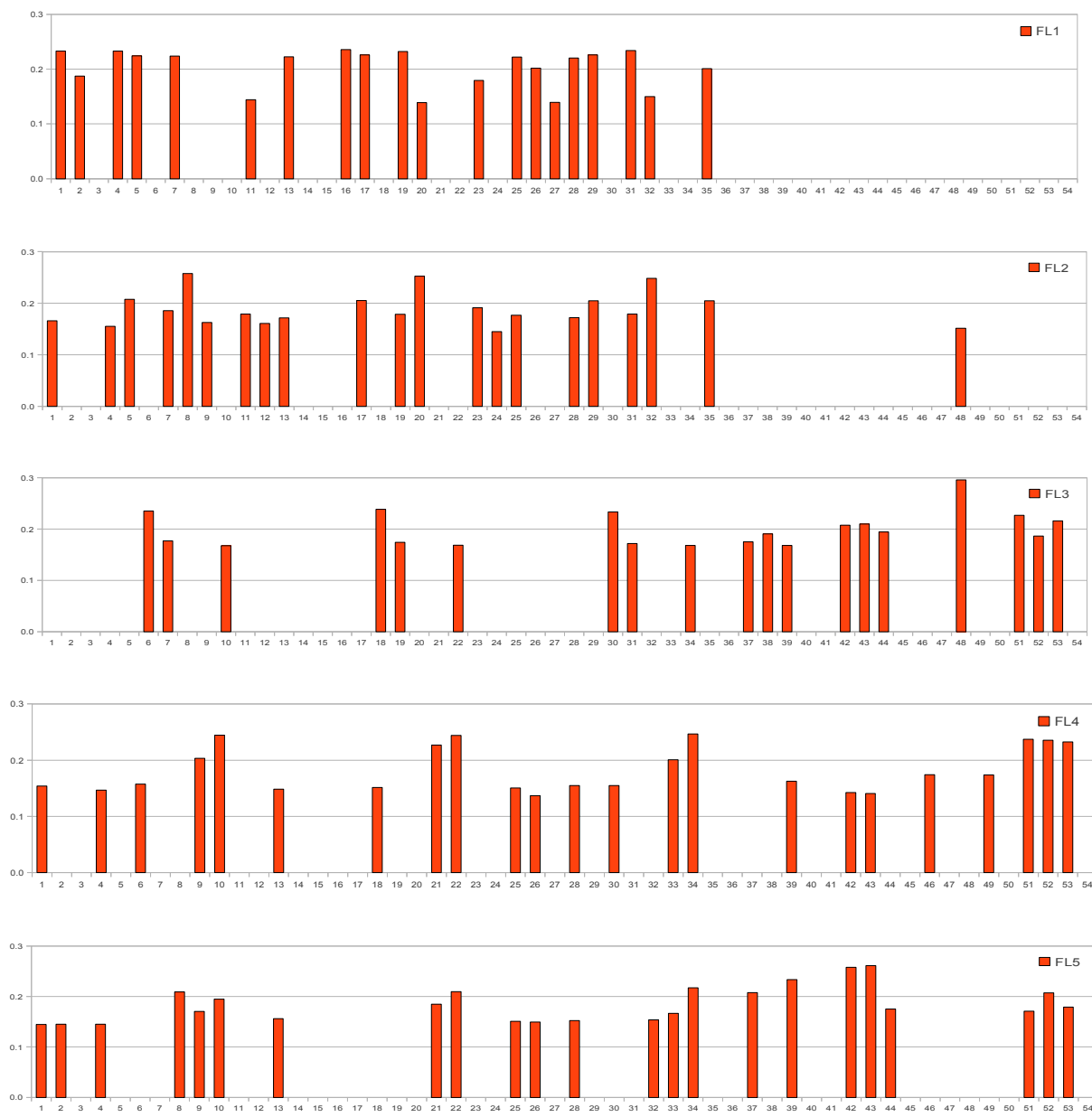


Figura 4.24 – Representação dos *loads* mais significativos (em valor absoluto) dos primeiros 5 factores latentes referentes ao sub-espaco das 54 respostas.

No que diz respeito à quantidade de informação contida nos factores latentes, verifica-se que o número de variáveis expressas sobre os cinco primeiros factores latentes mantém-se aproximadamente constante, sobretudo no que diz respeito ao sub-espço da resposta – por exemplo, sobre FL₁ estão representados 52 predictores e 20 respostas; sobre FL₂ existem contribuições significativas de 33 predictores e 21 respostas; sobre FL₃ encontra-se informação referente a 35 predictores e 19 respostas, em FL₄ estão 46 predictores e 23 respostas e sobre FL₅ estão representadas 37 predictores e 23 respostas.

Outra observação prende-se com o facto de a mesma variável surgir significativamente representada sobre diferentes factores latentes, por exemplo: dos 144 predictores originais apenas 126 variáveis estão significativamente representadas sobre os primeiros 5 factores latentes, entre as quais, 37 (29.4%) encontram-se repetidas sobre duas dessas componentes e outras 20 (15.9%) repetem-se por três vezes.

O mesmo acontece dentro do sub-espço das respostas – das 54 respostas estudadas apenas 44 estão representadas (com $f = 5$) entre as quais, 18 variáveis (40.9%) encontram-se duplicadas, 14 (31.8%) surgem triplicadas e 4 (9.0%) repetem-se ao longo de quatro factores latentes.

Esta sobreposição de variáveis transversalmente aos factores latentes pode significar que a modelação implícita através da presunção de modelo linear não é a descrição mais correcta, sendo necessário repetir as mesmas variáveis de forma a colmatar as dificuldades da ausência de linearidade.

A relação entre variáveis pode ser inferida pelas figuras 4.23 e 4.24, seguindo o mesmo factor latente. Contudo, sendo esta tarefa algo árdua, devido ao elevado número de variáveis, será avaliada a interdependência através da correlação directa entre variáveis dos sub-espços predictor e resposta na secção seguinte.

Na figura 4.25 encontra-se representados os fabricos nos sub-espços dos predictores (A) e de respostas (B).

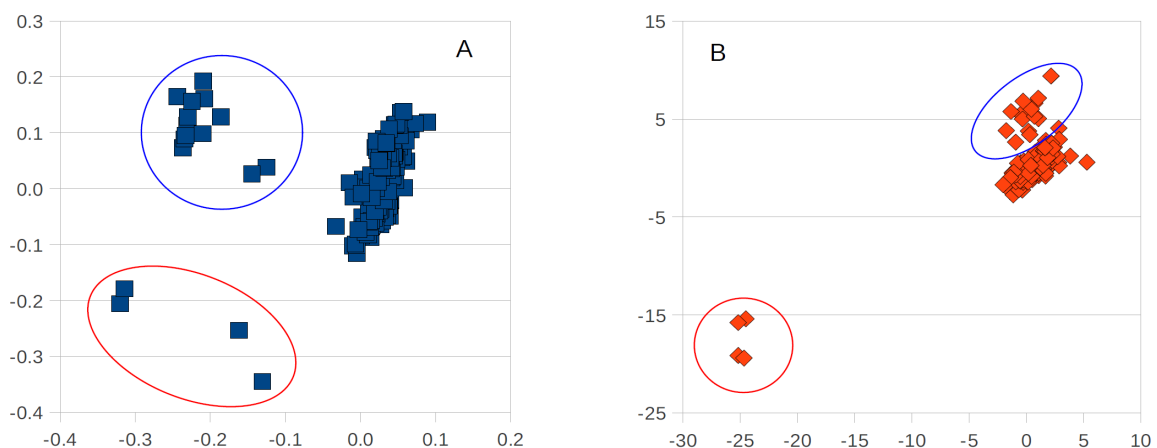


Figura 4.25 – Representação bidimensional dos fabricos nos sub-espços dos predictores (A) e das respostas (B).

Embora a realidade apresentada na figura 4.25 corresponda a uma descrição de apenas 40.5% e 18.7% da informação original de cada um dos sub-espacos predictores e respostas respectivamente, destaca-se a presença de algumas anomalias no fabrico – assinaladas com uma linha vermelha estão 4 situações anómalas que conduziram a produto em más condições; assinaladas a azul estão um conjunto de situações em que o produto ainda pode ser reutilizado após alguns acondicionamentos extraordinários. Já as situações de fabrico normal apenas estão representadas através dos respectivos símbolos.

Importa agora avaliar a interação entre as variáveis destes dois sub-espacos analisando um pouco mais em detalhe o que se passa ao nível da representação das variáveis sobre cada factor latente.

Em primeiro lugar importa saber reconhecer quais os factores latentes que são mais relevantes já que a análise anterior revelou apenas que os primeiros 5 factores latentes veiculam a maior parte desta informação. Trata-se de reconhecer dentre estes cinco factores onde é que a informação é mais relevante para a descrição da resposta. Para tal temos que atender ao modo como o algoritmo PLS processa a informação.

A modelação implícita por PLS assume apenas relações lineares entre predictores e respostas e, de forma iterativa, procura identificar em cada sub-espaco o conjunto de variáveis (vector latente) que mais se relaciona com o outro estabelecendo uma inter-relação designada de factor latente. Uma vez encontrada a relação externa entre espacos, o factor latente é removido – retira-se a respectiva informação em cada espaco de forma a ir sucessivamente evidenciando outras relações.

A sequência como o PLS vai procurando descrever a resposta com base nos predictores não é sempre a mais evidente já que os factores latentes que contêm informação mais relevante sobre o sistema não são descritos de forma ordenada, do mais relevante para o menos significativo.

Existe contudo maneiras de verificar e corrigir esta relevância atendendo não só à quantidade de informação descrita na resposta⁵ como ao valor do respectivo parâmetro⁶.

Assim sendo, os factores latentes ordenados por relevância decrescente são FL₃ (11.5%, $\beta_3 = 31.594$), FL₂ (10.8%, $\beta_2 = 30.630$), FL₄ (10.1%, $\beta_4 = 29.678$), FL₁ (7.9%, $\beta_1 = 26.248$) e FL₅ (4.1%, $\beta_5 = 18.841$).

O factor latente mais relevante é o terceiro factor latente (FL₃) através do qual é veiculada a informação referente a 33 predictores (**x020, x082, x134, x032, x016, x083, x113, x140, x108, x029, x046, x067, x057, x024, x031, x071, x022, x063, x073, x053, x087, x037, x055, x086, x043, x136, x132, x006, x089, x078, x065, x085, x010, x051 e x011**) e 19 respostas (**y53, y20, y07, y33, y57, y64, y48, y47, y49, y41, y58, y08, y40, y21, y34, y24, y37, y43 e y11**).

5 Quanto maior for a quantidade de resposta descrita pelo factor latente tanto mais relevante é este no estabelecimento do modelo.

6 O módulo do parâmetro do modelo PLS, também designado de coeficiente de sensibilidade, traduz o impacto do factor latente na equação do modelo.

Sobre o segundo factor latente (FL_2) estão representados 33 predictores ($x_{022}, x_{031}, x_{067}, x_{024}, x_{140}, x_{046}, x_{113}, x_{029}, x_{057}, x_{076}, x_{007}, x_{002}, x_{063}, x_{013}, x_{015}, x_{037}, x_{055}, x_{010}, x_{003}, x_{005}, x_{008}, x_{004}, x_{020}, x_{006}, x_{014}, x_{009}, x_{098}, x_{085}, x_{099}, x_{112}, x_{016}, x_{083}$ e x_{097}) que estão por sua vez relacionados com 21 respostas ($y_{09}, y_{22}, y_{35}, y_{06}, y_{19}, y_{38}, y_{32}, y_{25}, y_{08}, y_{34}, y_{12}, y_{21}, y_{27}, y_{31}, y_{14}, y_{01}, y_{10}, y_{13}, y_{05}, y_{53}$ e y_{26})

O terceiro factor latente mais relevante é FL_4 – contém informação de 46 predictores ($x_{071}, x_{002}, x_{077}, x_{136}, x_{132}, x_{022}, x_{031}, x_{138}, x_{026}, x_{070}, x_{007}, x_{029}, x_{024}, x_{055}, x_{059}, x_{009}, x_{037}, x_{046}, x_{135}, x_{067}, x_{134}, x_{113}, x_{140}, x_{108}, x_{057}, x_{078}, x_{017}, x_{053}, x_{073}, x_{030}, x_{042}, x_{028}, x_{063}, x_{008}, x_{036}, x_{044}, x_{014}, x_{060}, x_{142}, x_{033}, x_{144}, x_{034}, x_{143}, x_{025}$ e x_{058}) e 23 respostas ($y_{37}, y_{11}, y_{24}, y_{57}, y_{58}, y_{64}, y_{23}, y_{10}, y_{36}, y_{51}, y_{55}, y_{43}, y_{07}, y_{31}, y_{33}, y_{01}, y_{20}, y_{27}, y_{14}, y_{05}, y_{47}, y_{48}$ e y_{28}).

O quarto factor latente mais relevante é FL_1 – relaciona 52 predictores ($x_{080}, x_{041}, x_{047}, x_{054}, x_{122}, x_{062}, x_{107}, x_{118}, x_{120}, x_{092}, x_{125}, x_{056}, x_{124}, x_{106}, x_{129}, x_{052}, x_{114}, x_{072}, x_{100}, x_{099}, x_{111}, x_{097}, x_{093}, x_{074}, x_{075}, x_{098}, x_{095}, x_{110}, x_{128}, x_{094}, x_{096}, x_{141}, x_{101}, x_{123}, x_{068}, x_{121}, x_{112}, x_{103}, x_{127}, x_{139}, x_{126}, x_{133}, x_{081}, x_{142}, x_{065}, x_{131}, x_{137}, x_{040}, x_{051}, x_{116}, x_{077}$ e x_{058}) com 20 respostas ($y_{18}, y_{34}, y_{01}, y_{05}, y_{21}, y_{19}, y_{32}, y_{06}, y_{08}, y_{14}, y_{27}, y_{31}, y_{28}, y_{38}, y_{02}, y_{25}, y_{35}, y_{12}, y_{29}$ e y_{22}).

O quinto factor latente mais relevante coincide com FL_5 – contém informação de 37 predictores ($x_{084}, x_{034}, x_{033}, x_{025}, x_{130}, x_{085}, x_{044}, x_{053}, x_{073}, x_{087}, x_{009}, x_{027}, x_{039}, x_{014}, x_{038}, x_{032}, x_{078}, x_{050}, x_{028}, x_{018}, x_{059}, x_{004}, x_{036}, x_{109}, x_{083}, x_{021}, x_{138}, x_{030}, x_{108}, x_{017}, x_{049}, x_{135}, x_{045}, x_{103}, x_{023}, x_{003}$ e x_{090}) e 23 respostas ($y_{48}, y_{47}, y_{43}, y_{37}, y_{24}, y_{09}, y_{40}, y_{58}, y_{11}, y_{23}, y_{64}, y_{49}, y_{57}, y_{10}, y_{36}, y_{14}, y_{35}, y_{31}, y_{27}, y_{28}, y_{02}, y_{05}$ e y_{01}).

Esta informação mais pertinente será posteriormente analisada em maior detalhe.

4.3.2 Correlação

A análise da correlação entre variáveis de espaços diferentes – predictores e respostas – permite avaliar a sua inter-relação, segundo uma perspectiva linear.

Embora pareça uma tarefa fácil a de avaliar as correlações mais significativas entre variáveis, esta tarefa torna-se difícil devido à dimensão desta situação.

Este sistema em estudo contém 198 variáveis o que resulta numa matriz quadrada com as dimensões 198×198 . Por sua vez, esta matriz contém 198^2 (39204) elementos o que torna incomportável a sua listagem e representação.

Como estamos interessados em estabelecer a ponte entre predictores e respostas, podemos focar nossa atenção numa secção com as dimensões 144×54 – predictores versus respostas.

Sobre esta secção da matriz inspeccionou-se as maiores relações começando pelos valores ordenados (em módulo) das correlações respectivas.

Estes resultados obtidos foram deveras surpreendentes já que se verificou existir um número relativamente baixo de variáveis que parecem surgir com correlações muito acima do restante conjunto – estas parecem ser as variáveis mais significativas para o processo.

Num primeiro grupo de correlações mais elevadas, com valores absolutos entre 0.633 e 0.774, surgem 14 relações entre a resposta **y53** e 14 predictores (**x016, x020, x022, x024, x029, x031, x032, x046, x055, x057, x063, x067, x113 e x140**).

Antes de tirar mais ilações achou-se conveniente representar estas interdependências para avaliar a tendência desta interacção, não vá esta ser devida à presença de alguns valores singulares que já foram previamente detectados.

Na figura 4.26 encontram-se representadas estas relações que foram aqui identificadas.

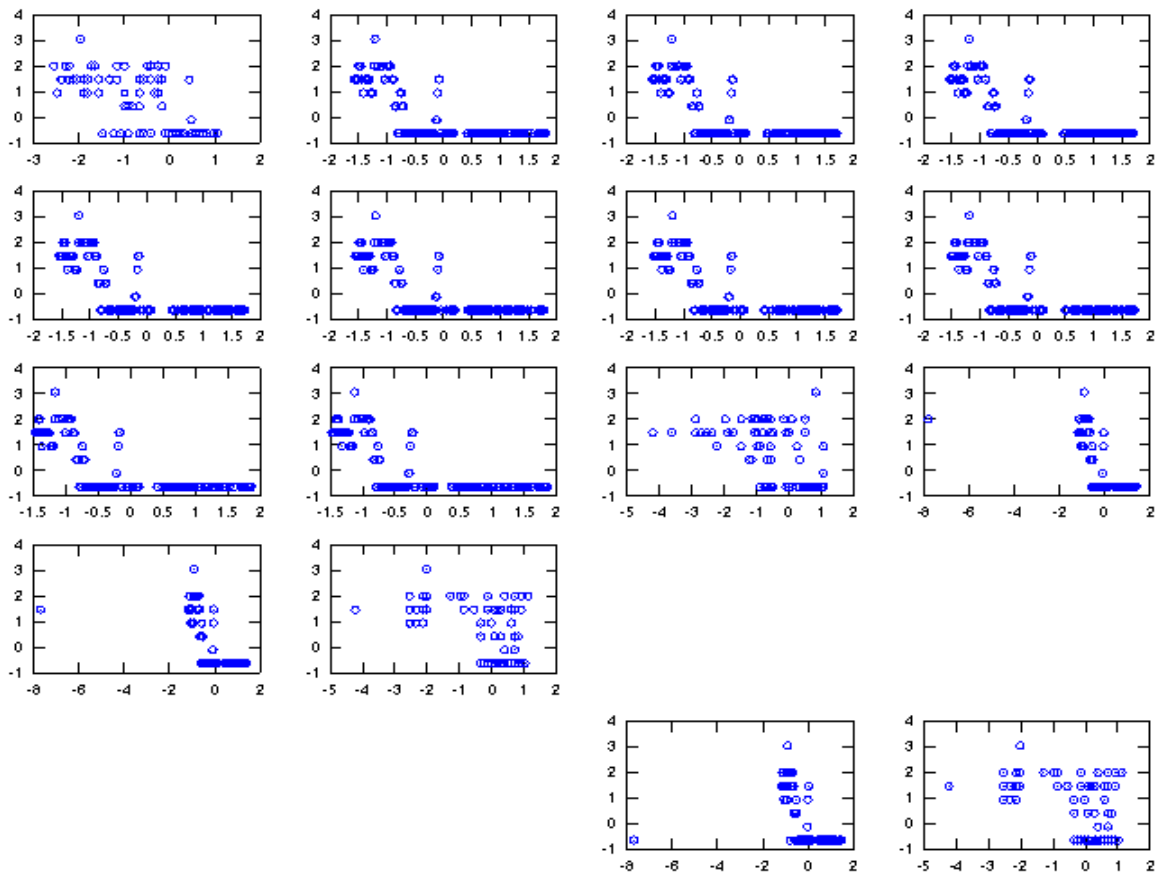


Figura 4.26 – Representação da inter-relação entre algumas das variáveis que apresentam maiores correlações com a resposta y53. Da esquerda para a direita e depois de cima para baixo estão x020, x029, x113, x140, x046, x024, x067, x057, x031, x022, x016, x063, x055 e x032. As duas últimas imagens apresentam correlações mais baixas (0.533 e 0.524) mas envolvem a relação da mesma resposta (y53) com x037 e x082, respectivamente.

Da figura 4.26 verifica-se que, em alguns casos (últimos gráficos) existem situações de anomalia (valores discrepantes) relacionadas com as situações previamente identificadas.

Contudo, a relação da resposta y53 (teor final de triglicérideos) com estas 16 variáveis revela como esta variável é muito sensível a este conjunto de variáveis. Estas relações serão posteriormente exploradas e analisadas em maior detalhe.

Convém ainda referir que neste conjunto de 14 relações mais elevadas estão representados 14 predictores distintos, todos referentes ao processo de transesterificação e nenhum referente à matéria-prima. Ordenadamente, estes são: **x016** (B0614.02-H₃PO₄, L614402 %), **x020** (B0614.03-NaOH, L614403 %), **x022** (NaOH dil. Water, total m³), **x024** (aggl.water cons., total m³), **x029** (washing water to R0614.04, total m³), **x031** (citric acid cons., total L), **x032** (B0614.06 citric acid, L614407 %), **x046** (Na-Methylat cons I, total ton), **x055** (Na-Methylat cons II, total ton), **x057** (Methanol consumption II, total ton), **x063** (process water III(S2), total m³), **x067** (process water II(S3), total m³), **x113** (methanol to tankfarm, total ton) e **x140** (glycerine to tankfarm, total ton). Todas estas estão relacionadas com **y53** (teor final de triglicérideos).

Depois deste conjunto coeso de correlações mais elevadas (valores situados entre 0.633 e 0.774) foi identificado um segundo conjunto mais difuso, com valores situados abaixo de 0.573 – trata-se de um

conjunto mais diversificado em termos de respostas aqui representadas.

Na figura 4.27 encontram-se representadas as 20 situações mais correlacionadas (valores compreendidos entre 0.573 e 0.500), encontradas neste conjunto.

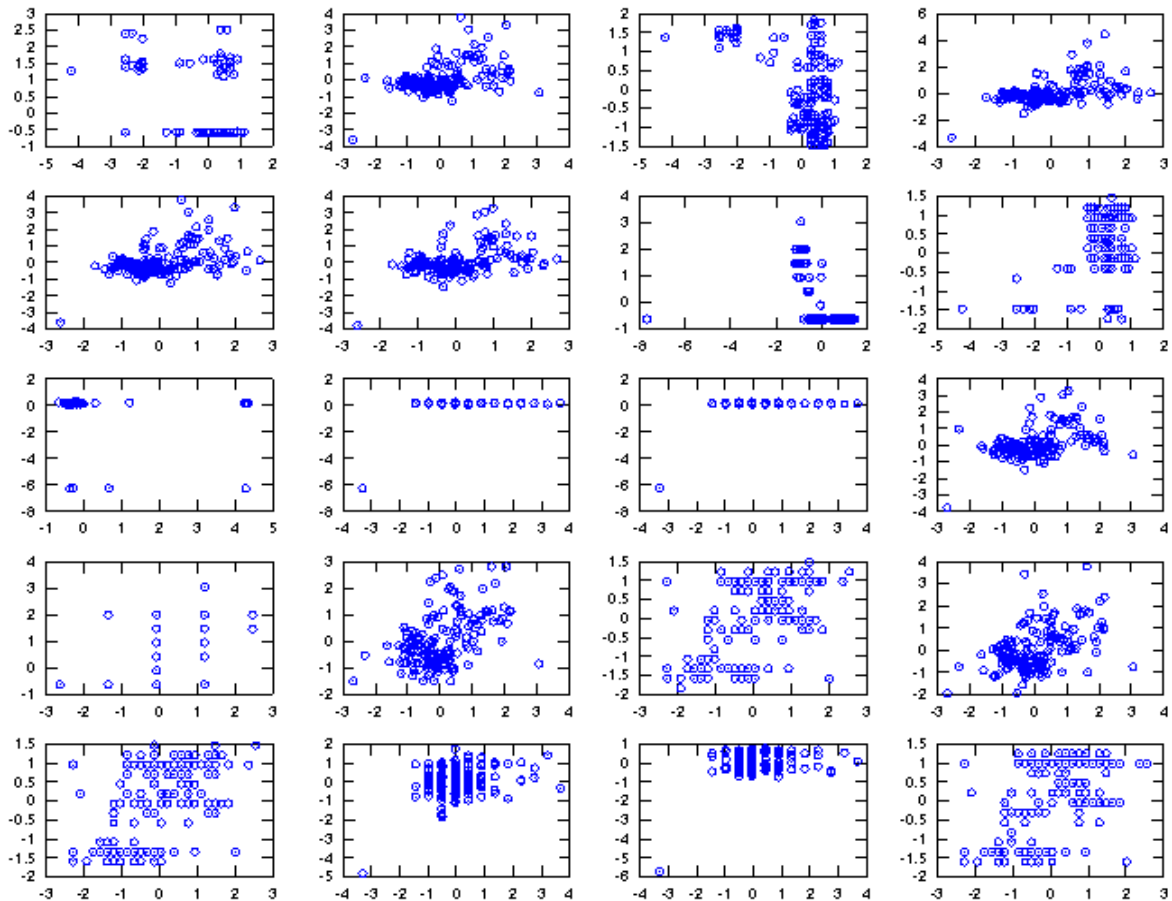


Figura 4.27 – Representação das 20 maiores inter-relações encontradas entre os valores 0.573 (imagem canto superior esquerda) e 0.500 (canto inferior direito).

Da figura 4.27 verifica-se que, nos gráficos representados, encontra-se em alguns casos, valores anómalos que distorcem as representações, sendo estes os maiores responsáveis pelos valores de correlação observados.

Neste conjunto de 20 relações apuradas há também algumas particularidades interessantes que convém referir:

- apenas estão presentes 7 predictores (**x002**, **x007**, **x013**, **x032**, **x037**, **x071** e **x082**) que se repetem algumas vezes. Esta representação neste conjunto assume a frequência respectiva de 4, 3, 5, 3, 1, 3 e 1.
- como respostas surgem 17 variáveis (**y06**, **y08**, **y09**, **y10**, **y11**, **y19**, **y22**, **y23**, **y24**, **y32**, **y35**, **y37**, **y38**, **y41**, **y53**, **y57** e **y64**)

Trata-se de um conjunto de variáveis mais diversificado tanto no sub-espço da resposta como no dos predictores, surgindo pela primeira vez três variáveis relacionadas com a matéria-prima.

4.4 Discussão

Neste trabalho foram exploradas três abordagens distintas para revelar alguma informação sobre este sistema: a análise PCA para inferir quais as variáveis mais relevantes em cada sector de dados, a correlação para avaliar as ligações mais relevantes e PLS para obter informação conjunta entre sub-espacos.

Importa agora reunir toda esta informação obtida para extrairmos mais informação sobre as metodologias de análise e sobre o sistema em causa.

4.4.1 PCA

A análise PCA permitiu verificar essencialmente que cada sub-espaco (matéria-prima, condições de produção e características do biodiesel) podem ser representados através de um número muito reduzido de componentes principais.

4.4.1.1 Matéria-prima

O sub-conjunto referente à matéria prima (x001 a x015) requer pelo menos 5 componentes para ser descrito a 68.8% contudo a informação mais relevante está contida em PC1, revelando a eventual existência de relações internas entre as variáveis x002, x003, x004, x005, x006, x007, x008, x009, x010, x013, x014 e x015. Contudo a análise de correlação apenas evidenciou uma relação muito significativa entre as variáveis x002 e x007.

Tabela 4.10 - Apresentação dos *loads* mais significativos referentes ao sub-espço da matéria-prima sobre as primeiras cinco componentes principais.

#	1	2	3	4	5
V%	33.61	11.58	8.83	7.90	6.89
$\Sigma V\%$	33.61	45.19	54.01	61.91	68.80
	PC1	PC2	PC3	PC4	PC5
x001			0.448		-0.424
x002	0.287			-0.495	-0.362
x003	0.279			0.263	
x004	0.292		0.429		
x005	0.264				
x006	0.269	-0.396			
x007	0.298		-0.456	-0.361	
x008	0.270		-0.419	0.424	
x009	0.270	0.418			
x010	0.287	-0.352			
x011		-0.391			
x012				-0.291	0.646
x013	0.294		-0.301	0.394	
x014	0.266	0.386			
x015	0.294	-0.332			

Desta tabela verifica-se que as variáveis evidenciadas através de PC1 são eventualmente cooperativas – evoluem em simultâneo no mesmo sentido.

4.4.1.2. Processo

No que diz respeito às condições de produção do processo fabril (x016 a x144), este espaço de 129 variáveis pode ser representado através de duas componentes principais ($p = 2$) que envolvem a participação conjunta de 80 variáveis e são capazes de descrever cerca de 49.2% da informação aí contida.

A primeira componente é a mais relevante (é capaz de descrever 36.4% da informação) e engloba a participação de 51 variáveis; a segunda componente descreve 12.8% e representa-se através de 29 variáveis.

Tabela 4.11 - Representação das variáveis do espaço processo ao longo das duas componentes principais.

Var.	PC1	PC2	Var.	PC1	PC2	Var.	PC1	PC2	Var.	PC1	PC2
x016		-0.161	x062	-0.138		x092	-0.141		x120	-0.140	
x020		-0.211	x063		-0.200	x093	-0.139		x121	-0.133	
x022		-0.222	x065	-0.110		x094	-0.138		x122	-0.142	
x024		-0.226	x067		-0.226	x095	-0.139		x123	-0.128	
x029		-0.227	x068	-0.130		x096	-0.138		x124	-0.140	
x031		-0.224	x071		-0.093	x097	-0.140		x125	-0.139	
x032		-0.160	x072	-0.137		x098	-0.142		x126	-0.123	
x037		-0.189	x073		0.091	x099	-0.142		x127	-0.126	
x040	-0.106		x074	0.137		x100	-0.140		x128	-0.136	
x041	-0.140		x075	-0.134		x101	-0.135		x129	-0.141	
x043		0.126	x076		-0.155	x103	-0.117		x131	0.110	
x046		-0.226	x077			x106	-0.142		x133	0.119	
x047	-0.140		x078		-0.088	x107	-0.141		x134		-0.155
x051	0.097		x080	-0.140		x108		-0.141	x137	-0.106	
x052	-0.141		x081	0.109		x110	-0.134		x138		
x053		0.091	x082		0.185	x111	-0.133		x139	-0.120	
x054	-0.139		x083		-0.173	x112	-0.129		x140		-0.227
x055		-0.181	x085		-0.123	x113		-0.227	x141	-0.134	
x056	-0.138		x086		0.126	x114	-0.135		x142	-0.117	
x057		-0.224	x087		-0.153	x116	-0.095				
x058	-0.088		x089		-0.091	x118	-0.140				

4.4.1.3. Predictores

O sub-espaço dos predictores engloba as matérias primas (x001 a x015) e as variáveis do processo (x016 a x144) num total de 144 variáveis. Como já foi anteriormente visto, este espaço pode ser descrito através de duas componentes principais ($p = 2$) – na primeira componente estão representadas 52 variáveis e esta é capaz de descrever 32.7% da informação total; sobre a segunda estão representadas outras 33 variáveis e esta consegue descrever 11.8% da informação. O impacto das variáveis sobre as componentes principais encontra-se evidenciado na tabela 4.12.

Tabela 4.12 - Impacto das variáveis do espaço predictor sobre as primeiras duas componentes principais.

Var.	PC1	PC2	Var.	PC1	PC2	Var.	PC1	PC2	Var.	PC1	PC2
x002		-0.083	x058	-0.088		x092	-0.141		x121	-0.133	
x007		-0.097	x062	-0.138		x093	-0.139		x122	-0.142	
x016		-0.155	x063		-0.197	x094	-0.138		x123	-0.128	
x020		-0.205	x065	-0.110		x095	-0.139		x124	-0.140	
x022		-0.220	x067		-0.223	x096	-0.138		x125	-0.139	
x024		-0.224	x068	-0.130		x097	-0.140		x126	-0.123	
x029		-0.224	x071		-0.088	x098	-0.142		x127	-0.126	
x031		-0.222	x072	-0.137		x099	-0.142		x128	-0.136	
x032		-0.152	x073		0.086	x100	-0.140		x129	-0.141	
x037		-0.185	x074	0.137		x101	-0.135		x131	0.110	
x040	-0.106		x075	-0.134		x103	-0.117		x132		-0.084
x041	-0.140		x076		-0.159	x106	-0.141		x133	0.119	
x043		0.119	x077	-0.083		x107	-0.141		x134		-0.151
x046		-0.223	x078		-0.088	x108		-0.138	x136		-0.084
x047	-0.139		x080	-0.139		x110	-0.134		x137	-0.106	
x051	0.097		x081	0.109		x111	-0.133		x139	-0.120	
x052	-0.140		x082		0.178	x112	-0.128		x140		-0.224
x053		0.086	x083		-0.167	x113		-0.224	x141	-0.134	
x054	-0.139		x085		-0.123	x114	-0.134		x142	-0.117	
x055		-0.176	x086		0.119	x116	-0.095				
x056	-0.138		x087		-0.149	x118	-0.140				
x057		-0.221	x089		-0.088	x120	-0.139				

4.4.1.3 Respostas

O sub-espço das respostas pode ser parcialmente descrito com duas componentes principais ($p = 2$), capazes de descrever 57% da informação contida neste sub-espço.

Sobre PC1 estão representadas 23 variáveis e esta é capaz de descrever 32.5% da informação; sobre PC2 encontram-se 19 variáveis e é capaz de descrever 24.5% do conjunto.

Tabela 4.13 - Relevância das respostas para as duas primeiras componentes principais.

Var.	PC1	PC2	Var.	PC1	PC2	Var.	PC1	PC2	Var.	PC1	PC2
y01	0.142		y15	0.180		y28	0.228		y46	0.228	
y02	0.228		y16	0.218		y29	0.148		y48		0.247
y05	0.174		y18	0.195		y31		0.190	y50	0.220	
y06	0.228		y19	0.218		y33		0.190	y52		-0.214
y07	0.182		y20		0.255	y34	0.218		y53		0.143
y10		-0.159	y21	0.149		y35	0.177		y55		0.139
y11		-0.256	y22		0.256	y36	0.228		y56		-0.215
y12		0.197	y23	0.221		y37		0.249	y64		-0.143
y13	0.153	-0.198	y26	0.193		y38	0.152	-0.199	y66	0.152	-0.199
y14		0.255	y27	0.228		y44		-0.204			

4.4.1.4 Processo global

A análise PCA ao conjunto global de dados permite representar os dados em 4 componentes principais de forma a descrever 51.9% dessa informação.

A componente PC1 repercute 54 variáveis de um total de 198, conseguindo descrever cerca de 25.1% da informação. As variáveis aqui reproduzidas são exclusivamente variáveis do processo e este conjunto tem muito em comum com a PC1 no caso da análise da variabilidade ao processo (4.4.1.2).

Sobre a segunda componente PC2, esta faz-se representar através de 59 variáveis e descreve 11.3% da informação total. As variáveis representadas pertencem aos predictores e respostas.

As componentes principais PC3 e PC4 representam-se por 55 e 52 variáveis e são responsáveis por descrever 9.6% e 5.9% da informação, respectivamente.

Tabela 4.14 - Impacto das variáveis sobre as componentes principais do sistema segundo uma perspectiva global.

Var.	PC1	PC2	PC3	PC4	Var.	PC1	PC2	PC3	PC4	Var.	PC1	PC2	PC3	PC4	Var.	PC1	PC2	PC3	PC4
x002		-0.116			x062	0.135				x113		-0.170		0.137	y18		-0.092	0.086	-0.129
x003		-0.074	0.091		x063		-0.148		0.114	x114	0.133				y19		-0.111	0.136	-0.108
x004			0.088		x065	0.106				x116	0.090				y20			0.200	0.092
x005		-0.075	0.081		x067		-0.169		0.138	x118	0.135				y21		-0.075	0.171	
x006			0.119		x068	0.126				x120	0.134				y22		-0.144		
x007		-0.119			x071		-0.076	-0.110		x121	0.128				y23		-0.118		-0.131
x008		-0.086	0.081		x072	0.132				x122	0.136				y24			-0.138	-0.193
x010			0.100		x073			0.096		x123	0.123				y25		-0.125	0.075	-0.083
x013		-0.085	0.098		x074	-0.132				x124	0.134				y27		-0.112	0.113	-0.139
x015		-0.075	0.081		x075	0.130				x125	0.135				y28		-0.088	0.097	-0.128
x016		-0.114	-0.085		x076		-0.142			x126	0.117				y29			0.105	
x020		-0.155	-0.100		x077	0.084			-0.085	x127	0.119				y31		-0.110	0.114	-0.141
x022		-0.165		0.151	x080	0.136				x128	0.130				y32		-0.111	0.136	-0.107
x024		-0.168		0.144	x081	-0.110				x129	0.134				y33			0.199	0.095
x026				0.096	x082		0.128	0.097		x131	-0.105				y34		-0.076	0.169	
x029		-0.168		0.144	x083		-0.119		0.073	x133	-0.113				y35		-0.151		
x030				0.089	x085		-0.094			x134		-0.114	-0.103		y36		-0.112		-0.116
x031		-0.166		0.148	x086		0.090			x135			-0.081	0.079	y37			-0.137	-0.195
x032		-0.122	-0.111		x087		-0.107			x137	0.102				y38		-0.129	0.078	-0.075
x037		-0.132		0.129	x092	0.135				x139	0.116				y40			-0.125	
x040	0.098		-0.073		x093	0.133				x140		-0.170		0.137	y41			0.115	
x041	0.137				x094	0.132				x141	0.128				y43			0.142	0.134
x043		0.090			x095	0.132				x142	0.111				y44				0.089
x046		-0.169		0.139	x096	0.132				y01		-0.111	0.105	-0.141	y47			0.141	0.106
x047	0.136				x097	0.134				y02		-0.078	0.081	-0.129	y48		0.071	0.141	0.105
x049	0.076				x098	0.135				y05		-0.106	0.107	-0.143	y49		-0.080	-0.098	
x050	-0.077				x099	0.135				y06		-0.112	0.135	-0.107	y51			-0.114	-0.122
x051	-0.094				x100	0.134				y07			0.199	0.096	y52				-0.102
x052	0.135				x101	0.129				y08		-0.077	0.172		y53		0.145	0.106	
x053			0.096		x103	0.113				y09		-0.152			y55			-0.083	-0.133
x054	0.136				x106	0.135				y10		-0.126		-0.111	y57			0.165	0.160
x055		-0.133		0.117	x107	0.135				y11			-0.137	-0.195	y58			-0.161	-0.175
x056	0.134				x108		-0.098	-0.097		y12		-0.115	0.082	-0.091	y64			-0.170	-0.163
x057		-0.169		0.131	x110	0.131				y13		-0.074							
x058	0.088				x111	0.130				y14		-0.115	0.105	-0.142					
x059				-0.097	x112	0.122				y15				-126					

4.4.2 PLS

Através da análise PLS foram considerados apenas os cinco factores latentes mais relevantes – FL₃ (11.5%, $\beta_3 = 31.594$), FL₂ (10.8%, $\beta_2 = 30.630$), FL₄ (10.1%, $\beta_4 = 29.678$), FL₁ (7.9%, $\beta_1 = 26.248$) e FL₅ (4.1%, $\beta_5 = 18.841$).

O factor latente mais relevante veicula a informação referente a 35 predictores e 19 respostas. Sobre o segundo factor latente mais relevante estão representados 33 predictores e 21 respostas. Embora estes dois factores latentes sejam os mais relevantes, considerando os primeiros cinco factores ($f = 5$) consegue-se ter a representação das variáveis todas do sistema à excepção de 13 predictores e 10 respostas.

As tabelas 4.15 e 4.16 revelam a relação das variáveis dos predictores e das respostas, respectivamente, com cada um destes quatro factores latentes mais relevantes.

Tabela 4.15 - Relevância dos preditores do sistema sobre os primeiros cinco factores latentes mais relevantes.

Var.	FL ₃	FL ₂	FL ₄	FL ₁	FL ₅	Var.	FL ₃	FL ₂	FL ₄	FL ₁	FL ₅	Var.	FL ₃	FL ₂	FL ₄	FL ₁	FL ₅
x002		0.160	-0.187			x049					-0.102	x097		-0.084			0.137
x003		0.137			0.084	x050					0.134	x098		-0.099			0.136
x004		0.134			0.123	x051	0.084			-0.096		x099		-0.087			0.138
x005		0.136				x052				0.139		x100					0.138
x006	-0.126	0.123				x053	-0.153		0.127		-0.193	x101					0.134
x007		0.169	-0.155			x054				0.142		x103				0.122	0.091
x008		0.134	-0.110			x055	0.143	0.142	0.152			x106					0.139
x009		0.111	-0.150		0.181	x056				0.140		x107					0.141
x010	-0.085	0.137				x057	0.173	0.186	0.135			x108	0.174		-0.136		0.110
x011	-0.084					x058			-0.083	0.088		x109					-0.121
x013		0.146				x059			-0.150		0.124	x110					0.136
x014		0.114	-0.100		0.154	x060			0.098			x111					0.138
x015		0.145				x062				0.141		x112		-0.087			0.123
x016	0.185	0.086				x063	0.161	0.158	0.112			x113	0.175	0.188	0.145		
x017			0.130		-0.102	x064						x114					0.138
x018					-0.132	x065	0.106			0.112		x116					0.092
x020	0.222	0.128				x066						x118					0.141
x021					-0.116	x067	0.174	0.188	0.147			x120					0.140
x022	0.161	0.191	0.171			x068				0.132		x121					0.129
x023					-0.085	x070			0.155			x122					0.141
x024	0.173	0.188	0.153			x071	0.163		-0.229			x123					0.134
x025			0.084		-0.213	x072				0.138		x124					0.140
x026			0.156			x073	-0.153		0.127		-0.193	x125					0.140
x027			0.083		-0.181	x074				-0.136		x126					0.118
x028			0.113		-0.132	x075				0.136		x127					0.120
x029	0.174	0.187	0.154			x076		0.183				x128					0.135
x030			0.126		-0.110	x077			-0.184	0.089		x129					0.139
x031	0.166	0.189	0.166			x078	0.110		-0.132		0.138	x130					0.205
x032	0.191				-0.142	x080				0.144		x131					-0.110
x033			0.092		-0.215	x081				-0.117		x132	0.128		-0.173		
x034			0.085		-0.220	x082	-0.207					x133					-0.118
x035						x083	0.179	0.084			0.117	x134	0.198		-0.145		
x036			0.106		-0.122	x084					0.277	x135			0.147		0.100
x037	0.148	0.145	0.149			x085	0.098	0.088			0.201	x136	0.133		-0.176		
x038					-0.148	x086	-0.142					x137					0.103
x039					-0.171	x087	0.151				0.185	x138			-0.158		0.112
x040				0.101		x088						x139					0.120
x041				0.143		x089	0.114					x140	0.175	0.188	0.144		
x042			0.120			x090					0.084	x141					0.134
x043	-0.141					x092				0.140		x142			0.096	0.116	
x044			-0.105		0.199	x093				0.137		x143			0.084		
x045					-0.096	x094				0.135		x144			0.088		
x046	0.174	0.188	0.148			x095				0.136							
x047				0.143		x096				0.135							

Tabela 4.16 - Relevância das respostas do sistema sobre os primeiros cinco factores latentes mais relevantes.

Var.	FL ₃	FL ₂	FL ₄	FL ₁	FL ₅	Var.	FL ₃	FL ₂	FL ₄	FL ₁	FL ₅	Var.	FL ₃	FL ₂	FL ₄	FL ₁	FL ₅
y01		0.166	-0.154	0.233	-0.145	y21	-0.174	0.178		0.232		y37	0.168		-0.247		-0.217
y02				0.187	-0.145	y22		0.253		0.139		y38		0.205		0.201	
y05		0.155	-0.147	0.233	-0.145	y23			-0.227		-0.185	y40	0.175				-0.207
y06		0.208		0.224		y24	0.168		-0.244		-0.209	y41	-0.191				
y07	-0.235		0.157			y25		0.191		0.180		y43	-0.168		0.163		0.233
y08	-0.177	0.185		0.224		y26		0.145				y47	-0.208		0.142		0.258
y09		0.258			-0.209	y27		0.177	-0.151	0.222	-0.151	y48	-0.210		0.141		0.261
y10		0.163	-0.203		-0.170	y28			-0.137	0.202	-0.149	y49	0.195				-0.175
y11	0.168		-0.244		-0.195	y29				0.139		y51			-0.174		
y12		0.179		0.144		y31		0.172	-0.155	0.220	-0.152	y53	-0.296	-0.152			
y13		0.161				y32		0.205		0.226		y55			-0.174		
y14		0.171	-0.148	0.222	-0.156	y33	-0.233		0.155			y57	-0.227		0.237		0.171
y18				0.236		y34	-0.172	0.179		0.234		y58	0.186		-0.235		-0.207
y19		0.205		0.226		y35		0.248		0.150	-0.154	y64	0.216		-0.232		-0.179
y20		0.239		0.152		y36				0.201	0.167						

O factor latente mais significativo FL₃ relaciona os predictores **x006** (-0.126), **x010** (-0.085), **x011** (-0.084), **x016** (0.185), **x020** (0.222), **x022** (0.161), **x024** (0.173), **x029** (0.174), **x031** (0.166), **x032** (0.191), **x037** (0.148), **x043** (-0.141), **x046** (0.174), **x051** (0.084), **x053** (-0.153), **x055** (0.143), **x057** (0.173), **x063** (0.161), **x065** (0.106), **x067** (0.174), **x071** (0.163), **x073** (-0.153), **x078** (0.110), **x082** (-0.207), **x083** (0.179), **x085** (0.098), **x086** (-0.142), **x087** (0.151), **x089** (0.114), **x108** (0.174), **x113** (0.175), **x132** (0.128), **x134** (0.198), **x136** (0.133), **x140** (0.175) com 19 respostas **y07** (-0.235), **y08** (-0.177), **y11** (0.168), **y20** (-0.239), **y21** (-0.174), **y24** (0.168), **y33** (-0.233), **y34** (-0.172), **y37** (0.168), **y40** (0.175), **y41** (-0.191), **y43** (-0.168), **y47** (-0.208), **y48** (-0.210), **y49** (0.195), **y53** (-0.296), **y57** (-0.227), **y58** (0.186), **y64** (0.216).

Esta informação deve ser posteriormente escrutinada no sentido de verificar e validar as inter-relações entre os sub-espacos dos predictores e das respostas.

4.4.3 Correlação

A análise de correlação entre os sub-espacos dos preditores e das respostas permitiu evidenciar dois conjuntos distintos em termos de grandezas de correlações.

Na análise de correlação foram identificados dois grupos de correlações mais elevadas – o primeiro grupo que envolve apenas uma resposta e outro grupo mais difuso que se estende a maior número de respostas.

Tabela 4.17 - Conjunto de correlações mais significativas entre preditores e respostas ($r > 0.500$).

	x002	x007	x013	x016	x020	x022	x024	x029	x031	x032	x037	x046	x055	x057	x067	x071	x082	x113	x140
y06			0.525																
y08			0.500																
y09	0.549	0.540																	
y10	0.508																		
y11																0.504			
y19			0.525																
y22		0.545																	
y23	0.504																		
y24																	0.506		
y32			0.525																
y35	0.524	0.537																	
y37																			0.500
y38			0.502																
y41										-0.573									
y53				-0.661	-0.774	-0.676	-0.709	-0.713	-0.691	-0.633	-0.533	-0.710	-0.635	-0.706	-0.709		0.524	-0.713	-0.712
y57										-0.547									
y64										0.529									

Importa agora fazer convergir toda a informação no sentido de tirar conclusões sobre o processo.

Começando a análise pela tabela mais simples sobre a correlação, verifica-se que a variável **y53** (teor final de triglicérideos) depende de 15 variáveis do processo de trans-esterificação: **x016** (-0.661), **x020** (-0.774), **x022** (-0.676), **x024** (-0.709), **x029** (-0.713), **x031** (-0.691), **x032** (-0.633), **x037** (-0.533), **x046** (-0.710), **x055** (-0.635), **x057** (-0.706), **x067** (-0.709), **x082** (0.524), **x113** (-0.713) e **x140** (-0.712). De salientar que y53 apenas apresenta uma dependência crescente com x082 sendo as restantes relações antagónicas.

Esta mesma relação vem também expressa, de uma forma mais global, através do impacto das variáveis representadas no factor latente mais relevante (FL₃) - neste factor surge **y53** (-0.296) relacionada com **x016** (0.185), **x020** (0.222), **x022** (0.161), **x024** (0.173), **x029** (0.174), **x031** (0.166), **x032** (0.191), **x037** (0.148), **x046** (0.174), **x055** (0.143), **x057** (0.173), **x067** (0.174), **x082** (-0.153), **x113** (0.175) e **x140** (0.175) – as mesmas interdependências em termos de sinal e grandeza relativa.

Passando agora ao lote de variáveis menos correlacionadas mas ainda assim relevantes, foram localizadas as relações de **x002** com **y09** (0.549), com **y10** (0.508), **y23** (0.504) e **y35** (0.524) – estas interdependências reflectem-se sobretudo através do segundo factor latente mais relevante (FL₂): **x002** (0.160) com **y09** (0.258), **y10** (0.163) e **y35** (0.248) mas ainda sobre FL₄ onde surge **x002** (-0.187) ligado a **y23** (-0.227). De salientar que as interdependências crescente/decrecente são também preservadas.

Se se atentar agora nas restantes relações verifica-se que **x007** relaciona-se também com **y09**, **y22** e **y35**; o predictor **x013** com **y06**, **y08**, **y19**, **y32** e **y38**. Já **x032** relaciona-se com **y42**, **y57** e **y64** enquanto que **x071** está associado a **y11**, **y24** e **y37**.

Este mesmo exercício de exploração de relações através dos factores latentes pode e deve ser estendido às restantes relações detectadas contudo este fastidioso trabalho será aqui encurtado por razões óbvias.

Vejamos agora o que se passa ao nível de PCA.

A análise PCA às matérias primas, às condições de fabrico, ao conjunto dos predictores e às respostas apenas são esclarecedoras quanto ao modo como a variabilidade interna de cada sub-espaço é descrita sobre as componentes principais.

Quanto às matérias primas, sobre a primeira componentes principal (PC1) surgem representadas as variáveis x002, x003, x004, x005, x006, x007, x008, x009, x010, x013, x014 e x015. Através da análise de correlação destacam-se as relações x002-x007 (0.818), x010-x015 (0.719), x009-x014 (0.687), x008-x013 (0.677), x006-x010 (0.553), x004-x014 (0.541), x004-x009 (0.540) e x006-x015 (0.522). Daqui se conclui que a componente PC1 está a dar indicação sobre estas relações internas de variáveis já que somente as variáveis x003, x005 e x008 não surgem interligadas.

Passando ao sub-espaço do processo (variáveis x016 a x144), resultados similares são observados – sobre a componente mais relevantes expressam-se 51 variáveis (x040, x041, x047, x051, x052, x054, x056, x058, x062, x065, x068, x072, x074, x075, x080, x081, x092, x093, x094, x095, x096, x097, x098, x099, x100, x101, x103, x106, x107, x110, x111, x112, x114, x116, x118, x120, x121, x122, x123, x124, x125, x126, x127, x128, x129, x131, x133, x137, x139, x141 e x142).

Inspeccionando as correlações mais significativas neste sub-espaço identificam-se elevadas correlações, com valores a rondar 0.999. Por exemplo, a variável x041 vem fortemente relacionada com quase todas as variáveis anteriores à excepção de x058 e x082.

No sub-espaço das respostas, a primeira componente principal transporta informação referente a 23

variáveis (y01, y02, y05, y06, y07, y08, y09, y12, y14, y15, y18, y19, y20, y21, y25, y27, y28, y31, y32, y33, y34, y35 e y38).

Neste sub-espço encontram-se também correlações muito elevadas (0.910 a 0.999). Se atentar por exemplo na primeira resposta (y01) são encontradas relações significativas (0.500 a 0.988) com todas as respostas evidenciadas sobre PC1 à excepção de y07, y20 e y33.

Embora estas análises PCA parciais não permitam obter inter-relações entre sub-espços elas são reveladoras, de uma forma simples, das eventuais inter-relações dentro do mesmo sub-espço o que é uma enorme vantagem já que reduz drasticamente o esforço para localizar interdependências.

Passemos agora à análise PCA globalizada, de onde contamos obter, de igual modo, a informação que relaciona predictores com respostas.

A primeira componente principal transporta informação relevante exclusiva ao processo sendo evidenciadas 54 variáveis (x040, x041, x047, x049, x050, x051, x052, x054, x056, x058, x062, x065, x068, x072, x074, x075, x077, x080, x081, x092, x093, x094, x095, x096, x097, x098, x099, x100, x101, x103, x106, x107, x110, x111, x112, x114, x116, x118, x120, x121, x122, x123, x124, x125, x126, x127, x128, x129, x131, x133, x137, x139, x141 e x142). Este conjunto inclui todas as variáveis evidenciadas no estudo PCA sobre o processo e acrescenta ainda as contribuições x044, x050 e x077.

Sobre **PC2** já se obtém informação referente a predictores e respostas – manifestam-se 59 variáveis das quais 7 referem-se às matérias primas (x002, x003, x005, x007, x008, x013 e x015), 25 ao processo (x016, x020, x022, x024, x029, x031, x032, x037, x043, x046, x055, x057, x063, x067, x071, x076, x082, x083, x085, x086, x087, x108, x113, x134 e x140) e 27 respostas (y01, y02, y05, y06, y08, y09, y10, y12, y13, y14, y18, y19, y21, y22, y23, y25, y27, y28, y31, y32, y34, y35, y36, y38, y48, y49 e **y53**).

Se se atender os resultados de PLS, o factor latente mais significativo **FL₃** relaciona 35 predictores (x006, x010, X011, x016, x020, x022, x024, x029, x031, x032, x037, x043, x046, x051, x053, x055, x057, x063, x065, x067, x071, x073, x078, x082, x083, x085, x086, x087, x089, x108, x113, x132, x134, x136 e x140) com 19 respostas (y07, y08, y11, y20, y21, y24, y33, y34, y37, y40, y41, y43, y47, y48, y49, **y53**, y57, y58 e y64).

Comparando a informação **PC2** e **FL₃** a componente principal difere do factor latente nos predictores referentes às matérias primas, coincide em grande parte dos predictores referentes ao processo (FL₃ apresenta os predictores x051, x053, x065, x073, x078, x089, x132, x136 que não constam em PC2; PC2 apresenta a variável adicional x076) mas difere bastante quanto às respostas (só coincidem em 6 variáveis).

Importa agora confrontar estes resultados PCA e PLS com as indicações pontuais manifestadas pelas correlações.

Anteriormente já tínhamos verificado a interdependência da resposta **y53** com 15 variáveis do processo de transesterificação: **x016, x020, x022, x024, x029, x031, x032, x037, x046, x055, x057, x067, x082, x113 e x140**. Estas relações estão ambas manifestadas em PCA (PC2) e em PLS (FL₃).

Passando ao grupo de correlações menos expressivas mas ainda assim relevantes, os predictores x002, x007, x013, x032 e x071 estão relacionados com as respostas y06, y08, y09, y10, y11, y19, y22, y23, y24, y32, y35, y37, y38, y41, y57 e y64.

Sobre PC2 encontram-se representadas os 5 predictores mas apenas 10 destas 16 respostas.

Sobre FL₃ só se encontram expressas x032 e x071 - x032 está correlacionada com y41, y53, y57 e y64 e surgem sobre o mesmo factor latente; x071 está correlacionada com y11, y24 e y37 que também se manifestam sobre este factor latente.

Os restantes predictores em falta x002, x007 e x013 surgem representados sobre o segundo factor latente mais relevante (FL₂) – a correlação entre x002 e y09, y10, y23 e y35 bem como as relações de x007 com y09, y22 e y35 e de x013 com y06, y08, y19, y32 e y38 surgem todas expressas nas componentes mais relevantes de FL₂ revelando uma concordância de 100% dos resultados da correlação com as indicações de PLS.

4.4.4. Análise do processo

Neste capítulo descrevem-se as interdependências do processo relacionando as relações encontradas por correlação definindo cada uma das variáveis e avançando razões / causas para o que foi observado.

Um dos dois métodos de análise exploratória mais utilizado é a análise de componentes principais (PCA). Tal como já foi referido, o objectivo da análise por PCA é reduzir a dimensionalidade do conjunto de dados, preservando ao mesmo tempo o máximo de informação. Isto é feito calculando-se combinações lineares das variáveis originais. A primeira componente principal PC1, é uma combinação linear de máxima variância, isto é, máxima informação. A PC2 é também de máxima variância, no entanto é ortogonal a PC1.

A razão da utilização do PLS tem vantagens inerentes ao uso das técnicas multivariadas, como o desenvolvimento de modelos que prevêm a detecção de “outliers”. Estes são tão importantes quanto a determinação do número de variáveis latentes.

Da análise dos resultados obtidos, há algumas relações que foram obtidas que são susceptíveis de serem analisadas em pormenor.

Para avaliar o efeito concreto de cada interdependência convém passar a uma análise mais minuciosa obtida através das correlações – esta segunda fase muito mais facilitada já que foi direccionada pela abordagem PLS prévia.

4.4.4.1. Análise da interdependência da resposta y53 com 15 variáveis do processo de transesterificação

A resposta **y53**, com a qual se pretende analisar a interdependência com 15 variáveis do processo, diz respeito ao teor de triglicéridos no biodiesel analisado para lote de expedição.

A variável **x016** (B0614.02; H₃PO₄; L614402 %) diz respeito ao nível do tanque pulmão de ácido fosfórico. Este tanque tem um tempo de “refill” aproximadamente meio dia o que significa que, usando *logsheets* de um dia os resultados vêm adulterados. Dada esta constatação o ideal seria tentar correlacionar com a variável mais próxima (FIC614302) o caudalímetro ou então com o totalizador (F614302). Uma outra variável que pode evidenciar uma correlação importante, é o nível do tanque de armazenagem do ácido fosfórico (L972421).

A quantidade de ácido fosfórico é importante pois é com este ácido que se consegue uma desgomagem química eliminando assim gomas que influenciam o processo de transesterificação e por consequência a respectiva reacção, tal como foi referido na introdução.

A variável **x020** (B0614.03-NaOH; L614403 %) indica o nível do tanque pulmão de NaOH. Este tanque tem um tempo de “refill” aproximadamente de duas horas o que significa que, usando *logsheets* de um dia os resultados vêm adulterados. O ideal também seria tentar correlacionar com a

variável mais próxima (FIC614303) o caudalímetro ou então com o totalizador (F614303). Para este caso a soda tem especial importância pois neutraliza o ácido fosfórico que se tem de introduzir em excesso para que a reacção se dê e ainda é introduzida para eliminar a acidez livre no óleo. Se esta neutralização não se der convenientemente e o óleo for com acidez para a transesterificação, reagem com o catalisador e por sua vez a reacção não é tão extensa.

A variável **x022** (NaOH dil. Water total m³) é a água de diluição que irá colocar a concentração de soda nos valores desejáveis, no entanto, dado que a quantidade de água é necessária para existir uma boa aglomeração esta é importante pois é aproximadamente duas vezes quando comparada com a água de aglomeração.

Assim, a variável **x024** (aggl. Water total m³) que é água de aglomeração, rectifica a necessidade de água em falta para a aglomeração. Tanto o **x22** como o **x24** vão ajudar na aglomeração das gomas e sabões. Quanto melhor decorrer este processo, melhor óleo neutro e desgomado teremos e por conseguinte melhor transesterificação/reacção.

A variável **x029** (washing water to R0614.04 total m³) corresponde à água que vai remover os sabões solúveis no óleo após a primeira separação das gomas. Esta água vai remover os sabões presentes no óleo.

A variável **x031** (ácido cítrico cons. Total L) é apenas um contador que tem em conta o ácido cítrico para a neutralização. O ácido cítrico promove uma diminuição de pH e devido à sua estrutura orgânica consegue com que seja mais fácil a remoção de sabões e algumas gomas que não saíram na primeira separação. Estas gomas e sabões seriam problemáticas para a transesterificação uma vez que podem provocar emulsões. Por outro ponto de vista, como evita a emulsão entre óleo e água de lavagem, a secagem dá-se com maior facilidade. Se assim não acontecer, o óleo leva maior humidade e ao entrar na transesterificação vai reagir com o catalisador e óleo formando sabões.

Quanto a **x032** (B0614.06 – L614407 %) indica o nível do tanque pulmão de cítrico. Este tanque tem um tempo de “refill” aproximadamente de três horas o que significa que, usando *logsheets* de um dia os resultados vêm adulterados. Deveria tentar correlacionar-se com a variável mais próxima (FIC614309) o caudalímetro ou então com o totalizador (F614309).

A variável **x037** (refined oil from deg. Total (ton)) trata-se do óleo que sai da neutralização e desgomagem. Por sua vez não se deveria analisar este nível de tanque, mas sim o totalizador colocado imediatamente após o tanque de armazenagem do óleo mas que não é visível nos *logsheets* que são impressos, pelo que não se teve acesso a esta informação.

Esta relação aparentemente não faz qualquer sentido com os triglicéridos do Biodiesel do lote.

No entanto, dado que em 2009 a Iberol não neutralizava para terceiros, a quantidade de saída deveria ser idêntica à de entrada na transesterificação. Se a quantidade à entrada da transesterificação estiver também relacionada, pode-se referir que um menor caudal faz com que o tempo de residência seja maior e daí o tempo de reacção aumenta pelo que o teor de triglicéridos diminui.

A variável **x046** (Na-Methylat – cons I – total em Ton) é um totalizador de quantidade de catalisador

na primeira reacção logo influencia directamente a velocidade de reacção, **x055** (Na-Methylat – com II – Total em Ton) é um totalizador de quantidade de catalisador na segunda reacção logo influencia directamente a velocidade de reacção e **x057** (Methanol cons. II – Total em Ton) o excesso de metanol aumenta também a velocidade de reacção (Princípio de Le Chatelier) com maior efeito na segunda reacção e não tanto na primeira. Verifica-se de uma análise do processo que **x046** se relaciona com o primeiro par de colunas da transesterificação e **x057** se relaciona com o segundo par de colunas da transesterificação o que implica que sempre que se verifica um aumento no gasto de catalisador, significa que essa informação foi dado no sistema de automatização por um operador para compensar uma anomalia no processo, isto é, no caso de haver um aumento no teor de glicéridos do biodiesel, essa informação é dado após análise no laboratório à UPB e cabe aos operadores na sala de comandos da UPB fazer os ajustes que considerem adequados para corrigir os glicéridos, nomeadamente dando a indicação de aumento do gasto de catalisador. No entanto, é sempre opção prioritária um aumento do metanol e não de catalisador. O metanol faz deslocar a reacção no sentido da formação de produtos e isso leva à correcção da anomalia mencionada anteriormente. As adições de metanol compensatórias são realizadas ao nível do 2º par de colunas da transesterificação.

A variável **x067** (Process water II (S3) Total m³) é a água utilizada para lavagem do biodiesel na 1ª etapa. Trata-se de uma água que irá remover metanol, catalisador, glicerol livre e outros contaminantes. Se a lavagem for mal efectuada pode existir deslocação do equilíbrio para os reagentes, isto é, se o glicerol não for removido, ao secarmos, a falta de metanol, o excesso de biodiesel e a elevada temperatura pode promover a reacção inversa da reacção esquematizada na equação seguinte,



Equação 4.1 – Reacção química representativa da produção de biodiesel.

A variável **x082** (B091201 meth. GW – L 912401 %) este nível, é o do tanque de reacção do desdobraimento dos ácidos gordos. É suposto ser constante uma vez que a transferência para a operação unitária seguinte é feita por gravidade (transbordo), esta relação não faz qualquer sentido, pois o nível deste tanque em nada influencia os triglicéridos do Biodiesel.

A variável **x113** (Methanol to tank farm – total em Ton) este metanol é o recuperado. Podemos ter dois pontos de vista:

- Em primeiro lugar uma maior recuperação implicou uma maior introdução de metanol para deslocar a reacção, o que poderá significar que anteriormente já terá havido uma tentativa de correcção por parte do operador no sentido de diminuir do glicéridos através da adição de metanol que constitui sempre uma opção prioritária em relação ao gasto de catalisador
- Em segundo lugar se a taxa de recuperação do metanol for elevada pode significar que o processo não apresenta problemas, correndo a reacção normalmente, pois o metanol desloca a reacção completamente no sentido da formação dos produtos e assim sendo a recuperação será

considerável.

No entanto é bom de referir que em 2009 aconteceu uma situação que se pode ter reflectido nos dados adquiridos e que pode eventualmente estar expressa nesta situação. O enchimento da coluna de metanol está relacionada com a eficiência da coluna e por conseguinte com a recuperação do metanol. Como a dada altura em 2009 o enchimento da coluna de metanol teve um problema, este foi completamente substituído. Assim, efectuou-se a substituição do enchimento da coluna de metanol por um outro tipo de enchimento com menor eficiência e que se refletiu numa menor % de recuperação de metanol e que não teve necessariamente relação directa com algum problema no produto final que fosse necessário corrigir.

A variável **x140** (Glycerine to tank farm – total Ton) a quantidade de glicerina produzida é uma medida directa da extensão da reacção, no entanto, é necessário ter em atenção que a humidade não é constante e pode influenciar o resultado. Esta humidade varia consoante tenhamos mais ou menos condições na parte de secagem da glicerina.

4.4.4.2. Relação entre x071 e y11, y24 e y37

A variável x071 refere-se à água de lavagem na transesterificação e relaciona-se com as respostas y11 (CP do 1º turno de biodiesel), y24 (CP do 2º turno de biodiesel) e y37 (CP do 3º turno de biodiesel).

A temperatura da água de lavagem (**x071**) é importante pois previne a emulsão e cria condições para que haja uma maior solubilidade dos contaminantes do biodiesel na parte aquosa. Se os contaminantes não forem removidos, estes podem tornar-se núcleos de cristalização e influenciar o CP (H₂O, glicerol livre, contaminação e/ou esteróis glucosídicos que normalmente se deslocam para a interface).

4.4.4.3. Relação entre x002 e y09, y10, y23 e y35

A variável x002 diz respeito à acidez do óleo desgomado do 1º turno e relaciona-se com o teor de água do 1º (y09) e 3º (y35) turnos de biodiesel, bem como, com a contaminação do 1º (y10) e 2º (y23) turnos de biodiesel.

A variável **x002** indica que a acidez de entrada da desgomagem tem influência na humidade do biodiesel e ainda na contaminação. Se realmente tivermos uma acidez muito elevada, a desgomagem e neutralização funciona mal, ficando um óleo neutro com acidez elevada, e sabões dissolvidos.

Isto acaba por destabilizar a transesterificação que por sua vez pode-se reflectir na secagem do biodiesel. Esta normalmente é afectada quando se introduz mais metanol necessário a compensar a fraca extensão da reacção. Este metanol acaba por se acumular no sistema e uma vez que as águas são reaproveitadas, afecta o sistema de vácuo que por sua vez não permite uma secagem eficiente.

Quanto à contaminação, se maiores contaminantes entram na transesterificação, a probabilidade de

aumentar a contaminação à saída pode-se tornar evidente. Quanto melhor correr a desgomagem, menor contaminação teremos à saída (antigamente neutralizava-se a oleína de palma RBD, mesmo esta já sendo neutra, para diminuir a contaminação final).

4.4.4.4. Relação entre x007 e y09, y22 e y35

A variável x007 é a acidez do óleo desgomado do 2º turno e relaciona-se com o teor de água do 1º (y09), 2º (y22) e 3º (y35) turnos de biodiesel.

Os motivos são os mesmos dos referidos no ponto 5.4 a acidez de entrada da desgomagem tem influência na humidade do biodiesel.

4.4.4.5. Relação de x013 com y06, y08, y19, y32 e y38

A variável x013 é a acidez do óleo neutro do 3º turno e relaciona-se com o teor de ésteres metílicos do 1º (y06), 2º (y19) e 3º (y32) turnos de biodiesel, bem como com o índice de iodo do 1º turno (y08) e teor de água na glicerina do 3º turno de biodiesel.

A acidez do óleo neutro relaciona-se com o teor de ésteres metílicos uma vez que este é o óleo de entrada na transesterificação e se tiver uma maior acidez vai destabilizar a transesterificação e consequentemente em vez de termos um teor elevado de ésteres metílicos (FAME), isto é, Biodiesel (MES \geq 96.5%) não o teremos.

Quanto à relação da acidez do óleo neutro com o índice de iodo, não parece haver uma explicação plausível uma vez que o índice de iodo que consiste no número de insaturações das cadeias dos ácidos gordos constituintes do FAME obtido, é directamente proporcional ao índice de iodo dos óleos usados na transesterificação, como tal a acidez interfere com outros parâmetros que não o índice de iodo.

A variável **x013** relaciona-se com **y38** na medida em que se a acidez à entrada for muito elevada, significa que teremos muitos sabões na transesterificação, o que implica maiores quantidades de água no processo para controlar a operação e por isso um maior caudal à entrada da etapa da secagem da glicerina. Visto esta etapa ter limitações no caudal à entrada para que a humidade seja reduzida, ao aumentarmos o caudal a tratar, implica um aumento na humidade à saída pois seca pior.

5. CONCLUSÕES

5 CONCLUSÕES

Este trabalho baseou-se na análise multivariada do diagnóstico da produção de biocombustíveis.

Como ponto de partida houve a necessidade de elaborar uma estratégia para a abordagem ao trabalho que se viria a realizar.

Assim, analisou-se o que se queria saber, quais as barreiras existentes para a realização deste trabalho, quais os factores preponderantes a ter em consideração e quais os verdadeiros objectivos do trabalho.

Dos objectivos seleccionaram-se três abordagens: como tornar a produção de Biodiesel cada vez mais produtora, quais os elementos que permitem/prejudicam a produção do Biodiesel e quais os elementos/factores/variáveis que favorecem o processo produtivo.

Neste trabalho houve algumas limitações relativamente à recolha de dados, isto é, problemas que tiveram que ser contornados de modo a garantir a rastreabilidade, recolha e relações obtidas no final.

A matriz de dados que foi obtida e que serviu de ponto de partida para a realização deste trabalho, continha informações relativas à matéria prima, dados da produção, dados da análise do produto obtido em laboração contínua, ou seja, análise de controlo diário, e por fim informação relativa à análise do produto final (Biodiesel).

Relativamente à informação sobre o produto final, este diz respeito à informação recolhida para o produto pronto para expedição e que no fim se caracteriza de produto conforme ou produto não conforme.

A rastreabilidade conseguiu-se por associação dos dados com o historial do processo produtivo e análise em laboratório relativos ao ano de 2009.

Da matéria-prima tinha-se informação diária em triplicado relativa ao controlo diário de três turnos de laboração: 1º Turno (08h-16h); 2º Turno (16h-00h) e 3º Turno (00h-08h, do dia seguinte).

Da análise do produto e sub-produtos da produção diária do biodiesel foram recolhidas também informações em triplicado relativas aos três turnos de produção 1º, 2º e 3º Turnos, que são analisados diariamente em laboratório.

Relativamente à informação recolhida para as variáveis do processo produtivo, foi tida em consideração a informação recolhida dos “*logsheets*” diários, um por cada dia de laboração, e associados às respectivas análises de produto final (três turnos de laboração).

No que respeita à informação relativa à análise completa do produto final – lote para expedição – a informação foi extrapolada para os dias de laboração que a antecederam e que são a imagem daquele produto final analisado combinando toda a informação anterior, uma vez que um lote de expedição corresponde a aproximadamente seis turnos de laboração.

De notar que foram tidas em consideração todas as paragens de laboração e as associações efectuadas com a maior credibilidade possível.

Após o tratamento dos dados recolhidos e depois de devidamente organizados, procedeu-se ao estudo por tratamento estatístico que fundamenta este estudo/trabalho.

No início deste estudo foram removidas as seguintes variáveis qualitativas, glicerol livre dos três turnos de laboração, cinzas sulfatadas do lote de expedição, ponto de inflamação, glicerol livre do lote, fósforo e metais analisados por ICP bem como a corrosão à lâmina de cobre.

Da remoção de dados resultou uma redução em cerca de 53.6 % da informação recolhida e que não foi processada.

Considerámos que dos predictores (x) fazem parte toda a informação relativa à matéria-prima, bem como todos os dados relativos ao processo produtivo. As respostas (y) dizem respeito a todos os parâmetros de análise de produto final (análise completa de lote de expedição).

A análise da correlação dá uma preciosa indicação directa entre a interacção de duas variáveis. Contudo, no caso da análise de sistemas multivariados esta tarefa é demasiado ingrata por ser difícil de realizar.

A análise PCA globalizada, incluindo os predictores e as respostas em simultâneo, pode eventualmente conseguir evidenciar algumas das relações mais evidentes entre variáveis se nos focarmos nos “loads” mais significativos que relacionam os dois sub-espacos. Contudo a informação que cada componente principal transporta não está somente relacionada com interacções mas com a descrição da variabilidade do sistema, razão pela qual esta abordagem é pouco selectiva.

A análise PLS além de ser a mais correcta e eficaz a evidenciar estas interacções entre sub-espacos de predictores e respostas, é muito vantajosa – com poucos factores latentes consegue extrair a informação mais relevante do sistema expondo, de uma forma global as interdependências entre os sub-espacos.

Contudo, os factores latentes exibem esta informação de uma forma globalizada não permitindo uma fácil relação pontual.

Da análise de algumas correlações e interdependências obtidas, chegou-se a conclusões que vão de encontro às expectativas e suspeitas que existem relativamente ao processo produtivo e que vêm confirmar factos concretos.

No entanto, foram obtidas correlações e interdependências que em nada se relacionam entre si, isto é, apesar de estatisticamente se relacionarem, de facto, no processo produtivo não são dependentes nem existe uma relação factual entre elas.

Tal constatação pode dever-se ao facto de existirem situações pontuais na matriz de dados que levaram ao estabelecimento de relações não coerentes.

Este trabalho foi em tudo de difícil abordagem e por vezes ingrato, uma vez que em nada é fácil abordar um processo fabril de laboração contínua num regime de 24 horas (um dia) sem acompanhamento presencial e sem registo das situações mais penosas para a análise multivariada.

Ficou comprometido um dos objectivos iniciais do trabalho que consistia na identificação dos factores e elementos que podem tornar o processo produtivo mais ou menos produtor.

Deste trabalho as conclusões que foram obtidas ficaram patentes no ponto 4.4. Discussão, em particular no ponto 4.4.4. Análise do processo, e que são constatações processuais.

Num comentário final, fica uma sugestão que se baseia no facto de que um trabalho deste género é

bastante complexo e de interessante desenvolvimento, mas que num outro contexto deverá ser executado com mais minúcia.

Creio que seria de interesse trabalhar melhor os dados obtidos, para num futuro, tentar estabelecer correlações entre respostas e correlações entre variáveis, separadamente.

Há relações muito interessantes e que não foram aqui abordadas por motivos já referidos anteriormente.

BIBLIOGRAFIA

Bibliografia

- [1] Uva, M., Relatório de estágio – Utilização de NIR na caracterização analítica de biodiesel, 2005.
- [2] Baptista, P., Relatório de estágio - Utilização de NIR na caracterização analítica de biodiesel, 2006.
- [3] Pedro Felizardo, Patricia Baptista, Margarida Sousa Uva, Jose C. Menezes, M. Joana Neiva Correia, J. Near Infrared Spectroscopy, 15, 2007, 97-105.
- [4] P. Felizardo, Patricia Baptista, Jose C. Menezes, M. Joana Neiva Correia, Multivariate near infrared spectroscopy models for predicting methanol and water content in biodiesel, Anal. Chim. Acta, 2007, doi:10.1016/j.aca.2007.02.05
- [5] Knothe, G., Dunn, R., Biodiesel, The use of Vegetable Oils and their derivatives as Alternative Diesel Fuels, National Center for Agricultural Utilization Research, U.S. Department of Agriculture.
- [6] Khan, A., Research into Biodiesel Kinetics & Catalyst Development; Brisbane – Australia, 2002.
- [7] Oils and Fats International, Julho de 2002.
- [8] Felizardo, P., Relatório de estágio - Produção de biodiesel a partir de óleos usados de fritura, 2003.
- [9] Cankci, M., VanGerpen, J., Biodiesel production via acid catalysis, Trans ASAE 42 (5), 1999, 1203-1210.
- [10] Nelson, L., Foglia, T., Lipase catalysed production of biodiesel, JAOCS 73, 1996, (8).
- [11] Abreu, F. et al, Utilization of metal complexes as catalysts in the transesterification, Journal of Molecular Catalysis A, Chemical 209, 2004, 29–33.
- [12] Bournay, L. et al, New heterogeneous process for biodiesel production: A way to improve the quality and the value of the crude glycerine produced by biodiesel plants, Catalysis Today 106, 2005, 190–192.
- [13] Knothe, G., VanGerpen, J., Krahl, J., The Biodiesel Handbook, AOCS Press, 2005.
- [14] VanGerpen, J., Biodiesel processing and production, Fuel Processing Technology 86, 2005, 1097-1107.
- [15] NQ.10/06/GQA/Produção.Biodiesel - Iberol
- [16] The Non Technical Barriers Network; Agence de l'Environnement et de la Maîtrise de l'Energie, Franca,1998.
- [17] Boletins do National Biodiesel Board; consultados na Internet em 21-07-2011.
- [18] Biocombustíveis, Conclusões do Fórum Energias Renováveis em Portugal, 2004.
- [19] Bozbas, K, Biodiesel as an alternative motor fuel: Production and policies in the European Union, Renewable and Sustainable Energy Reviews, 2005, 1–12.
- [20] Site da Internet www.biodiesel.org consultado em 08-06-2011.
- [21] Bondioli, P., Mariani, C., Vegetable oils derivatives as diesel fuel substitutes, analytical aspects – Determination of Methanol, Riv. Ital. Sostanze Grasse 69, 1992, 467-469.
- [22] Plank, C. et al, Simultaneous determination of glicerol and mono-, di- and Triglycerides in Vegetable oil methyl esters, J. Chromatography A, 697, 1995, 461-468.

- [23] Holeapek, M. et al, Analytical Monitoring of the production of biodiesel by HPLC with various detection methods, *J. Chromatography A*; 858, 1999, 13-31.
- [24] Gelbard, G. et al, ¹H Nuclear Magnetic Resonance Determination of the yield of the transesterification of rapeseed oil with methanol, *J. Am. Oil Chem. Soc.*; 72, 1995, 1239-1241.
- [25] Dimming, T. et al, ¹³C-NMR-Spektroskopie zur Bestimmung von Umsatz and ReaktionsKinetik der Umesterung von Triglyceriden zu Methylestern, *Chem. Tech.*; 51, 1999, 326-329.
- [26] Rodrigues, L., Marques, J., Cardoso, J., Menezes, J., Monitoring an Industrial Batch Crystallization Process by In-Situ Near Infrared Spectroscopy.
- [27] Reich, G., *Advanced Drug Delivery Reviews*, 57, 2005, 1109-1143.
- [28] Naes, T., Isakson, T., Fearn, T., Davies, T., *Multivariate Calibration and Classification*, NIR Publications, Chichester, UK, 2002;.
- [29] Otto, M., *Chemometrics*, Wiley-VCH, 1999.
- [30] Brown, S., *Chemical systems under indirect observation: latent properties and chemometrics*, *Focus on Analytical Spectrometry*, 208-225.
- [31] Knothe, G., *American Society of Agricultural Engineers*, Vol. 44(2), 2001, 193-200.
- [32] Knothe, G., *J. Am. Oil Chem. Soc.*, 78, no. 10, 2001.
- [33] Pimentel, M. et al, *Microchemical Journal*, Article in Press, 2006.
- [34] Knothe, G., Rapid Monitoring of Transesterification and Assessing Biodiesel Fuel Quality by Near-Infrared Spectroscopy Using a Fiber-Optic Probe, *J. Am. Oil Chem. Soc.* 76 (7), 1999, 795–800.
- [35] Guillen, M., *J. Sci. Food. Agric.*, 75, 1997, 1-11.
- [36] Y. B. Che Man, *Food Chemistry*, 67, 1999, 193-198.
- [37] Christy, A., *Chemometrics and Intelligent Laboratory Systems*, 82, 2006, 130-136.
- [38] Y. B. Che Man, *Food Chemistry*, 66, 1999, 109-114.
- [39] Hongfu, Y., *Fuel*, 85, 2006, 1720-1728.
- [40] Pimentel, M., *Microchemical Journal*, 82, 2006, 201-206.
- [41] Zagonel, G., *Talanta*, 63, 2004, 1021-1025
- [42] Adaptado de EN ISO 3675.
- [43] Analytix, *Advances in Analytical Chemistry*, Sigma-Aldrich, 3, 2000.
- [44] Adaptado de EN 14110.
- [45] Bondioli, P., Mariani, C., Fedeli, E., *La Rivista Italiana delle Sostanze Grasse*, Vol. LXIX; (1992) 467-469.
- [46] EN 14214, Automotive fuels – Fatty acid methyl esters (FAME) for diesel engines – Requirements and test methods, CEN – European Committee for Standardization, Brussels, Belgium.
- [47] Felicio, C., Bras, L., Lopes, J., Cabrita, L., Menezes, J., Comparison of PLS algorithms in gasoline and gas oil parameter monitoring with MIR and NIR, *Chemometrics and Intelligent Laboratory Systems*, 78,(2005), 74-80
- [48] P. Geladi, B. Kowalski, *Anal. Chim. Acta*, 185 (1986) 1-17

ANEXOS

Anexos

A1. Codificação das variáveis associadas às matérias primas

1º Turno Óleos				
Óleo desgomado		Óleo Neutro seco		
Humidade	Acidez	Acidez	Sabão	Humidade
x001	x002	x003	x004	x005
2º Turno Óleos				
Óleo desgomado		Óleo Neutro seco		
Humidade	Acidez	Acidez	Sabão	Humidade
x006	x007	x008	x009	x010
3º Turno Óleos				
Óleo desgomado		Óleo Neutro seco		
Humidade	Acidez	Acidez	Sabão	Humidade
x011	x012	x013	x014	x015

A2. Codificação das variáveis associadas ao processo de desgomagem e neutralização dos óleos (matérias primas) usados na reacção de transesterificação (produção de Biodiesel)

B0614.02 - H3PO4	B0972.04 - H3PO4	NaOH cons.		B0614.03-NaOH
L614402 %	L972421 %	F614303 (L/h)	total L	L614403 %
x016	x017	x018	x019	x020
NaOH dil. Water		aggl.water cons.		oil behind W0614.03
F614304 (L/h)	total m3	F614305 (L/h)	total m3	T614104(°C)
x021	x022	x023	x024	x025
wash water B0614.04		wash water to R0614.04		ác Citrico Cons.
L614404 %	T614108(°C)	F614306 (L/h)	total m3	F614309 (L/h)
x026	x027	x028	x029	x030
ác Citrico Cons.	B0614.06 . Ac Citrico	oil dryer T0614.01		
total L	L614407 %	L614408 %	T614105(°C)	P614204 (mbar)
x031	x032	x033	x034	x035
refined oil from deg.		oil behind W0614.04	oil in front of W0614.04	
F614310 (Kg/h)	total (ton)	T614106 °C	T614107 °C	
x036	x037	x038	x039	

A3. Codificação das variáveis do processo associadas à produção de Biodiesel (Etapa da Transesterificação)

Transesterification	B0911.08 - mixed oil	oil to Transesterification		B0911.06 Na-Methylat	B0911.07 Methanol	
	L911401 %	F911301 (Kg/h)	total ton	L911403 %	L911405 %	
	x040	x041	x042	x043	x044	
	Na-Methylat cons I		Methanol consumption I		oil behind W0911.01	
	F911303 (L/h)	total ton	F911304 (L/h)	total ton	T911103(°C)	
	x045	x046	x047	x048	x049	
	K0911.01	K0911.02	Oil in front of S0911.01	B0911.12 - recovery tank	Na-Methylat cons II	
	L911420 %	L911425 %	T911122(°C)	L911433 %	F911321 (L/h)	total ton
	x050	x051	x052	x053	x054	x055
	Methanol consumption II		oil behind W0911.04		K0911.03	K0911.04
	F911321 (L/h)	total ton	T911123(°C)	L911426 %	L911435 %	
	x056	x057	x058	x059	x060	
	Oil in front of S0911.02	process water III(S2)		B0911.05 collecting tank		
	T911127(°C)	F911324 (L/h)	total m3	L911431 %	T911128(°C)	
	x061	x062	x063	x064	x065	
	process water II(S3)		process water I(S4)		B0911.03 wash water tank	
	F911341 (L/h)	total m3	F911344 (L/h)	total m3	L911444 %	T911146(°C)
	x066	x067	x068	x069	x070	x071
	Oil in front of T0911.01	T0911.01 FAME dryer		Biodiesel to tankfarm		
	T911142(°C)	L911143 %	P911243 (mbar)	F911348 Kg/h	total ton	
x072	x073	x074	x075	x076		
B0911.101 feed tank clear		Biodiesel to S0911.105				
L9114101 (%)	T9111101(°C)	F9113101 (Kg/h)				
x077	x078	x079				

A4. Codificação das variáveis do processo associadas à produção de Biodiesel (Etapa da Recuperação do Metanol)

Methanol recovery	Meth.-GW to recovery	pH value meth. - GW	B0912.01 meth.GW		B0912.02 meth.GW	
	F912301 (Kg/h)	Q912501	L912401 %	T912101 (°C)	L912402 %	T912102 (°C)
	x080	x081	x082	x083	x084	x085
	B0912.02 - Fatty acid		K0912.02 - Fatty acid colum		meth. GW to K0912.01	
	L912403 %	T912103 °C	L912406 %	T912107 °C	P911205 mbar	F912302 (Kg/h)
	x086	x087	x088	x089	x090	x091
	meth. GW to W0912.01	K0912.01 methanol recovery				
	T912105 (°C)	T912111 (°C)	T912112 (°C)	T912113 (°C)	T912114 (°C)	T912115 (°C)
	x092	x093	x094	x095	x096	x097
	K0912.01 methanol recovery					
	T912116 (°C)	T912117 (°C)	T912118 (°C)	T912119 (°C)	T912120 (°C)	
	x098	x099	x100	x101	x102	
	waste air to K0912.03/04	waste air behind K0912.03/04		methanol to W0912.03	C. water to W0912.03	
	T912124 °C	P912210 mbar	P912211 mbar	T912123 (°C)	P912206 bar	T912127 °C
	x103	x104	x105	x106	x107	x108
	B0912.04 - Methanol	methanol back to K0912.01		methanol to tankfarm		
	L912408 %	F912304 (kg/h)	F912305 (kg/h)	T912126 (°C)	total ton	
	x109	x110	x111	x112	x113	
	methanol to tankfarm					
	steam to W0912.02	pH-value glycerine water		B0912.03-glycerine water		
P912204 bar	Q912502	L912405 %	T912122 °C			
x114	x115	x116	x117			

A5. Codificação das variáveis do processo associadas à produção de Biodiesel (Etapa da secagem da Glicerina)

glycerine water evaporation	glycerine to evaporation		steam to W0831.01
	F831301 (L/h)	total m3	P831201 (bar)
	x118	x119	x120
GW in front of W0831.04	GW in front of B0831.01	condensate	GW in front of B031.02
T831101 (°C)	T831102 (°C)	P831202 (bar)	T831103 (°C)
x121	x122	x123	x124
condensate	GW to W0831.03	GW in front of W0831.03	cond. In front of W0831.03
P831203 (bar)	T831104 (°C)	T831105 (°C)	T831106 (°C)
x125	x126	x127	x128
cond. In front of W0831.05	B0831.03 evaporator		waste air
T831110 (°C)	L831401 %	P831204 (mbar)	T831108 °C
x129	x130	x131	x132
waste air	CW in front of W0831.06	B0831.05 - condensate	
P831205 (mbar)	T831107 (°C)	L831403 %	T831111 °C
x133	x134	x135	x136
B0831.06 - Glycerine		glycerine to tankfarm	
L831404 %	T831112 °C	F831304 Kg/h	total ton
x137	x138	x139	x140
condensate to K0912.03/04		press top K0912.03/04	press bottom K0912.03/04
F831303 L/h	F831310 L/h	P912211 mbar	P912210 mbar
x141	x142	x143	x144

A6. Codificação das respostas obtidas para a análise do produto final do processo produtivo (Biodiesel)

Características Produto Final (Biodiesel)	1	1	1	1	1	1	1	1	1	1	1	1	1
	Biodiesel											Glicerina	Metanol
	Mono-	Di-	Tri-	G. livre	G. total	Ésteres	Ést 18:3	Ind. Iodo	Humidade	Contamin.	Cloud Point	Humidade	Humidade
	y01	y02	y03	y04	y05	y06	y07	y08	y09	y10	y11	y12	y13
	2	2	2	2	2	2	2	2	2	2	2	2	2
	Biodiesel											Glicerina	Metanol
	Mono-	Di-	Tri-	G. livre	G. total	Ésteres	Ést 18:3	Ind. Iodo	Humidade	Contamin.	Cloud Point	Humidade	Humidade
	y14	y15	y16	y17	y18	y19	y20	y21	y22	y23	y24	y25	y26
	3	3	3	3	3	3	3	3	3	3	3	3	3
	Biodiesel											Glicerina	Metanol
	Mono-	Di-	Tri-	G. livre	G. total	Ésteres	Ést 18:3	Ind. Iodo	Humidade	Contamin.	Cloud Point	Humidade	Humidade
	y27	y28	y29	y30	y31	y32	y33	y34	y35	y36	y37	y38	y39
	Análise Produto final - Expedição												
	Índice Cetan	Enxofre	Cinzas Sulf	Densidade	Viscosidade	P. Inflama.	Cont. total	I. iodo Ref	I. iodo Exp.	Estabilidade	Metanol	Mono	Di
	y40	y41	y42	y43	y44	y45	y46	y47	y48	y49	y50	y51	y52
Análise Produto final - Expedição													
Tri	Glicerol L	Glicerol T	T. ésteres	ster met.ác.l	CFPP	Fósforo	Sódio	Cálcio	Magnésio	Potássio	CP	Corrosão	
y53	y54	y55	y56	y57	y58	y59	y60	y61	y62	y63	y64	y65	
Análise Produto final - Expedição													
Res Carb													
y66													