



**THIAGO CARVALHO DE AVILA**

**ESPECTROSCOPIA RAMAN E QUIMIOMETRIA COMO FERRAMENTAS NO  
MONITORAMENTO *ON-LINE* DO PROCESSO FERMENTATIVO DA GLICOSE  
PELA *Saccharomyces cerevisiae***

**CAMPINAS  
2013**





**UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE QUÍMICA**

**THIAGO CARVALHO DE AVILA**

**ESPECTROSCOPIA RAMAN E QUIMIOMETRIA COMO FERRAMENTAS NO  
MONITORAMENTO *ON-LINE* DO PROCESSO FERMENTATIVO DA GLICOSE  
PELA *Saccharomyces cerevisiae***

**ORIENTADOR: PROF. DR. RONEI JESUS POPPI**

**DISSERTAÇÃO DE MESTRADO APRESENTADA  
AO INSTITUTO DE QUÍMICA DA UNICAMP PARA  
OBTENÇÃO DO TÍTULO DE MESTRE EM QUÍMICA NA  
ÁREA DE QUÍMICA ANALÍTICA**

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA  
POR THIAGO CARVALHO DE AVILA, E ORIENTADO PELO PROF.DR. RONEI JESUS POPPI.**

---

**Assinatura do Orientador**

**CAMPINAS  
2013**

FICHA CATALOGRÁFICA ELABORADA POR SIMONE LUCAS - CRB8/8144 -  
BIBLIOTECA DO INSTITUTO DE QUÍMICA DA UNICAMP

Av55e Avila, Thiago Carvalho de (1985-).  
Espectroscopia Raman e quimiometria como  
ferramentas no monitoramento *on-line* do processo  
fermentativo da glicose pela *Saccharomyces cerevisiae* /  
Thiago Carvalho de Avila. – Campinas, SP: [s.n.], 2013.

Orientador: Ronei Jesus Poppi.

Dissertação (mestrado) - Universidade Estadual de  
Campinas, Instituto de Química.

1. Fermentação - Monitoramento *on-line*. 2. Raman,  
espectroscopia. 3. Cartas de controle multivariadas.  
4. Mínimos quadrados parciais. 5. Análise de  
componentes principais multimodo. I. Poppi, Ronei  
Jesus. II. Universidade Estadual de Campinas. Instituto  
de Química. III. Título.

Informações para Biblioteca Digital

**Título em inglês:** Raman spectroscopy and chemometrics for *on-line* monitoring of glucose fermentation by *Saccharomyces cerevisiae*

**Palavras-chave em inglês:**

Fermentation - *on-line* monitoring  
Raman spectroscopy  
Multivariate control charts  
Partial least squares  
Multi-way principal component analysis

**Área de concentração:** Química Analítica

**Titulação:** Mestre em Química na área de Química Analítica

**Banca examinadora:**

Ronei Jesus Poppi [Orientador]  
Francisco Benedito Teixeira Pessine  
Edenir Rodrigues Pereira Filho

**Data de defesa:** 25/02/2013

**Programa de pós-graduação:** Química



*Dedico este trabalho em especial*

*a meus pais **FÁBIO C. DE ÁVILA** e **MARIA E. M. ÁVILA***

*e a meus irmãos **FÁBIO JR. DE ÁVILA** e **LUIZ O. DE ÁVILA**,*

*...  
dedico este trabalho também àquelas pessoas que sempre trazem  
**FELICIDADE**,*

*...  
àquelas pessoas que sempre que se fazem **FAMILIARES**,*

*...  
aos **AMIGOS** que sempre estão por perto, principalmente nas horas difíceis,*

*...  
às **PESSOAS QUE NÃO SE FAZEM MAIS PRESENTES** fisicamente, mais que  
sempre torceram e torcem por mim de onde estão,*

*...  
às pessoas que **AMO**,*

*...  
às pessoas que mesmo de longe sempre mandam mensagens de **OTIMISMO E  
CARINHO**,*

*...  
E enfim, às pessoas que ajudam ou ajudaram a construir **MINHA HISTÓRIA**.*



**“POSSO TER DEFEITOS, VIVER ANSIOSO E FICAR IRRITADO ALGUMAS VEZES, MAS NÃO ESQUEÇO DE QUE MINHA VIDA É A MAIOR EMPRESA DO MUNDO. E QUE POSSO EVITAR QUE ELA VÁ À FALÊNCIA.**

**SER FELIZ É RECONHECER QUE VALE A PENA VIVER, APESAR DE TODOS OS DESAFIOS, INCOMPREENSÕES E PERÍODOS DE CRISE.**

**SER FELIZ É DEIXAR DE SER VÍTIMA DOS PROBLEMAS E SE TORNAR AUTOR DA PRÓPRIA HISTÓRIA.**

**É ATRAVESSAR DESERTOS FORA DE SI, MAS SER CAPAZ DE ENCONTRAR UM OÁSIS NO RECÔNITO DA SUA ALMA.**

**É AGRADECER A DEUS A CADA MANHÃ PELO MILAGRE DA VIDA.**

**SER FELIZ É NÃO TER MEDO DOS PRÓPRIOS SENTIMENTOS. É SABER FALAR DE SI MESMO. É TER CORAGEM PARA OUVIR UM “NÃO”. É TER SEGURANÇA PARA RECEBER UMA CRÍTICA, MESMO QUE INJUSTA.”**

**DEZ LEIS PARA SER FELIZ, EDITORA SEXTANTE, 2003.**

**AUGUSTO CURY**



## À DESCOBERTA DO AMOR

“Ensaia um sorriso  
e oferece-o a quem não teve nenhum.

Agarra um raio de sol  
e desprende-o onde houver noite.

Descobre uma nascente  
e nela limpa quem vive na lama.

Toma uma lágrima  
e poussa-a em quem nunca chorou.

Ganha coragem  
e dá-a a quem não sabe lutar.

Inventa a vida  
e conta-a a quem nada compreende.

Enche-te de esperança  
e vive á sua luz.

Enriquece-te de bondade  
e oferece-a a quem não sabe dar.

Vive com amor  
e fá-lo conhecer ao Mundo.”

Mahatma Gandhi



## **Agradecimentos**

- ✓ A Deus, sempre, por ter guiado meu caminho até aqui.
- ✓ Ao orientador Prof. Dr. Ronei Jesus Poppi pela orientação, oportunidade, conhecimentos transmitidos, amizade, paciência e confiança.
- ✓ Ao grupo e amigos que fiz no LAQQA: André Souza, Guilherme Alexandrino, Guilherme Sabin, Humberto Machado, Laila Balbino, Luciana Assis Terra, Luciana Oliveira, Márcia Cristina, Mariana Almeida, Mónica Mamián Lopez, Paulo Filgueiras, Juliana, Guilherme Montovani, Mariana Baptistão, Fabiana e Robson pela amizade, prosas e conhecimentos compartilhados e transmitidos.
- ✓ À Inês Lunardi pela ajuda na parte experimental e análises no HPLC;
- ✓ Ao Pedro Augusto Tizei pelo auxílio no crescimento das células para os experimentos;
- ✓ Ao Instituto de Biologia da UNICAMP e em especial ao Laboratório de Genômica e Expressão, onde foi realizado o crescimento das células.
- ✓ À CAPES pelo financiamento do projeto e bolsa concedida.
- ✓ À Unicamp por fornecer toda estrutura física e tecnológica para a realização deste trabalho.
- ✓ Aos amigos: Geraldo, Ricardo, Rodrigo MRN, Rodrigo MG, Bruno, Laiane, Joel, Luiz, Icaro, Tito, Marcelo, Thiago, Thais, Daniel, Benedito, Aninha, Xênia, Jesiel, Reinaldo, Thamires, Marina, Giovanni, Nayanne, Letícia, Ayla, Helô, Amanda, Felipe, Camila, Guilherme, Isabela, Luiza e Suelen pela amizade, companheirismo, conversas e discussões.
- ✓ Aos familiares que sempre estão na torcida e me apoiando.



# **Curriculum Vitae**

## **DADOS PESSOAIS**

Nome: Thiago Carvalho de Avila  
Data de nascimento: 19/11/1985  
Naturalidade: Boa Esperança – MG  
Nacionalidade: Brasileira  
Endereço eletrônico: [thiago.avila.85@gmail.com](mailto:thiago.avila.85@gmail.com)

## **FORMAÇÃO ACADÊMICA**

- Graduação: Química Bacharelado com Atribuições Tecnológicas 16/08/2006-16/07/2010  
Instituição: Universidade Federal de Alfenas – UNIFAL-MG, Alfenas – MG.

## **ATIVIDADES ACADÊMICAS**

- Programa de Estágio Docente – 08/2012 - 12/2012  
Disciplina: QA-282 – Química Clássica  
Curso: Farmácia  
Instituto de Química – UNICAMP
- Iniciação Científica – 01/2008 - 12/2009  
Área: Química Analítica  
Orientador: César Ricardo Teixeira Tarley  
Título: Emprego de Sílica Gel Organicamente Modificada e Impressa Ionicamente para Pré-concentração Seletiva *On-line* de Ions Cobre.
- Monitoria – 2º semestre de 2008  
Disciplina: Química Inorgânica II  
Curso: Química  
Carga Horária: 60 horas  
Universidade Federal de Alfenas
- Monitoria – 1º semestre de 2007  
Disciplina: Química Geral  
Curso: Química  
Carga Horária: 80 horas  
Universidade Federal de Alfenas

## **PRODUÇÃO BIBLIOGRÁFICA**

- AVILA, T. C.; POPPI, R. J.; LUNARDI, I.; TIZEI, P. A. G.; PEREIRA, G. A. G. Raman Spectroscopy and Chemometrics for *On-Line* Control of Glucose Fermentation by *Saccharomyces cerevisiae*. *Biotechnology Progress*, vol. 28, nº 6, pp. 1598-1604, 2012.
- TARLEY, C. R. T.; AVILA, T. C.; SEGATELLI, M. G.; LIMA, G. F.; PEREGRINO, G. S.; SCHEEREN, C. W.; DIAS, S. L. P.; RIBEIRO, E. S. Silica-Alumina-Niobia ( $\text{SiO}_2/\text{Al}_2\text{O}_3/\text{Nb}_2\text{O}_5$ ) Matrix Obtained by the Sol-gel Processing Method: New Material for

*Online* Extraction of Zinc Ions. Journal of the Brazilian Chemical Society, v. 21, pp. 1106-1116, 2010.

- AVILA, T. C.; SEGATELLI, M. G.; TARLEY, C. R. T.; BEIJO, L. A.. Emprego de Sílica Gel Organicamente Modificada e Impressa Ionicamente para Pré-concentração Seletiva *On-line* de Íons Cobre. Química Nova, v. 33, pp. 301-308, 2010.

#### **TRABALHOS EM EVENTOS INTERNACIONAIS**

- AVILA, T. C.; POPPI, R. J.; LUNARDI, I.; TIZEI, P. A. G.; PEREIRA, G. A. G. Aplicação de Cartas de Controle Multivariadas no Monitoramento *On-line* da Fermentação de glicose pela *Saccharomyces cerevisiae* usando Espectroscopia Raman, 5º Congresso Ibero-americano de Química Analítica, Montevideu – Uruguai, 10/2012.

#### **TRABALHOS EM EVENTOS NACIONAIS**

- AVILA, T. C.; POPPI, R. J.; LUNARDI, I.; TIZEI, P. A. G. Monitoramento *on-line* de Produtos de Fermentação Empregando Espectroscopia Raman e Calibração Multivariada, 16º Encontro Nacional de Química Analítica, 2011, Campos do Jordão-SP.
- ÁVILA, T. C.; SEGATELLI, M. G.; TARLEY, C. R. T. Sistema de Pré-concentração em Fluxo de Íons  $\text{Cu}^{2+}$  Empregando Sílica Gel com Impressão Iônica, Jornada Científica da UNIFAL-MG, 2009, Alfenas-MG.
- AVILA, T. C.; SEGATELLI, M. G.; TARLEY, C. R. T. Avaliação do Desempenho Seletivo de Sílica Gel Organicamente Modificada e Ionicamente Impressa com Íons  $\text{Cu}^{2+}$  Empregando Sistema FIA, 32º Reunião Anual da SBQ, 2009, Fortaleza-CE.
- LIMA, G. F.; AVILA, T. C.; SEGATELLI, M. G.; PEREGRINO G. S.; DIAS S. L. P.; RIBEIRO E. S.; TARLEY, C. R. T. Avaliação do adsorvente baseado na mistura de óxido metálico misto ( $\text{Al}_2\text{O}_3/\text{Nb}_2\text{O}_5$ ) disperso em matriz de sílica porosa como material pré-concentrador de zinco, 15º Encontro Nacional de Química Analítica, 2009, Salvador-BA.

#### **PARTICIPAÇÃO EM EVENTOS**

- 5º Congresso Ibero-americano de Química Analítica, Montevideu – Uruguai, 2012.
- 16º Encontro Nacional de Química Analítica, Outubro de 2011, Campos do Jordão – SP.
- 15º Encontro Nacional de Química Analítica, Outubro de 2009, Salvador – BA.
- II Jornada Científica da UNIFAL-MG, Setembro de 2009, Alfenas – MG.
- XXII Encontro Regional da Sociedade Brasileira de Química, Novembro de 2008, Belo Horizonte – MG.
- I Jornada Científica da UNIFAL-MG, Setembro de 2008, Alfenas – MG.
- III Semana da Química da UNIFAL-MG, Setembro de 2007, Alfenas – MG.
- II Semana da Química da UNIFAL-MG, Setembro de 2006, Alfenas – MG.

## Resumo

ESPECTROSCOPIA RAMAN E QUIMIOMETRIA COMO FERRAMENTAS NO MONITORAMENTO *ON-LINE* DO PROCESSO FERMENTATIVO DA GLICOSE PELA *SACCHAROMYCES CEREVISIAE*.

Autor - Thiago Carvalho de Avila

Ronei Jesus Poppi - Orientador

Este trabalho visou o uso de Espectroscopia Raman e de Quimiometria para monitoramento e controle da fermentação de glicose por *Saccharomyces cerevisiae*. Na primeira etapa, foi utilizada calibração multivariada baseada no método dos Mínimos Quadrados Parciais (PLS) para quantificação de glicose, etanol, glicerol, ácido acético e células. Os modelos foram desenvolvidos baseados nos valores de concentração obtidos pelos métodos de referência, cromatografia líquida de alta eficiência – HPLC e espectrofotometria UV/Vis. Tanto na etapa de calibração quanto na de validação, a otimização foi realizada com eliminação de amostras anômalas, baseada nos valores de *leverage*, resíduos e escores. Na segunda etapa, cartas de controle multivariadas foram usadas para identificação de falhas em bateladas durante o processo de fermentação. Foram construídos modelos MPCA (*Análise de Componentes Principais Multimodo*) a partir de bateladas NOC (*Condições Normais de Operação*). As cartas de controle multivariadas foram aplicadas em dois modos de desdobramento dos dados obtidos durante o monitoramento, um preservando a direção das bateladas e outro a direção do tempo. As falhas estudadas foram temperatura, mudança no substrato e contaminação do sistema. No modo de desdobramento por bateladas, a carta de controle  $Q$  foi eficiente para detecção das falhas estudadas, fato comprovado pela classificação correta de três bateladas NOC como dentro de controle. No entanto, a carta de controle  $T^2$  não foi capaz de identificar as falhas estudadas corretamente como fora de controle. O modo de desdobramento pelo tempo também apresentou classificações corretas das falhas estudadas.

**Palavras Chave:** Monitoramento *On-line*, Espectroscopia Raman, Cartas de Controle Multivariadas, Mínimos Quadrados Parciais, *Análise de Componentes Principais Multimodo*.



## Abstract

RAMAN SPECTROSCOPY AND CHEMOMETRICS FOR *ON-LINE* MONITORING OF GLUCOSE FERMENTATION BY *SACCHAROMYCES CEREVISIAE*.

Author - Thiago Carvalho de Avila

Ronei Jesus Poppi - Adviser

This work aims the use of Raman Spectroscopy and Chemometrics in the monitoring and control in the fermentation of the glucose by *Saccharomyces cerevisiae*. In the first step, it was applied the multivariate calibration based on Partial Least Squares (PLS) for the quantification of glucose, ethanol, glycerol, acetic acid and cells. The developed of calibration models was performed against the concentration values obtained by the reference methods, High Performance Liquid Chromatography and UV/Vis spectrophotometer. The optimization of the calibration and validation steps, the elimination of outliers was performed based on the values of leverage, residues and scores. In the second step, multivariate control charts were used for identification of batch-fault during the fermentation process. Multi-way Principal Component Analysis (MPCA) models were developed from batch NOC (Normal Operation Conditions). The multivariate control charts were based on two modes of unfolding the multi-way data, obtained during monitoring, one preserving the direction of the batch and another the direction of time. The fault studied were temperature, changes in the substrate and contamination of the system. In unfolding batch mode, the chart  $Q$  was effective for detection of the faults studied, proven by the correctly classification of 3 NOC batches as in control. However, the chart  $T^2$  failed to identify faults studied. The unfolding in time mode, also presented correct classifications of the faults studied.

**Keywords:** *On-line* Monitoring, Raman Spectroscopy, Multivariate Control Charts, Partial Least Squares, Multi-way Principal Component Analysis.



## Lista de Siglas

**CCD** – DISPOSITIVOS DE CARGA ACOPLADA, (*CHARGE COUPLED DEVICES*)

**FDA** – FOOD AND DRUG ADMINISTRATION.

**MLR** – REGRESSÃO LINEAR MÚLTIPLA, (*MULTIPLE LINEAR REGRESSION*).

**MPCA** – ANÁLISES DE COMPONENTES PRINCIPAIS MULTIMODO, (*MULTI-WAY PRINCIPAL COMPONENT ANALYSIS*).

**MSPC** – CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS, (*MULTIVARIATE STATISTICAL PROCESS CONTROL*).

**NIPALS** – *NONLINEAR INTERACTIVE PARTIAL LEAST SQUARES*.

**NOC** – CONDIÇÕES NORMAIS DE OPERAÇÃO, (*NORMAL OPERATION CONDICTIONS*).

**PAT** – TECNOLOGIA DE ANÁLISES DE PROCESSOS, (*PROCESS ANALYTICAL TECHNOLOGY*).

**PC** – COMPONENTE(S) PRINCIPAL(S), (*PRINCIPAL(S) COMPONENT(S)*)

**PCA** – ANÁLISES DE COMPONENTES PRINCIPAIS, (*PRINCIPAL COMPONENT ANALYSIS*).

**PCR** – REGRESSÃO POR COMPONENTES PRINCIPAIS, (*PRINCIPAL COMPONENT REGRESSION*).

**PLS** – MÍNIMOS QUADRADOS PARCIAIS, (*PARTIAL LEAST SQUARES*).

**RC** – COEFICIENTE DE REGRESSÃO, (*REGRESSION COEFFICIENT*).

**RMSEC** – RAIZ DO ERRO QUADRÁTICO MÉDIO DE CALIBRAÇÃO (*ROOT MEAN SQUARE ERROR OF CALIBRATION*).

**RMSECV** – RAIZ DO ERRO QUADRÁTICO MÉDIO DE VALIDAÇÃO CRUZADA (*ROOT MEAN SQUARE ERROR OF CROSS VALIDATION*).

**RMSEP** – RAIZ DO ERRO QUADRÁTICO MÉDIO DE PREVISÃO (*ROOT MEAN SQUARE ERROR OF PREDICTION*).

**UV/Vis** – ULTRAVIOLETA/VISÍVEL.

**LV** – VARIÁVEL(S) LATENTE(S) (*LATENT VARIABLE*)



## Lista de Figuras

<b>FIGURA 1.1.</b> ESQUEMA DOS ESPALHAMENTOS RAYLEIGH E RAMAN STOKES E ANTI-STOKES.	6
<b>FIGURA 1.2.</b> ESQUEMA ILUSTRATIVO DE UM ESPECTRO RAMAN. -----	7
<b>FIGURA 1.3.</b> MODOS VIBRACIONAIS POSSÍVEIS PARA UMA MOLÉCULA NÃO LINEAR.-----	8
<b>FIGURA 1.4.</b> ESQUEMA TÍPICO DO MECANISMO DE FUNCIONAMENTO DE UM LASER.-----	10
<b>FIGURA 1.5.</b> ESQUEMA SIMPLIFICADO DA ROTA DE FERMENTAÇÃO DA GLICOSE POR SACCHAROMYCES CEREVISIAE. ADAPTADO DE [30]. -----	19
<b>FIGURA 1.6.</b> ESQUEMA DE COMO AS MATRIZES DE VARIÁVEIS (A) INDEPENDENTES E (B) DEPENDENTES SÃO ORGANIZADAS.-----	20
<b>FIGURA 1.7.</b> ESQUEMA DA DECOMPOSIÇÃO DA MATRIZ $X$ EM ESCORES E PESOS PELO MÉTODO PCA. -----	23
<b>FIGURA 1.8.</b> ESQUEMA DE UM ARRANJO TRIDIMENSIONAL DE DADOS. -----	25
<b>FIGURA 1.9.</b> ESQUEMA DA DECOMPOSIÇÃO DO ARRANJO TRIDIMENSIONAL DOS DADOS, $X$ , EM ESCORES E PESOS PELO MÉTODO DA MPCA. -----	25
<b>FIGURA 1.10.</b> OPERAÇÃO DE DESDOBRAMENTO PROPOSTO POR (A) NOMIKOS E MCGREGOR E POR (B) WORLD ET. AL.-----	27
<b>FIGURA 1.11.</b> REPRESENTAÇÃO ESQUEMÁTICA DA ROTAÇÃO NO EIXO DAS COMPONENTES PRINCIPAIS PARA OBTENÇÃO DE UMA MAIOR CORRELAÇÃO.-----	29
<b>FIGURA 1.12.</b> REPRESENTAÇÃO GEOMÉTRICA DE COMO A VARIÁVEL LATENTE MODELA OS DADOS PELO COEFICIENTE DE REGRESSÃO. -----	30
<b>FIGURA 1.13.</b> ESQUEMA DA DECOMPOSIÇÃO DAS MATRIZES $X$ E $Y$ EM ESCORES E PESOS PELO MÉTODO PLS.-----	30
<b>FIGURA 1.14.</b> REPRESENTAÇÃO ESQUEMÁTICA DAS CARTAS DE CONTROLE $Q$ E $T^2$ E SEUS LIMITES IDENTIFICANDO AMOSTRAS FORA DE CONTROLE.-----	35
<b>FIGURA 2.1.</b> ESQUEMA DO SISTEMA DO EXPERIMENTO ON-LINE PARA MONITORAMENTO DA FERMENTAÇÃO USANDO O ESPECTRÔMETRO RAMAN. -----	40
<b>FIGURA 2.2.</b> ESPECTROS RAMAN DOS PRODUTOS E SUBSTRATO PUROS DA FERMENTAÇÃO.	44

<b>FIGURA 2.3.</b> ESPECTROS RAMAN DA FERMENTAÇÃO DA GLICOSE PELA SACCHAROMYCES CEREVISIAE. -----	45
<b>FIGURA 2.4.</b> ESPECTROS RAMAN NO TEMPO INICIAL E FINAL DA FERMENTAÇÃO.-----	46
<b>FIGURA 2.5.</b> PERFIL DE CONCENTRAÇÃO OBTIDO PELA CALIBRAÇÃO E VALIDAÇÃO DO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA GLICOSE. -----	49
<b>FIGURA 2.6.</b> PERFIL DE CONCENTRAÇÃO OBTIDO PELA CALIBRAÇÃO E VALIDAÇÃO DO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA ETANOL. -----	49
<b>FIGURA 2.7.</b> PERFIL DE CONCENTRAÇÃO OBTIDO PELA CALIBRAÇÃO E VALIDAÇÃO DO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA GLICEROL. -----	50
<b>FIGURA 2.8.</b> PERFIL DE CONCENTRAÇÃO OBTIDO PELA CALIBRAÇÃO E VALIDAÇÃO DO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA ÁCIDO ACÉTICO. -----	50
<b>FIGURA 2.9.</b> PERFIL DE CONCENTRAÇÃO OBTIDO PELA CALIBRAÇÃO E VALIDAÇÃO DO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA CÉLULAS. -----	51
<b>FIGURA 2.10.</b> COEFICIENTES DE REGRESSÃO OBTIDOS PELOS MODELOS PLS1 CONSTRUÍDOS COM O PRIMEIRO GRUPO DE ESPECTROS PARA (A) GLICOSE, (B) ETANOL, (C) GLICEROL, (D) ÁCIDO ACÉTICO E (E) CÉLULAS. -----	52
<b>FIGURA 2.11.</b> REGRESSÃO LINEAR ENTRE OS COEFICIENTES DE REGRESSÃO (A) DA GLICOSE E ETANOL, (B) GLICOSE E GLICEROL, (C) GLICOSE E ÁCIDO ACÉTICO, (D) GLICOSE E CÉLULAS, (E) ETANOL E GLICEROL, (F) ETANOL E ÁCIDO ACÉTICO, (G) ETANOL E CÉLULAS, (H) GLICEROL E ÁCIDO ACÉTICO, (I) GLICEROL E CÉLULAS E (J) ÁCIDO ACÉTICO E CÉLULAS. -----	54
<b>FIGURA 2.12.</b> CORRELAÇÃO LINEAR ENTRE AS CONCENTRAÇÕES OBTIDAS PELOS MÉTODOS DE REFERÊNCIA, (A) GLICOSE E ETANOL, (B) GLICOSE E GLICEROL, (C) GLICOSE E ÁCIDO ACÉTICO, (D) ETANOL E GLICEROL, (E) ETANOL E ÁCIDO ACÉTICO, (F) GLICEROL E ÁCIDO ACÉTICO, (G) CÉLULAS E GLICOSE, (H) CÉLULAS E ETANOL, (I) CÉLULAS E GLICEROL E (J) CÉLULAS E ÁCIDO ACÉTICO. -----	56
<b>FIGURA 2.13.</b> PERFIL DE CONCENTRAÇÃO PREVISTO PELO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA GLICOSE. -----	59
<b>FIGURA 2.14.</b> PERFIL DE CONCENTRAÇÃO PREVISTO PELO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA ETANOL. -----	59

<b>FIGURA 2.15.</b> PERFIL DE CONCENTRAÇÃO PREVISTO PELO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA GLICEROL. -----	60
<b>FIGURA 2.16.</b> PERFIL DE CONCENTRAÇÃO PREVISTO PELO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA ÁCIDO ACÉTICO. -----	60
<b>FIGURA 2.17.</b> PERFIL DE CONCENTRAÇÃO PREVISTO PELO MODELO PLS E PELO MÉTODO DE REFERÊNCIA PARA CÉLULAS. -----	61
<b>FIGURA 2.18.</b> PERFIL DE CONCENTRAÇÃO PREVISTO PELO MODELO PLS2 E PELO MÉTODO DE REFERÊNCIA PARA GLICOSE.-----	63
<b>FIGURA 2.19.</b> PERFIL DE CONCENTRAÇÃO PREVISTO PELO MODELO PLS2 E PELO MÉTODO DE REFERÊNCIA PARA ETANOL.-----	63
<b>FIGURA 2.20.</b> PERFIL DE CONCENTRAÇÃO PREVISTO PELO MODELO PLS2 E PELO MÉTODO DE REFERÊNCIA PARA GLICEROL. -----	64
<b>FIGURA 2.21.</b> PERFIL DE CONCENTRAÇÃO PREVISTO PELO MODELO PLS2 E PELO MÉTODO DE REFERÊNCIA PARA ÁCIDO ACÉTICO. -----	64
<b>FIGURA 3.1.</b> (A) ESCORES DAS TRÊS COMPONENTES PRINCIPAIS DO MODELO MPCA E (B) ESCORES DA PRIMEIRA COMPONENTE PRINCIPAL VERSUS A SEGUNDA COMPONENTE PRINCIPAL DO MODELO MPCA. -----	73
<b>FIGURA 3.2.</b> CARTA DE CONTROLE Q BASEADA NA ANÁLISE MPCA (●) BATELADAS NOC; (●) BATELADAS FALHA-TEMPERATURA; (●) BATELADAS FALHA-SUBSTRATO; (●) BATELADAS FALHA-CONTAMINAÇÃO; (●) BATELADAS NOC-TESTE; (---) Q <sub>LIM</sub> COM NÍVEL DE CONFIANÇA DE 95%. -----	74
<b>FIGURA 3.3.</b> CARTA DE CONTROLE T <sub>2</sub> BASEADA NA ANÁLISE MPCA (●) BATELADAS NOC; (●) BATELADAS FALHA-TEMPERATURA; (●) BATELADAS FALHA-SUBSTRATO; (●) BATELADAS FALHA-CONTAMINAÇÃO; (●) BATELADAS NOC-TESTE; (---) T <sub>2LIM</sub> COM NÍVEL DE CONFIANÇA DE 95%. -----	75
<b>FIGURA 3.4.</b> GRÁFICOS DOS ESCORES NAS BATELADAS-NOC E DAS BATELADAS-FALHA. ---	76
<b>FIGURA 3.5.</b> ESPECTROS INICIAL E FINAL DE UMA BATELADA-NOC E DE UMA BATELADA-FALHA TEMPERATURA. -----	77
<b>FIGURA 3.6.</b> ESPECTROS INICIAL E FINAL DE UMA BATELADA-NOC E DE UMA BATELADA-FALHA SUBSTRATO.-----	77

<b>FIGURA 3.7.</b> ESPECTROS INICIAL E FINAL DE UMA BATELADA-NOC E DE UMA BATELADA-FALHA CONTAMINAÇÃO. -----	78
<b>FIGURA 3.8.</b> ESCORES DOS TEMPOS OBTIDOS DE CADA UMA DAS BATELADAS-NOC.-----	80
<b>FIGURA 3.9.</b> (A) AMPLIAÇÃO DA REGIÃO 1 DA FIGURA 3.8; (B) AMPLIAÇÃO DA REGIÃO 2 DA FIGURA 3.8, EM (ROSA) BATELADA 2, EM (AZUL) BATELADA 10 E EM (VERMELHO) BATELADA 11. -----	81
<b>FIGURA 3.10.</b> ESCORES DOS TEMPOS OBTIDOS DE CADA UMA DAS BATELADAS-NOC UTILIZADAS NA CONSTRUÇÃO DO MODELO MPCA.-----	82
<b>FIGURA 3.11.</b> (A) AMPLIAÇÃO DA REGIÃO 1 DA FIGURA 3.10 (BATELADA 7) E (B) AMPLIAÇÃO DA REGIÃO 2 DA FIGURA 3.10 (BATELADA 1). -----	83
<b>FIGURA 3.12.</b> ESCORES DA BATELADA-FALHA TEMPERATURA 1, (●) BATELADAS-NOC E (▼) BATELADA-FALHA.-----	84
<b>FIGURA 3.13.</b> ESCORES DA BATELADA-FALHA TEMPERATURA 2, (●) BATELADAS-NOC E (▼) BATELADA-FALHA.-----	85
<b>FIGURA 3.14.</b> ESCORES DA BATELADA-FALHA TEMPERATURA 3, (●) BATELADAS-NOC E (▼) BATELADA-FALHA.-----	85
<b>FIGURA 3.15.</b> ESCORES DA BATELADA-FALHA SUBSTRATO 1, (●) BATELADAS-NOC E (▼) BATELADA-FALHA.-----	86
<b>FIGURA 3.16.</b> ESCORES DA BATELADA-FALHA SUBSTRATO 2, (●) BATELADAS-NOC E (▼) BATELADA-FALHA.-----	87
<b>FIGURA 3.17.</b> ESCORES DA BATELADA-FALHA SUBSTRATO 3, (●) BATELADAS-NOC E (▼) BATELADA-FALHA.-----	87
<b>FIGURA 3.18.</b> ESCORES DA BATELADA-FALHA CONTAMINAÇÃO 1, (●) BATELADAS-NOC E (▼) BATELADA-FALHA.-----	88
<b>FIGURA 3.19.</b> ESCORES DA BATELADA-FALHA CONTAMINAÇÃO 2, (●) BATELADAS-NOC E (▼) BATELADA-FALHA.-----	89
<b>FIGURA 3.20.</b> ESCORES DA BATELADA-FALHA CONTAMINAÇÃO 3, (●) BATELADAS-NOC E (▼) BATELADA-FALHA.-----	89
<b>FIGURA 3.21.</b> ESCORES DA BATELADA NOC-TESTE 1, (●) BATELADAS-NOC E (▼) BATELADA-NOC-TESTE.-----	90

**FIGURA 3.22.** ESCORES DA BATELADA NOC-TESTE 2, (●) BATELADAS-NOC E (▼)  
BATELADA-NOC-TESTE. -----91

**FIGURA 3.23.** ESCORES DA BATELADA NOC-TESTE 3, (●) BATELADAS-NOC E (▼)  
BATELADA-NOC-TESTE. -----91



## Lista de Tabelas

<b>TABELA 1.1.</b> COMPARAÇÃO ENTRE OS TIPOS DE SISTEMAS DE MONITORAMENTO. -----	14
<b>TABELA 1.2.</b> ETAPAS DO ALGORITMO NIPALS PARA O MÉTODO PLS [38,39].-----	31
<b>TABELA 2.1.</b> RESULTADOS PARA OS MODELOS PLS1 DESENVOLVIDOS USANDO O GRUPO DE ESPECTROS OBTIDOS NO MONITORAMENTO ON-LINE. -----	48
<b>TABELA 2.2.</b> RESULTADOS PARA OS MODELOS PLS1 CONSTRUÍDOS PARA PREVISÃO DAS AMOSTRAS OBTIDAS NO MONITORAMENTO ON-LINE QUE NÃO APRESENTAM DADOS CORRESPONDENTES AO MONITORAMENTO OFF-LINE. -----	58
<b>TABELA 2.3.</b> VALORES DE RMSEC E RMSECV OBTIDOS PELO MODELO PLS2 PARA CADA ANALITO. -----	62
<b>TABELA 2.4.</b> TABELA COMPARATIVA DOS ERROS DE CALIBRAÇÃO E DE CALIBRAÇÃO CRUZADA DOS MODELOS PLS1 E PLS2. -----	65



# Sumário

<b>Prefácio</b>	<b>1</b>
<b>1. Capítulo I – Introdução Geral</b>	<b>4</b>
<b>1.1. Espectroscopia Raman</b>	<b>5</b>
1.1.1. Princípios da Espectroscopia Raman	5
1.1.2. Espectrômetros Raman	9
1.1.3. Aplicações, Vantagens e Desvantagens da Espectroscopia Raman	11
<b>1.2. Monitoramento de Bioprocessos</b>	<b>13</b>
<b>1.3. <i>Saccharomyces cerevisiae</i></b>	<b>17</b>
<b>1.4. Métodos Quimiométricos</b>	<b>19</b>
1.4.1. Pré-processamento dos dados	21
1.4.2. Análise de Componentes Principais	22
1.4.3. Análise de Componentes Principais Multimodo	24
1.4.4. Mínimos Quadrados Parciais	27
<b>1.5. Cartas de Controle Multivariadas baseadas em análise MPCA</b>	<b>32</b>
<b>1.6. Objetivos</b>	<b>36</b>
<b>2. Capítulo II – Quantificação de Substrato e Produtos de Fermentação</b>	<b>37</b>
<b>2.1. Introdução</b>	<b>39</b>
<b>2.2. Parte experimental</b>	<b>40</b>
2.2.1. Instrumentação	40
2.2.2. Micro-organismos	41
2.2.3. Processo de Monitoramento	42
2.2.4. Análise Quimiométrica	43
<b>2.3. Resultados e Discussão</b>	<b>44</b>
2.3.1. Espectros Raman	44
2.3.2. Análise dos Dados	47
2.3.2.1. PLS1	47

2.3.2.1.1. Coeficientes de Regressão do modelo PLS1 _____	51
2.3.2.1.2. Modelos para previsão das amostras obtidas no monitoramento on-line que não apresentam dados correspondentes ao monitoramento off-line _____	57
2.3.2.2. PLS2 _____	61
<b>2.4. Conclusões _____</b>	<b>65</b>
<b>3. Capítulo III – Aplicação de Cartas de Controle Multivariadas para Detecção de Falhas _____</b>	<b>67</b>
<b>3.1. Introdução _____</b>	<b>69</b>
<b>3.2. Parte Experimental _____</b>	<b>70</b>
3.2.1. Instrumentação e Micro-organismos _____	70
3.2.2. Processo de Monitoramento _____	70
3.2.3. Construção das Cartas de Controle Multivariadas baseadas na análise MPCA _____	71
<b>3.3. Resultados e Discussão _____</b>	<b>72</b>
3.3.1. Desdobramento em relação à batelada _____	72
3.3.1.1. Análise MPCA _____	72
3.3.1.2. Cartas de Controle $Q$ e $T^2$ _____	74
3.3.2. Desdobramento em relação ao tempo _____	79
3.3.2.1. Análise MPCA _____	79
3.3.2.2. Cartas de Controle $Q$ e $T^2$ _____	83
<b>3.4. Conclusões _____</b>	<b>92</b>
<b>4. Capítulo IV – Conclusões Gerais _____</b>	<b>95</b>
<b>Referências _____</b>	<b>99</b>

# Prefácio

---

A espectroscopia Raman juntamente com a Quimiometria vem se tornando uma ferramenta inovadora para a indústria no que diz respeito à sua aplicação no monitoramento e controle de processos, sendo que estes possibilitam acompanhamento em tempo real do perfil do processo, conseguindo dessa maneira manter o rendimento e a produtividade sob controle, além de diminuir os custos de produção e facilitar a identificação de eventuais falhas. Isso ajuda a tomada de decisões para uma correção ou até mesmo a paralização do processo para evitar perdas ainda maiores. Neste caso, o tempo de ação, que é o tempo gasto entre a amostragem e a realização da análise, é muito menor quando comparados a outros métodos, como por exemplo, cromatografia e análises via úmida, o que facilita e antecipa a tomada de decisões.

Um tipo de bioprocessamento muito utilizado na indústria hoje em dia é a fermentação na produção de bioetanol para uso como combustível. Os processos fermentativos, em sua maioria, são sensíveis às mudanças das condições normais de operação e para obtenção de um produto de qualidade e com bom rendimento é necessário manter o controle dessas condições normais de operação para que este processo não se torne pouco produtivo.

Tendo em vista o potencial da espectrometria Raman juntamente com a Quimiometria para monitoramento e controle de bioprocessos, este trabalho foi desenvolvido com intuito de realizar um monitoramento *on-line* da fermentação da glicose pela levedura *Saccharomyces cerevisiae* para quantificação das substâncias envolvidas na fermentação através da realização de uma calibração multivariada e da aplicação de cartas de controle multivariadas na detecção de bateladas-falha durante o processo. Dessa forma, para melhor entendimento, este trabalho foi dividido em 4 capítulos descritos a seguir.

O capítulo I apresenta uma introdução dos temas que foram abordados neste trabalho como a espectroscopia Raman e suas aplicações, fundamentos, aspectos instrumentais e monitoramento de bioprocessos. Além disso, é apresentada também uma breve introdução sobre a levedura *Saccharomyces cerevisiae* que foi utilizada nas fermentações e sobre os métodos quimiométricos empregados nas análises dos dados obtidos por espectroscopia Raman. Este capítulo é base para compreensão dos motivos da realização deste trabalho.

O capítulo II apresenta a primeira parte experimental do trabalho, que diz respeito ao monitoramento *on-line* do processo fermentativo da glicose pela *Saccharomyces cerevisiae*. Essa parte do trabalho teve como objetivo a quantificação de substrato e produtos de fermentação utilizando a calibração multivariada empregando o método dos mínimos quadrados parciais – PLS, uma calibração conhecida como de primeira ordem devido ao fato que a quantificação pode ser realizada na presença de interferentes, desde que estes estejam presentes tanto no conjunto de calibração quanto no de previsão.

O capítulo III apresenta a aplicação de cartas de controle multivariadas baseadas na análise de componentes principais multimodo – MPCA para identificação de bateladas-falha durante o processo de monitoramento da fermentação que está descrita no capítulo II. Para tal finalidade, foram empregadas as cartas de controle multivariadas  $Q$  e *Hotelling*  $T^2$  na identificação de falhas causadas por variações no processo devido à temperatura, mudança de substrato e contaminação. Foram avaliadas duas formas de identificação de falhas, uma considerando cada batelada separadamente e outra considerando os tempos de monitoramento.

O capítulo IV traz as conclusões gerais sobre os resultados obtidos neste trabalho. Esta Dissertação encerra com a apresentação das referências bibliográficas que foram úteis em sua elaboração.

# **Capítulo I**

---

## **Introdução Geral**



## 1.1. ESPECTROSCOPIA RAMAN

### 1.1.1. Princípios da Espectroscopia Raman

O princípio de toda espectroscopia óptica está no estudo das interações entre a radiação eletromagnética e a matéria. Essas interações são devidas à transmissão, absorção, reflexão, refração e espalhamento de uma radiação incidente na amostra. O efeito Raman é o espalhamento inelástico da radiação de comprimento de onda no visível ou infravermelho próximo que incide sobre a amostra. Foi primeiramente descrito por Raman e Krishnan em 1928 [1].

O estudo do espalhamento de luz realizado por Raman levou à constatação de que quando um feixe de radiação monocromática incide em uma amostra ocorre espalhamento dessa radiação. Na análise dessa radiação espalhada, utilizando um espectrógrafo, observou-se que além da linha (comprimento de onda) da radiação incidente haviam outras linhas com pequenos deslocamentos em relação ao comprimento de onda da radiação incidente. Esses diferentes tipos de espalhamentos são o espalhamento elástico, em que os fótons da radiação incidente mudam de direção e o espalhamento inelástico, que além da mudança de direção ocorre variação do comprimento de onda da radiação, ou seja, na energia da radiação espalhada [2].

A espectroscopia Raman se aplica à detecção da radiação espalhada por moléculas, resultantes de alteração na polarizabilidade em suas ligações causadas pelo campo elétrico da radiação incidente, o que leva a molécula a um estado conhecido como estado eletrônico virtual. Mas antes da descrição do funcionamento da espectroscopia Raman é necessário compreender os dois tipos de espalhamento de radiação, elástico e inelástico. No espalhamento elástico, também conhecido por espalhamento Rayleigh, a interação da molécula com o fóton não provoca mudanças nos níveis de energia vibracional e/ou rotacional da molécula fazendo com que a frequência da radiação incidente e da espalhada sejam as mesmas. Já no

espalhamento inelástico ou espalhamento Raman, a incidência da radiação altera a polarizabilidade da molécula alterando os níveis das energias vibracional e/ou rotacional por  $\pm \Delta E$ . A Figura 1.1 ilustra o esquema de espalhamento Raman considerando os estados eletrônicos virtuais. Podemos dizer, hipoteticamente, que a molécula absorve uma quantidade de energia  $\Delta E$  no espalhamento Stokes e perde uma quantidade de energia  $\Delta E$  no espalhamento anti-Stokes, uma vez que a radiação espalhada tem energia menor e maior que a da radiação incidente para os espalhamentos Stokes e anti-Stokes, respectivamente.

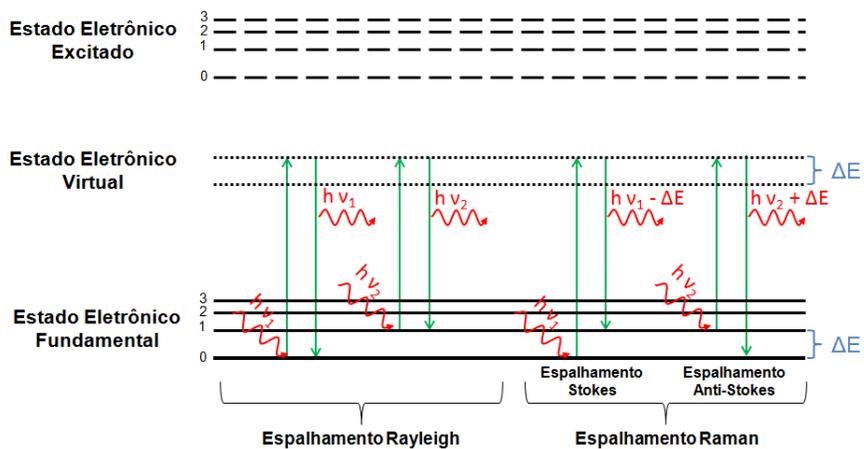


Figura 1.1. Esquema dos Espalhamentos Rayleigh e Raman Stokes e anti-Stokes.

Como pode ser visto na Figura 1.2, a comparação entre as linhas Stokes e anti-Stokes revela uma simetria entre elas em relação à linha Rayleigh, estando do lado de frequências mais baixas as linhas Stokes e do lado de frequências mais altas as linhas anti-Stokes [3]. As linhas anti-Stokes são menos intensas que suas correspondentes Stokes, fato que está ligado à probabilidade do fenômeno ocorrer,

já que a população dos estados excitados segue a distribuição de Boltzmann. Sendo assim, as linhas Stokes de um espectro são normalmente usadas para as análises, ficando a parte anti-Stokes do espectro descartada.

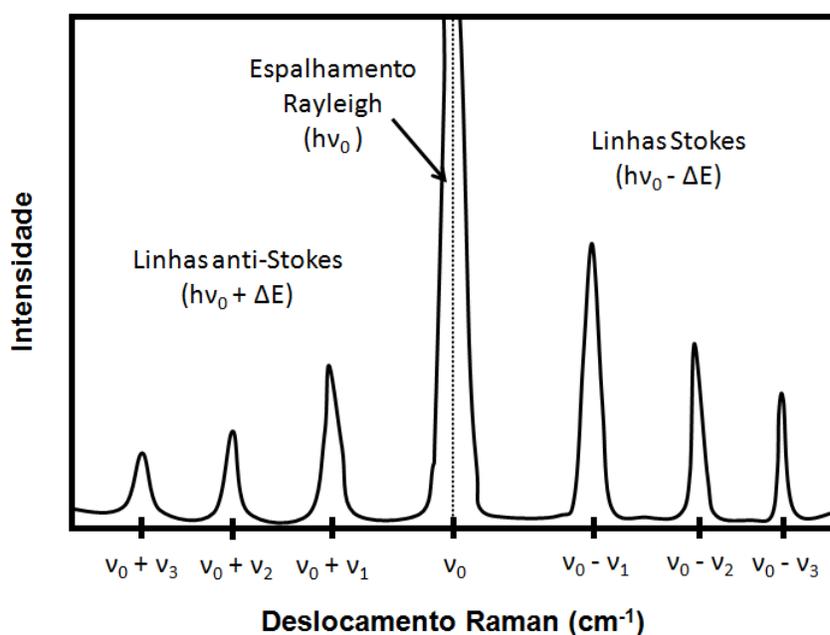


Figura 1.2. Esquema ilustrativo de um espectro Raman.

De forma geral, podemos dizer que um espalhamento Raman envolve a interação de dois fótons com o centro espalhador (amostra) ao mesmo tempo, sendo que um o fóton da radiação incidente e outro da radiação espalhada [4, 5]. Os movimentos vibracionais podem ocorrer de várias formas, Figura 1.3, para uma molécula não linear e são classificados como estiramento, tesoura (dobra), torção, balanço ou rotação. O estiramento é a variação da distância média entre dois átomos que formam uma ligação química e pode ser de dois tipos, simétrico ou assimétrico.

A tesoura refere-se à vibração que causa mudança no ângulo de ligação entre três átomos. Na rotação os átomos se movem na mesma direção sem alterar o ângulo da ligação no plano. E o movimento dos átomos entrando e saindo do plano que contém a molécula é a torção, quando o movimento dos átomos ocorre em direções opostas e o balanço, quando o movimento dos átomos ocorre na mesma direção [6].

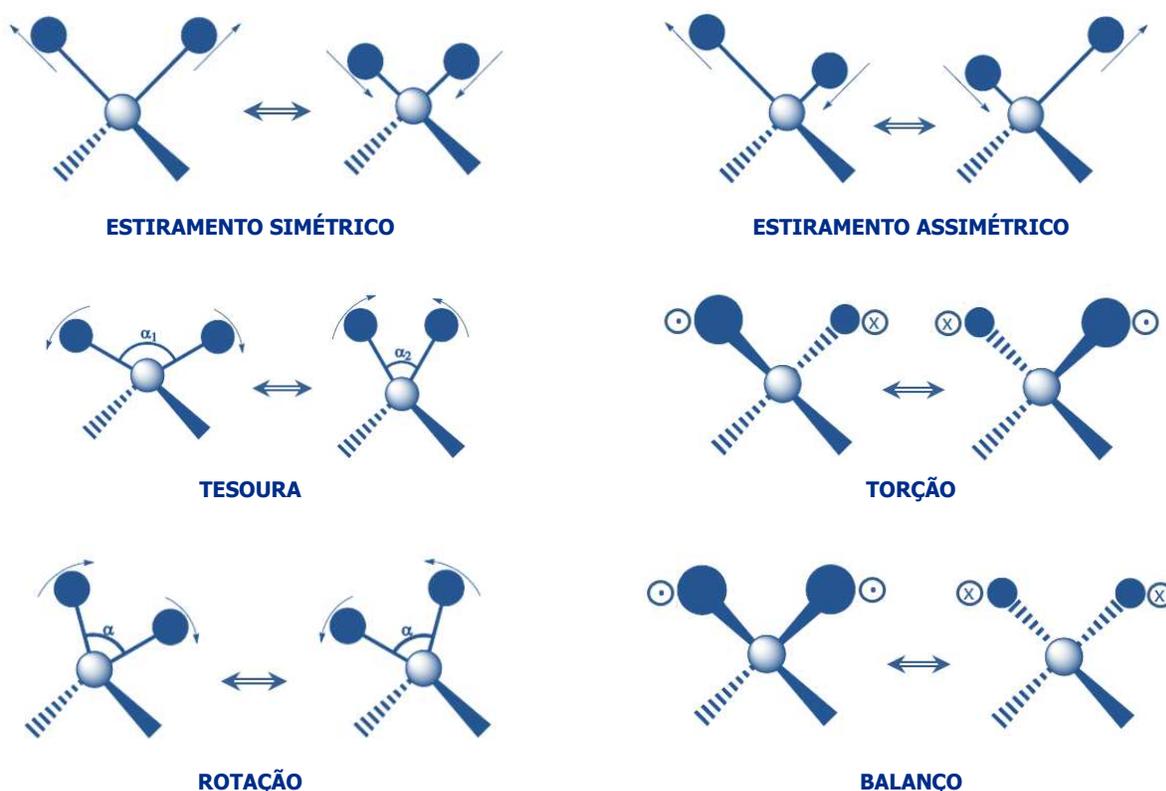


Figura 1.3. Modos Vibracionais possíveis para uma molécula não linear.

A atividade de um modo vibracional de uma ligação no infravermelho é devido à variação do momento de dipolo elétrico durante a vibração. Uma ligação que apresenta dipolo elétrico é dita ser polar. No entanto, considerando uma molécula

homonuclear, esta não irá apresentar dipolo elétrico, porém irá apresentar atividade Raman. Na espectroscopia Raman é considerado o momento de dipolo induzido causado pelo campo elétrico da radiação incidente e não o momento de dipolo intrínseco da ligação, ou seja, a presença do campo elétrico irá causar um deslocamento da nuvem eletrônica em relação aos núcleos. Esse deslocamento dos centros das cargas positivas e negativas, leva à formação de dipolo induzido. A polarizabilidade de uma molécula diz respeito à propriedade dessas cargas se reorganizarem sob a ação de um campo elétrico e depende da facilidade com que ocorre a redistribuição da nuvem eletrônica da molécula [2].

A intensidade de uma banda Raman é bastante complexa de ser entendida devido à sua dependência em relação à polarizabilidade da molécula, da intensidade da fonte e da concentração do grupo ativo da molécula. A intensidade da emissão Raman é influenciada pela presença de absorção (da amostra), em que na sua ausência a intensidade de emissão pode aumentar com a quarta potência da frequência da radiação incidente, entretanto, raramente se faz uso dessa propriedade, uma vez que a radiação ultravioleta pode causar foto-decomposição e fluorescência. Geralmente, as intensidades das bandas Raman são diretamente proporcionais à concentração do analito na amostra, ao contrário do que ocorre na espectroscopia de absorção, na qual a relação concentração/intensidade é logarítmica [3].

### 1.1.2. Espectrômetros Raman

Atualmente as fontes de radiação mais empregadas na espectroscopia Raman, e que ajudou na ampliação do uso da técnica, são os lasers, cujo significado é Amplificação de Luz por Emissão da Radiação Estimulada (*Light Amplification by Stimulated Emission of Radiation*), devido às suas características, como alta intensidade, bandas estreitas e coerência. O mecanismo de ação dos lasers é o bombeamento de um meio ativo, que pode ser constituído de um cristal, um

semicondutor, uma solução ou um gás, pela radiação de uma fonte externa de forma que poucos fótons de energia apropriada engatilham a emissão de uma cascata de fótons com a mesma energia. O bombeamento pode ser realizado por meio de descarga, corrente elétrica ou luz. A amplificação de luz se dá pela passagem da radiação repetidas vezes através do meio ativo pela ação de espelhos como mostrado na Figura 1.4 [3].

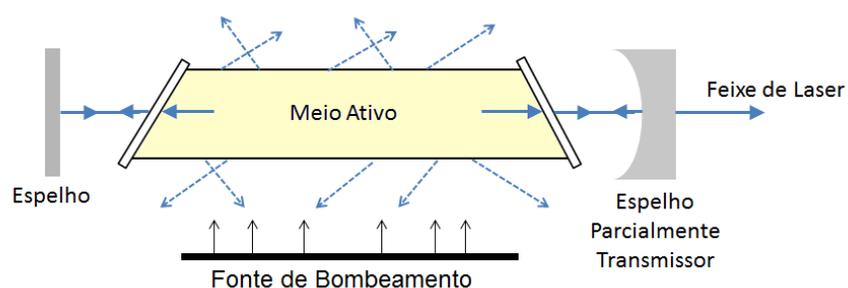


Figura 1.4. Esquema típico do mecanismo de funcionamento de um laser.

Para que ocorra amplificação de luz é necessário uma inversão de população, ou seja, a quantidade de espécies no estado de maior energia deve ser maior que a quantidade de espécies no estado de menor energia.

Além das fontes de lasers um espectrômetro Raman requer um bom dispositivo de seleção de comprimento de onda que consiga separar as linhas fracas do espalhamento Raman da radiação provinda do espalhamento Rayleigh que ocorre com maior intensidade. Os dispositivos mais empregados são os monocromadores e as redes holográficas. Os transdutores mais empregados são os Dispositivos de Carga Acoplada (*Charge Coupled Devices - CCD*) que geralmente operam em torno de  $-50\text{ }^{\circ}\text{C}$  para melhorar a razão sinal/ruído [3].

Os transdutores CCD, conectados a tubos de fotomultiplicadora, são sensíveis à radiação em 783 nm, produzida por lasers de diodo, que promovem excitação de muitos compostos e são utilizados na espectroscopia Raman sem interferência significativa de fluorescência. Já os CCD's não são sensíveis à radiação em 1.064 nm produzida por lasers de Nd:YAG, que é formado por íons neodímio em um cristal hospedeiro de ítrio-alumínio. A aplicação desse laser é útil em sistemas biológicos, uma vez que o comprimento de onda de 1.064 nm elimina a interferência devido à fluorescência [3, 5].

### **1.1.3. Aplicações, Vantagens e Desvantagens da Espectroscopia Raman**

A espectroscopia Raman apresenta vasta aplicação em Química, auxiliando tanto em análises qualitativas como quantitativas de sistemas inorgânicos, orgânicos e biológicos, sendo possível análises de amostras em qualquer estado físico, sejam elas gasosas, líquidas ou sólidas, semi-sólidas, pastas ou géis. Esta técnica pode ser aplicada em controle de qualidade, controle de processos industriais agrícolas, farmacêuticos, petroquímicos e alimentícios.

Geralmente os espectros Raman apresentam menos bandas que os espectros no infravermelho para uma mesma amostra; dessa forma, a espectroscopia Raman sofre menos problemas com a presença de bandas de sobretons e de combinação, tornando mais simples as medidas quantitativas. Essa vantagem faz com que a espectroscopia Raman seja empregada para análises quantitativas; no entanto, o alto custo dos equipamentos comparados a equipamentos de absorção faz com que isso não ocorra. O alto custo de espectrômetros Raman é devido aos lasers de diodo, mas à medida que o custo desses equipamentos diminui o uso da espectroscopia Raman vem se tornando mais popular [3, 5].

A aplicação da espectroscopia Raman em sistemas inorgânicos, na maioria das vezes é superior à do infravermelho, devido ao fato desses sistemas se

apresentarem aquosos. A água apresenta banda intensa nos espectros de infravermelho e não interfere nos espectros Raman. Outro fato importante é que as energias vibracionais de ligações metal-ligante se apresentam geralmente na região de 100 a 700  $\text{cm}^{-1}$ , região mais facilmente observada na espectroscopia Raman que no infravermelho [3].

A principal desvantagem da espectroscopia Raman é a baixa sensibilidade comparada a métodos de absorção, exigindo para as análises, altas concentrações, alto tempo de aquisição de dados ou o uso de uma superfície específica [7]. Em geral, a espectroscopia Raman é mais frequentemente utilizada em análises qualitativas que quantitativas; porém, a quantificação pode ser bastante simples quando a intensidade de uma banda de algum analito for proporcionalmente linear à sua concentração. Por outro lado, é preciso estar ciente dos fatores não ideais que podem prejudicar uma quantificação exata, como a fluorescência, a turbidez da amostra e a deposição dos produtos de reação em paredes ou no fundo da cela [8].

Nos últimos 20 anos, avanços na instrumentação, com lasers de diodo mais estáveis e detectores CCD mais sensíveis, fizeram com que a espectroscopia Raman fosse mais explorada. Esses dois fatores, detector e laser, contribuíram com a espectroscopia Raman para que esta seja utilizada para estudos de monitoramento de processos ou reações. O uso de um sistema óptico combinado com um detector CCD permite a obtenção de vários espectros em curto espaço de tempo, o que possibilita a detecção de produtos em um monitoramento [7].

Entre as espectroscopias vibracionais, a espectroscopia Raman é uma técnica promissora para aplicações no monitoramento de bioprocessos, devido à baixa interferência de água nos espectros, ampla faixa de comprimento de onda, disponibilidade de sondas de fibras ópticas acopladas ao espectrômetro [9], não ser destrutiva, não requer pré-tratamento das amostras e poder determinar mais de um componente ao mesmo tempo [10].

A espectroscopia Raman tem sido muito utilizada para monitoramento e análise de processos por oferecer algumas vantagens sobre outras técnicas

espectroscópicas em análises químicas. Os métodos de infravermelho (IR) apresentam limitações no caso de monitoramento de bioprocessos devido à forte banda de adsorção de água, enquanto os espectros Raman são ricos em informações que permite a análise de um amplo número de espécies orgânicas não cromóforas no UV/Vis, podendo ser usados para identificar espécies desconhecidas, sendo dessa forma, uma técnica apropriada para análises de meios de reações em sistemas aquosos [7].

A espectroscopia Raman aplicada ao monitoramento de bioprocessos produz dados multivariados, no qual muitas vezes o número de variáveis é maior que o número de observações, podendo haver colinearidade dos dados, o que requer o uso de ferramentas de estatística multivariada para visualizar e identificar possíveis correlações relevantes no conjunto de dados.

## 1.2. MONITORAMENTO DE BIOPROCESSOS

O mercado de bioprocessos tem se mostrado altamente competitivo e aquecido nos últimos anos, o que faz necessário o investimento em novas tecnologias com finalidade de aumentar a produtividade e melhorar a qualidade de seus produtos. Um monitoramento eficaz de um bioprocessos é uma necessidade crescente devido à importância dessas indústrias, exigências em redução dos custos de produção e garantia da qualidade do produto final.

O monitoramento de um bioprocessos pode ser classificado de acordo com o sistema de medição como *off-line*, *at-line*, *on-line*, *in-line* e *non-invasive*. No monitoramento *off-line* a amostragem é realizada manualmente através de válvulas de amostragem no reator, podendo a amostra sofrer pré-tratamento ou não, e transportada até o equipamento analisador. Uma tentativa de melhora nos sistemas *off-line* deu origem aos sistemas *at-line*, em que o instrumento de análise é posicionado próximo ao ponto de amostragem, a fim de diminuir o tempo e análise.

Os sistemas *in-line* são aqueles em que o sensor de medida está em contato direto com a amostra, eliminando dessa forma a etapa de amostragem. Um sistema *on-line* é aquele automatizado, onde um fluxo é gerado através de um duto que é utilizado para extrair a amostra. Os sistemas *on-line* são divididos em dois tipos: intermitentes, onde a amostra é transferida do processo para o instrumento analisador e contínuos, onde a amostra passa através de uma cela de medição podendo retornar ou ser descartada do processo. Já os sistemas *non-invasive* têm as vantagens dos sistemas *in-line*, mas sem o contato direto do sensor com a amostra [11]. A Tabela 1.1 apresenta uma comparação das vantagens e desvantagens de cada um desses sistemas.

Tabela 1.1. Comparação entre os tipos de sistemas de monitoramento.

<b>Sistema</b>	<b>Vantagens</b>	<b>Desvantagens</b>
<b><i>Off-line</i></b>	Facilidade de desenvolvimento de métodos e de manutenção.	Tempo de amostragem e de reportagem dos resultados.
<b><i>At-line</i></b>	Maior rapidez na obtenção dos resultados (comparado ao <i>off-line</i> ).	Tempo de amostragem e de reportagem dos resultados.
<b><i>In-line</i></b>	Não tem etapas de amostragem e possui medidas mais representativas.	Desgastes e obstrução dos sensores devido ao contato direto com a amostra.
<b><i>On-line</i></b>	Sistema automatizado de amostragem e análise.	Processo de amostragem separa uma linha analítica do processo.
<b><i>Non-invasive</i></b>	Não tem etapas de amostragem e sistema analisador não entra em contato direto com o processo.	Difícil encontrar sensores adequados para esse tipo de sistema.

O monitoramento *off-line* de bioprocessos é realizado em amostras retiradas do meio de cultura no biorreator e suas análises são realizadas após algum tempo da amostragem. Embora esta abordagem permita medições de alta precisão, há uma desvantagem óbvia quando se compara o tempo de amostragem e de análise com os obtidos por um monitoramento *on-line*. O monitoramento *off-line* normalmente é

utilizado para o desenvolvimento de modelos matemáticos que podem ser usados para o controle de processos futuros. Em contraste, as técnicas de monitoramento em tempo real têm a vantagem de fornecer informações durante o processo, fornecendo uma abordagem do estado do bioprocessamento, sendo essencial para detecção precoce de eventuais problemas, permitindo a ação imediata para resolver a situação enquanto o processo está sendo realizado [12]. Um método *on-line* ideal para o monitoramento de bioprocessos deve ser rápido, não requerer a remoção da amostra do biorreator e ser capaz de quantificar diferentes analitos simultaneamente.

Bioprocessos são usados na indústria para fabricação de diversos produtos, como por exemplo, etanol, ácidos orgânicos, aminoácidos, insulina, enzimas e antibióticos. A capacidade de monitoramento das variáveis e a composição de seus produtos é importante para supervisão do controle de qualidade e otimização das condições de operação. No desenvolvimento de um bioprocessamento, as aquisições de dados *on-line* e a flexibilidade de análise de substratos e metabólitos são importantes para obtenção de informações sobre o desempenho do processo e identificação precoce de condições adversas ao crescimento microbiano, possibilitando uma ação corretiva para evitar uma queda na produtividade [9]. Para que um método de monitoramento *on-line* seja eficiente este deve ser capaz de prever as concentrações de substratos e produtos e avaliar outras características de qualidade dos produtos finais, tais como impurezas e contaminantes. Todavia, a análise destes compostos por métodos tradicionais, além de serem demorados, requerem equipamentos de análises específicas e pré-tratamento de amostras.

A crescente demanda por etanol produzido via fermentação e a grande variedade de coprodutos e metabólitos intermediários obtidos durante a fermentação alcoólica leva à necessidade de métodos analíticos capazes de fornecer informações sobre as condições dos processos de fermentação em tempo real [13]. Os métodos ópticos como os de fluorescência de multi-comprimento de onda [14-16] e espectroscopia de infravermelho próximo [13, 17-20] têm sido amplamente utilizadas para desenvolvimento de novas técnicas para monitoramento de bioprocessos. Ainda dentro dos métodos ópticos a espectroscopia Raman é uma técnica promissora para

aplicações em monitoramento de bioprocessos, o que faz com que tenham surgido, ultimamente, várias aplicações desta no monitoramento de bioprocessos, como por exemplo, usando a bactéria *Escherichia coli* [9], a levedura *Saccharomyces cerevisiae* [21, 22] e a bactéria *Lactobacillus casei* [23].

Para garantir a qualidade de um produto é necessário um controle dos parâmetros químicos, físicos e biológicos que envolvem os bioprocessos através da otimização destes. Por isso, nos últimos anos o foco em técnicas *on-line* e *in-line* para monitoramento de bioprocessos vem aumentando, impulsionado pela necessidade incessante da indústria na otimização de processos. O FDA (*Food and Drug Administration*), órgão do governo americano, tem estabelecido metas para que se possa aplicar a tecnologia de processos analíticos (*Process Analytical Technology - PAT*) em monitoramento de bioprocessos com objetivo de melhorar a compreensão, controle e análise, enfatizando que a qualidade de um produto deve ser inerente ao processo de fabricação. O PAT é tido como um sistema que visa projetar, analisar e controlar a fabricação através de medições em tempo real, objetivando a qualidade do produto final através de um controle do processo [11, 24].

A aplicação do PAT visa possibilitar a obtenção de informações, quantitativas e qualitativas seja o processo químico, físico ou biológico, para que possam ser utilizadas a fim de otimizar as condições de processo, tempo e matéria-prima e não apenas como forma de controle do processo. Se o sistema é suficientemente rápido, este permite ao analisador um tempo de ação que possibilite um reajuste das variáveis do processo para retornar às condições normais em caso de alguma variação prejudicial ao processo ser detectada, embora sempre exista um intervalo entre tempo de medida e tempo de ação, que determina a robustez de um monitoramento frente a perturbações [11]. A leitura direta de variáveis fundamentais de processos biológicos, como por exemplo, biomassa, ainda não foi alcançada, apesar da disponibilidade de sistemas *on-line* com sensores que oferecem diferentes tipos de sinais [24].

Os resultados de monitoramento de bioprocessos ajudam proporcionar maior eficiência, produtividade, reprodutibilidade, redução de custos, melhoria de controle de qualidade e redução da poluição do meio ambiente. O composto ativo de todos os bioprocessos, os micro-organismos, são muito sensíveis às condições de processos, e pequenas alterações, como por exemplo, pH, temperatura ou pressão podem alterar o metabolismo celular e mudar radicalmente a eficiência do processo e assim a produtividade, podendo até mesmo tornar o processo inútil [12].

Em análises laboratoriais a manipulação das amostras é realizada sob condições estabelecidas por normas, podendo sofrer algum tipo de pré-tratamento a fim de aumentar a seletividade e/ou sensibilidade. Além disso, os instrumentos que ficam em laboratórios estão mais seguros contra efeitos de variações instrumentais devido a transporte, manuseio, variação da luminosidade ambiente do que equipamentos de campo, que estão mais expostos. As análises realizadas em laboratórios analíticos são utilizadas como forma de garantir a qualidade do produto e não como forma de controlar o processo. O controle de qualidade não reduz os custos de produção, uma vez que as informações sobre o processo são obtidas após o fim do processo e revelam somente a aceitabilidade ou não do produto [11].

A principal tendência e importância da implementação de monitoramentos *on-line* e em tempo real na indústria é permitir a redução do uso de reagentes e diminuição de resíduos (química verde), além de menor custo de produção e maior sustentabilidade.

### **1.3. SACCHAROMYCES CEREVISIAE**

*Saccharomyces cerevisiae* é o micro-organismo, não patogênico, mais utilizado em fermentações alcoólicas, que é a transformação de açúcares em etanol e gás carbônico, em escala industrial. Esse micro-organismo possui metabolismo anaeróbico facultativo, realizando a fermentação na presença de algum açúcar.

Apesar do etanol ser o principal produto de fermentação, o processo envolve vias metabólicas adicionais, onde a mais importante é a fermentação glicerol-pirúvico, que produz glicerol e ácido pirúvico e, partindo desse último, a produção de vários compostos, incluindo ácidos succínico e lático, acetona e 2,3-butanodiol. No entanto algumas modificações genéticas podem ser realizadas nessas leveduras para produção industrial de outros produtos comerciais [13] ou para uma maior produção de etanol inviabilizando outras vias metabólicas.

*S. cerevisiae* é um micro-organismo que apresenta alta eficiência na fermentação de açúcares, alta taxa de crescimento, metabolismo eficiente, alta conversão de açúcares em etanol, tolerância a altas concentrações de etanol, baixa concentração de oxigênio e a variações de temperatura além de poderem realizar a fermentação em meios ácidos [25].

*S. cerevisiae* é capaz de fazer a fermentação com uma grande variedade de carboidratos como fontes de energia, como por exemplo, glicose, sacarose e maltose, esses dois últimos após uma etapa de hidrólise [26]. A produção de bioetanol por esta levedura é atualmente, em volume, o maior processo de fermentação industrial [27]. O etanol celulósico é uma opção de futuro, mas o amido de milho ainda é a matéria-prima mais empregada na produção de etanol em muitos países, incluindo Estados Unidos, Canadá e China. O procedimento convencional para produzir etanol a partir de amido de milho requer duas etapas: primeiro a hidrólise do amido pela amilase e posterior fermentação por *S.cerevisiae* [28]. No Brasil o processo é realizado utilizando como matéria-prima a cana-de-açúcar, fonte de sacarose, que pode ser convertida em etanol diretamente pela levedura *S. cerevisiae* sem a necessidade da etapa enzimática [29]. Na Figura 1.5 está apresentado o metabolismo para a fermentação da glicose para produção de etanol pela *S. cerevisiae*.

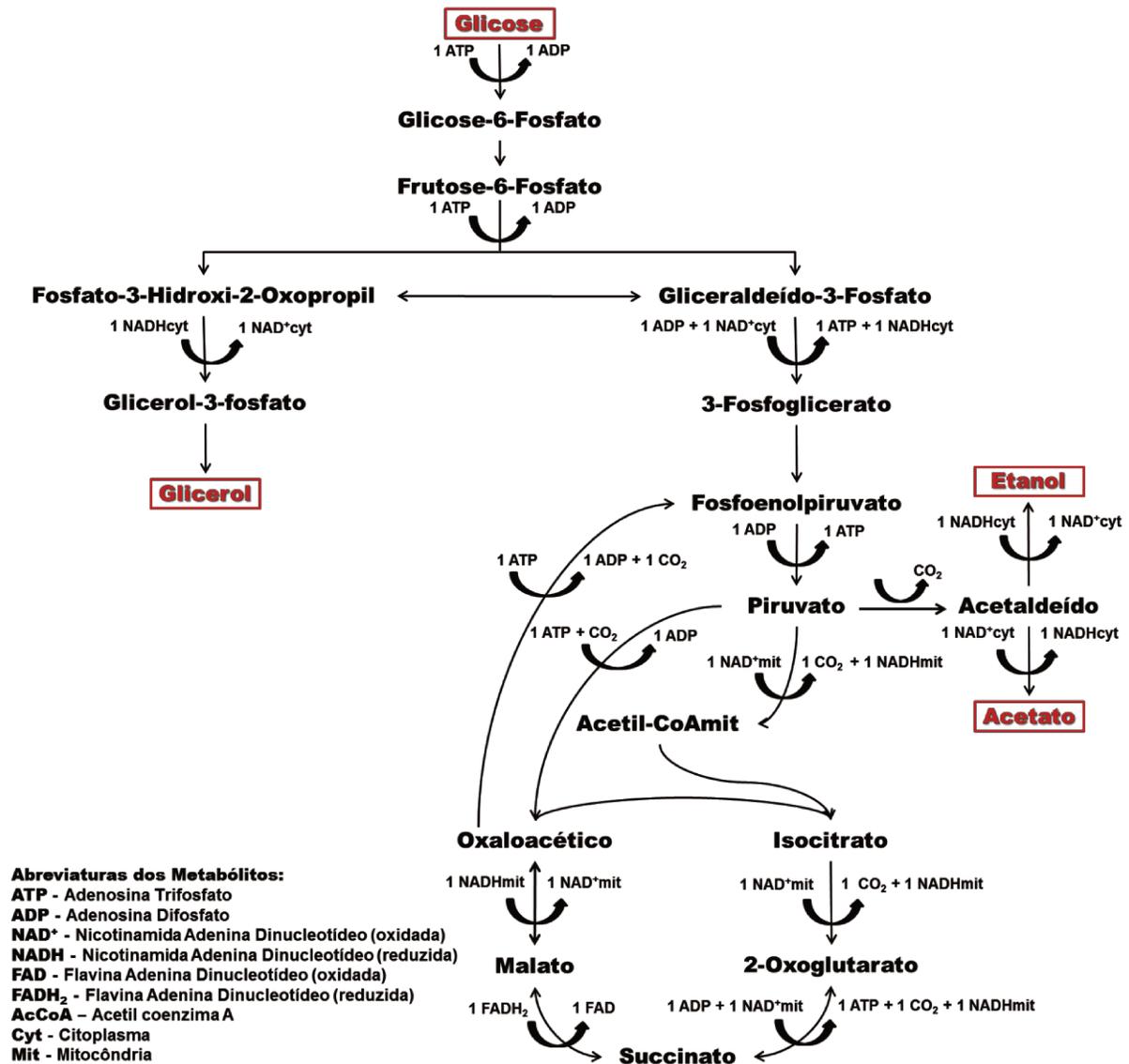


Figura 1.5. Esquema simplificado da rota de fermentação da glicose por *Saccharomyces cerevisiae*. Adaptado de [30].

## 1.4. MÉTODOS QUIMIOMÉTRICOS

Quimiometria envolve a aplicação de métodos matemáticos e estatísticos para solucionar e entender problemas químicos. Em técnicas como Ressonância

Magnética Nuclear, Espectroscopia no Infravermelho, Espectroscopia UV/Vis, Espectrometria de Massas e Espectroscopia Raman, os resultados não estão explícitos nos espectros e requerem tratamentos dos dados, complexos do ponto de vista matemático e estatístico, que possam, por exemplo, correlacionar intensidade do sinal obtido com concentração. Dessa forma, grande quantidade de sinais, curvas e picos podem ser tratados para uma possível quantificação de espécies presentes no sistema.

Geralmente na Quimiometria se faz o uso de matrizes para organizar os dados e, na maioria dos casos, quando se trata de dados espectroscópicos, as matrizes de dados são formadas pelas intensidades das variáveis espectrais dispostas em cada coluna e pelo número de amostras dispostos nas linhas, como representado na Figura 1.6. Dados espectrais são tratados como variáveis independentes por necessitarem de processamento para que se possa obter alguma informação química. Por outro lado, valores em matrizes que contém dados que são conhecidos e determinados por métodos de referência são chamadas de variáveis dependentes, que em sua maior parte, na química analítica, se trata de valores de concentração. A Figura 1.6 também traz um esquema de como as matrizes de variáveis dependentes são dispostas.

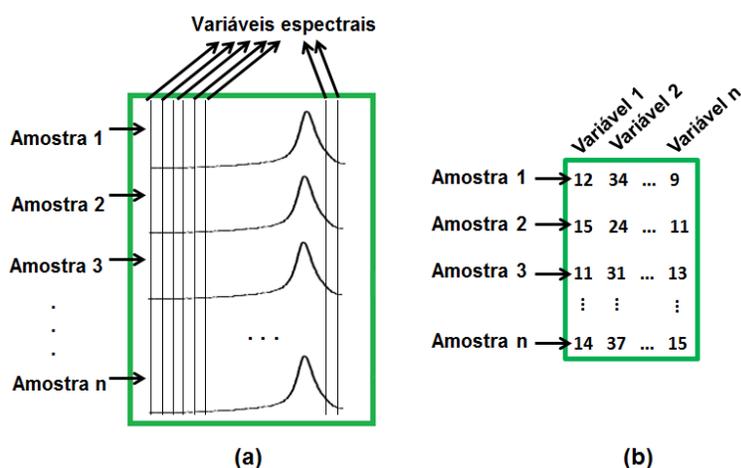


Figura 1.6. Esquema de como as matrizes de variáveis (a) independentes e (b) dependentes são organizadas.

### 1.4.1. Pré-processamento dos dados

A etapa de pré-processamento dos dados é muito importante em uma análise quimiométrica, já que em técnicas espectroscópicas, na maioria das vezes, o sinal analítico obtido pode se apresentar com intensidades diferentes devido à presença de ruído instrumental, interferentes e/ou variações sistemáticas da linha de base. Dessa forma, o pré-processamento é utilizado para eliminar ou diminuir a influência desses efeitos na análise dos dados. No entanto, deve-se ter atenção e conhecimento químico nesta etapa, pois um pré-processamento dos dados de forma incorreta pode levar o analista a obter falsas conclusões.

Geralmente os tipos de pré-processamentos mais utilizados para dados químicos são:

- Centralização dos dados na média: subtração de cada valor de uma matriz de dados em uma linha pela da média dos valores de sua coluna.
- Escalonamento: cada variável de uma linha de uma matriz de dados é dividida pelo desvio padrão de cada variável (coluna).
- Auto-escalonamento: aplicação dos métodos de centralização dos dados e escalonamento, só que neste caso, além das variáveis exercerem o mesmo peso na construção do modelo, elas terão média igual a zero e desvio padrão igual a um.
- Alisamento (suavização): utilizada para diminuir os efeitos de ruídos aumentando a razão sinal/ruído dos espectros. O alisamento pode ser realizado com o algoritmo desenvolvido por Savitzky-Golay [31], ajustando um polinômio de ordem baixa e uma janela, número de pontos (variáveis), através do método dos mínimos quadrados. O número de pontos da janela deve ser adequado, já que um número

elevado pode levar à perda de informações espectrais e um número reduzido à não eliminação dos ruídos.

- 1ª e 2ª derivada: utilizadas para correção de linha base e para facilitar a visualização dos picos presentes do espectro original. Um efeito negativo dos pré-processamentos com 1ª e 2ª derivada é que ocorre uma intensificação dos ruídos presentes, principalmente quando a relação sinal/ruídos é baixa.

Existem muitos outros tipos de pré-processamentos que podem ser utilizados em dados de espectroscopia; além disso, a combinação de alguns desses métodos também é possível.

### 1.4.2. Análise de Componentes Principais

A Análise de Componentes Principais (*Principal Component Analysis – PCA*) é a decomposição de uma matriz de dados através de uma combinação de variáveis, ou fatores, que irá descrever a maior variabilidade nos dados. O primeiro passo para uma análise PCA é a formação de uma matriz de variância/covariância dos dados ( $\mathbf{Z}$ ) que irá isolar a fonte de variação dos dados através de ortogonalização baseada na mudança de base vetorial de acordo com a equação 1.1 [32].

$$\mathbf{Z} = \mathbf{X}^T \mathbf{X} \quad 1.1$$

A matriz  $\mathbf{Z}$  deve apresentar uma simetria para que possa ser diagonalizada, de forma que  $\mathbf{Z} = \mathbf{Z}^T$  e, por uma transformação unitária, a matriz de covariância pode ser diagonalizada pela equação 1.2 [32].

$$\mathbf{\Lambda} = \mathbf{P}^T \mathbf{Z} \mathbf{P} \quad 1.2$$

onde  $\mathbf{\Lambda}$  é a matriz diagonal que representa os autovalores de  $\mathbf{Z}$  e  $\mathbf{P}$  é a matriz de autovetores, também chamada de matriz peso. Basicamente essa matriz peso forma uma nova base ortonormal que explica a variância de  $\mathbf{X}$  e a projeção de  $\mathbf{X}$  em  $\mathbf{P}$  gera os escores,  $\mathbf{T}$ . Sendo assim, a matriz de dados é decomposta nas matrizes escores e pesos, denominadas componentes principais (PC), e que pode ser representada pela equação 1.3 [32].

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T = \sum_{i=1}^I \mathbf{t}_i \mathbf{p}_i^T \quad 1.3$$

Na PCA, algebricamente, os PC são combinações lineares ortogonais dos dados originais e pode ser representado geometricamente pela Figura 1.7, em que “i” é o número de PC.

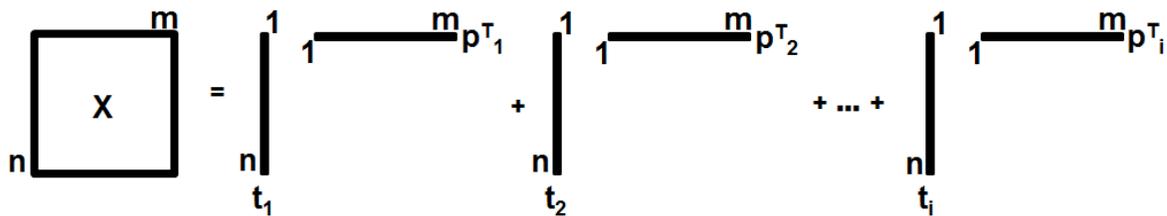


Figura 1.7. Esquema da decomposição da matriz  $\mathbf{X}$  em escores e pesos pelo método PCA.

PCA é uma ferramenta exploratória que através dos componentes principais fornece uma visão geral dos dados obtidos em uma representação gerada em um novo espaço. Assim, o primeiro componente principal representa a direção de maior variabilidade, o segundo componente principal representa a direção de segunda maior variabilidade e assim por diante para os componentes de ordem inferior.

O número máximo de PC's de um modelo é igual ao número de variáveis presentes nos dados. No entanto, nem todas PC's trazem informações úteis ao modelo; geralmente as primeiras PC's explicam quase toda variância dos dados e as últimas PC's geralmente modelam ruídos inerente aos dados [32-34].

PCA é um método que sofre forte influência quando o conjunto de dados apresenta amostras anômalas, sendo necessário um pré-processamento dos dados a fim de encontrar e eliminar essas amostras anômalas antes da construção do modelo.

### **1.4.3. Análise de Componentes Principais Multimodo**

A análise de componentes principais multimodo (*Multi-way Principal Component Analysis – MPCA*) é estatisticamente consistente com a Análise de Componentes Principais – PCA. No entanto, para o caso de monitoramento de processos, a organização dos dados é diferente daquele da análise PCA. Neste caso as amostras são bateladas, ou seja, uma amostra que contém os dados de todo um processo que foi monitorado em intervalos de tempos definidos. Cada batelada é representada por uma matriz que contém os espectros tomados a cada tempo e o arranjo tridimensional é formado pelo concatenamento dessas matrizes. Dessa forma, o arranjo tridimensional dos dados contém as variáveis espectrais na coordenada x, os tempos de monitoramento na coordenada y e o número de bateladas na coordenada z, como representado na Figura 1.8.

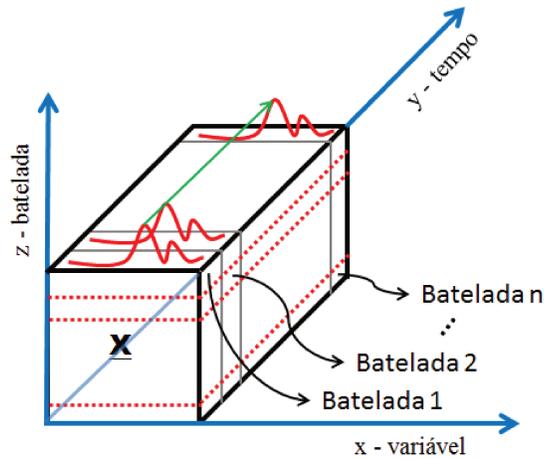


Figura 1.8. Esquema de um arranjo tridimensional de dados.

Como na análise por PCA, na análise por MPCA o arranjo tridimensional é decomposto em uma matriz de escores, um arranjo tridimensional de pesos e mais um arranjo tridimensional de resíduos, de acordo com a equação 1.4 e esquematizado na Figura 1.9.

$$\underline{X} = \underline{T}_r \otimes \underline{P}_r^T + \underline{E} \quad 1.4$$

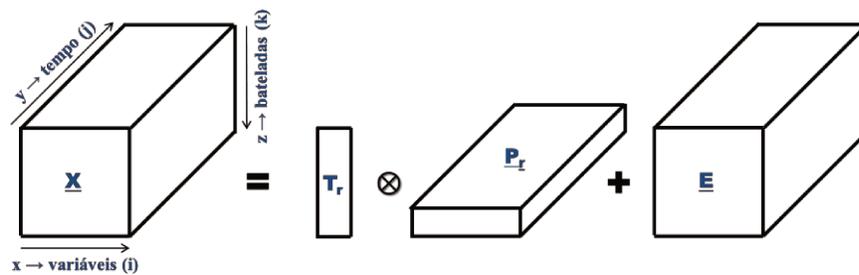


Figura 1.9. Esquema da decomposição do arranjo tridimensional dos dados,  $\underline{X}$ , em escores e pesos pelo método da MPCA.

O arranjo tridimensional,  $\underline{\mathbf{X}}$ , é decomposto de forma a obter o menor resíduo possível. O resultado da decomposição pode ser dividido em duas partes, sistemática e não-sistemática. A parte sistemática, representada pelas matrizes de escores,  $\underline{\mathbf{T}}_r$ , e pesos,  $\underline{\mathbf{P}}_r$ , que expressam toda informação sobre  $\underline{\mathbf{X}}$ , representando a variação determinística. Os escores estão relacionados somente com as amostras e os pesos relacionados com as variáveis e seus tempos de variação. A parte não sistemática é a parte residual. Essa parte tende a não apresentar nenhuma informação sobre as amostras, estando associada à parte não determinística nos dados [35].

A aplicação da análise PCA no arranjo tridimensional  $\underline{\mathbf{X}}$  desdobrado gera um resultado semelhante à aplicação da análise MPCA neste arranjo tridimensional. O desdobramento do arranjo tridimensional pode ser realizado de três modos, sendo que dois deles apresentam maior aplicação. Um desses modos foi proposto por Nomikos e McGregor [36], ocorrendo uma vetorização da batelada e o desdobramento dos dados resulta em uma matriz bidimensional, com dimensões de  $K \times JI$  ( $K$  bateladas,  $J$  tempos e  $I$  variáveis espectrais), preservando a direção das bateladas, tornando possível a detecção de variações entre as bateladas. O outro modo de desdobramento foi proposto por Wold *et al.* [37], resultando em uma matriz bidimensional de dimensões  $J \times KI$ , onde modelos locais podem ser desenvolvidos para cada tempo. Neste caso, cada matriz é processada por PCA, preservando a direção das variáveis, tornando possível identificar qual variável é mais importante na construção do modelo PCA. Os dois tipos de desdobramentos são demonstrados na Figura 1.10.

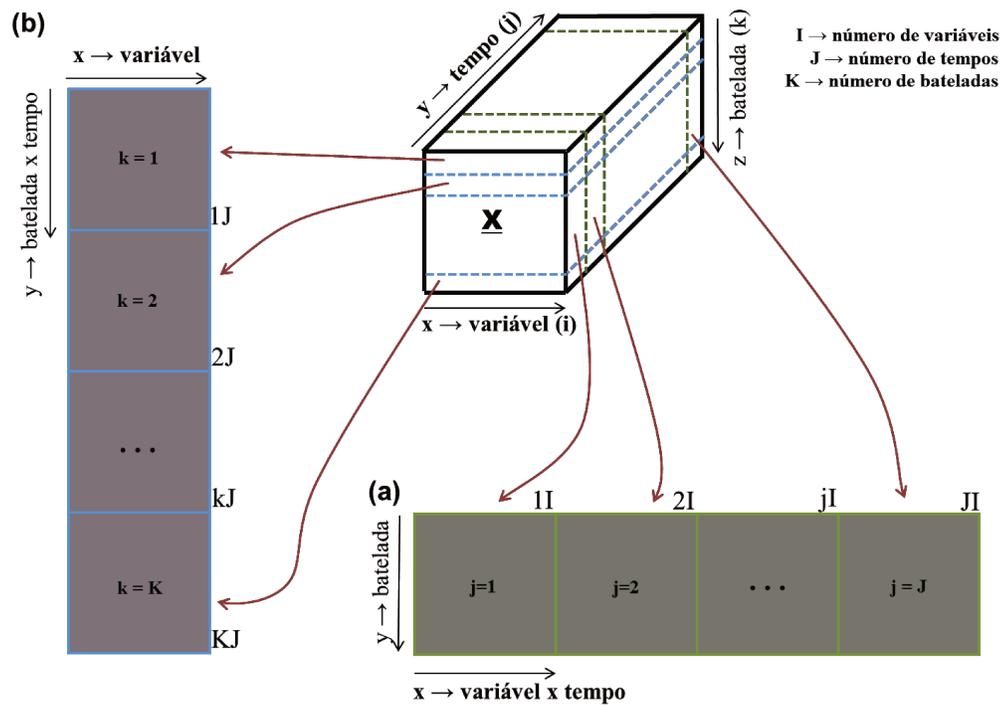


Figura 1.10. Operação de desdobramento proposto por (a) Nomikos e McGregor e por (b) World *et. al.*

Como na PCA, na análise MPCA também é importante uma seleção do número ideal de PC's, a fim de eliminar PC's que modelam ruídos inerentes aos dados.

#### 1.4.4. Mínimos Quadrados Parciais

O método dos Mínimos Quadrados Parciais (*Partial Least Squares – PLS*) é o método mais utilizado na literatura para o desenvolvimento de modelos para a calibração multivariada, oferecendo a possibilidade da análise de dados como espectros, que apresentam sinais sobrepostos e determinações simultâneas de

vários constituintes [10]. O PLS foi desenvolvido na década de 60 por Wold para a aplicação na área de economia, conhecida hoje como econometria [38].

No PLS, ao contrário da Regressão Linear Múltipla (*Multiple Linear Regression – MLR*), pode-se analisar dados com forte colinearidade (correlacionados), com ruídos e grande número de variáveis e também responder a variáveis distintas simultaneamente. O PLS não necessita de um conhecimento detalhado dos constituintes presentes na amostra, desde que os mesmos estejam presentes na calibração do modelo [33], fato conhecido como “vantagem de primeira ordem”.

A Regressão por Componentes Principais (*Principal Component Regression – PCR*) é um método que utiliza o resultado da aplicação da PCA na matriz  $\mathbf{X}$  (a matriz de escores,  $\mathbf{T}$ ) para obter uma relação linear com a matriz de resposta  $\mathbf{Y}$ . No entanto, mesmo o PCR fazendo uma correlação linear com a matriz  $\mathbf{Y}$ , o método não é capaz de garantir a melhor covariância entre as matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ . Devido a esse fato foi desenvolvido o método PLS. Ele é baseado na decomposição das duas matrizes,  $\mathbf{X}$  que contém as variáveis independentes e  $\mathbf{Y}$  que contém as variáveis dependentes, em escores,  $\mathbf{T}$  e  $\mathbf{U}$ , e pesos,  $\mathbf{P}$  e  $\mathbf{Q}$ , além de uma matriz de resíduos,  $\mathbf{E}$  e  $\mathbf{F}$ , de acordo com as equações 1.5 e 1.6.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{i=1}^I \mathbf{t}_a \mathbf{p}_i^T + \mathbf{E}_X \quad 1.5$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y = \sum_{i=1}^I \mathbf{u}_i \mathbf{q}_i^T + \mathbf{F}_Y \quad 1.6$$

Durante a decomposição da matriz  $\mathbf{X}$ , as informações da matriz  $\mathbf{Y}$  são levadas em consideração, com objetivo de obter maior correlação entre  $\mathbf{X}$  e  $\mathbf{Y}$  [32]. Como a decomposição das matrizes não ocorre de forma independente existe uma relação linear (interna) entre os escores de  $\mathbf{X}$  e  $\mathbf{Y}$  para cada componente principal de acordo

com a equação 1.7, onde  $b_i$  representa para cada componente principal o coeficiente de regressão do modelo linear.

$$\mathbf{u}_i = \mathbf{b}_i \mathbf{t}_i \quad 1.7$$

sendo,

$$\mathbf{b}_i = \frac{\mathbf{u}_i^T \mathbf{t}_i}{\mathbf{t}_i^T \mathbf{t}_i} \quad 1.8$$

e “i” representa cada componente principal.

No PLS essa maior correlação linear entre as matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  causa uma leve variação nos valores dos escores para garantir a melhor relação entre  $\mathbf{X}$  e  $\mathbf{Y}$ . Isso é obtido através de uma leve rotação no eixo das Componentes Principais (*Principal Components – PC’s*), como representado na Figura 1.11. Após essa rotação essas PC’s passam a serem chamadas de Variáveis Latentes (*Latent Variables – LV’s*). A Figura 1.12 mostra uma representação geométrica de como a LV modela os dados das matrizes de variáveis dependentes e independentes.

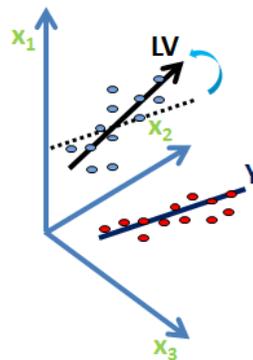


Figura 1.11. Representação esquemática da rotação no eixo das componentes principais para obtenção de uma maior correlação.

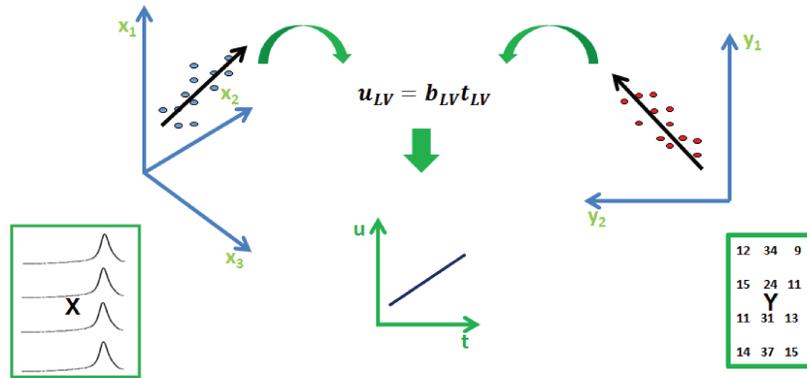


Figura 1.12. Representação geométrica de como a variável latente modela os dados pelo coeficiente de regressão.

A Figura 1.13 apresenta o esquema da decomposição das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$ , de dimensões  $n \times m$  e  $n \times r$ , respectivamente nas matrizes escores e pesos. Essa decomposição pode ser realizada empregando o algoritmo NIPALS (Nonlinear Iterative Partial Least Squares) [38] que é o mais utilizado para tal; porém, existem diversos algoritmos que podem realizar essa decomposição obtendo resultados semelhantes [34].

$$\begin{matrix}
 \begin{matrix} m \\ \mathbf{X} \\ n \end{matrix} & = & \begin{matrix} LV \\ \mathbf{T} \\ n \end{matrix} & \begin{matrix} LV \\ \mathbf{P}' \\ m \end{matrix} & + & \begin{matrix} m \\ \mathbf{E}_X \\ n \end{matrix} \\
 \\
 \begin{matrix} r \\ \mathbf{Y} \\ n \end{matrix} & = & \begin{matrix} LV \\ \mathbf{U} \\ n \end{matrix} & \begin{matrix} LV \\ \mathbf{Q}' \\ r \end{matrix} & + & \begin{matrix} r \\ \mathbf{E}_Y \\ n \end{matrix}
 \end{matrix}$$

Figura 1.13. Esquema da decomposição das matrizes X e Y em escores e pesos pelo método PLS.

O algoritmo NIPALS clássico leva  $\mathbf{u}$  e  $\mathbf{W}$  a ter uma norma de valor unitário, onde  $\mathbf{u}$  é igual à matriz  $\mathbf{Y}$  caso haja somente uma variável dependente, ou será a primeira coluna de  $\mathbf{Y}$  caso haja mais de uma variável dependente, dessa forma,  $\mathbf{u}$  é uma matriz 1x1. A matriz  $\mathbf{W}$  é uma matriz de pesos que substitui a matriz  $\mathbf{P}$  para garantir ortogonalidade de  $\mathbf{T}$  e para que se encontre a melhor covariância entre  $\mathbf{X}$  e  $\mathbf{Y}$ . Somente após a convergência é que  $\mathbf{P}$  será calculado. A Tabela 1.2 explica as etapas do algoritmo NIPALS.

Tabela 1.2. Etapas do algoritmo NIPALS para o método PLS [38,39].

<b>Etapa</b>	<b>Descrição</b>	<b>Equação</b>
1	Escalonar e Centrar na média as matrizes $\mathbf{X}$ e $\mathbf{Y}$ .	
2	Encontrar $\mathbf{u}$ , caso $\mathbf{Y}$ tenha mais de uma variável.	
3	Calcular $\mathbf{W}$ .	$\mathbf{W} = (\mathbf{u}^T \cdot \mathbf{X}) / (\mathbf{u}^T \cdot \mathbf{u})$
4	Normalizar $\mathbf{W}$ .	$\mathbf{W} = (\mathbf{W}) / (\ \mathbf{W}\ )$
5	Calcular $\mathbf{T}$ .	$\mathbf{T} = (\mathbf{X} \cdot \mathbf{W}) / (\mathbf{W}^T \cdot \mathbf{W})$
6	Calcular $\mathbf{q}$ .	$\mathbf{q}^T = (\mathbf{T}^T \cdot \mathbf{Y}) / (\mathbf{T}^T \cdot \mathbf{T})$
7	Normalizar $\mathbf{q}$ .	$\mathbf{q} = (\mathbf{q}) / (\ \mathbf{q}\ )$
8	Calcular $\mathbf{U}$ .	$\mathbf{U} = (\mathbf{Y} \cdot \mathbf{q}) / (\mathbf{q}^T \cdot \mathbf{q})$
9	Testar convergência de $\mathbf{U}$ : se convergiu vá para etapa 9, se não volte para 2.	
10	Calcular a matriz de pesos $\mathbf{P}$ .	$\mathbf{P}^T = (\mathbf{T}^T \cdot \mathbf{X}) / (\mathbf{T}^T \cdot \mathbf{T})$
11	Encontrar a Correlação entre $\mathbf{T}$ e $\mathbf{U}$ .	$\mathbf{b} = (\mathbf{T}^T \cdot \mathbf{U}) / (\mathbf{T}^T \cdot \mathbf{T})$
12	Calcular $\mathbf{E}$ .	$\mathbf{E} = \mathbf{X} - \mathbf{T} \cdot \mathbf{P}^T$
13	Calcular $\mathbf{F}$ .	$\mathbf{F} = \mathbf{Y} - \mathbf{b} \cdot \mathbf{T} \times \mathbf{q}^T$
14	Se o número de LV não é suficiente, substituir $\mathbf{X}$ e $\mathbf{Y}$ por $\mathbf{E}$ e $\mathbf{F}$ , respectivamente, e repetir o procedimento.	

PLS pode ser aplicado para a matriz de variáveis dependentes contendo somente uma ou mais de uma variável. Quando se utiliza uma matriz  $\mathbf{Y}$  contendo somente uma variável dependente (uma única coluna na Figura 1.13) o modelo é dito PLS1, em que o modelo irá refletir apenas a covariância entre as variáveis independentes e a única variável dependente presente em  $\mathbf{Y}$ . Já quando se utiliza uma matriz contendo mais de uma variável dependente gera-se um modelo único, PLS2, onde todas as colunas (variáveis) contribuem para os pesos do modelo. Isso pode melhorar o desempenho do modelo, principalmente quando os dados nas diferentes colunas de  $\mathbf{Y}$  estão correlacionados. No entanto, a vantagem de utilizar o modelo PLS1 é que se pode avaliar cada parâmetro obtido para cada variável separadamente.

## **1.5. CARTAS DE CONTROLE MULTIVARIADAS BASEADAS EM ANÁLISE MPCA**

Uma carta de controle multivariada é uma ferramenta utilizada para análise e identificação de amostras, ou na maioria dos casos bateladas, fora de controle, através do Controle Estatístico Multivariado de Processos (*Multivariate Statistical Process Control – MSPC*) desenvolvido por Nomikos e MacGregor [36, 40], baseado na análise MPCA. As cartas de controle são usadas para detectar mudanças na média ou na relação (covariância) entre os parâmetros relacionados com uma ideia central de monitorar simultaneamente as variáveis que envolvem o processo, controlando somente algumas poucas combinações lineares independentes entre essas variáveis.

A filosofia do MSPC é desenvolver um modelo empírico que possa caracterizar o processo e através da comparação de novos processos com este modelo realizar um monitoramento (acompanhamento). A construção de uma carta de controle é dividida em duas fases distintas: a primeira consiste na elaboração dos parâmetros da carta de controle utilizando as amostras sob condições normais de

operação (NOC). Esses parâmetros obtidos caracterizam as amostras de referência e serão usados para comparações com futuros processos. Assim as amostras NOC devem ser obtidas sem variações estatísticas que possam causar grande alargamento dos limites das cartas; a segunda fase consiste da análise de novas amostras, que serão avaliadas de acordo com os parâmetros estimados na primeira fase [41].

O monitoramento *on-line* de bioprocessos é realizado na forma de bateladas em vários tipos de indústrias, como por exemplo, farmacêuticas, alimentícias, bioquímicas e energéticas e, devido à importância dessas indústrias hoje em dia, são necessárias metodologias para identificação de falhas durante o monitoramento como forma de melhorar a qualidade do produto final. Em geral, os dados obtidos para um monitoramento *on-line* usando espectrômetro Raman podem ser arranjados em um arranjo tridimensional,  $\mathbf{X}$  ( $I \times J \times K$ ), onde na maioria dos casos  $I$ ,  $J$  e  $K$  são números de variáveis espectrais, tempo e de bateladas (Figura 1.10). O controle dessas bateladas é realizado por cartas de controle conhecidas como estatística  $Q$  e estatística *Hotelling's*  $T^2$  obtidas a partir da análise MPCA. A estatística  $Q$  é definida como o quadrado da estimativa do erro do modelo e a estatística *Hotelling's*  $T^2$  definida como a soma dos escores quadráticos e normalizados. Eles propuseram um sistema com base na análise MPCA a partir de modelos construídos com monitoramento *on-line* de bateladas realizadas sob NOC para identificar estatisticamente variações significativas no processo, onde novas bateladas podem ser avaliadas como normais ou falhas usando as cartas de controle.

Antes do desenvolvimento das cartas de controle é necessário realizar análise MPCA e a partir dos resultados obtidos as cartas de controle  $Q$  e  $T^2$  podem ser calculadas. A literatura reporta muitos exemplos do uso de cartas de controle multivariadas na detecção de falhas em bateladas de processos [42-46].

A carta de controle  $T^2$  é obtida de acordo com a equação 1.9 [47], onde " $\mathbf{t}$ " é a matriz de escores e " $\mathbf{S}$ " é a matriz covariância do modelo construído com as bateladas NOC. O limite de controle superior  $T_{lim}$  segue uma distribuição  $F$  e pode

ser calculado de acordo com a equação 1.10 [48], onde "a" é o número de componentes principais no modelo MPCA, "N" o número de amostras e "F" é a distribuição F com N-1 graus de liberdade ao nível de significância  $\alpha$ .

$$T^2 = \mathbf{t}_i \mathbf{S}^{-1} \mathbf{t}_i^T = \mathbf{x}_i \mathbf{P} \mathbf{S}^{-1} \mathbf{P}^T \mathbf{x}_x^T \quad 1.9$$

$$T_{lim} = \frac{a(N-1)}{N-a} F_{(a, N-1; \alpha)} \quad 1.10$$

A carta de controle  $Q$  é obtida a partir dos resíduos das amostras no modelo MPCA e é definida como o somatório dos resíduos ao quadrado, como demonstrado na equação 1.11, seguindo uma distribuição  $\chi^2$ , onde  $x_n$  e  $\hat{x}_n$  são os valores medidos e previstos respectivamente, resultando no valor residual [36, 40]. O limite de controle superior para a carta  $Q$ ,  $Q_{lim}$ , pode ser obtido de acordo com a equação 1.12 [49] seguindo uma distribuição normal, onde  $z_\alpha$  é o desvio padrão com um nível de confiança superior de  $(1-\alpha)$ ,  $h_0 = 1 - 2\theta_1\theta_3/3\theta_2^2$  e  $\theta_i = \sum \lambda_i$ , com  $\lambda_i$  sendo o autovalor associado a cada vetor peso da covariância dos dados.

$$Q = \sum_{n=1}^N (x_n - \hat{x}_n)^2 = \sum_{n=1}^N \mathbf{e}_n^2 = \mathbf{e}_n^T \mathbf{e}_n \quad 1.11$$

$$Q_{lim} = \theta_1 \left[ 1 - \frac{\theta_2 h_0 (1 - h_0)}{\theta_1^2} + \frac{z_\alpha (2\theta_2 h_0^2)^{1/2}}{\theta_1} \right]^{1/h_0} \quad 1.12$$

Usando as equações de 1.9 a 1.12 é possível dizer se uma determinada batelada foi ou não identificada como uma falha. Quando os valores  $T^2$  e  $Q$  obtidos para cada batelada, a partir do modelo MPCA, estiverem acima dos limites de controle,  $T_{lim}$  e  $Q_{lim}$ , encontrados, estas bateladas serão identificadas como falha pelas cartas de controle, como representado esquematicamente na Figura 1.14. Um monitoramento efetivo é realizado empregando as duas cartas de controle para identificar a presença ou não de possíveis alterações indesejáveis durante o processo, onde a carta  $T^2$  é utilizada para controlar a variabilidade das componentes principais selecionadas e a carta  $Q$  para monitorar os resíduos obtidos pelo modelo.

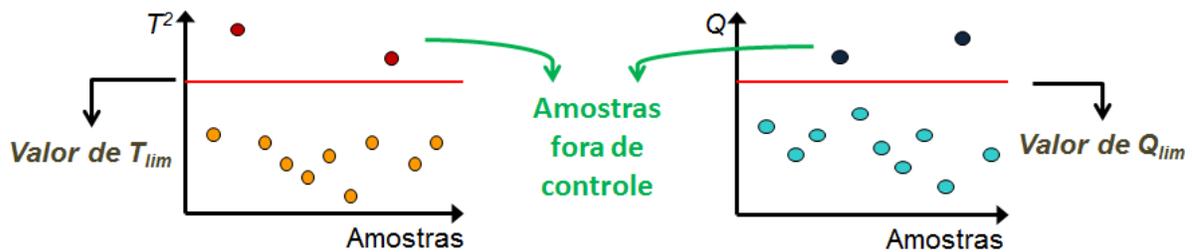


Figura 1.14. Representação esquemática das cartas de controle  $Q$  e  $T^2$  e seus limites identificando amostras fora de controle.

As variações entre bateladas podem ser ocasionadas devido a vários fatores ou combinação deles, como por exemplo, contaminação, matéria-prima inadequada, impurezas e desvios das variáveis de processo em relação à sua trajetória padrão [40]. O controle tem como objetivo manter a menor variação possível entre as bateladas para que os produtos finais estejam dentro dos padrões de qualidade. Caso ocorra alguma variação significativa entre as bateladas NOC e teste, esta estará fora de controle e o produto estará fora dos limites do controle de qualidade.

## 1.6. OBJETIVOS

Este trabalho teve como objetivo principal a utilização da espectroscopia Raman juntamente com métodos quimiométricos no monitoramento *on-line* da fermentação da glicose pela levedura *Saccharomyces cerevisiae*. Para isso o trabalho foi dividido em duas partes:

- Primeira: Quantificação de substrato e produtos de fermentação utilizando calibração multivariada empregando o método dos Mínimos Quadrados Parciais (PLS).
  
- Segunda: Aplicação de cartas de controle multivariadas baseadas na Análise de Componentes Principais Multimodo (MPCA) para identificação de bateladas-falha durante o processo de monitoramento.

## **Capítulo II**

---

# **Quantificação de Substrato e Produtos de Fermentação**



## 2.1. INTRODU  O

O etanol como combust vel tem sido considerado como alternativa para o problema energ tico que envolve o mundo no  mbito do desenvolvimento sustent vel, da escassez de combust veis f sseis e de quest es ambientais. Devido ao grande impacto que a ind stria alcooleira tem tido no mercado, tornou-se fundamental a necessidade de metodologias anal ticas sofisticadas, avan adas e eficientes para melhorar o entendimento e a condu  o de processos fermentativos.

Os monitoramentos *on-line* de processos tem sido bastante explorados em ind strias, uma vez que eliminam etapas de amostragem, que na maioria das vezes s o demoradas e apresentam as maiores fontes de erros. Uma t cnica que pode ser utilizada para o monitoramento de fermenta  es   a espectroscopia Raman, que permite a an lise da amostra (obten  o de espectros) em um curto per odo de tempo quando comparado a outras t cnicas, como por exemplo, a cromatografia e a Resson ncia Magn tica Nuclear.

Devido   import ncia de processos fermentativos na ind stria, a espectroscopia Raman juntamente com a Quimiometria tornou-se uma importante ferramenta para ser aplicada em sistemas de monitoramento, uma vez que h  possibilidade de obten  o de informa  es em tempo real sobre o processo utilizando a calibra  o multivariada. A calibra  o multivariada permite a obten  o de um perfil das concentra  es dos componentes envolvidos na fermenta  o atrav s da constru  o de modelos matem ticos com os dados de espectroscopia Raman quando estes s o calibrados contra os dados de concentra  o obtidos pelos m todos de refer ncia de cada componente.

Tendo em vista a relev ncia dos processos fermentativos no mercado hoje em dia, nesta parte do trabalho foi desenvolvido um monitoramento *on-line* da fermenta  o da glicose para a quantifica  o de substrato e produtos da fermenta  o utilizando a espectroscopia Raman e a calibra  o multivariada baseada no m todo dos M nimos Quadrados Parciais.

## 2.2. PARTE EXPERIMENTAL

### 2.2.1. Instrumenta o

Os instrumentos utilizados no monitoramento da fermenta o foram: reator de 2 L, banho termost tico, espectr metro Dispersivo RamanStation 400F (Perkin Elmer) equipado com laser em 785 nm de 250 mW e detector CCD operando a -50  C na faixa de 3050 a 200  $\text{cm}^{-1}$ . Foi utilizada uma cubeta de quartzo para a leitura em fluxo e para propuls o dos fluidos uma bomba perist ltica Ismatec (Zurique-Su a) de oito canais, contendo um tubo de Tygon  com 1,5 mm de di metro interno e tubo de polietileno 0,8 mm de di metro interno. O esquema do sistema montado para o monitoramento *on-line*   apresentado na Figura 2.1.

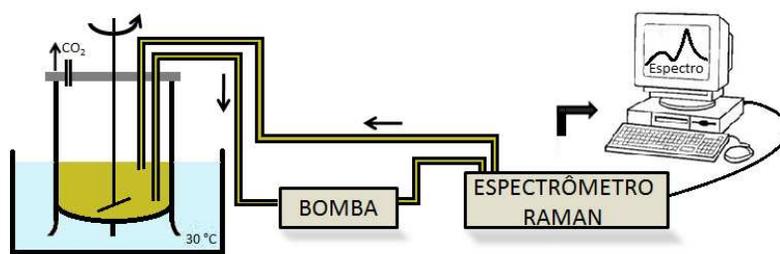


Figura 2.1. Esquema do sistema do experimento *on-line* para monitoramento da fermenta o usando o Espectr metro Raman.

Para o monitoramento *off-line* foi utilizado um cromat grafo HPLC Agilent 1200 Infinity Series LC utilizando uma coluna BIO-RAD Aminex HPX-87H (300 x 7.8

mm) e como fase m vel uma solu o de H<sub>2</sub>SO<sub>4</sub> em pH 2,82 e detec o por  ndice de refra o. Foram empregados filtros da Millipore® de 22  m para filtrar as al quotas para as an lises utilizando HPLC para a quantifica o de glicose, etanol, glicerol e  cido ac tico. J  para a quantifica o de c lulas foi utilizado Espectrofot metro Cary 50 UV-Vis (Varian) operando em 600 nm.

### **2.2.2. Micro-organismos**

Os micro-organismos utilizados s o de uma linhagem isolada pelo Laborat rio de Gen mica e Express o do Instituto de Biologia da Universidade Estadual de Campinas, chamada de JAY270, que ficam armazenados a - 80  C. Na literatura [50] s o relatados o isolamento de v rias linhagens de micro-organismos utilizados em processos fermentativos em usinas de  lcool do Brasil, incluindo a PE-2, que deu origem   JAY270, que   derivada de uma levedura usada por muitas usinas, por produzirem alto teor de etanol e resistir a todos os estresses causados por um processo fermentativo em escala industrial. Na literatura [29] encontra-se a caracteriza o gen tica da linhagem JAY270 que ajuda entender o que a torna t o adaptada aos processos industriais das usinas.

Os reagentes utilizados para o preparo do meio de cultura para o crescimento e fermenta o foram  gua Milli-Q da Millipore®, D-glicose Anidra (Synth), extrato de levedura em p  (Himedia) e peptona bacteriol gica (Himedia).

O crescimento dos micro-organismos foi realizado em meio contendo D-glicose 2%, peptona 2% e extrato de levedura 1% durante dois dias a 30  C em rota o de 250 rpm em um shaker, sendo que a cada 12 horas o meio de cultura de crescimento foi trocado. Ap s o crescimento das c lulas, estas foram centrifugadas a 5000 rpm e as c lulas  midas foram ressuspensas no meio de cultura de fermenta o e utilizadas para o experimento.

### 2.2.3. Processo de Monitoramento

O sistema de monitoramento consistiu na fermenta o da glicose pela levedura *Saccharomyces cerevisiae* produzindo, como maiores constituintes, etanol,  cido ac tico e glicerol. Para o experimento, 800 mL de solu o contendo D-glicose 10%, peptona 2%, extrato de levedura 1% e 9,7 gramas de c lulas de  midas foram colocados no reator em banho termost tico a 30  C, sob agita o mec nica de 100 rpm. Antes do processo todo material utilizado foi esterilizado em autoclave.

A fermenta o foi monitorada durante 14 horas a uma vaz o de 10 mL min<sup>-1</sup>. Os espectros foram obtidos a cada 15 minutos usando um espectr metro Raman na faixa de 1800 a 300 cm<sup>-1</sup> com resolu o de 4 cm<sup>-1</sup>. Os espectros foram obtidos com 100 exposi es do laser de 1 segundo cada, sendo o espectro final a m dia das 100 exposi es realizadas.

No desenvolvimento dos modelos de calibra o multivariada e valida o foi realizado um monitoramento *off-line* utilizando HPLC para quantifica o de glicose, etanol,  cido ac tico e glicerol. Neste procedimento,  lquotas de 1,0 mL foram retiradas a cada 30 minutos e armazenadas a -20  C. Ap s esse procedimento, as  lquotas foram descongeladas e centrifugadas durante 15 minutos a 5000 rpm e em seguida o sobrenadante foi filtrado em filtros Millipore® 22  m para a an lise em HPLC. Outro monitoramento *off-line* foi realizado utilizando o espectrofot metro UV-Vis para quantifica o de c lulas. Neste caso,  lquotas de 1,0 mL foram retiradas a cada 60 minutos e efetuada as leituras em seguida. Para ambos os m todos de refer ncia, foram constru das curvas de calibra o para encontrar a concentra o do analito.

Dessa forma, foram obtidos 57 espectros para o monitoramento *on-line*, 29  lquotas para o monitoramento *off-line* usando o HPLC e 15  lquotas para o monitoramento *off-line* usando o espectrofot metro UV-Vis.

#### 2.2.4. An lise Quimiom trica

O tratamento dos dados foi realizado em ambiente MATLAB R2011b, usando o PLSToolbox 6.2.1. para desenvolver os modelos de calibra o multivariada usando o PLS.

No desenvolvimento dos modelos PLS para quantifica o de glicose, etanol,  cido ac tico e glicerol os espectros obtidos no monitoramento *on-line* foram divididos em 2 grupos. O primeiro grupo apresentou os espectros obtidos no mesmo tempo de aquisi o das al quotas para o monitoramento *off-line*, totalizando 29 espectros, onde 19 foram utilizados no conjunto de calibra o e 10 no conjunto de valida o para glicose, etanol,  cido ac tico e glicerol. Para quantifica o de c lulas, foram 15 espectros, sendo 10 no conjunto de calibra o e 5 no conjunto de valida o. Assim, os valores obtidos pelos m todos de refer ncia, HPLC e espectr metro UV/Vis, foram utilizados no desenvolvimento dos modelos de calibra o. O segundo grupo apresentou os espectros obtidos no monitoramento *on-line* em tempos onde as al quotas para o monitoramento *off-line* n o estavam dispon veis, contendo 28 espectros para previs o de glicose, etanol,  cido ac tico e glicerol e 42 espectros para previs o da concentra o de c lulas. Estes espectros formaram o conjunto de previs o, utilizados para acompanhar a fermenta o. A escolha das amostras foi realizada de forma que as amostras de calibra o e de previs o estivessem tanto no in cio como no meio e fim da faixa de calibra o.

Nos modelos desenvolvidos, o n mero ideal de vari veis latentes foi selecionado considerando os valores da raiz do erro quadr tico m dio de valida o cruzada - RMSECV (*Root Mean Square Error of Cross Validation*) obtidos por valida o cruzada dos dados de calibra o. As amostras identificadas como an malas na calibra o dos modelos foram descartadas, j  que essas amostras podem produzir modelos de previs o err neos, e tanto na calibra o quanto na valida o   necess rio descart -las de seus respectivos conjuntos. Al m dos valores

de RMSECV, tamb m foram calculados os valores da raiz do erro quadr tico m dio de calibra o (*Root Mean Square Error of Calibration – RMSEC*) e de previs o (*Root Mean Square Error of Prediction – RMSEP*, respectivamente),.

## 2.3. RESULTADOS E DISCUSS O

### 2.3.1. Espectros Raman

A Figura 2.2 mostra os espectros Raman para os componentes puros presentes no monitoramento. Como pode ser observado, os espectros apresentam muitas bandas que se sobrep em e por isso ha necessidade do uso de m todos quimiom tricos no tratamento dos dados.

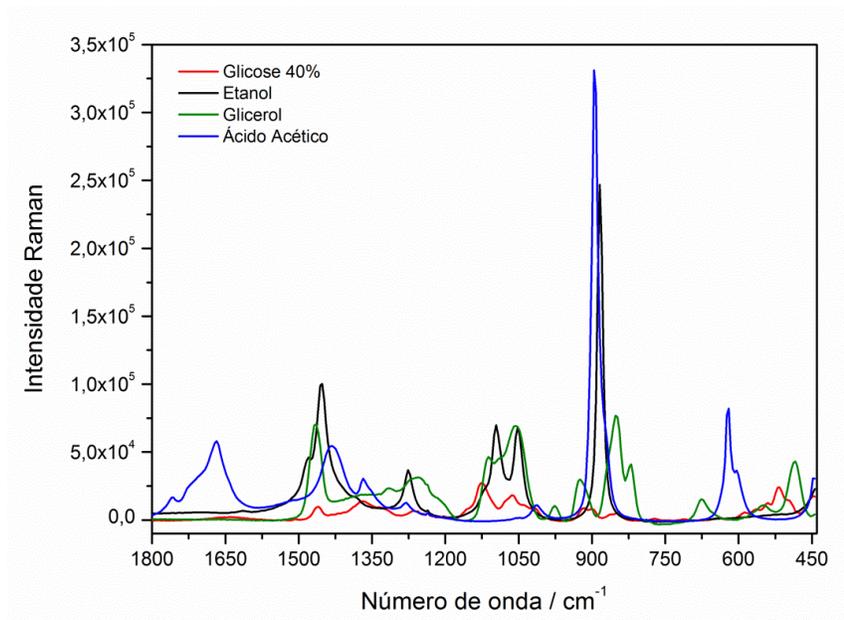


Figura 2.2. Espectros Raman dos produtos e substrato puros da fermenta o.

Os 57 espectros obtidos durante a fermenta  o est o apresentados na Figura 2.3. Nesta figura podemos identificar a mudan a de intensidade dos picos, alguns se tornam mais intensos e outros menos. J  a Figura 2.4 apresenta os espectros obtidos no tempo inicial e final para monitoramento da fermenta  o, sendo n tida a percep  o de mudan a entre esses espectros obtidos nos tempos de 0 e 14 horas.

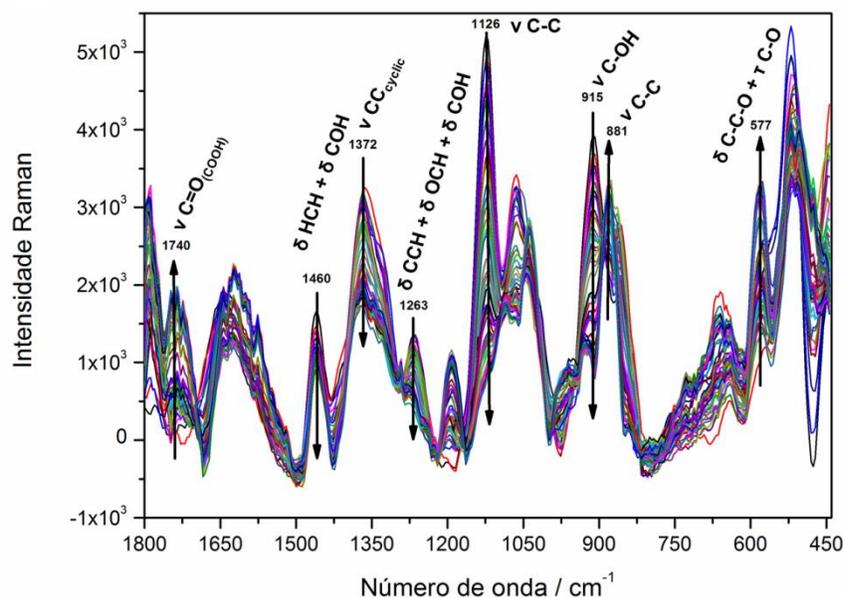


Figura 2.3. Espectros Raman da fermenta  o da glicose pela *Saccharomyces cerevisiae*.

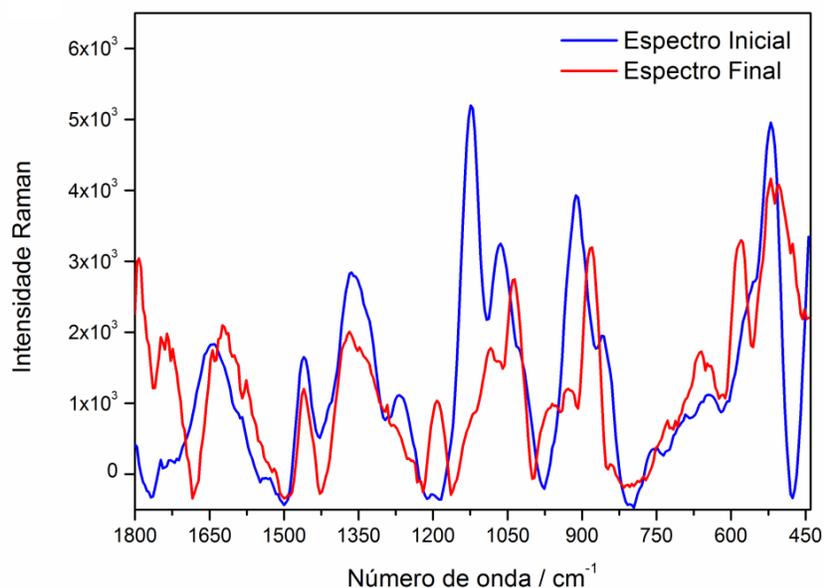


Figura 2.4. Espectros Raman no tempo inicial e final da fermenta o.

Na an lise das Figuras 2.3 e 2.4 podemos identificar a diminui o de intensidade dos picos em 1126 e 915  $\text{cm}^{-1}$  e o aumento de intensidade em 881  $\text{cm}^{-1}$ . Os picos em 1126 e 915  $\text{cm}^{-1}$  s o, respectivamente, caracter sticos dos estiramentos CC [21] e COH [51] da glicose. O pico em 881  $\text{cm}^{-1}$    caracter stico para etanol, um estiramento da liga o CC [21]. Para o glicerol os picos em 1254, 1109 e 1049  $\text{cm}^{-1}$  s o para a tor o da liga o  $\text{CH}_2$  e estiramentos COH prim rio e secund rio, respectivamente [52], e para o  cido ac tico o pico em 1666  $\text{cm}^{-1}$    devido ao estiramento C=O, que em solu es dilu das tendem a se deslocar para 1710  $\text{cm}^{-1}$  [53]. No entanto, como o glicerol e o  cido ac tico est o presentes em pequenas quantidades, seus picos n o s o t o intensos para serem identificados nos espectros como para os picos da glicose e o etanol. Tamb m na Figura 2.3 podemos observar a exist ncia de outros picos importantes, como por exemplo, em 1740  $\text{cm}^{-1}$  devido ao estiramento  $\text{C}=\text{O}_{(\text{COOH})}$  [54], 1460  $\text{cm}^{-1}$  devido   dobra das liga es HCH e COH [51, 55], 1372  $\text{cm}^{-1}$  devido ao estiramento  $\text{CC}_{\text{c clico}}$  [55], 1263  $\text{cm}^{-1}$  devido   deforma o

das liga es CCH, OCH e COH [10], e  $577\text{ cm}^{-1}$  devido   deforma o da liga o CCO e   tor o da liga o CO [10].

### 2.3.2. An lise dos Dados

A regi o do espectro utilizada para constru o do modelo PLS foi de 1800 a  $440\text{ cm}^{-1}$ , por apresentar os picos mais intensos. O pr -processamento empregado foi alisamento empregando o algoritmo de Savitzky-Golay [31] com janela de 15 pontos e um polin mio de segunda ordem. Esse pr -processamento foi usado para diminuir os ru dos dos dados. Outros tipos de pr -processamentos foram testados; no entanto n o apresentaram melhorias aos resultados dos modelos de regress o PLS. Os modelos de regress o PLS foram desenvolvidos realizando centragem na m dia e valida o cruzada, retirando uma amostra por vez (*leave one out*), para a previs o das quantidades de glicose, etanol,  cido ac tico, glicerol e c lulas.

A previs o de amostras obtidas no monitoramento *on-line* que n o apresentam dados correspondentes ao monitoramento *off-line* foi realizada para comparar os perfis de concentra o obtidos com aqueles onde se conhecia os valores obtidos pelos m todos de refer ncia. Foram constru dos tamb m modelos de regress o PLS1 e PLS2 a fim de comparar os resultados obtidos pelas duas metodologias.

#### 2.3.2.1. PLS1

Os modelos para quantificar o substrato e os produtos foram constru dos usando o primeiro grupo de espectros, em que todas as amostras foram obtidas em tempos que apresentam dados correspondentes ao m todo de refer ncia. Durante a constru o dos modelos para glicose, etanol, glicerol e  cido ac tico, algumas amostras foram identificadas como an malas no conjunto de dados de calibra o e

de valida o. Essas amostras podem ter sido identificadas como an malas, pois foram obtidas durante o in cio do monitoramento, nos tempos 0 e 0,5 horas, respectivamente para a calibra o e valida o, devido   problemas de homogeneidade no in cio do monitoramento ap s a mistura do substrato e in cio do processo fermentativo. Ap s a exclus o dessas amostras os modelos foram constru dos. O modelo constru do para quantifica o de c lulas n o identificou a presen a de amostras an malas. Na Tabela 2.1 est o apresentados os resultados para os modelos PLS1 constru dos para quantifica o de glicose, etanol, glicerol,  cido ac tico e c lulas, os modelos apresentaram coeficientes de determina o acima de 96,77 % e baixos valores de erros, tanto para a etapa de calibra o quanto para a etapa de valida o. As Figuras 2.5 a 2.9 mostram que os perfis de concentra o para calibra o e valida o obtidos pelos modelos est o condizentes ao perfil de concentra o obtido pelos m todos de refer ncia.

Tabela 2.1. Resultados para os modelos PLS1 desenvolvidos usando o grupo de espectros obtidos no monitoramento *on-line*.

Modelo PLS	N� de amostras		N� de LV	% de Vari�ncia Explicada	R <sup>2</sup>	RMSEC*	RMSECV*	RMSEP*	Intervalo de Calibra�o*
	Calibra�o	Valida�o							
Glicose	18	9	2	96,77	0,974	0,69	0,84	0,43	0,06 – 10,01
Etanol	18	9	2	98,95	0,981	0,14	0,20	0,25	0,48 – 4,28
Glicerol	18	9	2	99,52	0,993	0,013	0,017	0,018	0,11 – 0,65
�cido Ac�tico	18	9	2	98,87	0,983	0,0021	0,0031	0,0032	0,02 – 0,08
C�lulas	10	5	2	97,93	0,979	1,28	2,06	1,14	8,19 – 30,84

\* % para Glicose, Etanol,  cido Ac tico e Glicerol e g L<sup>-1</sup> para C lulas.

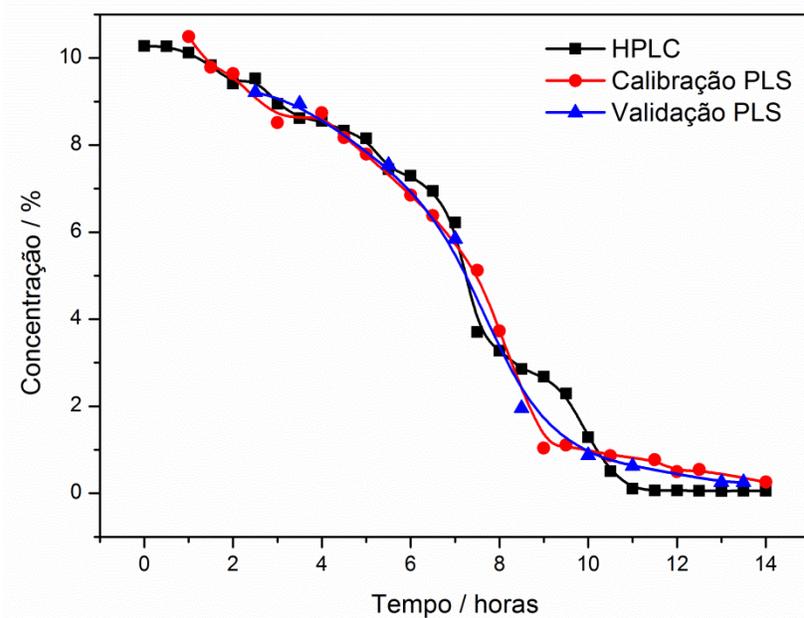


Figura 2.5. Perfil de concentra o obtido pela calibra o e valida o do modelo PLS e pelo m todo de refer ncia para glicose.

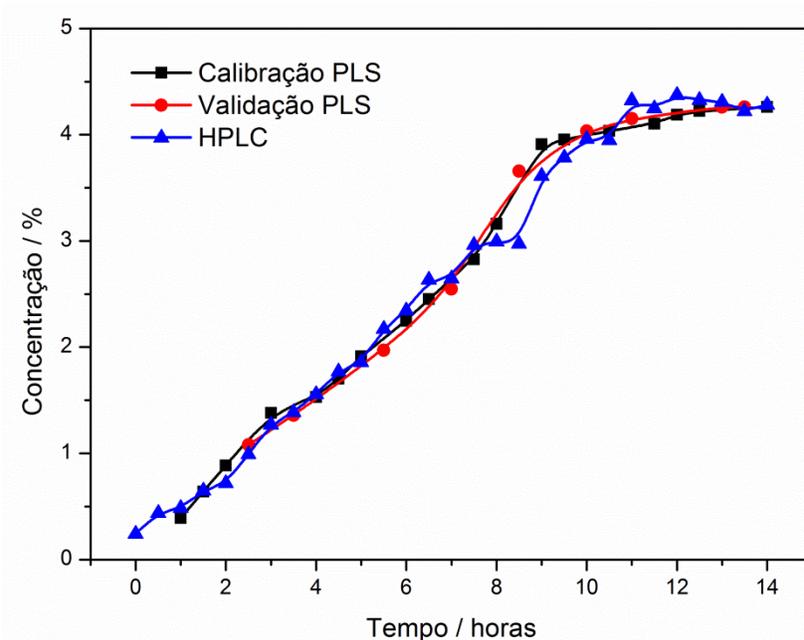


Figura 2.6. Perfil de concentra o obtido pela calibra o e valida o do modelo PLS e pelo m todo de refer ncia para etanol.

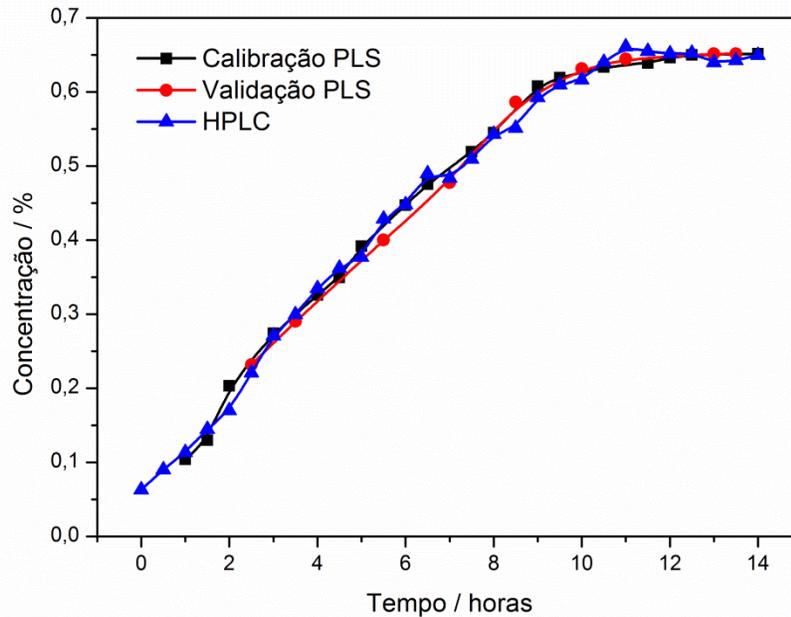


Figura 2.7. Perfil de concentra o obtido pela calibra o e valida o do modelo PLS e pelo m todo de refer ncia para glicerol.

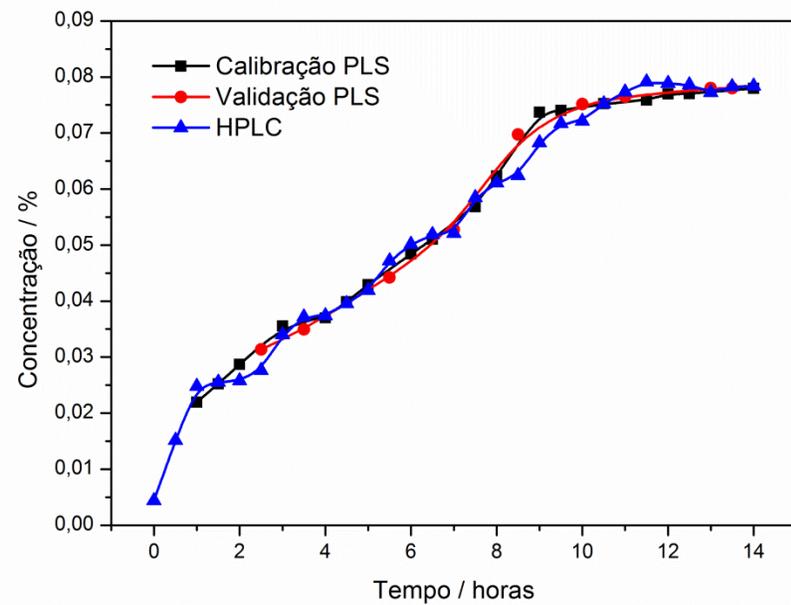


Figura 2.8. Perfil de concentra o obtido pela calibra o e valida o do modelo PLS e pelo m todo de refer ncia para  cido ac tico.

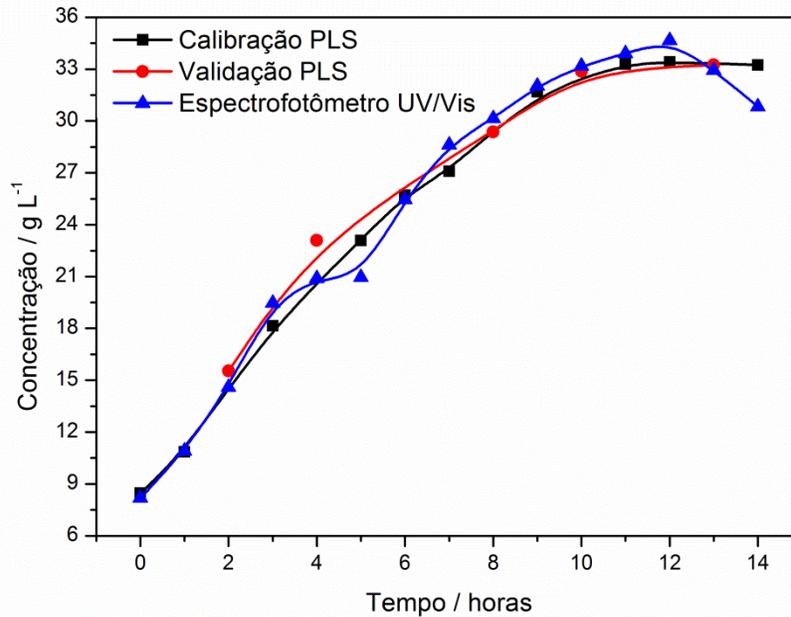


Figura 2.9. Perfil de concentra  o obtido pela calibra  o e valida  o do modelo PLS e pelo m todo de refer ncia para c elulas.

### 2.3.2.1.1. Coeficientes de Regress o do modelo PLS1

Na Figura 2.10 est o apresentados os coeficientes de regress o obtidos para os modelos desenvolvidos com o primeiro grupo de espectros. A an lise dos coeficientes de regress o revela as vari veis espectrais mais importantes do espectro para cada componente. Quanto maior o valor para o coeficiente de regress o maior   sua sensibilidade para aquela determinada vari vel, no caso n mero de onda. Dessa forma, a Figura 2.10 revela que a regi o do espectro entre 1180 e 800 cm<sup>-1</sup> foi importante para todos os componentes estudados e que os n meros de onda de 1725, 1630, 585 e 480 cm<sup>-1</sup> tamb m se destacaram em todos os modelos constru dos.

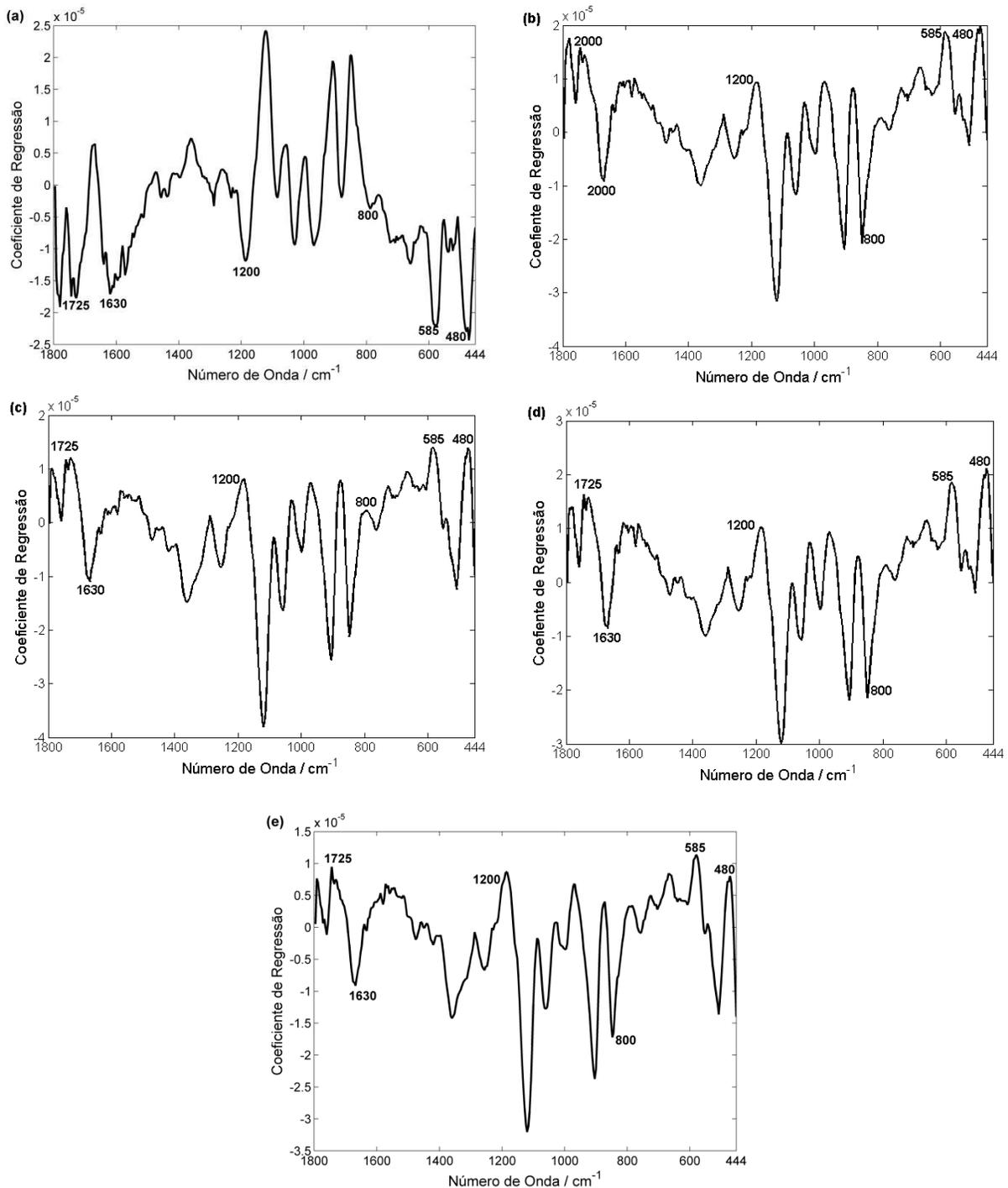


Figura 2.10. Coeficientes de Regress o obtidos pelos modelos PLS1 constru dos com o primeiro grupo de espectros para (a) glicose, (b) etanol, (c) glicerol, (d)  cido ac tico e (e) c lulas.

O fato dos coeficientes de regress o (RC's) serem diferentes para cada componente revela que os modelos foram eficientes na identifica o dos constituintes, mas a Figura 2.10 revelou que os RC's dos componentes estudados s o semelhantes e que as vari veis mais importantes na constru o dos modelos s o as mesmas. Uma prova dessa semelhan a pode ser explicada pela regress o linear feita entre os RC's e apresentada na Figura 2.11. Houve uma alta correla o linear entre os RC's, tendo em vista os valores dos coeficientes de regress o,  $R^2$ , acima de 0,70. No entanto, n o podemos dizer que os modelos constru dos n o est o corretos ou superajustados uma vez que se tem o conhecimento de que em um processo fermentativo temos o consumo de um substrato para forma o de produtos e que a taxa de consumo do substrato est  relacionada   taxa de forma o dos produtos.

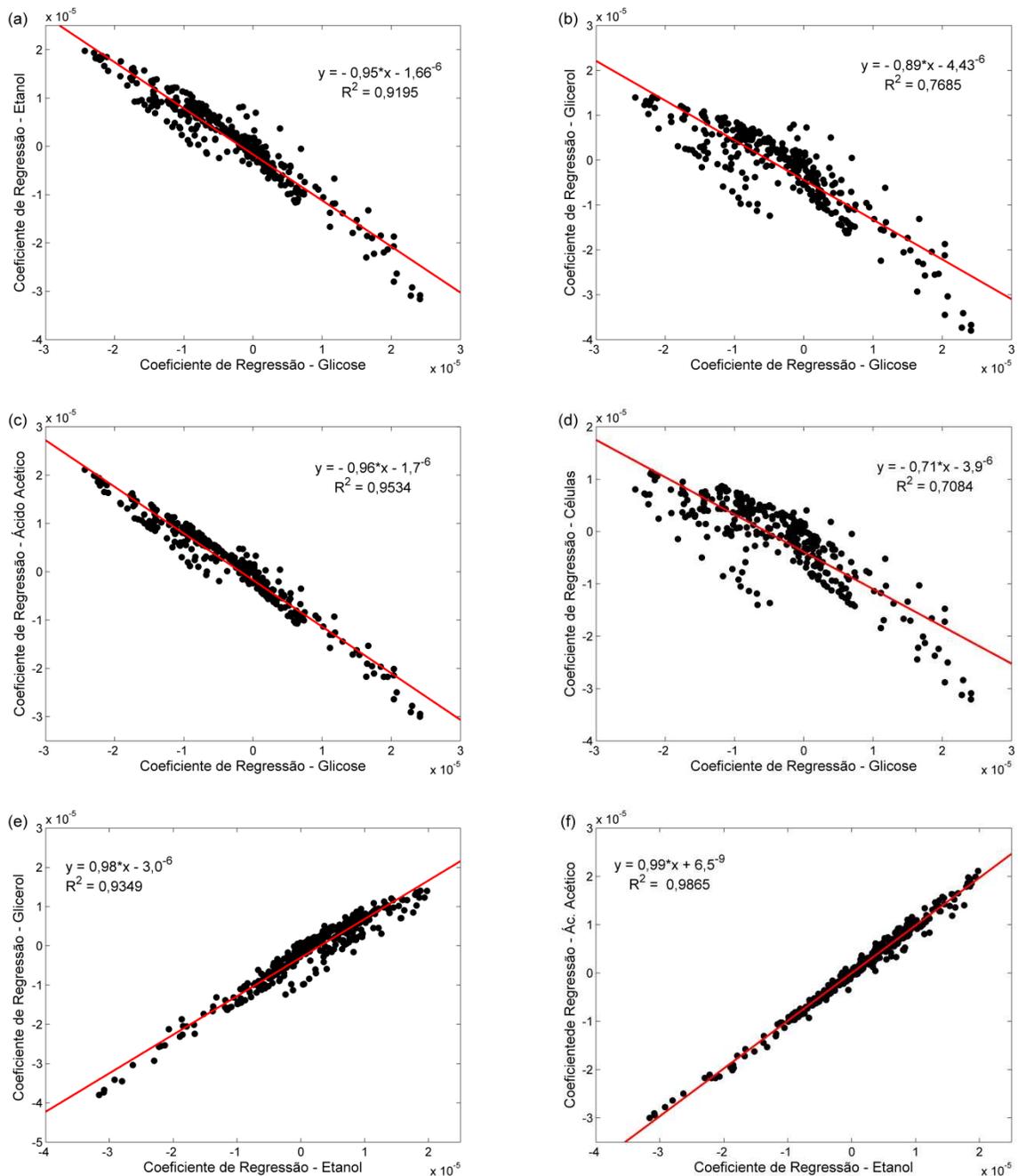


Figura 2.11. Regress o linear entre os coeficientes de regress o (a) da glicose e etanol, (b) glicose e glicerol, (c) glicose e  cido ac tico, (d) glicose e c lulas, (e) etanol e glicerol, (f) etanol e  cido ac tico, (g) etanol e c lulas, (h) glicerol e  cido ac tico, (i) glicerol e c lulas e (j)  cido ac tico e c lulas.

Continua o da Figura 2.11.

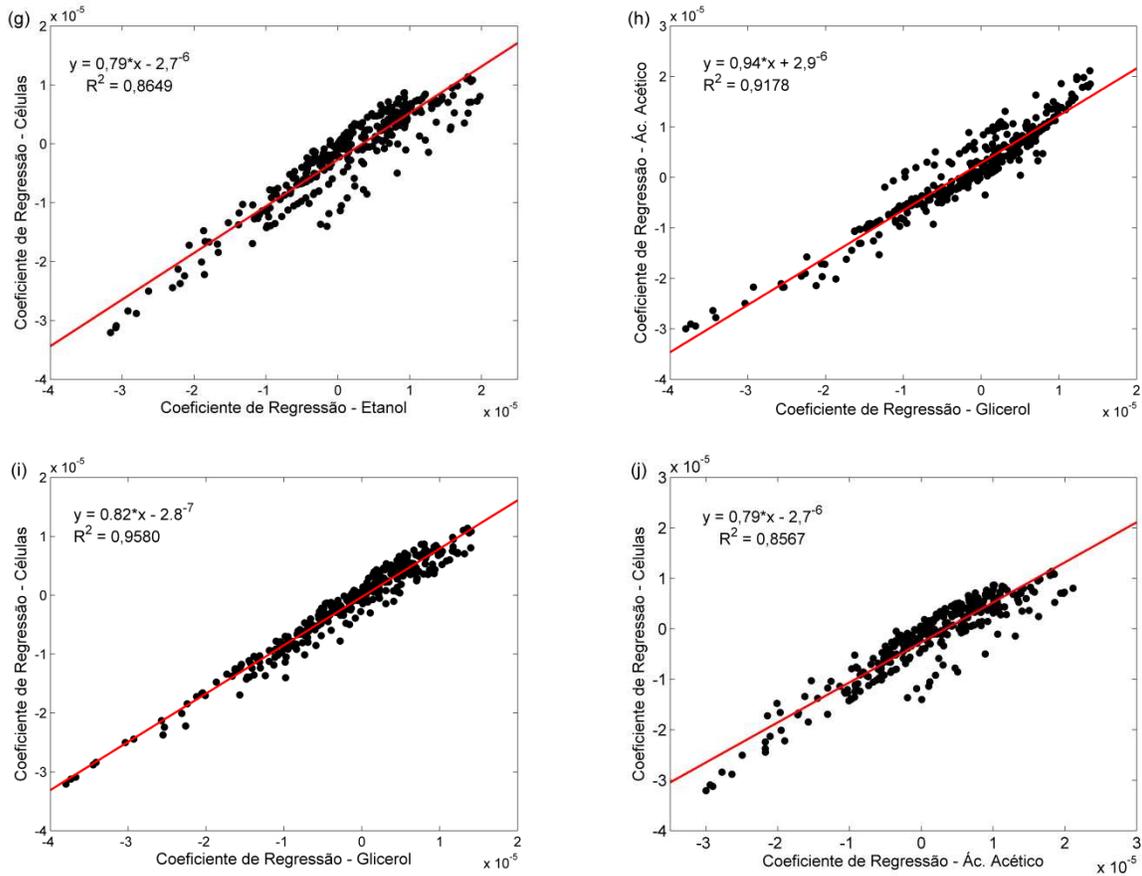


Figura 2.11. Regress o linear entre os coeficientes de regress o (a) da glicose e etanol, (b) glicose e glicerol, (c) glicose e  cido ac tico, (d) glicose e c lulas, (e) etanol e glicerol, (f) etanol e  cido ac tico, (g) etanol e c lulas, (h) glicerol e  cido ac tico, (i) glicerol e c lulas e (j)  cido ac tico e c lulas.

Uma regress o linear entre as concentra es obtidas pelos m todos de refer ncia (Figura 2.12) comprova a correla o entre substrato e produtos do processo fermentativo. Sendo assim pode-se dizer que os modelos PLS1 s o v lidos, uma vez que essa correla o entre os componentes estudados s o intr secas ao processo fermentativo monitorado. Mas um cuidado especial deve ser tomado nesses casos, uma vez que uma varia o desconhecida ou inesperada da concentra o do substrato ou nas condi es de processos acarretar  uma mudan a no perfil do processo tornando os modelos inv lidos para a quantifica o dos componentes.

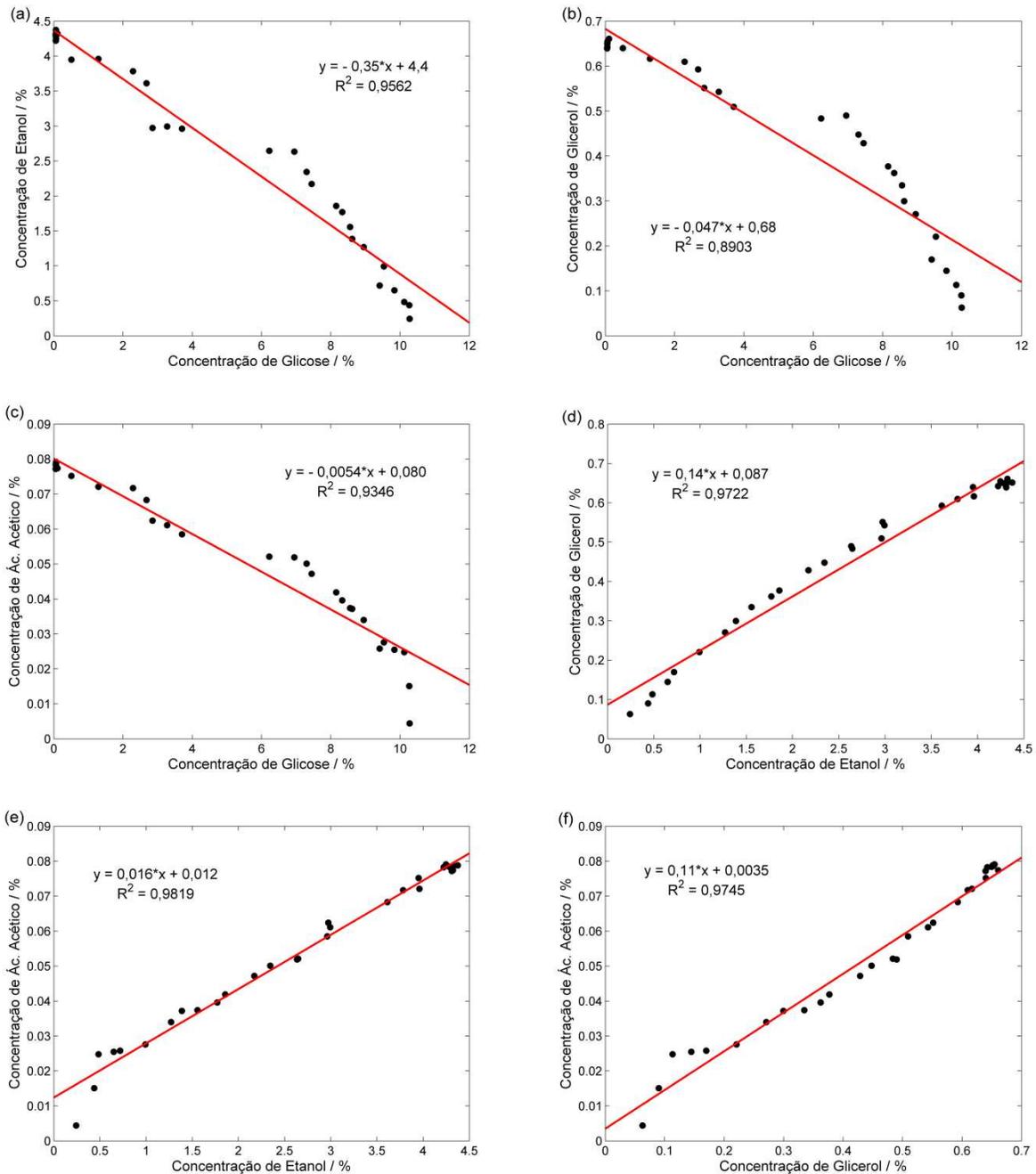


Figura 2.12. Correla o linear entre as concentra es obtidas pelos m todos de refer ncia, (a) glicose e etanol, (b) glicose e glicerol, (c) glicose e  cido ac tico, (d) etanol e glicerol, (e) etanol e  cido ac tico, (f) glicerol e  cido ac tico, (g) c lulas e glicose, (h) c lulas e etanol, (i) c lulas e glicerol e (j) c lulas e  cido ac tico.

Continua o da Figura 2.12.

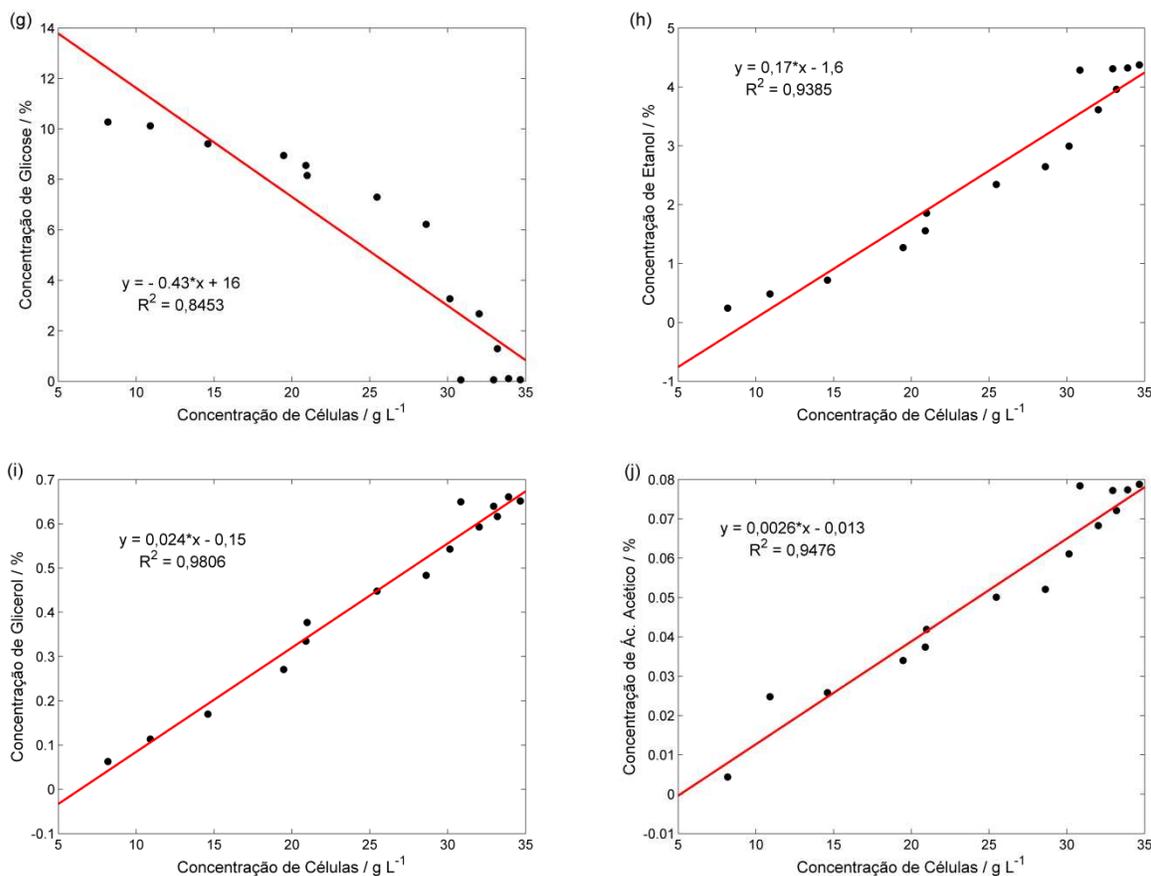


Figura 2.12. Correla o linear entre as concentra es obtidas pelos m todos de refer ncia, (a) glicose e etanol, (b) glicose e glicerol, (c) glicose e  cido ac tico, (d) etanol e glicerol, (e) etanol e  cido ac tico, (f) glicerol e  cido ac tico, (g) c elulas e glicose, (h) c elulas e etanol, (i) c elulas e glicerol e (j) c elulas e  cido ac tico.

### 2.3.2.1.2. Modelos para previs o das amostras obtidas no monitoramento on-line que n o apresentam dados correspondentes ao monitoramento off-line.

A previs o para as amostras com espectros sem dados correspondentes ao monitoramento *off-line*, segundo grupo de espectros, foi realizada utilizando todas as amostras que apresentam dados correspondentes ao monitoramento *off-line* para o

desenvolvimento dos modelos (fase de calibra o), primeiro grupo de espectros. Dessa forma, o n mero de amostras na matriz de calibra o e de previs o foram 29 e 28, respectivamente nos modelos para glicose, etanol, glicerol e  cido ac tico, e de 15 e 42 no modelo para quantifica o de c lulas. Na constru o dos modelos, no conjunto de calibra o, as amostras obtidas nos tempos de 0 e 0,50 horas para glicose, etanol e glicerol e no conjunto de previs o, as amostras de 0,25; 0,75 e 1,25 horas para glicose, etanol e glicerol foram identificadas como an malas e foram desconsideradas. No modelo constru do para quantifica o de  cido ac tico e c lulas n o foram identificadas a presen a de amostras an malas. Os modelos de previs o foram constru dos e os dados para estes est o apresentados na Tabela 2.2. A an lise dos valores de RMSEC e RMSECV obtidos pelos modelos de previs o s o semelhantes  queles obtidos pelos modelos de calibra o, apresentados na Tabela 2.1, evidenciando que os modelos de previs o podem ser empregados para previs o de novas amostras com baixo erro. As Figuras 2.13 a 2.17 ajudam a visualizar este fato, pois os perfis de concentra es previstos gerados pelo modelo est o coerentes com os perfis de concentra o obtidos pelos m todos de refer ncia.

Tabela 2.2. Resultados para os modelos PLS1 constru dos para previs o das amostras obtidas no monitoramento *on-line* que n o apresentam dados correspondentes ao monitoramento *off-line*.

Modelo PLS	N�mero de amostras		N�mero de LV	% de vari�ncia explicada	RMSEC*	RMSECV*
	Calibra�o	Valida�o				
Glicose	27	25	2	97,47	0,61	0,69
Etanol	27	25	2	98,44	0,16	0,21
Glicerol	27	25	2	99,39	0,014	0,017
�cido Ac�tico	29	28	2	98,96	0,0022	0,0026
C�lulas	15	42	2	98,27	1,11	1,57

\* % para Glicose, Etanol,  cido Ac tico e Glicerol e g L<sup>-1</sup> para C lulas.

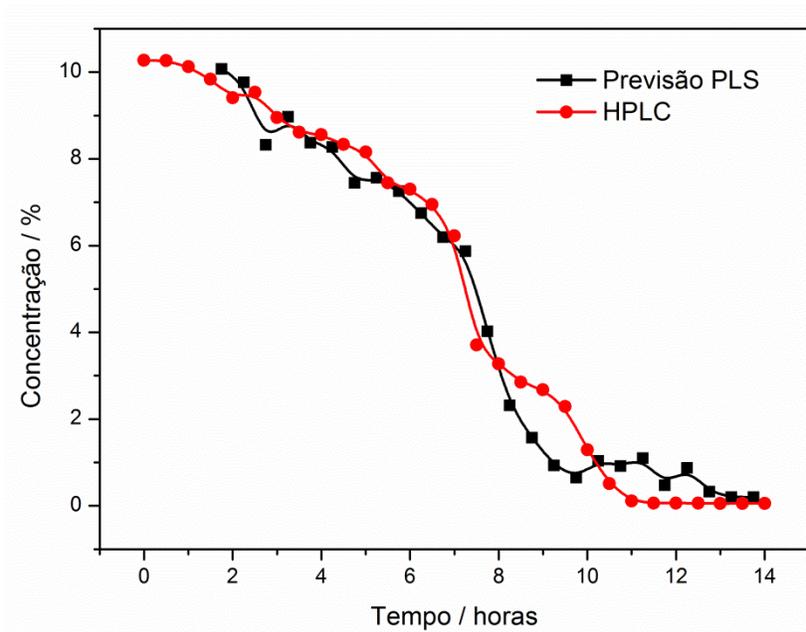


Figura 2.13. Perfil de concentra o previsto pelo modelo PLS e pelo m todo de refer ncia para glicose.

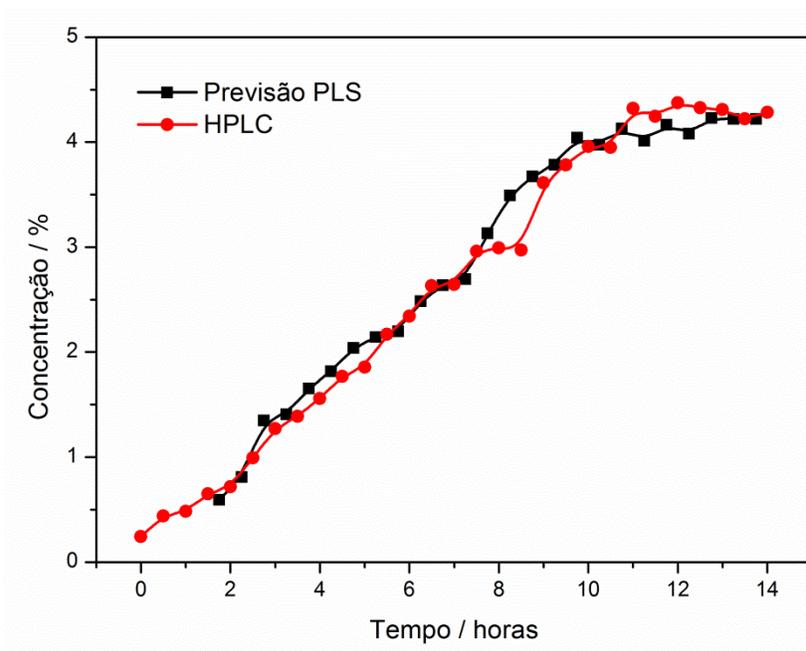


Figura 2.14. Perfil de concentra o previsto pelo modelo PLS e pelo m todo de refer ncia para etanol.

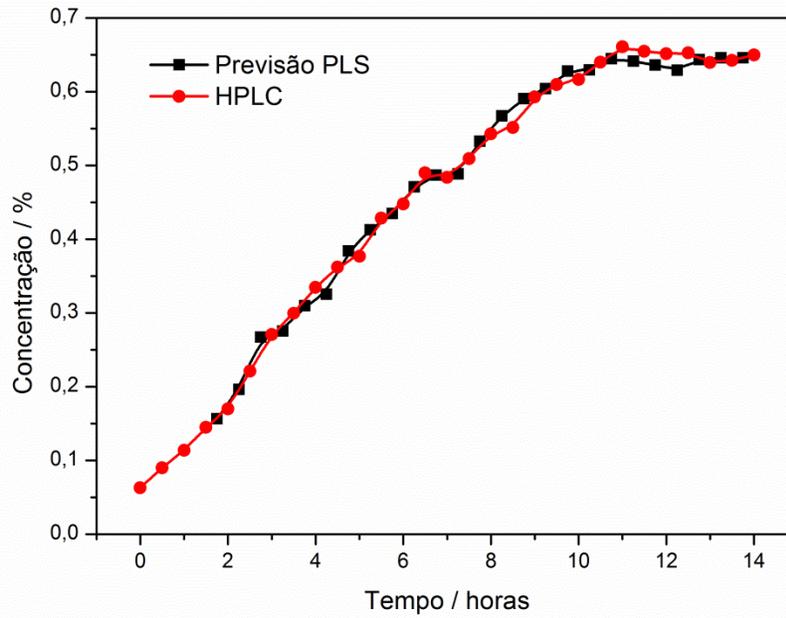


Figura 2.15. Perfil de concentra o previsto pelo modelo PLS e pelo m todo de refer ncia para glicerol.

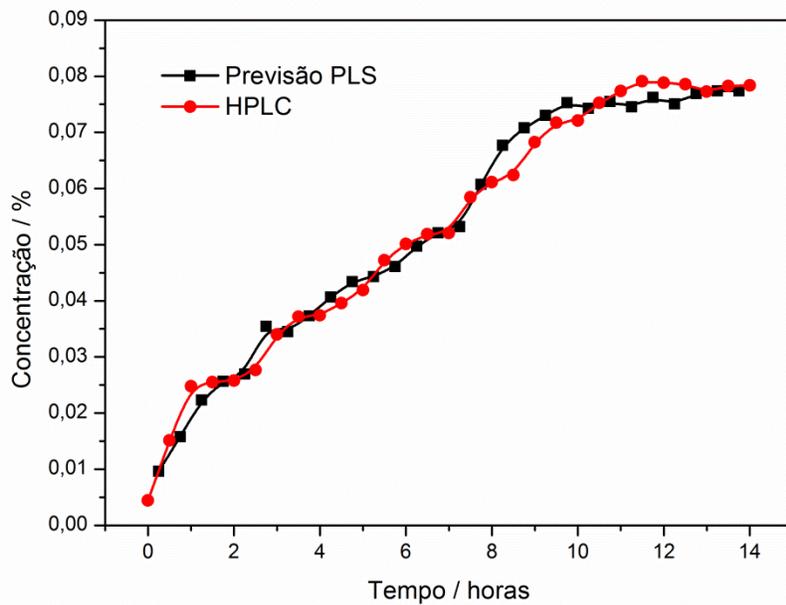


Figura 2.16. Perfil de concentra o previsto pelo modelo PLS e pelo m todo de refer ncia para  cido ac tico.

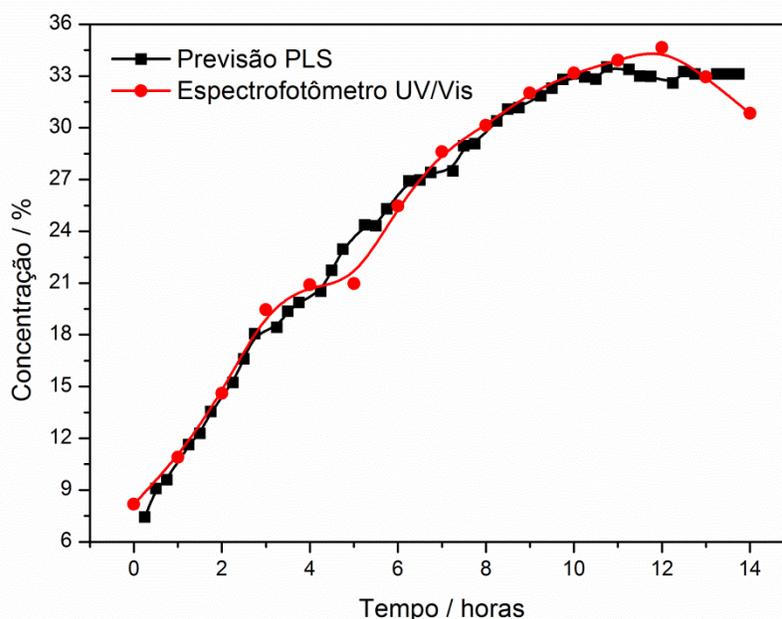


Figura 2.17. Perfil de concentra o previsto pelo modelo PLS e pelo m todo de refer ncia para c lulas.

### 2.3.2.2. PLS2

Como foi discutido na sess o 2.3.2.1.1, a correla o existente entre os componentes estudados no monitoramento *on-line* permite o desenvolvimento de um modelo PLS2, j  que este se aplica muito bem em sistemas onde as vari veis dependentes apresentam correla o entre si. Sendo assim um modelo PLS2 foi constru do a fim de conhecer e comparar os perfis de concentra o obtidos pelos modelos PLS1 e PLS2. No entanto nesta etapa considerou-se apenas os compostos glicose, etanol, glicerol e  cido ac tico, descartando a previs o da quantidade de c lulas devido ao reduzido n mero de amostras dispon veis para este analito, j  que para glicose, etanol, glicerol e  cido ac tico foram obtidos 29 amostras com dados correspondentes ao monitoramento *off-line* e para c lulas apenas 15. Para que o modelo PLS2 n o fosse constru do apenas com 15 amostras decidiu-se, ent o, descartar a quantifica o de c lulas nesta etapa.

As amostras obtidas nos tempos de 0 e 0,5 horas no conjunto de calibra o e nos tempos de 0,25; 0,75 e 1,25 horas no conjunto de previs o foram identificadas como an malas e removidas do modelo. O modelo PLS2 constru do sem essas amostras e com 2 vari veis latentes explicou 98,16 % da vari ncia de **Y** (vari veis dependentes). A compara o dos valores de RMSEC e RMSECV obtidos pelo modelo PLS2, apresentados na Tabela 2.3, com os valores obtidos pelos modelos PLS1, apresentados na Tabela 2.2, revela que, neste caso, o uso do modelo PLS2   adequado e traz vantagens como o menor n mero de modelos, uma vez que o modelo PLS2   constru do para todas as vari veis dependentes simultaneamente, sem necessidade da constru o de um modelo para cada vari vel independente e que os pesos trazem informa o de todas vari veis j  que estes s o influenciados igualmente por todas vari veis na constru o do modelo. Outro fato que contribui para esta conclus o   a coer ncia entre os perfis de concentra o obtidos pelo modelo PLS2 e os perfis de concentra o obtidos pelos m todos de refer ncia, apresentados nas Figuras 2.18 a 2.21.

Tabela 2.3. Valores de RMSEC e RMSECV obtidos pelo modelo PLS2 para cada analito.

<b>Analito</b>	<b>RMSEC / %</b>	<b>RMSECV / %</b>
Glicose	0,63	0,71
Etanol	0,19	0,21
Glicerol	0,015	0,018
�cido Ac�tico	0,0026	0,0029

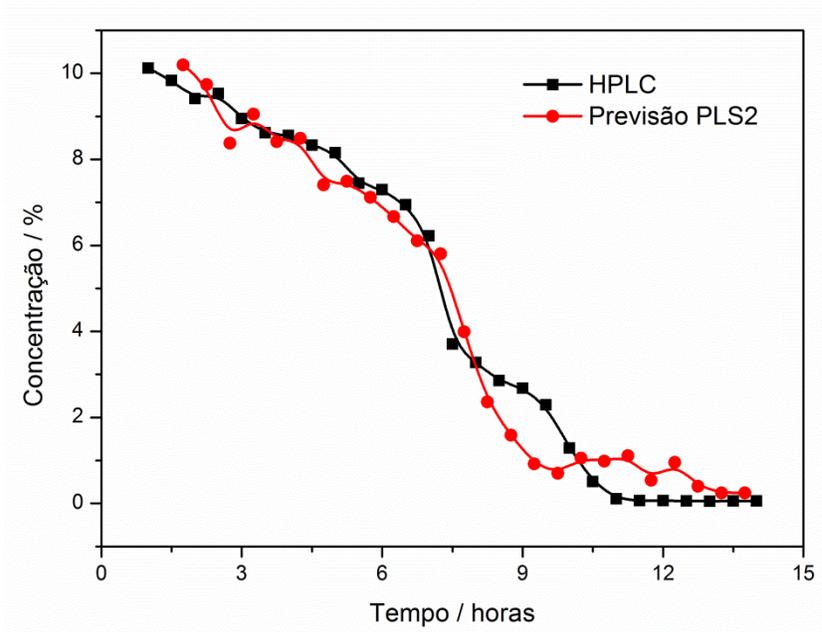


Figura 2.18. Perfil de concentra o previsto pelo modelo PLS2 e pelo m todo de refer ncia para glicose.

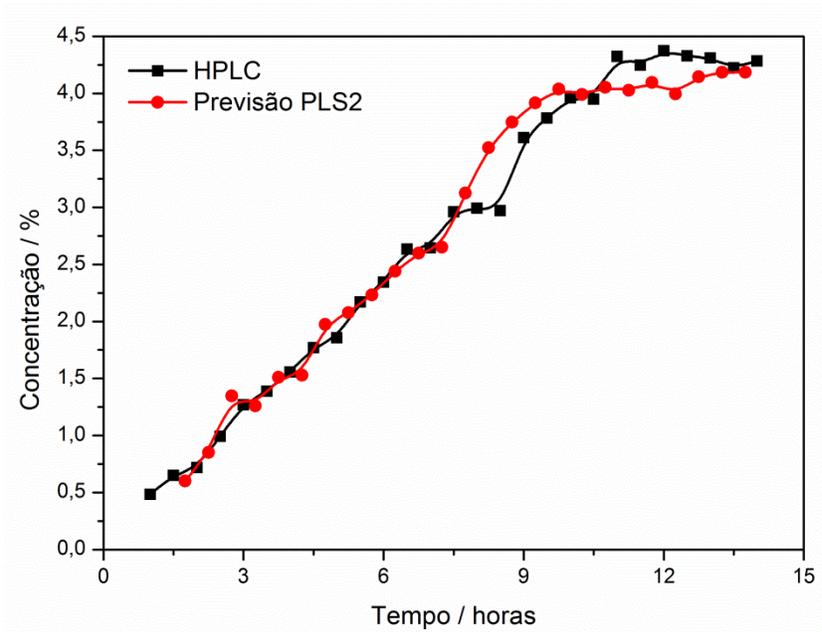


Figura 2.19. Perfil de concentra o previsto pelo modelo PLS2 e pelo m todo de refer ncia para etanol.

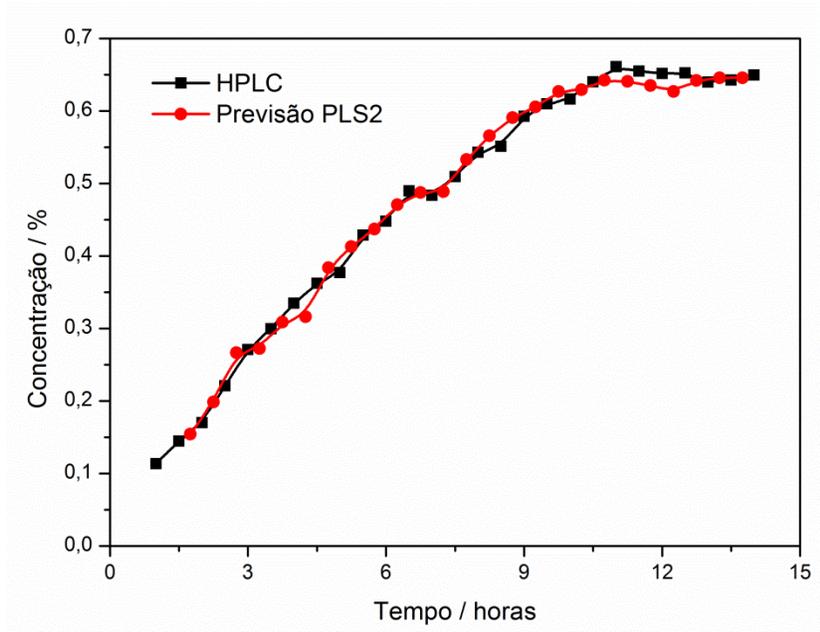


Figura 2.20. Perfil de concentra o previsto pelo modelo PLS2 e pelo m todo de refer ncia para glicerol.

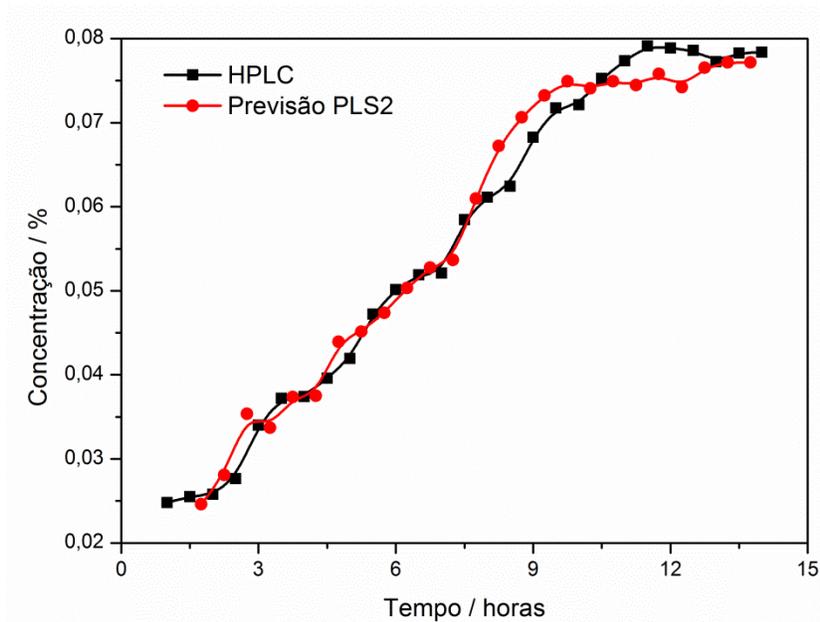


Figura 2.21. Perfil de concentra o previsto pelo modelo PLS2 e pelo m todo de refer ncia para  cido ac tico.

## 2.4. CONCLUS ES

Neste cap tulo utilizou-se a calibra  o multivariada para estimar as concentra  es dos constituintes da fermenta  o da glicose pela *Saccharomyces cerevisiae*, e como pode ser notado, os resultados obtidos pela aplica  o da calibra  o multivariada em um monitoramento *on-line* permitem o acompanhamento do perfil do processo apesar da correla  o existente entre os constituintes. O fato da correla  o permitiu a utiliza  o do m todo de PLS2, uma vez que os erros apresentados por este m todo n o diferem daqueles obtidos pelo m todo de PLS1, como pode ser notado na Tabela 2.4.

Tabela 2.4. Tabela comparativa dos erros de calibra  o e de calibra  o cruzada dos modelos PLS1 e PLS2.

	Glicose		Etanol		Glicerol		�cido Ac�tico	
	PLS1	PLS2	PLS1	PLS2	PLS1	PLS2	PLS1	PLS2
<b>RMSEC / %</b>	0,69	0,63	0,14	0,19	0,013	0,015	0,0021	0,0026
<b>RMSECV / %</b>	0,84	0,71	0,20	0,21	0,017	0,018	0,018	0,0029
<b>RMESP / %</b>	0,43	-	0,25	-	0,0031	-	0,0032	-

Outro fator que nos leva a esta conclus o   a compara  o entre os perfis de concentra  o com o tempo para glicose, etanol, glicerol e c lulas obtidos pelo m todo PLS, pois estes apresentam concord ncia com aqueles obtidos pelos m todos de refer ncia, permitindo, dessa forma, um acompanhamento em tempo real do processo fermentativo sem a necessidade das an lises pelos m todos de refer ncia, que s o demoradas e muitas vezes requerem pr -tratamento das amostras.

No entanto,   importante frisar que devido   exist ncia de correla  o entre os constituintes do processo fermentativo, deve-se ter um minucioso conhecimento do

processo para aplica o da metodologia desenvolvida em escala industrial, j  que pequenas mudan as nas vari veis de processo, como por exemplo, na mat ria-prima, temperatura e/ou pH, tornam os modelos desenvolvidos inadequados uma vez que a varia o das condi oes alteram o perfil do processo. Assim   necess rio que o processo seja controlado e reprodu vel.

## **Capítulo III**

---

# **Aplicação de Cartas de Controle Multivariadas para Detecção de Falhas**



### 3.1. INTRODUÇÃO

O controle *on-line* de uma batelada é realizado por meio de seu monitoramento em intervalos de tempo definidos. Todo processo apresenta uma variabilidade natural devido a causas comuns intrínsecas ao processo. Se essa variabilidade é considerada aceitável diz-se que o processo está sob controle. Quando variações no processo são originadas de causas desconhecidas ou parâmetros de processos mal ajustados, é necessária uma correção para que este processo volte às condições otimizadas, ou mesmo uma interrupção do processo para evitar perdas ainda maiores. É neste sentido que as cartas de controle foram desenvolvidas, visando o controle estatístico do processo a fim de identificar falhas através de monitoramento *on-line*.

Atualmente muitos produtos industriais são obtidos por processos realizados em bateladas, como por exemplo, a fermentação alcoólica. Esses processos geram grande quantidade de dados, já que são monitorados analisando algumas variáveis ao longo do tempo. Por esse motivo existe uma crescente busca no desenvolvimento de cartas de controle multivariadas que possam ser aplicadas a monitoramentos *on-line*.

As cartas de controle  $Q$  e  $T^2$  são cartas de controle multivariadas baseadas na análise MPCA, que reduz a dimensionalidade do conjunto de dados através das componentes principais sem que haja perda da representatividade dos dados, permitindo uma avaliação das variáveis de processo a cada instante. A partir da análise MPCA pode-se aplicar as estatísticas  $Q$  e *Hotelling*  $T^2$  para detecção de falhas em bateladas. As cartas de controle multivariadas são indicadas em situações em que as variáveis monitoradas são correlacionadas.

A identificação de falhas em bateladas ajuda no aumento da produtividade, no controle de qualidade tanto dos produtos quanto de matéria-prima, na diminuição os custos do produto final e na tomada de decisões em tempo real. A diminuição do tempo de ação também é um dos motivos pela qual as cartas de controle

multivariadas tem tido grande destaque, pois elimina a necessidade das análises realizadas em monitoramento *off-line*, que geralmente são demoradas e requerem pré-tratamento das amostras. Sendo assim, nesta parte do trabalho foi realizada a aplicação de cartas de controle multivariadas para detecção de bateladas falha no monitoramento *on-line* da fermentação da glicose pela levedura *Saccharomyces cerevisiae*.

## **3.2. PARTE EXPERIMENTAL**

### **3.2.1. Instrumentação e Micro-organismos**

A instrumentação, os reagentes, o sistema de monitoramento *on-line* e os micro-organismos utilizados nesta parte do trabalho são os mesmos descritos nos itens 2.2.1 e 2.2.2 do Capítulo II - Quantificação de Substrato e Produtos de Fermentação.

### **3.2.2. Processo de Monitoramento**

No desenvolvimento das Cartas de Controle Multivariada, 13 bateladas NOC (Normal Operation Condition) foram monitoradas para definição dos limites de controle. As bateladas foram conduzidas e monitoradas da mesma forma como descrito no item 2.2.3. do Capítulo II - Quantificação de Substrato e Produtos de Fermentação; assim, as condições normais de operação são: temperatura 30 °C, agitação mecânica de 100 rpm, concentração do meio de cultura de 10 % de glicose, 2 % de peptona e 1 % de extrato de levedura. Os monitoramentos foram realizados durante 13 horas, em espectrômetro Raman na faixa de 1800 a 300  $\text{cm}^{-1}$  com resolução de 4  $\text{cm}^{-1}$ , com intervalos de medição de 15 minutos, o que resultou em um

total de 53 espectros para cada batelada. Dessa forma, para cada batelada foi obtida uma matriz 376 x 53 (número de variáveis espectrais x tempo). Os dados dos monitoramentos foram organizados em um arranjo tridimensional de dimensão 376 x 53 x 13.

Três tipos de falhas foram induzidas no sistema a fim de testar as cartas de controle multivariadas. As condições usadas para gerar as três falhas foram temperatura de 25 °C, substituição do substrato glicose por sacarose e contaminação da fermentação com outro tipo de levedura. Todas as falhas foram realizadas em triplicata. Além disso, mais três bateladas NOC foram realizadas a fim de provar a eficiência das cartas de controle, formando as bateladas teste.

### **3.2.3. Construção das Cartas de Controle Multivariadas baseadas na análise MPCA**

As análises das cartas de controle multivariadas,  $Q$  e  $T^2$ , foram obtidas a partir da análise MPCA aplicada aos dois modos de desdobramento, propostos por Nomikos e MacGregor [36], que preserva a direção das bateladas, e por Wold *et. al* [37], que preserva a direção das variáveis. A partir dos valores de escores, pesos e resíduos obtidos para cada modelo, os valores de  $Q$  e  $T^2$  podem ser calculados para cada batelada ou tempo, dependendo do desdobramento e os valores de  $Q_{lim}$  e  $T_{lim}$  podem ser estimados para identificar os limites de cada carta de controle, a partir dos quais as amostras serão consideradas falhas.

### **3.3. RESULTADOS E DISCUSSÃO**

#### **3.3.1. Desdobramento em relação à batelada**

##### *3.3.1.1. Análise MPCA*

A análise MPCA foi realizada usando a região do espectro de 1800 a 556  $\text{cm}^{-1}$ , devido a esta região apresentar menor ruído entre as bateladas, o que poderia causar influência nos resultados das cartas de controle. Este procedimento reduziu a dimensão do arranjo dos dados das bateladas NOC de 376 x 53 x 13 para 312 x 53 x 13. Vários tipos de pré-processamento dos dados foram realizados; no entanto, o que melhor se adequou foi o alisamento usando o algoritmo Savitzky-Golay [31] com janela de 15 pontos e um polinômio de segundo grau. O modelo MPCA foi construído usando 3 componentes principais que juntos explicaram 97,58 % da variância dos dados. A Figura 3.1 apresenta os escores das três componentes principais, onde é possível observar que duas das amostras se mostraram diferentes das outras dentro do conjunto de dados. A projeção em duas dimensões utilizando as duas primeiras componentes principais facilita a visualização dessa diferença entre as bateladas. Isso pode ter ocorrido devido a problemas durante o processo de fermentação, como por exemplo, uma contaminação ou o não crescimento dos micro-organismos. Dessa forma essas duas amostras foram descartadas e um novo modelo foi desenvolvido com as 11 bateladas restantes usando 3 componentes principais que explicaram 98,14 % da variância dos dados. A partir do modelo MPCA construído, foram estimados os limites das cartas de controle.

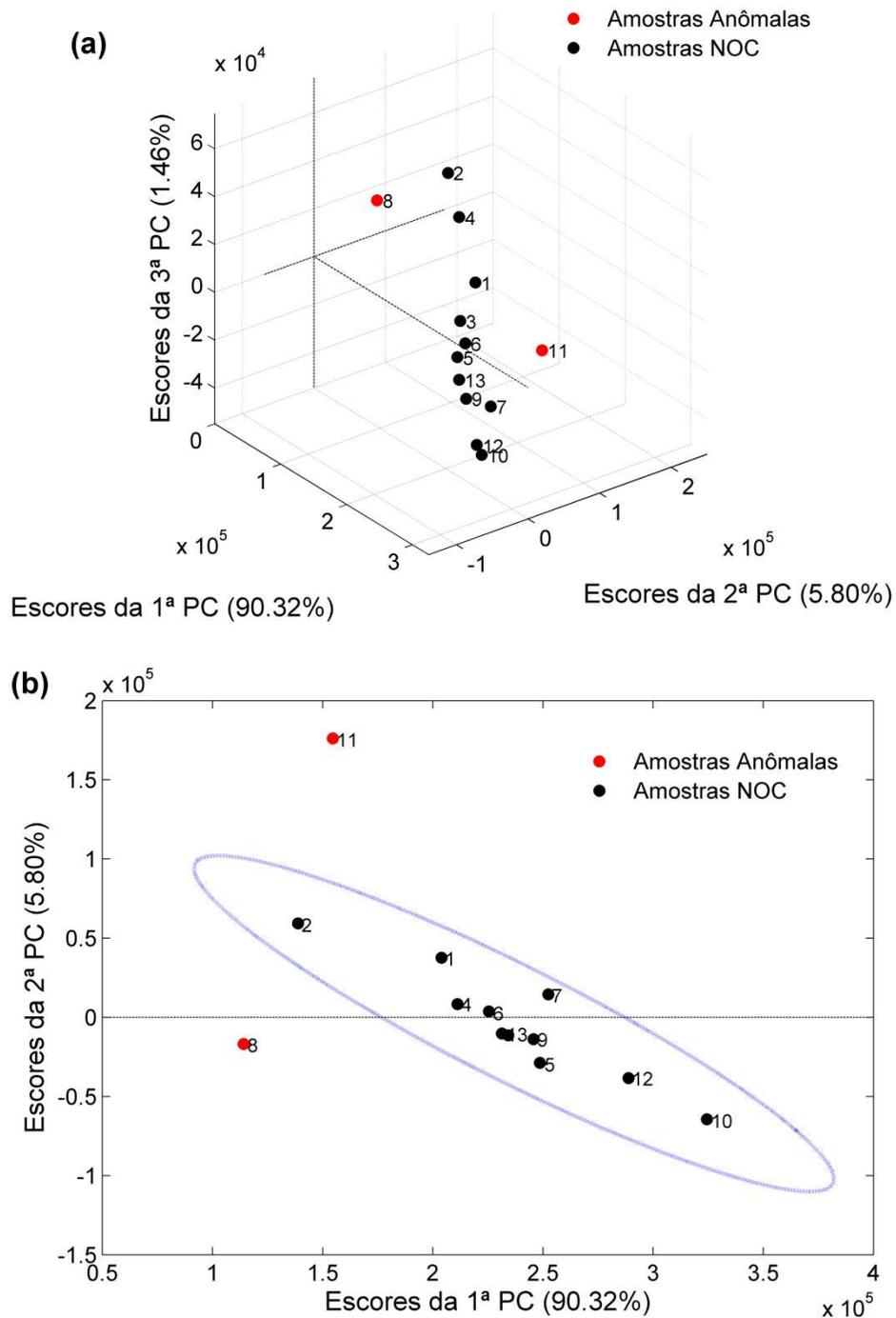


Figura 3.1. (a) Escores das três componentes principais do modelo MPCA e (b) Escores da primeira componente principal versus a segunda componente principal do modelo MPCA.

### 3.3.1.2. Cartas de Controle Q e $T^2$

A carta de controle Q foi eficiente para detecção das bateladas-falha no processo, o que pode ser observado na Figura 3.2. Ao nível de confiança de 95 % todas as falhas estudadas foram corretamente identificadas, mostrando a eficiência da carta Q na detecção de falhas, fato comprovado pela correta classificação das 3 bateladas NOC usadas como conjunto de teste. No entanto a carta de controle  $T^2$ , apresentada na Figura 3.3, não foi eficiente para detectar a presença de falhas no monitoramento ao nível de confiança de 95%.

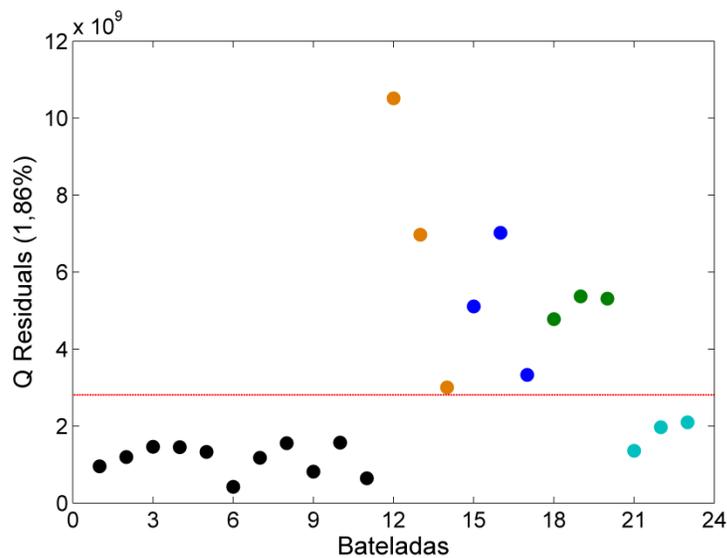


Figura 3.2. Carta de Controle Q baseada na análise MPCA (●) Bateladas NOC; (●) Bateladas Falha-Temperatura; (●) Bateladas Falha-Substrato; (●) Bateladas Falha-Contaminação; (●) Bateladas NOC-teste; (---)  $Q_{lim}$  com Nível de Confiança de 95%.

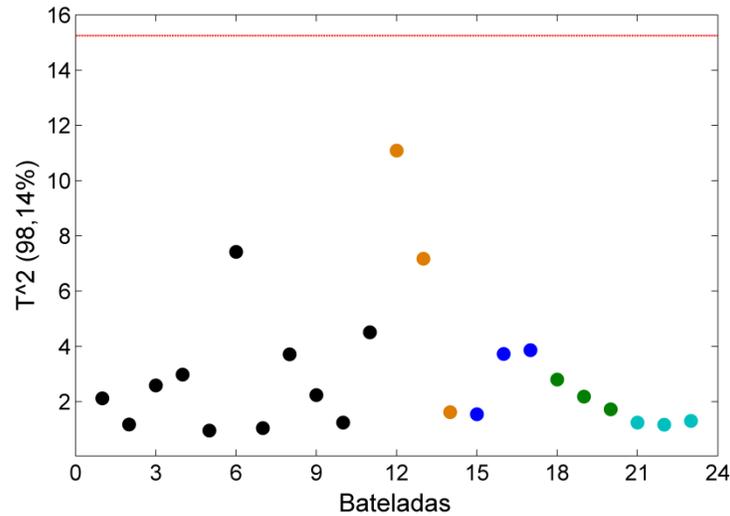


Figura 3.3. Carta de Controle  $T^2$  baseada na análise MPCA (●) Bateladas NOC; (●) Bateladas Falha-Temperatura; (●) Bateladas Falha-Substrato; (●) Bateladas Falha-Contaminação; (●) Bateladas NOC-teste; (---)  $T^2_{lim}$  com Nível de Confiança de 95%.

A diferença de resultados entre as cartas  $Q$  e  $T^2$  pode ser explicada pelo fato da carta  $Q$  levar em conta em seus cálculos os valores residuais obtidos pelo modelo MPCA, enquanto a carta  $T^2$  considera em seus cálculos os valores de escores obtidos pelo modelo, revelando que os escores das bateladas NOC são semelhantes aos escores das bateladas- falha, fato que pode ser comprovado na Figura 3.4.

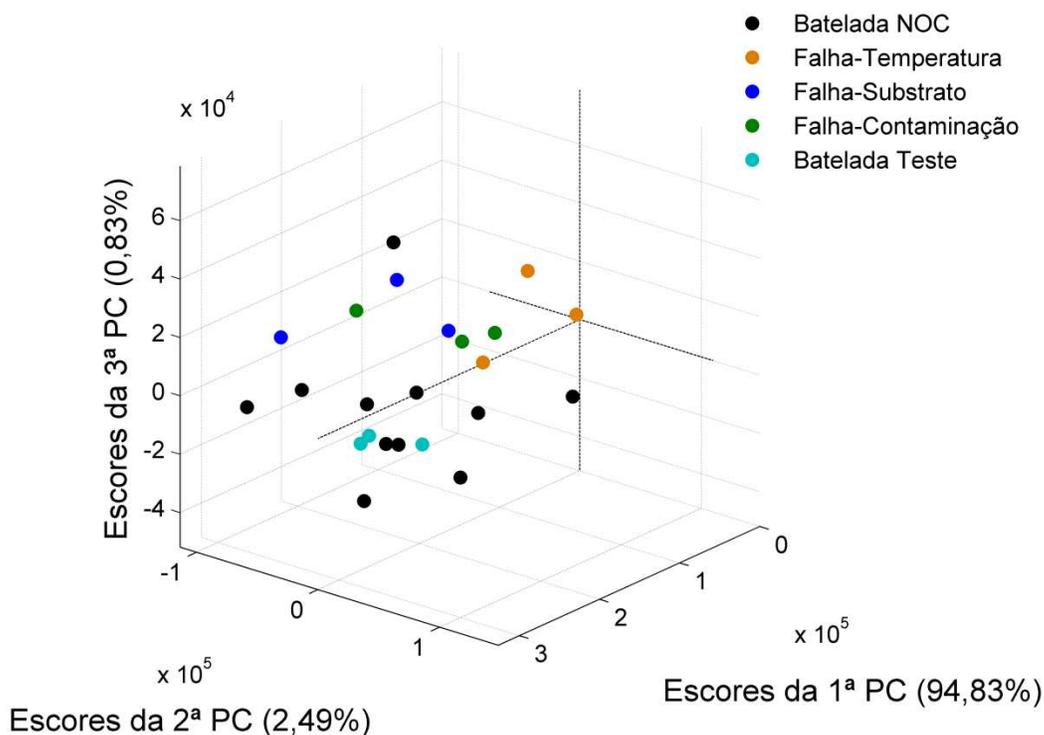


Figura 3.4. Gráficos dos escores nas bateladas-NOC e das bateladas-falha.

As Figuras 3.5 a 3.7 apresentam os espectros inicial e final de uma batelada NOC e de cada uma das bateladas-falha. A análise das figuras revela a não existência de diferenças de linha base entre os espectros e que os espectros não apresentam mudanças significativas entre bateladas NOC e bateladas-falha para causar grandes variações nos escores destas para que possam ser identificadas pela carta  $T^2$ . Por outro lado, as mudanças causadas nos espectros com tempo, na cinética do processo, causaram variações nos resíduos, de bateladas NOC e bateladas falha, suficientes para diferenciação dessas bateladas pela carta  $Q$ . Nesse caso pode-se dizer que as bateladas-falha não são propriamente modeladas e que altos valores de resíduos são observados.

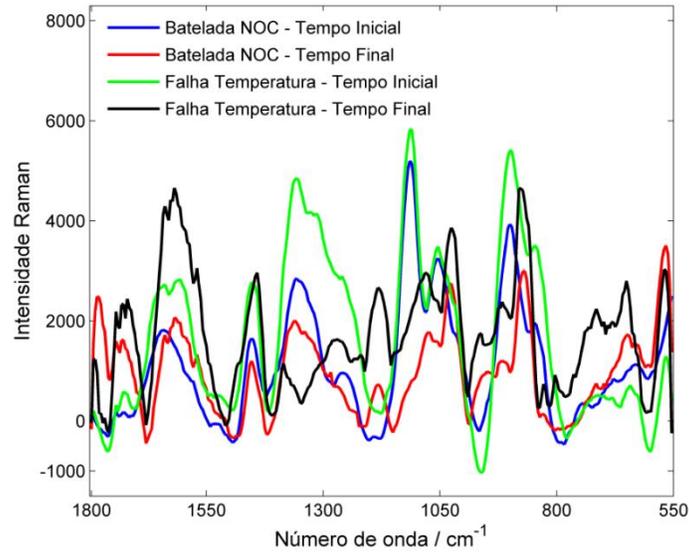


Figura 3.5. Espectros inicial e final de uma Batelada-NOC e de uma Batelada-Falha Temperatura.

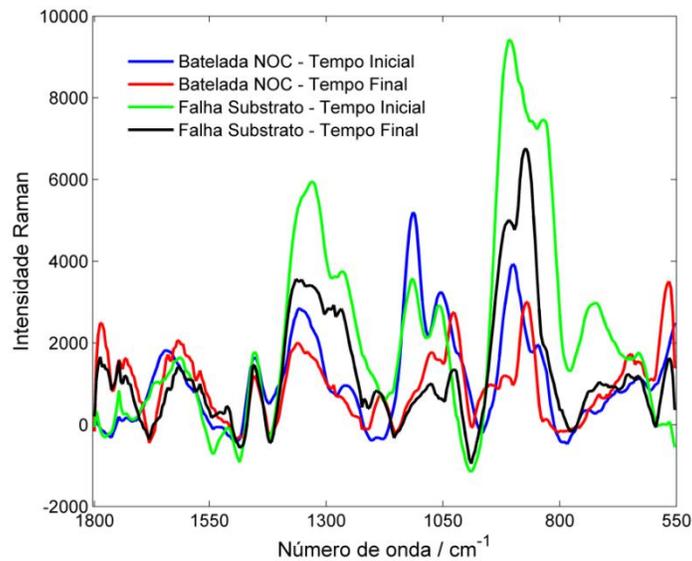


Figura 3.6. Espectros inicial e final de uma Batelada-NOC e de uma Batelada-Falha Substrato.

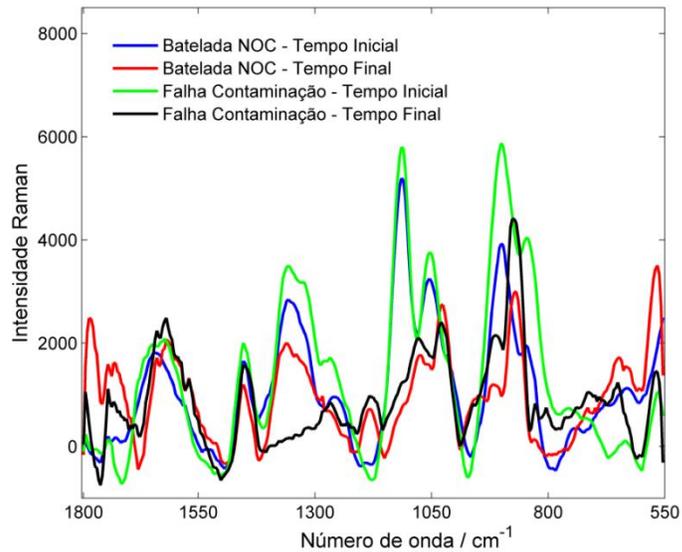


Figura 3.7. Espectros inicial e final de uma Batelada-NOC e de uma Batelada-Falha Contaminação.

O método MPCA e o desdobramento proposto por Nomikos e McGregor não permite a obtenção dos resultados em tempo real, uma vez que há a necessidade de se ter uma batelada completa para a análise pelo modelo MPCA. No entanto, uma tática pode ser utilizada para permitir a utilização deste método na identificação de possíveis falhas; uma substituição dos tempos restantes por uma média ou por algum dado obtido de bateladas NOC podem ser utilizados para completar a batelada em análise e assim ter o conhecimento ou não se esta é uma falha.

### 3.3.2. Desdobramento em relação ao tempo

#### 3.3.2.1. Análise MPCA

A análise MPCA foi realizada usando a mesma região do espectro e o mesmo pré-processamento realizado na etapa anterior. No entanto neste caso o arranjo dos dados foi disposto de forma diferente, 13 x 53 x 312. O modelo MPCA foi construído usando 3 componentes principais que juntas explicaram 96,69 % da variância dos dados. O resultado para a análise MPCA está apresentado na Figura 3.8, em que a cada 53 pontos tem-se uma batelada, já que cada batelada foi monitorada em 53 tempos. Assim os pontos de 1 a 53 representam a primeira batelada, de 54 a 106 a segunda batelada e assim sucessivamente. Como pode ser notado na Figura 3.8 que mostra um gráfico dos valores de  $T^2$  contra  $Q$  para todos os tempos, três bateladas apresentaram muitos pontos consecutivos fora dos limites das cartas de controle  $Q$  e  $T^2$ . Assim, essas bateladas foram descartadas do modelo, pois poderiam ocasionar classificações errôneas das falhas analisadas. A Figura 3.9 mostra a ampliação das regiões 1 e 2 da Figura 3.8, onde é possível notar que a bateladas 2, 10 e 11 apresentaram respectivamente 8, 6 e 27 pontos consecutivos fora dos limites das cartas. Essas bateladas são representadas pelos pontos de 54 a 106, 478 a 530 e 531 a 583, respectivamente.



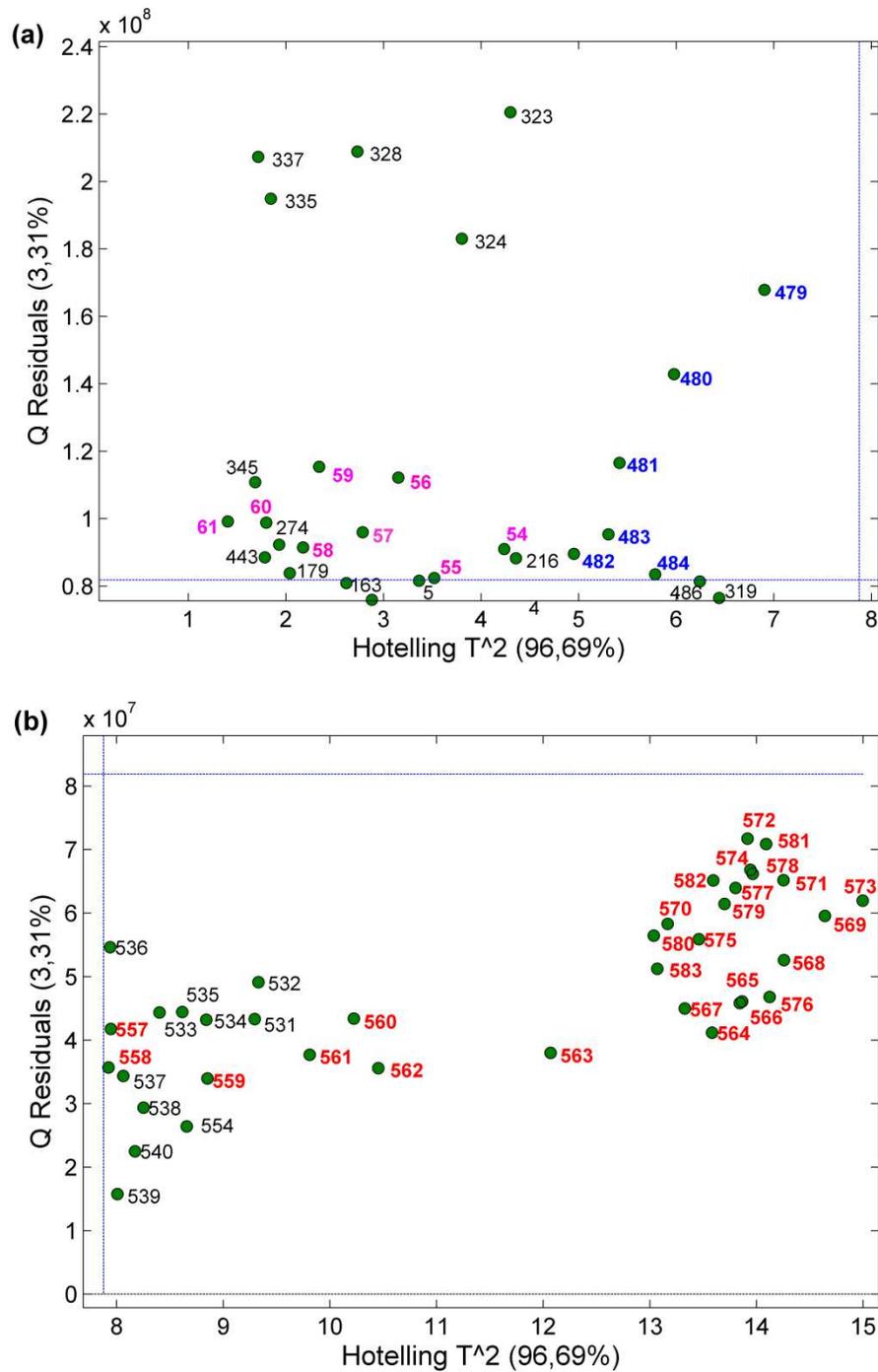


Figura 3.9. (a) Ampliação da região 1 da Figura 3.8; (b) Ampliação da região 2 da Figura 3.8, em (rosa) batelada 2, em (azul) Batelada 10 e em (vermelho) Batelada 11.



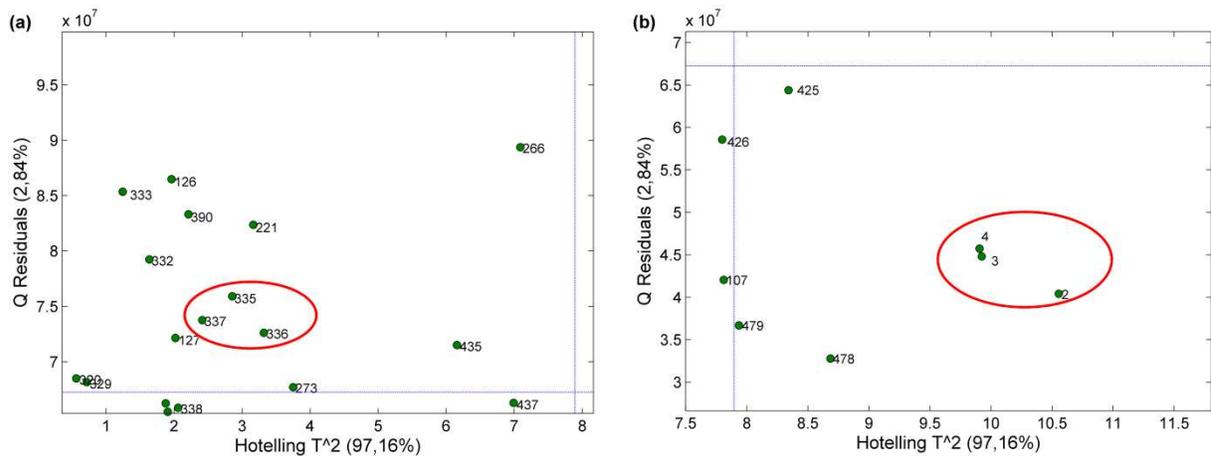


Figura 3.11. (a) Ampliação da região 1 da Figura 3.10 (batelada 7) e (b) Ampliação da região 2 da Figura 3.10 (batelada 1).

### 3.3.2.2. Cartas de Controle Q e $T^2$

No desdobramento do arranjo tridimensional dos dados preservando a direção das variáveis, as cartas de controle multivariadas, Q e  $T^2$ , foram utilizadas simultaneamente para a classificação de falhas, ou seja, uma batelada será considerada falha se apresentar 4 tempos consecutivos fora de qualquer um dos limites das cartas de controle. Todos os limites das cartas de controle foram calculados com um nível de confiança de 95 %. Nas análises das falhas, cada falha foi representada pelos 53 tempos enumerados de 531 a 583.

As três bateladas-falha temperatura foram classificadas corretamente como fora de controle, duas das triplicatas, mostradas nas Figuras 3.12 e 3.13, apresentaram mais de quatro pontos consecutivos acima dos limites das duas cartas de controle. No entanto, uma das triplicatas, Figura 3.14, se apresentou de forma diferente das outras, pois a maioria dos tempos ficou abaixo dos limites das cartas, mesmo assim 5 pontos consecutivos ficaram acima dos limites de controle, sendo dessa forma considerada como fora de controle. Essa diferença pode ter sido



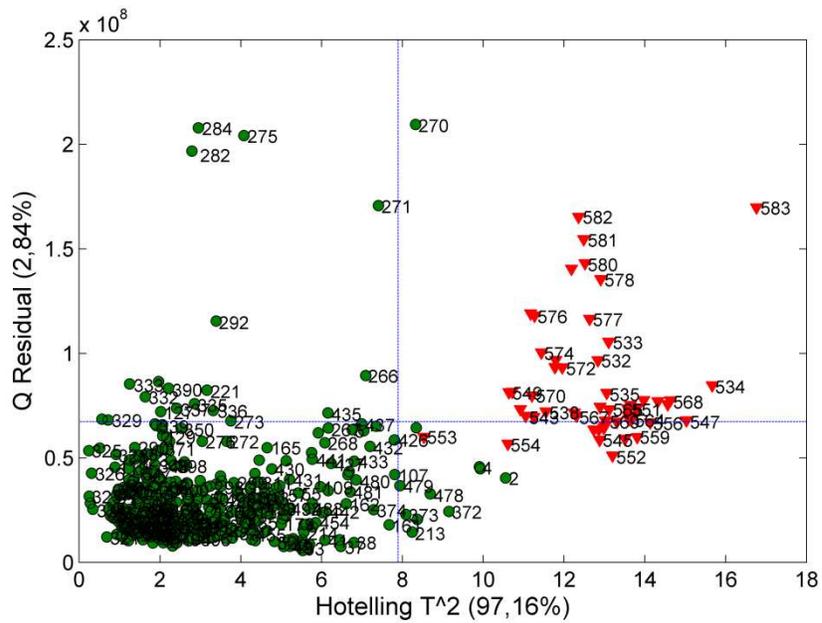


Figura 3.13. Escores da Batelada-falha Temperatura 2, (●) Bateladas-NOC e (▼) Batelada-falha.

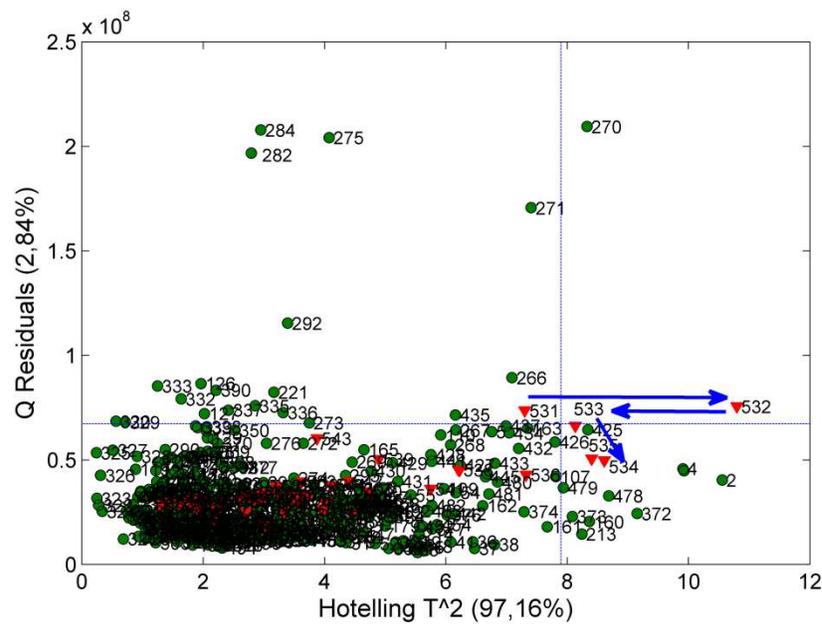


Figura 3.14. Escores da Batelada-falha Temperatura 3, (●) Bateladas-NOC e (▼) Batelada-falha.

Para as bateladas-falhas relacionadas ao substrato, as cartas de controle foram eficientes na classificação destas como fora de controle. No entanto, somente a carta Q apresentou mais de 4 tempos consecutivos acima do limite, a carta  $T^2$ , não foi eficiente na identificação da falha substrato. As bateladas-falha estão apresentadas nas Figuras 3.15 a 3.17.

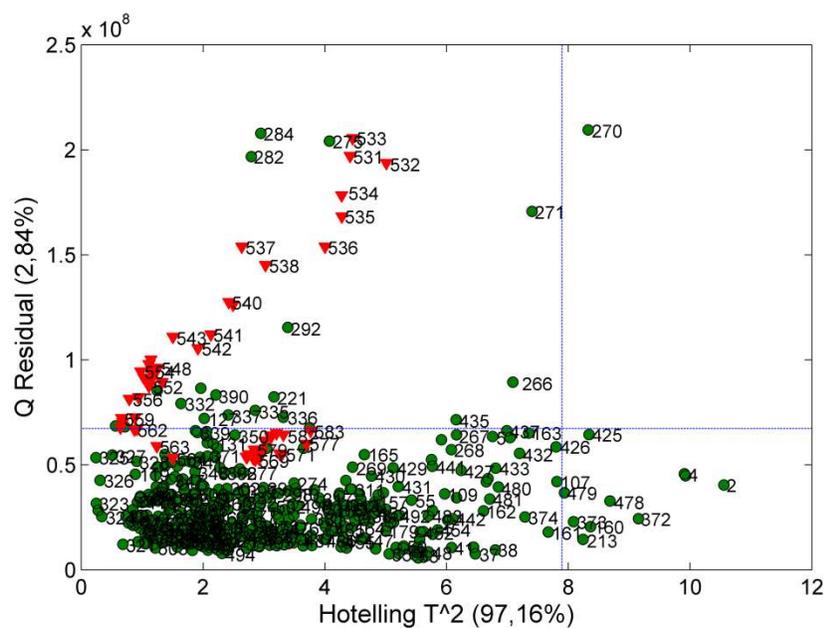


Figura 3.15. Escores da Batelada-falha Substrato 1, (●) Bateladas-NOC e (▼) Batelada-falha.

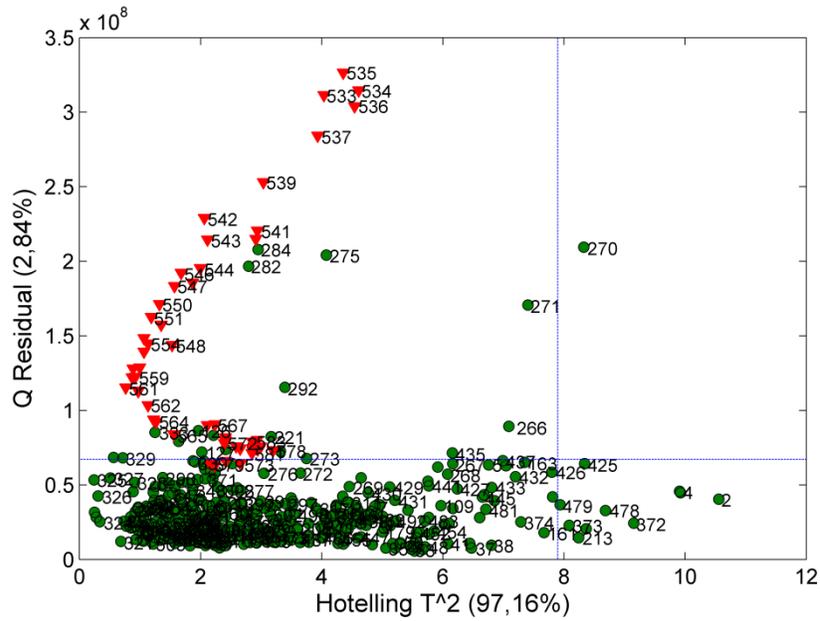


Figura 3.16. Escores da Batelada-falha Substrato 2, (●) Bateladas-NOC e (▼) Batelada-falha.

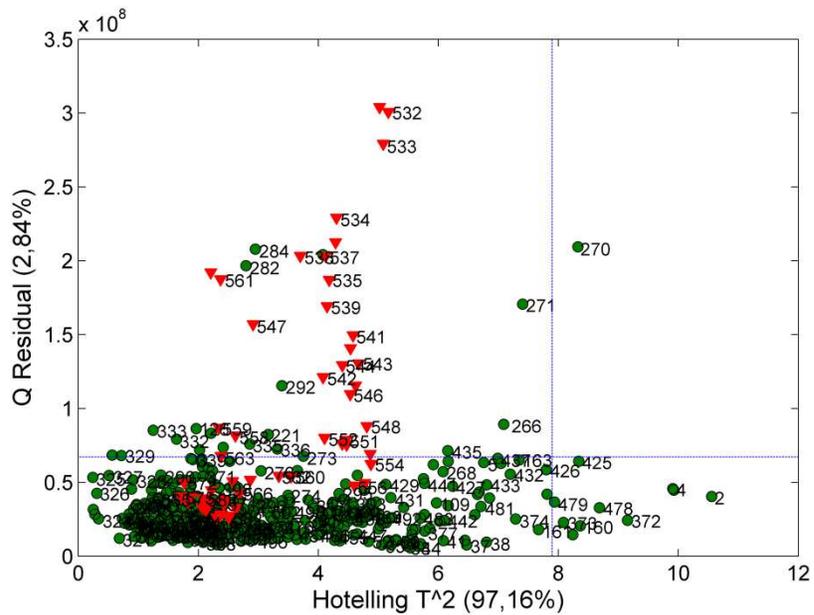


Figura 3.17. Escores da Batelada-falha Substrato 3, (●) Bateladas-NOC e (▼) Batelada-falha.

Já para as bateladas-falha contaminação, as cartas de controle também foram eficientes; neste caso, todas as triplicatas foram consideradas como fora de controle, como apresentado nas Figuras 3.18 a 3.20. Para duas das bateladas ambas as cartas de controle conseguiram classificação correta, mas para uma, somente a carta Q foi eficiente na classificação da falha.

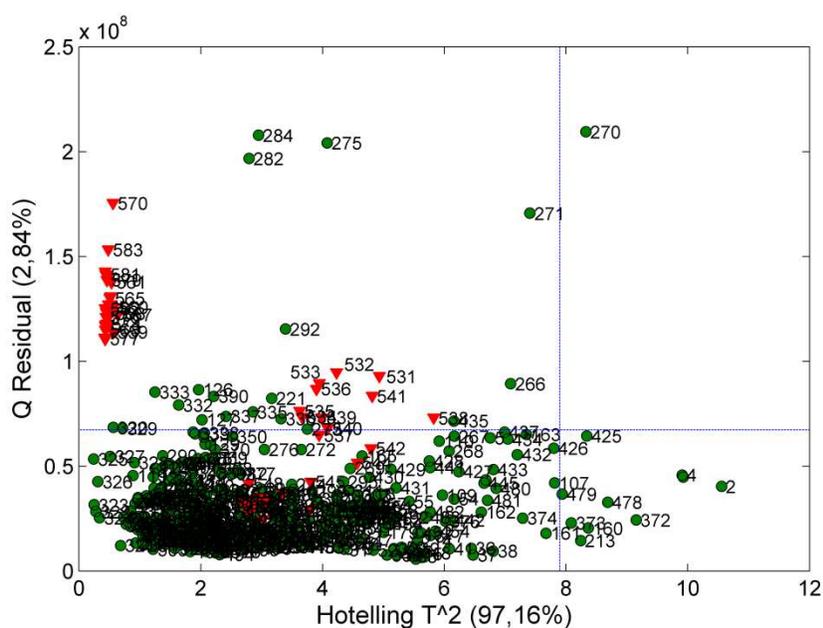


Figura 3.18. Escores da Batelada-falha Contaminação 1, (●) Bateladas-NOC e (▼) Batelada-falha.

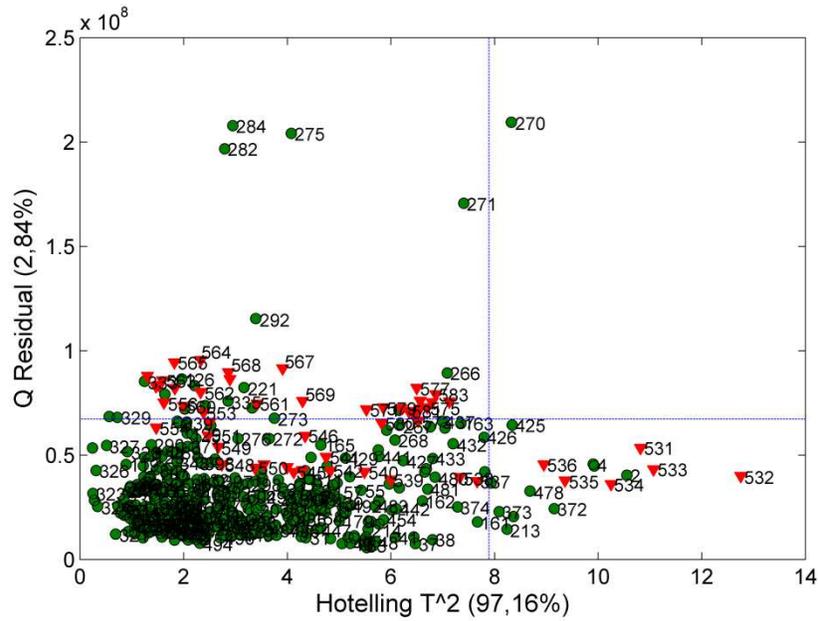


Figura 3.19. Escores da Batelada-falha Contaminação 2, (●) Bateladas-NOC e (▼) Batelada-falha.

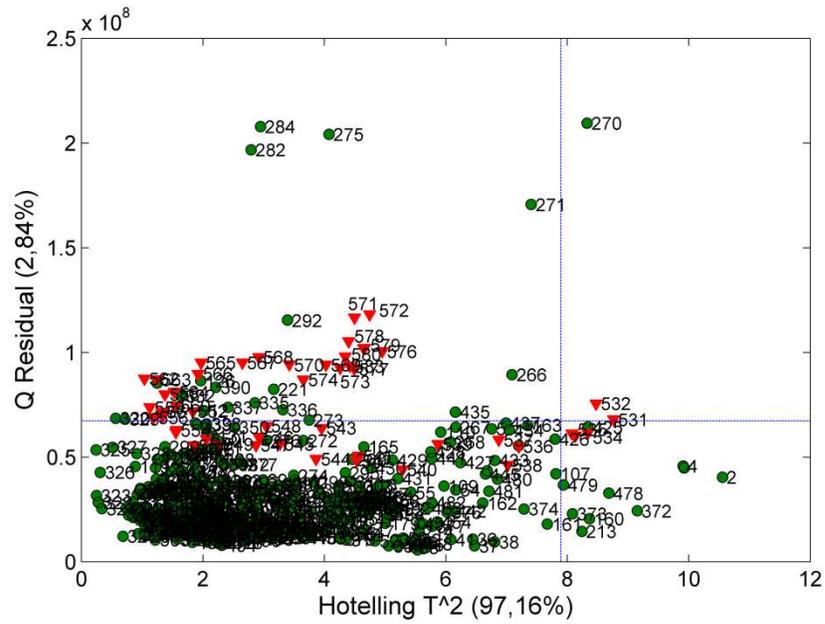


Figura 3.20. Escores da Batelada-falha Contaminação 3, (●) Bateladas-NOC e (▼) Batelada-falha.



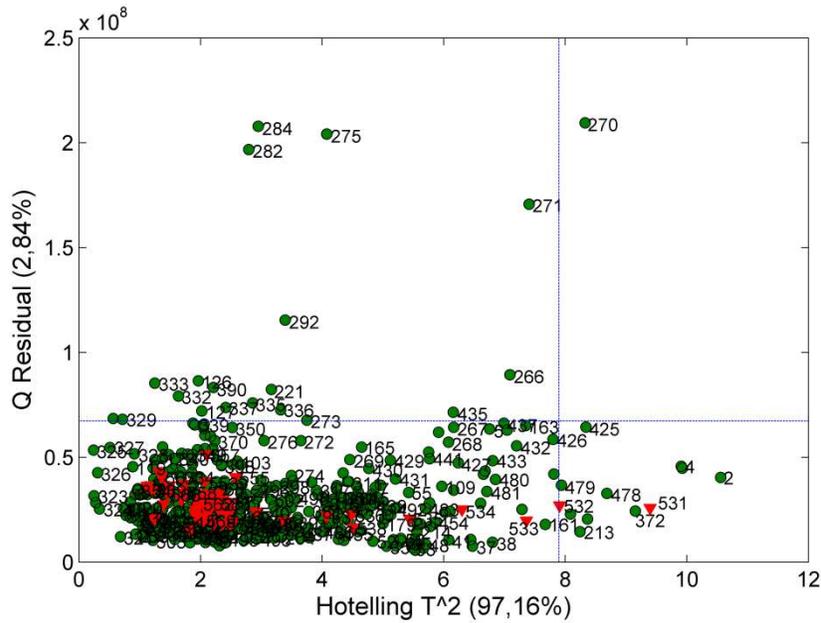


Figura 3.22. Escores da Batelada NOC-teste 2, (●) Bateladas-NOC e (▼) Batelada-NOC-teste.

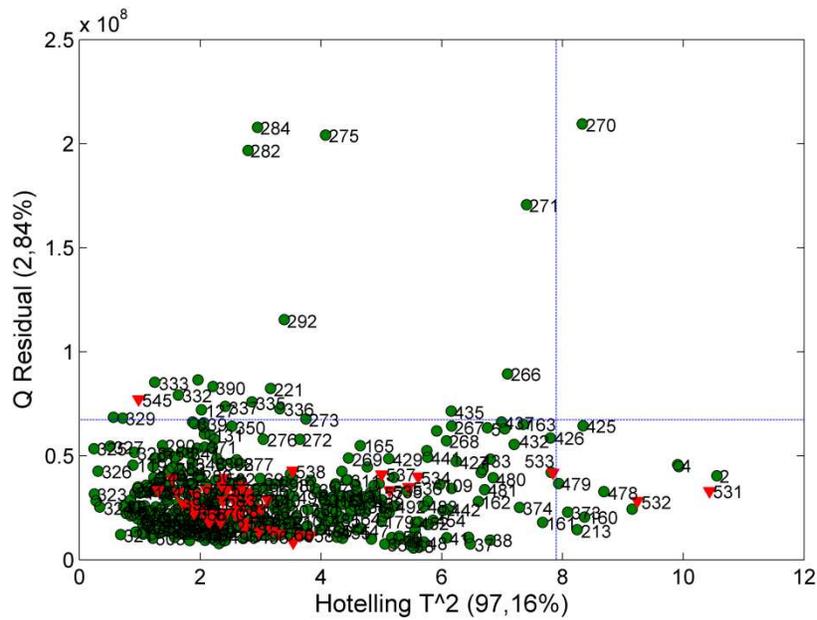


Figura 3.23. Escores da Batelada NOC-teste 3, (●) Bateladas-NOC e (▼) Batelada-NOC-teste.

De modo geral, a estratégia proposta por Wold e colaboradores, permitiu a identificação de bateladas-falha considerando a presença de 4 tempos consecutivos fora dos limites das cartas de controle  $Q$  e  $T^2$ . O tempo necessário para que os 4 tempos fossem medidos é de uma hora, uma vez que cada medida é realizada a cada 15 minutos. A proposta de Wold pode ser aplicada de outra forma, a qual seria mais adequada para uma aplicação industrial em um sistema automatizado, que seria a construção de modelos PCA temporais, ou seja, para cada tempo seria construído um modelo e avaliado a existência ou não de falhas, essa tática poderia eliminar a necessidade de se esperar o término da batelada para construção de um modelo, dando uma visualização geral dos escores das bateladas ao longo do tempo de monitoramento, permitindo um método não apenas para monitoramento e controle e sim de identificação e correção de eventuais erros.

### 3.4. CONCLUSÕES

No método proposto por Nomikos e Mcgregor foi possível concluir que a carta de controle  $Q$  foi eficiente para detecção de possíveis falhas durante a fermentação, causadas por variações de temperatura, contaminação e mudanças no substrato, enquanto a carta de controle  $T^2$  não foi eficiente na identificação dessas falhas. A explicação para a diferença de resultados obtidos pelas duas cartas de controle está na base utilizada para os cálculos, uma vez que a estatística  $Q$  utiliza os valores de resíduos e a estatística *Hotelling*  $T^2$  utiliza os valores de escores. Baseado nessa informação foi possível concluir que as falhas estudadas apresentaram espectros que não são diferentes dos espectros obtidos pelas bateladas NOC, mas que as diferenças causadas por modificações induzidas no processo foram capazes de causar variações na cinética do processo, o que gerou mudanças significativas na matriz de resíduos que foram identificadas pela carta de controle  $Q$ , podendo esta

ser empregada na identificação de falhas em bateladas de processos fermentativos da glicose pela levedura *Saccharomyces cerevisiae*.

Já para o método proposto por Wold e colaboradores as cartas de controle classificaram corretamente todas as falhas estudadas, apresentando erro de classificação somente de uma das bateladas teste realizada sob condições normais de operação. Mesmo com uma classificação errada de uma batelada teste, o método pode ser considerado eficiente para aplicação na identificação de bateladas-falha, uma vez que a classificação desta batelada como fora de controle ficou no limite de pontos consecutivos considerados como falha para uma batelada e estes pontos são os primeiros 4 tempos obtidos, evidenciando que uma variação dos parâmetros ou alguma instabilidade no início do processo podem ter causado este erro.



## **Capítulo IV**

---

### **Conclusões Gerais**



Neste trabalho foi desenvolvida uma metodologia para o monitoramento *on-line* da fermentação da glicose pela levedura *Saccharomyces cerevisiae* a fim de estimar as concentrações de substrato, produtos e células por espectroscopia Raman e calibração multivariada com PLS, e também uma metodologia para o uso de cartas de controle multivariadas para identificação de falhas nas bateladas.

Na primeira parte do trabalho, que se refere ao uso da calibração multivariada para estimar as concentrações dos constituintes da fermentação, os resultados obtidos pelos modelos de calibração PLS para previsão da concentração de glicose, etanol, glicerol, ácido acético e células foram bastante satisfatórios e com valores de erros baixos, quando comparados aos obtidos pelos métodos de referência. Os perfis de concentração obtidos pelos modelos PLS desenvolvidos foram coerentes com aqueles obtidos pelos métodos de referência, indicando que a calibração multivariada com PLS pode ser aplicada para o acompanhamento de perfis de processos de monitoramentos *on-line* de fermentações.

Já na segunda parte, foi proposta a identificação de falhas em um monitoramento *on-line* com auxílio de cartas de controle multivariadas. Para tal, foi utilizada a análise MPCA e as estatísticas  $Q$  e *Hotelling*  $T^2$ . A análise MPCA foi importante, pois ajudou na identificação de amostras anômalas antes da aplicação das cartas de controle, amostras essas que poderiam levar a classificações errôneas das bateladas. Este estudo de aplicação de cartas de controle foi realizado utilizando dois modos de desdobramento dos dados, um que preserva a direção das bateladas e outro que preserva a direção das variáveis. Cada um dos métodos de desdobramento utilizados apresentaram suas vantagens, o proposto por Nomikos e McGregor permite a análise de uma batelada como um todo, ou seja, dentro do modelo MPCA cada batelada apresenta um único valor de escore, e o método proposto por Wold permite a análise das bateladas a cada tempo, onde um valor de escore é obtido para cada tempo, sendo possível a construção de modelos temporais e a obtenção de resultados a cada tempo analisado. A desvantagem desse método aplicado à fermentação é devido à necessidade de 4 tempos consecutivos fora dos limites das cartas o que requer um tempo de monitoramento de 1 hora.

Finalmente, é importante ressaltar que a metodologia desenvolvida pode ser útil no controle de bioprocessos industriais, auxiliando nas adições de substratos, quantidade e pureza de produtos formados e verificação da presença de contaminantes. Outro fato relevante é que a metodologia desenvolvida não requer o pré-tratamento das amostras, o que resulta em maior frequência analítica com análises simultâneas de vários analitos e sem consumo de reagentes.

# Referências

---

1. Raman, C. V. e Krishnan, K. S.; A new type of secondary radiation. *Nature*, **1928**, Vol. 121, pp. 501 - 502.
2. Sala, O.; I<sub>2</sub><sup>-</sup> Uma molécula didática. *Química Nova*, **2008**, Vol. 31, pp. 914 – 920.
3. Skoog, D. A., Holler, F. J e Crouch, S. R.; *Princípios de análise instrumental 5ª edição*, São Paulo: Bookman, **2009**.
4. Couto, R. M. O.; Espalhamento Raman eletrônico do Sm<sup>2+</sup> em CaF<sub>2</sub>. Tese de Doutorado, Campinas: UNICAMP, **1981**.
5. Schrader, B. e Bougeard, D.; *Infrared and Raman Spectroscopy: Methods and Applications*. Weinheim : VCH, **1995**.
6. Sala, O.; *Fundamentos da Espectroscopia Raman e no Infravermelho*, São Paulo: UNESP, **1996**.
7. Fletcher, P. D. I., Haswell, S. J. e Zhang, X.; Monitoring of chemical reactions within microreactors using an inverted Raman microscopic spectrometer, *Electrophoresis*, **2003**, Vol. 24, pp. 3239 – 3245.
8. Aarnoutse, P. J. e Westerhuis, J. A.; Quantitative Raman reaction monitoring using the solvent as internal standard, *Analytical Chemistry*, **2005**, Vol. 77, pp. 1228 – 1236.
9. Lee, H. L. T., Boccazzi P., Gorret, N., Ram, R. J. e Sinskey, A. J.; *In situ* bioprocess monitoring of *Escherichia coli* bioreactions using Raman spectroscopy, *Vibrational Spectroscopy*, **2004**, Vol. 35, pp. 131 – 137.
10. Almeida, M. R., Alves, R. S., Nascimbem, L. B. L. R., Stephani, R., Poppi, R. J. e Oliveira, L. F. C.; Determination of amylose content in starch using Raman spectroscopy and multivariate calibration analysis, *Analytical and Bioanalytical Chemistry*, **2010**, Vol. 397, pp. 2693 – 2701.
11. Trevisan, M. G. e Poppi, R. J.; Química analítica de processos, *Química Nova*, **2006**, Vol. 29, pp. 1065-1071.

12. Vojinovic, V., Cabral, J. M. S. e Fonseca, L. P.; Real-time bioprocess monitoring Part I: *In situ* sensors, *Sensors and Actuators B*, **2006**, Vol. 114, pp. 1083 – 1091.
13. Blanco, M., Peinado, A. C. e Mas, J.; Analytical monitoring of alcoholic fermentation using NIR spectroscopy, *Biotechnology and Bioengineering*, **2004**, Vol. 88, pp. 536 – 542.
14. Skibsted, E., Lindemann, C., Roca, C. e Olsson, L.; *On-line* bioprocess monitoring with a multi-wavelength fluorescence sensor using multivariate calibration, *Journal of Biotechnology*, **2001**, Vol. 88, pp. 47-57.
15. Clementschitsch, F., Jürgen C., K., Florentina, P. e Bayer, K.; Sensor combination and chemometric modelling for improved process monitoring in recombinant *E. coli* fed-batch cultivations, *Journal of Biotechnology*, **2005**, Vol. 120, pp. 183 – 196.
16. Ödman, P., Johansen, L. C., Olsson, L., Gernaey, K. V. e Lantz, A. E.; *On-line* estimation of biomass, glucose and ethanol in *Saccharomyces cerevisiae* cultivations using *in-situ* multi-wavelength fluorescence and software sensors, *Journal of Biotechnology*, **2009**, Vol. 144, pp. 102 – 112.
17. Cozzolino, D., Parker, M., Damberg, R. G., Herderich, M. e Gishen, M.; Chemometrics and visible-near infrared spectroscopic monitoring of red wine fermentation in a pilot scale, *Biotechnology and Bioengineering*, **2006**, Vol. 95, pp. 1001 – 1007.
18. Schuster, K. C., Mertens, F. e Gapes, J. R.; FTIR spectroscopy applied to bacterial cells as a novel method for monitoring complex biotechnological processes, *Vibrational Spectroscopy*, **1999**, Vol. 19, pp. 467 – 477.
19. Egidio, V., Sinelli, N., Giovanelli, G., Moles, A. e Casiraghi, E.; NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation, *European Food Research and Technology*, **2010**, Vol. 230, pp. 947 – 955.
20. Finn, B., Harvey, L. M. e McNeil B.; Near-infrared spectroscopic monitoring of biomass, glucose, ethanol and protein content in a high cell density baker's yeast fed-batch bioprocess, *John Wiley & Sons Ltd*, **2006**, Vol. 23, pp. 507 – 517.
21. Picard, A., Daniel, I., Montagnac, G. e Oger, P.; *In situ* monitoring by quantitative Raman spectroscopy of alcoholic fermentation by *Saccharomyces cerevisiae* under high pressure, *Extremophiles*, **2007**, Vol. 11, pp. 445 – 452.

22. Sivakesava, S., Irudayaraj, J. e Demirci, A. Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman, *Journal of Industrial Microbiology & Biotechnology*, **2001**, Vol. 26, pp. 185 – 190.
23. Sivakesava, S., Irudayaraj, J. e Demirci A.; Simultaneous determination of multiple components in lactic acid fermentation using FT-MIR, NIR, and FT-Raman spectroscopic techniques, *Process Biochemistry*, **2001**, Vol. 37, pp. 371 – 378.
24. Clementschitsch, F. e Bayer, K.; Improvement of bioprocess monitoring: development of novel concepts, *Microbial Cell Factories*, **2006**, Vol. 5, pp. 19 – 29.
25. Andrietta, M. G. S., Andrietta, S. R., Steckelberg C. e Stupiello E. N. A.; Bioethanol - Brazil, 30 years of Proálcool, *International Sugar Journal A*, **2007**, Vol. 109, pp. 195 – 200.
26. Yang, S., Jia, N., Li, M. e Wang, J.; Heterologous expression and efficient ethanol production of a *Rhizopus glucoamylase* gene in *Saccharomyces cerevisiae*, *Molecular Biology Reports*, **2011**, Vol. 38, pp. 59 – 64.
27. Medina, V. G., Almeringet, M. J. H. Maris, A. J. A. e Pronk, J. T.; Elimination of Glycerol Production in Anaerobic Cultures of a *Saccharomyces cerevisiae* Strain Engineered To Use Acetic Acid as an Electron Acceptor, *Applied and Environmental Microbiology*, **2010**, Vol. 76, pp. 190 – 195.
28. Guo, Z., Qiu, C., Zhang, L., Ding, Z., Wang, Z. e Shi, G.; Expression of aspartic protease from *Neurospora crassa* in industrial ethanol-producing yeast and its application in ethanol production, *Enzyme and Microbial Technology*, **2011**, Vol. 48, pp. 148 – 154.
29. Argueso, J. L., Carazzolle, M. F., Mieczkowski, P. A., Duarte, F. M., Netto, O. V. C., Missawa, S. K., Galzerani, F., Costa, G. G. L., Vidal, R. O., Noronha, M. F., Dominska, M., Andrietta, M. G. S., Andrietta, S. R., Cunha, A. F., Gomes, L. H., Tavares, F. C. A., Alcarde, A. R., Dietrich, F. S., McCusker, J. H., Petes, T. D. e Pereira, G. A. G.; Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production, *Genome Research*, **2010**, Vol. 19, pp. 2258 – 2270.
30. Xiaojing, X., Limin, C. e Xun, C.; Elementary flux mode analysis for optimized ethanol yield in anaerobic fermentation of glucose with *Saccharomyces cerevisiae*, *Chinese Journal of Chemical Engineering*, **2008**, Vol. 16, pp. 135 – 142.

31. Savitzky, A. e Golay, M. J. E.; Smoothing and differentiation of data by Simplified Least Squares procedures, *Analytical Chemistry*, **1964**, Vol. 36, pp. 1627 – 1639.
32. Otto, M.; *Chemometrics – Statistics and computer application in Analytical Chemistry*, Weinheim: Wiley, **1999**.
33. Beebe, K. R., Pell, R. J. e Seasholtz, M. B.; *Chemometrics: A practical guide*, New York: Wiley, **1998**.
34. Beebe, K. R. e Kowalski, B. R.; An introduction to multivariate calibration and analysis, *Analytical Chemistry*, **1987**, Vol. 59, pp. 1007A – 1017A.
35. Marçõ, P. H.; Estudo da influência da radiação e pH no comportamento cinético de antocianinas de plantas do gênero *hibiscus* por métodos quimiométricos. Tese de doutorado, Campinas: UNICAMP, **2009**.
36. Nomikos, P. e MacGregor, J. F.; Monitoring batch processes using multiway Principal Component Analysis, *AIChE Journal*, **1994**, Vol. 40, pp. 1361 – 1375.
37. Wold, S., Kettaneh, N., Fridén, H. e Holmberg, A.; Modelling and diagnostics of batch processes and analogous kinetic experiments, *Chemometrics and Intelligent Laboratory Systems*, **1998**, Vol. 44, pp. 341 – 340.
38. Geladi, P. e Kowalski, B. R.; Partial least squares regression: A tutorial, *Analytical Chemical Acta*, **1986**, Vol. 185, pp. 1 – 17.
39. Wold, H.; Nonlinear estimation by iterative least squares procedures. In F. David, *Research papers in statistics, Festschrift for J. Neyman*. New York: Wiley, **1966**, pp. 411 – 444.
40. Nomikos, P. e MacGregor, J. F.; Multivariate SPC charts for monitoring batch process, *Technometrics*, **1995**, Vol. 37, pp. 41 – 59.
41. Rosa, A. F. P.; Método para controle estatístico multivariado de processo em batelada. Dissertação de Mestrado, Porto Alegre: UFRGS, **2001**.
42. Aguado, D., Ferrer, A., Ferrer, J. e Seco, A.; Multivariate SPC of a sequencing batch reactor for wastewater treatment, *Chemometrics and Intelligent Laboratory Systems*, **2007**, Vol. 85, pp. 82 – 93.

43. Yu, J.; Fault Detection using Principal Components-based gaussian mixture model for semiconductor manufacturing processes, *IEEE Transactions on Semiconductor Manufacturing*, **2011**, Vol. 24, pp. 432 – 444.
44. Lee, J., Yoo, C. e Lee, I.; Fault detection of batch processes using Multiway Kernel Principal Component Analysis, *Computers and Chemical Engineering*, **2004**, Vol. 28, pp. 1837 – 1847.
45. Flores-Cerrillo, J. e MacGregor, J. F.; Multivariate monitoring of batch processes using batch-to-batch information, *AIChE Journal*, **2004**, Vol. 50, pp. 1219 – 1228.
46. Lee, D. S. e Vanrolleghem, P. A.; Adaptive consensus Principal Component Analysis for on-line batch process monitoring, *Environmental Monitoring and Assessment*, **2004**, Vol. 92, pp. 119 – 135.
47. Hotelling, H.; *Multivariate quality control*, in *Techniques of Statistical Analysis*, New York: McGraw-Hill, **1947**.
48. Jackson, J. E.; *A user's guide to principal components*, New York: John Wiley and Sons, **1991**.
49. Jackson, J. E. e Mudholkar, G. S.; Control procedures for residuals associated With Principal Component Analysis, *Technometrics*, **1979**, Vol. 21, pp. 341 – 349.
50. Basso, L. C., Amorim, H. V., Oliveira, A. J. e Lopes, M. L.; Yeast selection for fuel ethanol production in Brazil, *Federation of European Microbiological Societies*, **2008**, Vol. 8, pp. 1155 – 1163.
51. Gussem, K. D., Vandenabeele, P., Verbeken, A. e Moens, L.; Raman spectroscopic study of *Lactarius* spores (Russulales, Fungi), *Spectrochimica Acta Part A*, **2005**, Vol. 61, pp. 2896 – 2908.
52. Mudalige, A. e Pemberton, J. E.; Raman spectroscopy of glycerol/D<sub>2</sub>O solutions, *Vibrational Spectroscopy*, **2007**, Vol. 45, pp. 27 – 35.
53. Nakabayashi, T. e Nishi, N.; States of molecular associates in binary mixtures of acetic acid with protic and aprotic polar solvents: a Raman spectroscopic study, *Journal of Physical Chemistry A*, **2002**, Vol. 106, pp. 3491 – 3500.
54. Synytsya, A., Copíková, J., Matejka, P. e Machovic, V.; Fourier transform Raman and Infrared spectroscopy of pectins, *Carbohydrate Polymers*, **2003**, Vol. 54, pp. 97 – 106.

55. Mokrane, A., Friant-Michel, P., Cartie, A. e Rivail, J.; Scaled semiempirical method for the calculation of vibrational spectra Molecular vibrational frequencies of monosaccharides and disaccharides by PM3 method, *Journal of Molecular Structure (Theochem)*, **1997**, Vols. 395-396, pp. 71 – 80.