UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PROCESSOS

MODELAGEM DE PROCESSOS ENZIMÁTICOS E FERMENTATIVOS USANDO OTIMIZAÇÃO POR ENXAME DE PARTÍCULAS

DISSERTAÇÃO DE MESTRADO

Christian Luiz da Silveira

Santa Maria, RS, Brasil

2015

MODELAGEM DE PROCESSOS ENZIMÁTICOS E FERMENTATIVOS USANDO OTIMIZAÇÃO POR ENXAME DE PARTÍCULAS

Christian Luiz da Silveira

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Engenharia de Processos da Universidade Federal de Santa Maria (UFSM, RS) como requisito parcial para obtenção do grau de **Mestre em Engenharia de Processos**.

Orientadora: Prof[a]. Dr[a]. Nina Paula Gonçalves Salau
Coorientador: Prof. Dr. Marcio Antonio Mazutti

Santa Maria, RS, Brasil
2015

Ficha catalográfica elaborada através do Programa de Geração Automática
da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

da Silveira, Christian Luiz
MODELAGEM DE PROCESSOS ENZIMÁTICOS E FERMENTATIVOS USANDO OTIMIZAÇÃO POR ENXAME DE PARTÍCULAS / Christian Luiz da Silveira.-2015.
154 p.; 30cm

Orientadora: Nina Paula Gonçalves Salau
Coorientador: Marcio Antonio Mazutti
Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Processos, RS, 2015

1. Modelagem e simulação 2. Estimação de parâmetro 3. Bioprocessos 4. Engenharia Química I. Salau, Nina Paula Gonçalves II. Mazutti, Marcio Antonio III. Título.

Universidade Federal de Santa Maria

Centro de Tecnologia
Pós-Graduação em Engenharia de Processos

A Comissão Examinadora, abaixo assinada, aprova a Dissertação de Mestrado

MODELAGEM DE PROCESSOS ENZIMÁTICOS E FERMENTATIVOS USANDO OTIMIZAÇÃO POR ENXAME DE PARTÍCULAS

Elaborada por
Christian Luiz da Silveira

como requisito parcial para obtenção do grau de
Mestre em Engenharia de Processos

COMISSÃO EXAMINADORA

_____
Nina Paula Gonçalves Salau
(Presidente/Orientadora)

_____
Débora Jung Luvizetto Faccin, Dra. (UFRGS)

_____
Fernanda de Castilhos, Dra. (UFSM)

Santa Maria, 20 de Fevereiro, 2015.

Agradecimentos

Encerramentos de ciclos, como o marcado por este trabalho, tornam o ato de agradecer uma tarefa difícil de ser cumprida. Ora, agradecer pela conclusão do mestrado e elaboração desta dissertação implica em agradecer a quem me ajudou na formação como Engenheiro Químico. Por conseguinte, obviamente que o que veio antes disso importa, deveria, portanto, agradecer aos que me ofereceram conhecimentos e educação antes disso, nos ensinos médio e fundamental.

Poderia seguir esse trem das causas *ad infinitum*. Contudo, não vou além do que me construiu desde os primeiros anos, do mais importante: a família. Agradeço veementemente, incansavelmente e profundamente a meus pais irmãos. Pai, obrigado por ter sempre me mostrado a sensatez, a sobriedade e a fortaleza. Mãe, obrigado pela força e carinho, e pelo temperamento enérgico que, muitas vezes, se faz tão necessário quanto a calma mais firme. Mano, muito obrigado por todas as lições, desde sempre, por toda a ajuda que tu deste, que eu sei que foram, muitas vezes, além do que seria simplesmente viável. Mana, muito obrigado por todo carinho, pelas simulações de provas na infância e pela companhia por anos em que moramos juntos. A vocês, que são minha verdadeira e única família, muito obrigado.

Agradeço a Bibiana, pela ajuda, pelo carinho, pela compreensão e pela paciência em me ouvir incessantemente falando do trabalho. Sei que sou enjoativo quando coloco alguma coisa na cabeça, mas a minha persistência se mescla à teimosia de uma maneira única. Obrigado por tudo, este trabalho tem o toque do teu amor também.

A Nina, muito obrigado pelo apoio e conhecimento oferecidos desde os tempos da graduação. Obrigado pela paciência e dedicação. Espero pelo menos ter correspondido ao esperado e estar deixando esse grau um pouco melhor, mais apto e mais sábio do que quando entrei.

Ao Mazutti, obrigado pelo apoio, tanto na graduação quanto no mestrado. Obrigado por, gentilmente, ter cedido os dados experimentais usados aqui neste trabalho, espero ter feito um bom trabalho com eles.

Agradeço ao Schwaab por todo o conhecimento que compartilhou e pela ajuda dada nesse tempo todo. Com certeza aprendi muito contigo. Aos demais colegas e professores do

departamento, um muito obrigado por terem confiado em mim como mestrando e colega docente neste último ano.

“Of all that is written, I love only

What a person has written

With his own blood.”

Friederich Nietzsche

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia de Processos
Universidade Federal de Santa Maria

MODELAGEM DE PROCESSOS ENZIMÁTICOS E FERMENTATIVOS USANDO OTIMIZAÇÃO POR ENXAME DE PARTÍCULAS

AUTOR: Christian Luiz da Silveira
ORIENTADORA: Nina Paula Gonçalves Salau
COORIENTADOR: Marcio Antonio Mazutti
Data e Local de Defesa: Santa Maria, 20 de Fevereiro, 2015.

A modelagem e a simulação de processos consistem em um recurso de grande importância para diversos processos químicos e biotecnológicos. A simulação de processos nos permite predizer o comportamento das variáveis de estado do processo, levando-nos a vantagens técnicas e econômicas, como, por exemplo, a prevenção de perdas de tempo e insumos por não conhecer particularidades do processo, a garantia de segurança, a qualidade do produto e, principalmente, a otimização do processo, permitindo estudar e alcançar as melhores condições para o referido processo, o que deve culminar em mais produto produzido com melhor qualidade e com menos esforços e custos. Neste trabalho, a modelagem e a simulação de dois processos biotecnológicos – hidrólise enzimática e fermentação em estado sólido – foram feitas com a validação dos modelos propostos com dados experimentais através da técnica de estimação de parâmetros para permitir ao engenheiro prever o comportamento do processo e tomar decisões. O procedimento de modelagem também envolve a avaliação de equações diferenciais e de equações algébricas, dessa forma, o engenheiro deve estar apto a usar diferentes métodos de integração numérica. Dois procedimentos principais de estimação de parâmetros foram utilizados – Otimização por Enxame de Partículas e Levenberg-Marquardt -, e se lançou mão de dois métodos de integração numérica, Runge-Kutta e Dormand-Prince. Dados experimentais de trabalhos anteriores foram utilizados para realizar diversos testes para assegurar a precisão dos modelos em predizer as variáveis de estado do processo e, portanto, serem modelos confiáveis e úteis. No primeiro artigo apresentado, diversos modelos mecanicísticos e empíricos foram testados para se ajustarem aos dados experimentais da hidrólise enzimática; testes estatísticos foram realizados para verificar qual dos modelos melhor descreveria o processo, de forma que o melhor modelo se mostrou ser um modelo totalmente empírico não-autônomo. Os demais artigos tratam da modelagem de um processo de fermentação em estado sólido. Verificou-se que o modelo é bastante preciso e adequado para o uso em predições, principalmente para o perfil de temperatura no leito do biorreator, uma vez que o modelo prevê os gradientes de temperatura ao longo do tempo e da altura do leito. Também, procedimentos numéricos, tais como a análise de identifiabilidade dos parâmetros, para a percepção de quais são os parâmetros mais importantes para a estimação, e a reparametrização do modelo, para reduzir o número total de parâmetros a serem estimados e evitar problemas de magnitude no modelo, foram empregadas com sucesso. Este trabalho mostrou que a modelagem e a simulação de processos possuem enorme importância para a indústria, e diferentes técnicas podem ser aplicadas com maior ou menor esforço e sucesso. Além disso, espera-se que o trabalho tenha contribuído para o estado da arte em modelagem, de uma maneira geral, na área de bioprocessos.

Palavras-chave: bioprocessos, modelagem e simulação, estimação de parâmetros, integração numérica.

ABSTRACT

Thesis for the degree of Master of Science
Post-Graduation Program in Process Engineering
Universidade Federal de Santa Maria

ENZYMATIC AND FERMENTATIVE PROCESSES MODELING USING PARTICLE SWARM OPTIMIZATION

AUTHOR: Christian Luiz da Silveira
ADVISOR: Nina Paula Gonçalves Salau
COADVISOR: Marcio Antonio Mazutti
Date and Local: Santa Maria, February 20th, 2015.

The process modeling and simulation is a greatly important procedure for many chemical and biotechnological processes. The process simulation allows to predict elementary behavior of the state variables of the process, leading to many economical and process advantages, such as the avoidance of losses of time and materials for not knowing the process particularities, the safety guarantee, the product quality, and, mostly the process optimization, permitting to study and to reach the best conditions of a process, which shall yield in more products with quality produced with less effort and expenditures. In this work, the modeling and simulation of two biological processes – enzymatic hydrolysis and solid state fermentation – were performed in order to develop models and estimate parameters that enable an engineer to predict the process behavior and to make decisions about the process. The modeling procedure also involves the computing of differential equations, and algebraic-differential equations; in this manner, the engineer must be able to use different numerical integration methods. Mainly two parameters estimation procedures were used – Particle-Swarm Optimization and Levenberg-Marquardt -, and two numerical integration methods were also resorted – Runge-Kutta and Dormand-Prince. Experimental data from previous works were used to perform several tests in order to assure that the models were predicting correctly the state variables of the process and, in this manner, were reliable and useful. In the first paper, several mechanistic and empirical models are tested to fit the enzymatic hydrolysis experimental data; statistical tests were performed to verify which of those models would best describe the process, which was found to be and entirely empirical non-autonomous model.
The following papers are about the modeling of the solid-state fermentation process. The model was found to be very accurate and adequate to be used for predictions, mainly for the bed temperature of the packed-bed bioreactor, since it could predict the temperature gradients along the time and height of the bed. Also, some numerical procedures such as parameters identifiability, to realize which were the most important parameters to be estimated, and model reparametrization, to reduce the total number of parameters to be estimated and avoid magnitude problems of the model, were successfully performed. This work has shown that the modeling and simulation of processes holds an enormous importance for industry, and different techniques can be applied with more or less effort and success. Further, hopefully, this work has contributed to the state of the art of modeling, in a general way, for biological processes.

Keywords: biological process, modeling and simulation, parameters estimation, numerical integration.

TABELAS

CAPÍTULO 2 - REVISÃO

CAPÍTULO 3

CAPÍTULO 4

CAPÍTULO 5

CAPÍTULO 6

FIGURAS

CAPÍTULO 2 - REVISÃO

CAPÍTULO 3

CAPÍTULO 4

CAPÍTULO 5

CAPÍTULO 6

APÊNDICES

ANEXOS

ABREVIAÇÕES

ANN	Rede Neuronal Artifical (*Artificial Neural Network*)
CAE	Engenharia Assistida por Computador (*Computer Aided Engineering*)
CFD	Fluidodinâmica Computacional (*Computational Fluid Dynamics*)
DE	Equação Diferencial (*Diferential Equation*)
EH	Hidrólise Enzimática (*Enzymatic Hydrolysis*)
EVOC	Computação Evolucionária (*Evolutionary Computation*)
EVOP	Programação Evolucionária (*Evolutionary Programming*)
FAMEs	Ácidos Graxos Metil-Éster (*Fatty Acid Methyl Esters*)
FEA	Análise de Elementos Finitos (*Finite Element Analysis*)
GA	Algoritmo Genético (*Genetic Algorithm*)
LS	Mínimos Quadrados (*Least Squares*)
MBD	Dinâmica de Múltiplos Corpos (*Multibody Dynamics*)
NE	Número de Experimentos (*Number of experiments*)
ODE	Equação Diferencial Ordinária (*Ordinary Differential Equation*)
OF	Função Objetivo (*Objective Function*)
PDE	Equação Diferencial Parcial (*Partial Differential Equation)*
PSO	Otimização por Enxame de Partículas (*Particle Swarm Optimization)*
SA	Arrefecimento Simulado (*Simulated Annealing*)
SmF	Fermentação Submersa (*Submerged Fermentation*)
SSF	Fermentação em Estado Sólido (*Solid State Fermentation*)
TRS	Açúcares Totais Liberados (*Total Released Sugar*)

SUMÁRIO

1 INTRODUÇÃO

1.1 Motivação

A Engenharia Assistida por Computador (CAE) é um ramo crescente da Engenharia Química mundial que compreende a Análise de Elementos Finitos (FEA), a Fluidodinâmica Computacional (CFD), a Dinâmica de Múltiplos Corpos (MBD), a Otimização, entre outros. A otimização, por sua vez, compreende a modelagem de processos e a estimação de parâmetros. Nos últimos anos, com a melhoria do processamento computacional, o uso de ferramentas computacionais é crescente e está se tornando cada vez mais eficiente.

A modelagem de processos tem se tornado uma prática bastante comum na Engenharia Química. Como reportado por Dobre e Marcano (2007), o principal propósito pelo qual um modelo é criado é o de projetar e de desenvolver uma planta de escala industrial usando dados de escala laboratorial. Além disso, Pinto e Lage (2001) descrevem modelos matemáticos nas Engenharias Química e de Processos como ferramentas para tornar projetos e suas otimizações possíveis. Também, na ausência de modelos matemáticos estaríamos fadados a repetir o que já foi feito ou a gastar tempo e dinheiro tentando alcançar processos ou condições novas e melhores (Pinto e Lage, 2001).

Muitos exemplos de modelagem, simulação e controle em bioprocessos estão disponíveis na literatura. Vats e Negi (2013) usaram uma Rede Neuronal Artificial (ANN) para a modelagem e o controle da liberação de açúcar redutor e de polifenóis em folhagem de pinheiro; Meng *et al*. (2014) usaram ANN para predições da viscosidade cinemática do biodiesel. Gianfrancesco *et al*. (2010) usaram a CFD para simular o processo de *spray drying* de uma solução de maltodextrina em um atomizador rotatório; Elqotbi *et al*. (2013) usaram a CFD para modelar e simular um biorreator agitado com duas fases usando equações de Navier-Stokes. Rodríguez-Fernández *et al.* (2011) usaram a modelagem matemática para determinar parâmetros cinéticos relacionados ao processo de fermentação em estado sólido (SSF) e Hashemi *et al*. (2011) usaram modelagem matemática para predizer a produção de α-amilase por SSF usando *Bacillus sp*.

O processo de SSF é definido como uma fermentação cujas condições não permitem água livre no substrato; portanto, o teor de umidade deve ser o suficiente apenas para possibilitar o crescimento e manutenção do metabolismo dos microrganismos (Pandey, 2003). Diversos produtos são obtidos por SSF; Han *et al*. (2001) apresentaram um produto semelhante ao queijo,

chamado *sufu*, pela fermentação da soja; além disso, SSF pode produzir enzimas, tais como α-amilase, frutosil-transferase, lipase e pectinase, ácidos orgânicos como ácidos lático e cítrico, e goma xantana (Couto e Sanromán, 2006).

Algumas diferenças entre SSF e Fermentação submersa (SmF) podem ser salientadas. Hölker *et al*. (2004) apresentaram algumas vantagens da SSF comparada à SmF, tais como a menor demanda de água e de esterilização, a fermentação de substrato sólido insolúvel em água e menos demanda de energia para aquecimento; além disso, Viniegra-González *et al*. (2003) relataram maior produção enzimática em SSF que em SmF. No entanto, o escalonamento do processo ainda é prejudicado por problemas como, por exemplo, grandes gradientes de temperatura no biorreator, causados pelo aquecimento gerado pelo metabolismo celular (Hölker and Lenz, 2005).

A modelagem de processos aplicada à SSF tem um importante papel, podendo ser determinante para o escalonamento e otimização do processo. Modelos matemáticos que permitem a descrição do crescimento dos microrganismos e a transferência de calor nos fermentadores têm sido desenvolvidos, sendo tais modelos cruciais para o entendimento do reator e, dessa forma, para o projeto de *layouts* e estratégias de controle para a otimização do processo (Lenz *et al*., 2004).

Contudo, para alcançar um modelo confiável para ser utilizado na otimização do processo, alguns passos precisam ser tomados. Primeiramente, os fatos que cercam o processo devem ser estudados; então, algumas hipóteses são feitas, por exemplo, se o processo pode ser considerado isotérmico, ou se os gases de uma mistura podem ser considerados ideais; depois, o sistema de equações obtido é concatenado em um só modelo completo; por fim, o modelo deve ser testado e melhorado. Quanto melhor o modelo for em descrever a realidade, melhor o modelo (Pinto e Lage, 2001).

Melhorar um modelo matemático pode significar assumir outra hipótese, incluir outro fator ou estimar os parâmetros do modelo até que este alcance o seu melhor desempenho. De acordo com Schwaab e Pinto (2007), estimar parâmetros significa usar um modelo de referência e variar seus parâmetros até que as predições do modelo cheguem o mais próximo possível dos dados experimentais. Portanto, estar apenas próximo dos dados experimentais não é o suficiente, o modelo precisa estar o mais próximo possível dos dados experimentais; então isso é, também, um problema de otimização, ou de CAE.

Além disso, reparametrizar um modelo pode ser útil para melhorá-lo. A reparametrização pode evitar altas correlações no modelo e diminuir o custo computacional para resolver o problema (Schwaab *et al.*, 2008b; Espie e Macchietto, 1988).

A definição de uma função objetivo (OF) se torna importante para expressar o quão bom está o modelo ajustado. Minimizar a OF é maximizar a probabilidade de obter os resultados experimentais. Depois que a OF é definida, o próximo passo é alcançar o seu mínimo; isso pode ser feito por métodos derivativos, de busca direta ou ainda por métodos heurísticos de otimização para a estimação de parâmetros (Schwaab *et al*., 2008a).

A fim de levar o bioprocesso a melhores condições operacionais e para obter melhores resultados, a modelagem e a simulação de processos são tópicos extremamente importantes. O processo de modelagem, que contém diversas etapas, uma vez feito com sucesso pode resultar num modelo bom o suficiente para ser utilizado na otimização de um processo, escalonamento, orientações para um novo *layout* ou para controle, podendo ser útil especialmente para a pesquisa e a indústria.

1.2 Objetivos

O objetivo deste trabalho é estudar a modelagem de dois bioprocessos diferentes, a saber, a hidrólise enzimática (EH) e a SSF. A estimação de parâmetros dos modelos propostos para ambos os processos é estudada, usando dois tipos de estimadores: um de método derivativo e outro heurístico.

Os principais objetivos são:

- Estudar o processo de EH, selecionar modelos utilizados na literatura e propor o melhor modelo para descrever os dados experimentais do processo;
- Estudar o processo de SSF, propor um modelo que descreva o processo e validar com dados experimentais;
- Estimar parâmetros usando tanto um método derivativo quanto um método heurístico;
- Utilizar técnicas de análise de identifiabilidade e reparametrização para melhorar o processo de estimação de parâmetros.

1.3 Estrutura do Trabalho

Esta dissertação está dividida em 7 capítulos, conforme descrições a seguir:

Capítulo 1 – INTRODUÇÃO

Este capítulo consiste em uma breve introdução como uma orientação relativa à motivação e objetivos do trabalho.

Capítulo 2 – REVISÃO

Neste capítulo, uma revisão sobre os aspectos gerais do trabalho será apresentada. Haverá uma introdução aos conceitos dos dois bioprocessos estudados, EH e SSF. Também, conceitos de modelagem serão apresentados, e métodos numéricos frequentemente utilizados para simulação serão explicados. Por último, alguns métodos de otimização que podem ser utilizados em estimações de parâmetros serão apresentados.

Capítulo 3 - ENZYMATIC HYDROLYSIS MODELS SELECTION USING PARTICLE SWARM OPTIMIZATION (Submitted)

Este capítulo é composto pelo primeiro trabalho desenvolvido neste mestrado. Esta é a primeira abordagem dos conceitos de modelagem e, principalmente, do uso da Otimização por Enxame de Partículas (PSO). Os dados da EH foram obtidos por Collares *et al*. (2012) e utilizados neste trabalho para a escolha dos modelos utilizando o PSO para a estimação de parâmetros.

Capítulo 4 - MODELING THE MICROBIAL GROWTH AND TEMPERATURE PROFILE IN A FIXED-BED BIOREACTOR (Published)

O quarto capítulo contém o primeiro trabalho utilizando SSF e é, também, a primeira publicação desta dissertação. Neste trabalho, um modelo simples de crescimento celular e perfil de temperatura no reator é proposto. Os parâmetros foram estimados usando um método derivativo e o modelo validado com dados experimentais obtidos por Mazutti *et al*. (2010).

Capítulo 5 - IDENTIFIABILITY MEASURES TO SELECT THE PARAMETERS TO BE ESTIMATED IN A SOLID-STATE FERMENTATION MODEL (Submitted)

O capítulo 5 é composto por um aprimoramento do modelo anterior. Além do crescimento celular e do perfil de temperatura no reator, o modelo agora compreende também o consumo de substrato, a produção de etanol, CO_2 e O_2, e uma nova variável de estado chamada fator fisiológico foi adicionada. Ainda, uma análise de identifiabilidade foi feita para que se verificasse quais são os parâmetros mais importantes para serem estimados. Neste trabalho o método derivativo de Levenberg-Marquardt foi utilizado para a estimação de parâmetros.

Capítulo 6 - SOLID-STATE FERMENTATION PROCESS MODEL REPARAMETRIZATION PROCEDURE FOR PARAMETERS ESTIMATION USING PARTICLE SWARM OPTIMIZATION (Accepted)

O capítulo 6 é composto pelo último trabalho desta pesquisa. Procurando diminuir o número de parâmetros a serem estimados e suas correlações, um procedimento de reparametrização foi feito. Uma vez que o modelo assumiu uma nova forma, os parâmetros sofreram alterações e o PSO foi utilizado outra vez para encontrar a região mais promissora dos parâmetros. Então, um método derivativo foi utilizado para assegurar que o mínimo da OF foi alcançado.

Capítulo 7 – DISCUSSÃO

O capítulo 7 contém a discussão sobre os resultados obtidos. Este capítulo busca unir as informações dos quatro capítulos anteriores em uma breve discussão.

Capítulo 8 – CONCLUSÕES

Este último capítulo apresenta as principais conclusões da pesquisa, uma vez que as conclusões específicas foram apresentadas em cada um dos capítulos, contendo também sugestões para futuros trabalhos nesta área.

2 REVISÃO

O procedimento de modelagem requer alguns passos que devem ser cuidadosamente tomados. Primeiramente, um processo precisa ser escolhido para o desenvolvimento do modelo, neste trabalho dois bioprocessos diferentes foram abordados: um processo de hidrólise enzimática (EH) e um de fermentação em estado sólido (SSF).

Tendo o processo escolhido, este deve ser estudado para a construção correta do modelo que o descreve fenomenologicamente ou empiricamente. Este modelo deve ser desenvolvido para que possa ser simulado. Modelos dinâmicos, são, geralmente, compostos por equações diferenciais ordinárias (ODE) dependentes do tempo, o que quer dizer que devem ser integradas usando um método numérico específico. Existem diferentes métodos numéricos que podem ser usados, cada um apresentando suas próprias vantagens e desvantagens.

Se o modelo pode ser integrado por algum método, o próximo passo é ajustá-lo ou otimizá-lo. Neste trabalho, a otimização consiste basicamente em estimação de parâmetros, usando algoritmos heurísticos e métodos derivativos, que serão discutidos adiante. O procedimento de estimação de parâmetros consiste em minimizar uma função objetivo (OF), de forma que se obtenha o dado simulado o mais próximo possível do dado experimental.

2.1 Bioprocessos

Atualmente existem diversos bioprocessos conhecidos. Todos esses processos têm seus próprios modelos e bases teóricas. Neste trabalho, apenas os casos específicos de EH e SSF serão discutidos por serem o foco deste trabalho de dissertação.

2.1.1 Hidrólise Enzimática

Russell (1981) define a hidrólise como a reação entre um íon e a água. Exemplificando com uma hidrólise envolvendo um ânion:

$$A^- + H_2O \leftrightarrow HA + OH^-$$

Portanto, esta reação remove prótons das moléculas da água para formar *HA* e íons de hidróxido. Um exemplo comum de hidrólise é a dissociação da molécula de sacarose em glicose e frutose na presença de água, como pode ser visto na Figura 1.

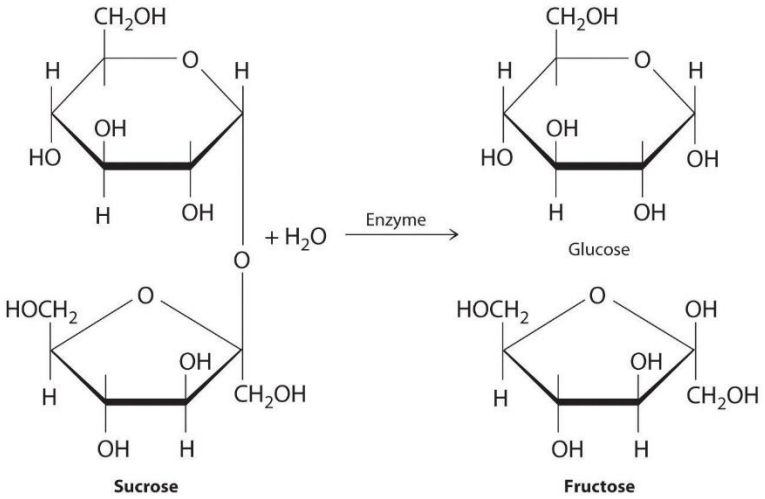

Figura 1. Dissociação hidrolítica da sacarose em glucose e frutose.
Fonte: AVERILL e ELDREDGE, 2014.

A hidrólise pode ser realizada tanto pelo uso de ácidos quanto pelo uso de enzimas. A EH é um processo biotecnológico muito comum, geralmente presente na indústria de alimentos. Algumas das enzimas mais utilizadas são:

i. α-Amilase: esta enzima é amplamente encontrada na natureza e age sobre moléculas de amido. A ação da α-amilase sobre a amilopectina resulta em glicose e maltose. Essa enzima pode ser obtida pelo uso do microrganismo *Aspergillus oryzae* (Aquarone *et al.*, 2001; Souza e Magalhães, 2010).
ii. Pectinase: esta enzima hidrolisa a pectina presente em plantas. Pectinases têm um papel no enfraquecimento do tecido de algumas plantas, auxiliando um futuro processo de hidrólise (Sharma *et al.*, 2013).
iii. Amiloglucosidase: também conhecida como glucoamilase, é uma exoenzima que age como catalisadora de reações de hidrólise, principalmente de ligações α(1-6). Esta é a única enzima de sacarificação capaz de hidrolisar completamente o amido em glicose (Risso, 2014).

Rattanachomsri *et al.* (2009) usaram uma hidrólise multi-enzimática para obter açúcar redutor. As enzimas utilizadas em seu trabalho foram celulase, pectinase, β-glucosidase, α-

amilase e glucoamilase. A eficiência máxima alcançada em sua hidrólise foi de 85,3% em base de açúcar fermentescível obtido (glicose e xilose). Por outro lado, Paolucci-Jeanjean *et al*. (2000) usaram apenas α-amilase na hidrólise enzimática do amido da mandioca e alcançaram uma conversão para açúcares redutores totais (TRS) de 25 g/L em 1 hora de reação.

Collares *et al.* (2012) apresentaram o uso da α-amilase, da pectinase e da amiloglucosidase na hidrólise enzimática do amido da mandioca para obter açúcares fermentescíveis. Seus resultados mostraram que a concentração de 0,52 a 0,84 mg de amiloglucosidase/g de amido levou a uma concentração máxima de TRS e à eficiência máxima da hidrólise. Além disso, o uso da pectinase aumentou a eficiência da hidrólise e da liberação de açúcares totais.

A fim de predizer a quantidade de TRS obtida pela hidrólise, é importante desenvolver um modelo matemático que descreva o processo com grande fidelidade.

Diversas hipóteses foram propostas até agora para predizer completamente o processo de hidrólise enzimática, como a hipótese da inibição das celulases pela celubiose e a hipótese da desnaturação enzimática. Bezerra *et al.* (2011) apresentaram alguns dos modelos propostos:

i. Modelo de adsorção de Langmuir explicando a catálise enzimática da celulose;
ii. Modelos empíricos baseados em Michaelis-Menten;
iii. Concentração de substrato constante;
iv. Cinéticas fractais.

Polakovic e Bryjak (2004) apresentaram algumas equações de taxa para descrever o substrato consumido e o produto obtido num processo de hidrólise enzimática como função da concentração de enzimas, substrato, produto e tempo.

Brown *et al*. (2010) compararam seis diferentes modelos mecanísticos. Em seu trabalho é reportado que a hidrólise de substratos lignocelulósicos depende de algumas características enzimáticas, tais como a adsorção da enzima na biomassa lignocelulósica antes da reação, inibição por produto competitiva ou não-competitiva, sinergia de componentes enzimáticos, transferência de enzima para substrato, tamanho de partículas, cristalinidade, entre outros fatores. Apesar de todos estes fatores serem importantes para o processo, ainda se pretende englobar tudo em um só modelo simplificado.

Ruffell *et al*. (2010) propuseram um modelo empírico baseado no número de Kappa, na carga enzimática e na concentração inicial de fibra recuperada para predizer hexose obtida de celulose. Park *et al*. (2002) usaram um modelo empírico para predizer a conversão de celulose

de papel usado em açúcar redutor através da hidrólise enzimática, este modelo previu com precisão bastante satisfatória os dados experimentais. Vázquez *et al*. (2000) aproximaram a cinética de uma EH pela soma de duas reações de primeira ordem, como um modelo pseudo-empírico, a fim de usar um modelo simples que descrevesse todo o processo. Os termos utilizados representariam as proporções de celulose facilmente hidrolisada e celulose resistente à hidrólise.

2.1.2 Fermentação em estado sólido

O processo de fermentação em estado sólido (SSF) é definido como uma fermentação cujas condições não permitem água livre no substrato; portanto, o teor de umidade deve ser suficiente apenas para suportar o crescimento e a manutenção do metabolismo microbiano (Pandey, 2003). Reid (1989) reportou que se a umidade na SSF for muito baixa, o crescimento do microrganismo é impedido, porém se for muito alta, os espaços entre as partículas podem ser obstruídos, impedindo a circulação de gases e favorecendo a compactação do substrato.

Em outro trabalho, Pandey (1992) reporta que a SSF tem sido usada desde a antiguidade, provavelmente há mais de 3000 anos pelos egípcios para a fabricação de pães, mais tarde na Ásia para a produção de queijo pelo uso da *Penicillium roqueforti*. Na China e no Japão, a SSF começou a ser utilizada há cerca de 2000 anos para a produção de molho de soja *koji* e *miso*.

Na verdade, diversos produtos podem ser obtidos pela SSF; Han *et al*. (2001) apresentaram um produto similar ao queijo, chamado *sufu*, que é um fermentado de soja obtido pela SSF; além disso, SSF pode produzir enzimas como a α-amilase, a fructosil transferase, a lipase e a pectinase, ácidos orgânicos como o lático e o cítrico, e a goma xantana (Couto e Sanromán, 2006). Pandey (1992) apresentou em seu trabalho uma tabela listando diversos microrganismos, seus produtos e o substrato utilizado para sua obtenção.

Ademais, algumas diferenças entre a SSF e a fermentação submersa (SmF) podem ser apontadas. Por exemplo, alguns produtos de fungos, como enzimas e metabólitos secundários, são desenvolvidos para uso em substrato sólido úmido, não em substrato líquido, e o cultivo de microrganismos em suspensões aquosas pode diminuir a eficiência de sua obtenção (Hölker *et al*., 2004).

Viniegra-González *et al*. (2003) compararam a produção de três enzimas diferentes – a invertase, a pectinase e a tanase – por *Aspergillis niger* obtido por SSF e SmF. Os resultados

mostraram que o substrato foi consumido e a biomassa produzida mais rápido por SSF, além disso, as conversões por enzima foram maiores para a SSF.

Ainda a respeito das vantagens e das desvantagens, Hölker *et al*. (2004) apresentaram uma tabela que compreende algumas vantagens da SSF sobre a SmF, tais como: menor demanda por água, alta concentração de produto final, menos demanda de esterilização, alta produtividade, menor demanda energética e fácil aeração. Couto e Sanromán (2006) também apresentaram algumas vantagens, como por exemplo: a tecnologia simples requerida e a melhor circulação de oxigênio. Contudo, os autores também apresentaram algumas desvantagens da SSF sobre a SmF, como dificuldades de escalonamento e difícil controle de parâmetros do processo (pH, calor, umidade, entre outros). As vantagens apresentadas por ambos, Hölker *et al*. (2004) e Couto e Sanromán (2006), foram comparadas de forma resumida na Tabela 1.

Tabela 1. Vantagens da SSF e SmF comparadas uma à outra.

SSF	SmF
Maior produtividade	Menores gradientes (pH, temperatura, oxigênio)
Melhor circulação de oxigênio	Mais fácil de controlar as variáveis do processo
Substrato de baixo custo	Menor dificuldade para escalonar
Baixa demanda de água	Maior efetividade de mistura
Alta concentração de produto final	
Baixa/inexistente repressão por catabólito	
Baixa demanda por esterilização	
Simula o ambiente natural	
Baixa demanda energética	

Apesar das vantagens da SSF sobre a SmF, algumas desvantagens da SSF ainda dificultam o escalonamento do processo. Um problema fundamental é a grande quantidade de calor produzido pelo metabolismo celular durante o crescimento, o qual causa significantes gradientes de temperatura no leito do reator (Hölker e Lenz, 2005).

Para evitar problemas como diferenças significantes de transferência de calor local e gradientes de temperaturas acentuados é importante desenvolver um bom *layout* do reator e boas estratégias de controle para a otimização do processo. Essas ações estão intimamente ligadas ao modelo matemático do processo, o qual pode descrever a cinética do crescimento

microbiano e a transferência de calor nos fermentadores. Dessa maneira, a modelagem do processo de SSF tem sua importância para alcançar o escalonamento e a otimização do processo (Lenz *et al*., 2004).

2.2 Engenharia Assistida por Computadores (CAE)

Conhecendo o processo que será estudado e de posse de seus dados experimentais, o próximo passo é lançar mão de estratégias computacionais para monitorar, controlar e otimizar as condições operacionais. Tudo isso pode ser feito sob o uso e análise de modelos e simulação por ferramentas computacionais.

Nisto reside a importância da Engenharia Assistida por Computadores (CAE), um ramo crescente da Engenharia Química que compreende a Análise de Elementos Finitos (FEA), a Fluidodinâmica Computacional (CFD), a Dinâmica de Múltiplos Corpos (MBD) e a Otimização. Como uma tendência resultante do crescimento da capacidade computacional, mais atenção tem sido dada para tais ferramentas computacionais nas Engenharias Química (EQ) e de Processos.

No início dos anos 70 foram publicados os primeiros trabalhos usando FORTRAN. Na época os computadores eram muito inferiores em capacidade e velocidade. Com o desenvolvimento de melhores computadores, as linguagens de programação também passaram por modificações e melhorias. Muitas linguagens, sistemas operacionais e softwares estão disponíveis atualmente, cada um compreendendo certos aspectos da CAE com mais ênfase.

2.2.1 Modelagem e simulação

A modelagem de processos tem se tornado uma prática comum na EQ. Dobre e Marcano (2007) definem o primeiro objetivo da modelagem como o desenvolvimento de um modelo que possa ser usado para investigar certo problema. Esse modelo pode compreender um processo inteiro em partes, i.e., pode ser subdividido para cada parte ou equipamento de um processo e unido para formar o modelo do processo completo. Pode ser utilizado para desenvolver uma planta industrial baseado em dados obtidos em laboratório ou escala piloto.

Existem diversos tipos de modelos matemáticos usados para descrever processos de EQ. Geralmente, modelos matemáticos podem ser divididos em empíricos e fenomenológicos. Modelos empíricos utilizam equações para descrever dados experimentais, não sendo apoiados por preceitos teóricos. Um modelo empírico famoso é o de Verhulst (1838), que utilizou

estatísticas populacionais para propor um modelo de crescimento populacional, também chamado de Equação Logística, cf. Eq. 1.

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \quad (1)$$

Onde *N(t)* é o número populacional no tempo *t*, *r* é a taxa de crescimento e *K* é o máximo populacional permitido.

Modelos fenomenológicos, contudo, são suportados por bases teóricas, tais como a termodinâmica, a conservação de energia e massa e a transferência de massa e calor. Estes modelos não podem ser considerados melhores que os empíricos, contudo, estes são válidos para uma faixa muito maior de condições operacionais de um processo. Um modelo empírico é válido somente até onde se tem dados experimentais, não podendo ser extrapolado para as mais diversas condições operacionais. Um modelo fenomenológico, porém, pode ser válido para um processo completo (Pinto e Lage, 2001).

Existe um gênero de modelos chamado híbrido. Este tipo de modelo matemático pode combinar equações empíricas e fenomenológicas.

Outra distinção útil é definir modelos a parâmetros concentrados ou distribuídos. Modelos a parâmetros concentrados são modelos cujas variáveis não mudam ao longo do espaço, como na hipótese de um CSTR perfeitamente agitado. Modelos a parâmetros distribuídos são mais comuns na modelagem de reatores PFR, quando concentração, por exemplo, varia com o tempo e com o comprimento do reator.

Pinto e Lage (2001) enfatizam que modelos matemáticos são ferramentas para desenvolver e otimizar projetos. Sem modelagens, provavelmente estaríamos fadados a repetir o que já foi feito, ou a perder tempo e dinheiro tentando alcançar novos e melhores processos ou condições operacionais de um processo.

Existem diversos exemplos de uso de CAE na EQ. Vats e Negi (2013) usaram uma Rede Neuronal Artificial (ANN) para a modelagem e o controle da liberação de açúcar redutor e de polifenóis de folhagem de pinheiro. Eles utilizaram diferentes percentagens e cargas de NaOH, de potência de microondas (Watts) e de volume de água como dados de entrada na ANN para predizer o total de açúcares e polifenóis liberados (saída/resposta).

Meng *et al*. (2014) utilizaram ANN para predizer a viscosidade cinemática do biodiesel a 313 K com uma única entrada equivalente à fração mássica de 19 tipos de ácidos graxos metil-éster (FAMEs). Eles utilizaram esta abordagem para evitar o uso individual das viscosidades cinemáticas dos FAMEs, necessários para as equações de mistura.

Outra ferramenta CAE útil são softwares de CFD. Gianfrancesco *et al*. (2010) utilizaram a CFD para simular o processo de *spray drying* de uma solução de maltodextrina com um atomizador rotatório. A CFD é útil para a resolução das diversas Equações Diferenciais Parciais (PDE) que compõem o modelo, com variáveis em função do tempo e do espaço.

Ainda sobre o uso de CFD, Elqotbi *et al*. (2013) usaram esta abordagem para modelar e simular um reator agitado bifásico utilizando equações de Navier-Stokes. O modelo foi usado para melhor entendimento do mecanismo do reator, seu comportamento sob mudanças na viscosidade e a ação dos impelidores na cinética das reações.

A modelagem matemática também foi utilizada por Rodríguez-Fernández *et al*. (2011) para determinar os parâmetros cinéticos relacionados à SSF. Diferentes fluxos de ar foram usados para a estimação da taxa de crescimento microbiana específica, do coeficiente de manutenção das células e do coeficiente de conversão biomassa/oxigênio.

A fim de predizer a produção de α-amilase num processo de SSF utilizando *Bacillus sp.*, Hashemi *et al*. (2011) utilizaram um modelo matemático fenomenológico baseado em balanços de massa. Os autores reportaram terem descrito com sucesso a produção de α-amilase no biorreator em batelada.

Como a importância da modelagem da SSF já foi estabelecida, é importante salientar os passos do processo para a construção de um modelo. Pinto e Lage (2001) descreveram alguns passos que devem ser tomados no desenvolvimento de um modelo matemático. Primeiramente, os fatos devem ser estudados, ou seja, determinar quais são os fenômenos e as relações constitutivas que compõem o processo. Depois, algumas hipóteses devem ser feitas como, por exemplo, se o processo pode ser considerado isotérmico ou se a mistura de gases pode ser considerada ideal. Então, o modelo é criado, as equações são unidas em um sistema para formar um único modelo matemático. Por fim, o modelo deve ser testado e aperfeiçoado. Quanto mais precisamente o modelo descrever a realidade, melhor é o modelo.

Portanto, estar apenas próximo dos dados experimentais, ou da realidade, não é o suficiente para um modelo, ele deve estar o mais próximo possível da realidade. Dessa maneira, a modelagem também é um problema de otimização; portanto, um problema de CAE. Aprimorar um modelo matemático envolve assumir hipóteses, incluir fatores diferentes e estimar parâmetros que compõem o modelo até que este alcance o seu melhor desempenho (Schwaab e Pinto, 2007).

Obviamente, todas essas hipóteses e fatores a serem incluídas no modelo dependem do tipo de modelo matemático. Modelos fenomenológicos, geralmente, terão seus fatores e parâmetros adicionados como hipóteses teóricas, como taxa constante ou inibição por substrato

para o perfil de crescimento microbiano no caso de modelos de bioprocessos. Modelos apoiados por planejamento experimental, por exemplo, podem ter fatores lineares, quadráticos ou ainda de múltiplas interações a serem analisados.

Em alguns modelos, parâmetros e variáveis podem estar fortemente correlacionados, ou ter magnitudes muito diferentes uns dos outros. Isso pode prejudicar a resolução do problema, requerendo muito esforço computacional para tal. Para superar este problema, às vezes um procedimento muito comum que pode ser utilizado é a reparametrização do modelo. O procedimento de reparametrização pode evitar alta correlação paramétrica e diminuir a demanda computacional para resolver o problema, sendo útil para melhorar os resultados (Schwaab *et al*., 2008b; Espie e Macchietto, 1988). Além disso, Espie e Macchietto (1988) reportam que diferentes transformações podem ser escolhidas, podendo ou não ser adequadas, afetando a eficiência e a robustez da função objetivo de mínimos quadrados em algoritmos de minimização.

Para conduzir bioprocessos em melhores condições operacionais e obter melhores resultados, a modelagem de processos se torna muito importante. A modelagem, que compreende diversos passos, uma vez feita com sucesso pode resultar em um modelo bom o suficiente para ser utilizado na otimização, escalonamento, orientações para um novo layout e controle do processo, com resultados subsequentes favoráveis para a indústria e a pesquisa. Dessa forma, modelos para processos de SSF e de EH foram desenvolvidos.

O crescimento microbiano em SSF tem sido descrito pelo modelo empírico desenvolvido por Verhulst (1838). Na época, o matemático belga Verhulst desenvolveu a equação logística para descrever sua ideia de que o crescimento exponencial de populações não duraria para sempre, mas apenas até enfrentar dificuldades como a escassez de comida e de espaço, alcançando um chamado estado estacionário populacional (Bacaër, 2011).

O modelo de Verhulst tem sido usado e ajustado com sucesso para dados experimentais de crescimento microbiano para SSF (Hamidi-Esfahani *et al.*, 2004; Mitchell *et al.*, 1991; Lagemaat e Pyle, 2005). O modelo referido é descrito pela Eq. 1, porém, para o processo de SSF ele é representado por:

$$\frac{dX}{dt} = \mu X \left(1 - \frac{X}{X_{max}}\right) \tag{2}$$

Onde X representa a concentração de biomassa, X_{max} a concentração máxima de biomassa e μ a taxa de crescimento celular.

A taxa de crescimento celular pode ser obtida assumindo algumas hipóteses. Geralmente, uma das três hipóteses abaixo é adotada:

i. Taxa de crescimento específica constante:

$$\mu = constante \tag{3}$$

ii. Taxa de crescimento específica baseada em Monod:

$$\mu = \frac{\mu_{max} S}{K_S + S} \tag{4}$$

iii. Taxa de crescimento específica com inibição por substrato:

$$\mu = \frac{\mu_{max} S}{K_S + S + k_1 S^2} \tag{5}$$

Onde μ_{max} é a taxa de crescimento específico máxima, S é a concentração de substrato, K_S é a constante de meia-velocidade e k_1 é a constante de dissociação.

A equação do balanço de energia da SSF precisa considerar as densidades e os calores específicos do leito, do substrato e da umidade do ar (Fanaei e Vaziri, 2008; Silveira *et al.*, 2014). Também, deve considerar o calor metabólico gerado pelo crescimento microbiano, o calor latente de evaporação da água e a capacidade do ar de carregar a água, resultando na Eq. 6.

$$\rho_b C_{pb} \frac{\partial T}{\partial t} = \rho_s (1 - \varepsilon) Y_Q \frac{dX}{dt} + \rho_a C_{pa} V_z \left(\frac{\partial T}{\partial z}\right) + \rho_a \mathrm{f} \lambda V_z \left(\frac{\partial T}{\partial z}\right) \tag{6}$$

Onde ρ_s, ρ_a, e ρ_b são as densidades do substrato, do ar e do leito, respectivamente; C_{pa} e C_{pb} são, respectivamente, os calores específicos do ar e do leito; ε representa a fração de vazio; Y_Q, o coeficiente de calor metabólico; V_z, a velocidade de alimentação de ar úmido; f, a capacidade do ar de carregar água; e λ, o calor latente de evaporação da água.

A Eq. 6 necessita de duas Equações Algébricas para estar completamente definida. Estas duas equações são utilizadas para a obtenção da densidade do leito, Eq. 7, e do calor específico do leito, Eq. 8.

$$\rho_b = \varepsilon \rho_a + (1 - \varepsilon)\, \rho_s \tag{7}$$

$$C_{pb} = \frac{\varepsilon \rho_a (C_{pa} + \mathrm{f}\lambda) + (1 - \varepsilon) \rho_s C_{ps}}{\rho_b} \tag{8}$$

Conforme pode ser visto na Eq. 7, a densidade do leito é calculada com base na sua composição dividida em uma fração sólida (substrato) e uma fração de vazios (ar). O calor específico do leito leva em consideração os calores específicos do ar e do substrato e também a umidade do ar.

O processo de SSF envolve, além disso, o consumo de substrato e a obtenção de um produto final (etanol), do dióxido de carbono e do oxigênio. Todas essas variáveis de estado podem ser relacionadas ao processo de crescimento microbiano e seu metabolismo por coeficientes de conversão (Mitchell *et al*., 2004), expressas pelas Equações 9-12.

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}}\frac{dX}{dt} \tag{9}$$

$$\frac{dP}{dt} = Y_{P/X}\frac{dX}{dt} \tag{10}$$

$$\frac{dCO_2}{dt} = \frac{1}{Y_{CO_2/X}}\frac{dX}{dt} \tag{11}$$

$$\frac{dO_2}{dt} = \frac{1}{Y_{O_2/X}}\frac{dX}{dt} \tag{12}$$

Onde $Y_{S/X}$, $Y_{P/X}$, $Y_{CO_2/X}$, and $Y_{O_2/X}$ são os coeficiente de conversão de substrato a células, de células a produto, de células a CO_2 e de células a O_2, respectivamente.

2.2.2 Integração numérica

Nas Engenharias Química e de Processos frequentemente encontramos problemas que envolvem equações diferenciais. Equações diferenciais (DE) são equações que compreendem uma relação entre uma função desconhecida e uma ou mais de suas derivadas. Se a equação inclui apenas a derivada relativa a uma variável independente ela é chamada de Equação Diferencial Ordinária (ODE); se existem mais derivadas relativas a outras variáveis independentes, esta é chamada de Equação Diferencial Parcial (PDE) (Davis, 1984).

Nem todas as ED com significado prático possuem solução analítica. Dessa forma, é importante saber como aplicar métodos numéricos para a resolução de equações que,

eventualmente, apareçam num problema de engenharia. Frequentemente esses problemas surgem de fenômenos ou processos transientes, onde a variável independente é unicamente o tempo.

2.2.2.1 Métodos de Euler Explícito e Implícito

A primeira publicação do que viria a ser chamado de Método de Euler está no livro de Leonhard Euler, intitulado "*Introductio in Analysin Infinitorum*" (1748), o qual possui uma versão na língua inglesa (2013). No entanto, esse livro foi escrito em latim, e algumas de suas notações são diferentes atualmente.

Diversos autores descrevem o Método de Euler, incluindo Davis (1984), Ramirez (1989), Parker e Chua (1989) e Bequette (1998).

O Método de Euler Explícito consiste na divisão do intervalo de integração [x_0,x_N] em *N* subintervalos, o que resulta numa variável de passo *h*, Eq. 13.

$$h = \frac{x_N - x_0}{N} \tag{13}$$

Uma vez que o método é utilizado para resolver problemas de valor inicial, ele pode ser usado com uma diferenciação para frente para obter uma discretização conforme Eq. 14.

$$\frac{dx}{dt} = \frac{x_{k+1} - x_k}{t_{k+1} - t_k} = f(x) \tag{14}$$

Como o intervalo de tempo Δt é constante para este método, considera-se que:

$$h = \Delta t = t_{k+1} - t_k \tag{15}$$

Então podemos resolver a Eq. 14 para x_{k+1} para obter o valor da variável de estado para o próximo passo no tempo (Eq. 16).

$$x_{k+1} = x_k + \Delta t\, f(x_k) \tag{16}$$

Uma abordagem similar é utilizada para o método de Euler Implícito; porém, com uma diferenciação para trás. Consequentemente, uma equação implícita é obtida (Eq. 17). Para resolver esta equação, geralmente se requer um método de resolução de equações algébricas, tal como o Método de Newton.

$$x_{k+1} = x_k + \Delta t\, f(x_{k+1}) \tag{17}$$

O Euler Explícito é mais fácil de resolver ou implementar computacionalmente; no entanto, o Euler Implícito permite utilizar intervalos de tempo maiores.

2.2.2.2 Métodos Runge-Kutta

O Método Runge-Kutta (RK) foi proposto por Carl Runge (1895) e, posteriormente, desenvolvido por Wilhelm Kutta (1901).

Os métodos Runge-Kutta são divididos em diferentes subseções, podendo ser de segunda ordem, quarta ordem, quarta-quinta ordem (Fehlberg, 1970), implícito, entre outras abordagens que já foram publicadas em diversos trabalhos.

Os métodos RK clássicos são o de segunda (RK2) e quarta ordem (RK4), os quais são baseados no método de Euler. A seguir o RK4 será descrito. O RK2 é muito similar ao RK4, porém, com menos passos (incrementos).

Quatro coeficientes são calculados a cada passo no tempo no RK4: m_1, m_2, m_3 e m_4. Os coeficientes são funções da variável de estado no tempo *k*. Uma vez que os coeficientes foram calculados, eles são utilizados para computar a variável de estado no tempo *k+1*. O procedimento é feito até o fim do intervalo de integração.

O bloco executado nas iterações é definido abaixo como Equação 18.

$$\begin{aligned} m_1^k &= f(t^k, x^k) \\ m_2^k &= f\left(t^k + \frac{\Delta t}{2}, x^k + m_1^k \frac{\Delta t}{2}\right) \\ m_3^k &= f\left(t^k + \frac{\Delta t}{2}, x^k + m_2^k \frac{\Delta t}{2}\right) \\ m_4^k &= f\left(t^k + \Delta t, x^k + m_3^k \Delta t\right) \\ x^{k+1} &= x^k + \left[\frac{m_1^k}{6} + \frac{m_2^k}{3} + \frac{m_3^k}{3} + \frac{m_4^k}{6}\right]\Delta t \end{aligned} \tag{18}$$

O procedimento para o RK2 usa apenas o primeiro (m_1^k) e o segundo (m_2^k) coeficiente e a variável de estado no tempo *k+1* é dada pela Eq. 19.

$$x^{k+1} = x^k + m_2^k \Delta t \tag{19}$$

Portanto, a média ponderada apresentada na Eq. 18, que representa o método RK4, fornece uma solução melhor, uma vez que ela utiliza quatro coeficientes angulares para calcular o valor do próximo passo. A Figura 2 mostra o comportamento dos métodos de Euler, RK2 e

RK4, onde os coeficientes angulares foram desenhados e a solução do passo pode ser vista e comparada com a solução analítica.

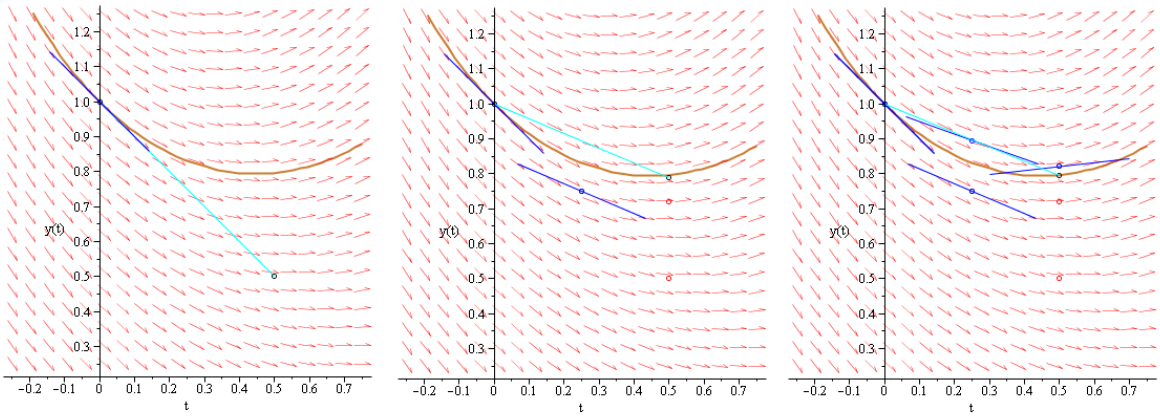

Figura 2. Comparação entre as aproximações por Euler (a), RK2 (b) e RK4 (c) (HARDER, 2012).

2.2.2.3 Método Dormand-Prince

O Método Dormand-Prince também é um método Runge-Kutta, por este motivo é frequentemente chamado RKDP. Este é um método de resolução de ODE muito eficiente e robusto, razão pela qual é utilizado como método padrão em softwares como Matlab e GNU Octave. Além disso, RKDP pode ser encontrado com código aberto para Fortran, Python, C/C++ e Java.

O Método RKDP foi apresentado por Dormand e Prince (1980) e, um ano depois, um aprimoramento com ordens superiores foi proposto pelos autores (1981). A ideia principal do método é que, assim como no RK, a cada passo de integração são utilizadas médias ponderadas; porém, este possui duas médias diferentes, uma para cada ordem. Enquanto o RK4 utiliza 4 parâmetros *k*, o RKDP utiliza 7.

Além disso, o método RKDP é adaptativo, o que determinará qual ordem este irá utilizar, 4ª ou 5ª. Essa decisão é feita de acordo com o fator de escala adaptativo *s*, calculado pela Eq. 20.

$$s = \sqrt[4]{\frac{h\varepsilon_{abs}}{2(t_f - t_0)|y_{n+1} - z_{n+1}|}} \tag{20}$$

Onde s é o fator de escala, h é o tamanho do passo, ε_{abs} representa a tolerância absoluta, t_f e t_0 são os tempos final e inicial, respectivamente, y_{n+1} é a aproximação de 5ª ordem e z_{n+1} representa a aproximação de 4ª ordem.

O fator de escala é calculado para a decisão do tamanho do passo, cf. Eq. 21, se o fator de escala for maior que 2, o passo é o dobro do passo anterior, se o fator estiver entre 1 e 2 o passo é mantido, e se o fator de escala for menor do que 1 o passo é diminuído pela metade:

$$\begin{cases} s \geq 2, h \leftarrow 2h \\ 1 \leq s \leq 2, h \leftarrow h \\ s < 1, h \leftarrow \frac{h}{2} \; e \; repetir \; iteração \end{cases} \tag{21}$$

Depois de calcular os parâmetros *k*, a média ponderada é calculada como no método RK. O procedimento para obter k_i para funções autônomas, as quais são as mais comuns em problemas de Engenharia Química, é descrito pela Eq. 22.

$$k_i = f\left(y_k + a_{i1}k_1 + a_{i2}k_2 + a_{i3}k_3 + \cdots + a_{i,i-1}k_{i-1}\right), i = 1..7 \tag{22}$$

Onde os coeficientes *a* são obtidos pela *tableau* de Butcher para o método de Dormand-Prince (Apêndice) (Ascher e Petzold, 1998; Chapra e Canale, 2009). Também, os pesos para a média ponderada variam de acordo com a ordem de aproximação e são utilizados de acordo com a Eq. 23, semelhante ao último termo da Eq. 18. Esses pesos também podem ser encontrados na *tableau* de Butcher, que consta no apêndice, assim como um algoritmo disponível para GNU Octave® pode ser visto na seção de Anexos.

$$x^{k+1} = x^k + \left[b_{i,1}k_1^k + b_{i,2}k_2^k + b_{i,3}k_3^k + b_{i,4}k_4^k + b_{i,5}k_5^k + b_{i,6}k_6^k + b_{i,7}k_7^k\right]\Delta t \tag{23}$$

2.3 Estimação de parâmetros e otimização

Como já mencionado, um passo importante a ser dado na modelagem de processos, além de boas hipóteses, é estimar um conjunto de parâmetros ótimos para serem utilizados nas equações do modelo, mantendo assim a precisão deste modelo em predizer dados experimentais ou o comportamento do processo (Schwaab e Pinto, 2007).

A estimação de parâmetros é um problema de otimização por si só. O procedimento de estimação de parâmetros visando ajustar um modelo à realidade busca a minimização de uma função objetivo (OF), a qual deve servir como finalidade métrica para o procedimento numérico. Por exemplo, uma métrica muito comum onde se busca a minimização é usar a simples soma de resíduos como OF:

$$F_{obj,residual} = \sum\left|y_{exp} - y_{mod}\right| \tag{24}$$

Onde y_{exp} representa os dados experimentais e y_{mod} representa os dados obtidos pela simulação do modelo. Nesta OF é computado apenas um erro simples, que é a distância entre

os dados experimentais e os do modelo; assim, para o procedimento de estimação de parâmetros, esta função precisa ser minimizada.

Outra OF muito comum, porém mais útil, de ser minimizada em problemas de otimização é a função dos mínimos quadrados (LS):

$$F_{obj,LS} = \sum\left(y_{exp} - y_{mod}\right)^2 \tag{25}$$

Esta última OF apresentada, a função dos mínimos quadrados, satisfaz alguns axiomas que precisam ser satisfeitos (Schwaab e Pinto, 2007). Por exemplo, dados y_{exp}, y_{mod}, e $d\left(y_{exp}, y_{mod}\right)$ como uma função da distância entre os dados experimentais e do modelo:

i. $d\left(y_{exp}, y_{mod}\right)$ precisa ser real e positiva:

$$d\left(y_{exp}, y_{mod}\right) \in \Re\,;\; d\left(y_{exp}, y_{mod}\right) \geq 0 \tag{26}$$

ii. $d\left(y_{exp}, y_{mod}\right)$ precisa ser igual a zero se, e somente se, $y_{exp} = y_{mod}$:

$$d\left(y_{exp}, y_{mod}\right) = 0 \;\Leftrightarrow y_{exp} = y_{mod} \tag{27}$$

iii. $d\left(y_{exp}, y_{mod}\right)$ é comutativa:

$$d\left(y_{exp}, y_{mod}\right) = d\left(y_{mod}, y_{exp}\right) \tag{28}$$

iv. $d\left(y_{exp}, y_{mod}\right)$ satisfaz a desigualdade triangular:

$$d\left(y_{exp}, y_{mod}\right) \leq\; d\left(y_{exp}, z\right) + d\left(z, y_{mod}\right) \tag{29}$$

De fato, os quatro axiomas são satisfeitos pela função LS. Sendo esta uma função quadrática, seu resultado sempre será positivo, satisfazendo o primeiro axioma; também, a única forma da LS ser igual a zero é quando $y_{exp} = y_{mod}$, de nenhuma outra forma isso ocorre, satisfazendo o segundo axioma; além disso, como consequência do caráter quadrático da função, a LS é comutativa, satisfazendo o terceiro axioma; e, para o último axioma, deve-se notar que $\sqrt{\left(y_{exp} - y_{mod}\right)^2} = \left|y_{exp} - y_{mod}\right|$, e a desigualdade triangular pode ser facilmente mostrada como satisfeita pelo teorema de Pitágoras.

Outros tipos de métrica podem ser utilizadas neste contexto, por exemplo:

$$d\left(y_{exp}, y_{mod}\right) = \left[\sum_{i=1}^{NE}\left(y_{exp,i} - y_{mod,i}\right)^N\right]^{1/N} \tag{30}$$

$$d(y_{exp}, y_{mod}) = \left[\sum_{i=1}^{NE} w_i \left(y_{exp,i} - y_{mod,i}\right)^N\right]^{1/N} \quad (31)$$

$$d(y_{exp}, y_{mod}) = exp\left[\sum_{i=1}^{NE} \left(y_{exp,i} - y_{mod,i}\right)^2\right] - 1 \quad (32)$$

Onde NE representa o número de experimentos realizados e N é um número par e inteiro.

Depois de estabelecida a função objetivo, seu mínimo precisa ser alcançado. Isso pode ser feito usando diferentes tipos de métodos de otimização para estimar os parâmetros do modelo, estes métodos podem ser derivativos, de busca direta ou ainda métodos heurísticos de otimização (Schwaab *et al*., 2008a).

2.3.1 Métodos de otimização

Como anteriormente mencionado, os métodos de otimização para estimação de parâmetros podem ser divididos em três grupos: derivativos, de busca direta e heurísticos.

2.3.1.1 Métodos de busca direta

Em seu livro "Optimization of Chemical Processes", Edgar e Himmelblau (1988) fazem uma boa descrição de alguns métodos de busca direta. Alguns desses métodos serão brevemente explicados, uma vez que eles não serão utilizados no trabalho.

Métodos de busca direta têm sua vantagem em serem simples de entender e aplicar. Estes são adequados para a otimização sem restrição considerando duas variáveis, no entanto, para um número maior de variáveis estes não são tão robustos, sendo os métodos indiretos, ou derivativos, os mais aconselháveis para este caso.

O método de busca das grades geralmente é utilizado com o apoio de planejamento experimental para ser aplicado na minimização de uma OF. Uma série de pontos é avaliada em torno de um ponto de referência (*k*) de acordo com o plano experimental escolhido (a Figura 3 apresenta alguns exemplos de planos experimentais). No próximo passo (*k+1*), a busca continua em direção ao ponto cujo valor foi melhor para a OF.

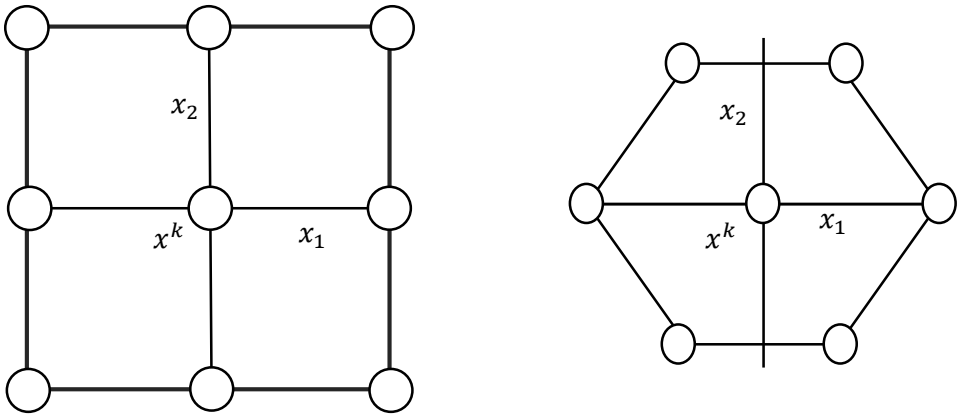

Figura 3. Esquemas para o uso do vetor *x* na busca do mínimo da função *f(x)*.

Outro método é o de Busca Univariável, descrito a seguir. Sendo a OF uma função de *n* variáveis, $OF = f(x_1, x_2, \dots, x_n)$, existiriam *n* direções de busca fixas para a procura do mínimo da OF. A função é minimizada sequencialmente, em cada direção de busca, buscando uma dimensão. Apesar da simplicidade, este método pode se tornar muito lento ao se aproximar do mínimo, sobrecarregando o procedimento de busca.

O Método Simplex Sequencial usa uma figura geométrica regular e avalia a função em cada vértice. A cada iteração, um vértice novo substitui o pior vértice da avaliação anterior, afastando-se do pior vértice. Portanto, a direção de busca muda, mas o tamanho do passo de busca é fixo.

Mais tarde, este método foi modificado por Nelder e Mead (1965) para uma versão mais sofisticada e mais eficiente.

2.3.1.2 Métodos derivativos

Métodos derivativos, também chamados de métodos indiretos, fazem uso das derivadas para determinar a direção de busca da otimização. Dessa forma, a direção de busca é adequada e a avaliação da OF a cada iteração deverá ser melhor que a anterior. Portanto, para um problema de minimização:

$$f(x^{k+1}) < f(x^k) \tag{33}$$

A direção de busca, para ser ótima, precisa satisfazer o seguinte requisito:

$$\boldsymbol{\nabla}^T f(\boldsymbol{x})\boldsymbol{s} < 0 \tag{34}$$

Este requisito é suportado pela seguinte relação:

$$\boldsymbol{\nabla}^T f(\boldsymbol{x})\boldsymbol{s}^k = |\boldsymbol{\nabla} f(\boldsymbol{x}^k)||\boldsymbol{s}^k| cos\theta \quad (35)$$

Como pode ser visto da Eq. 35, para um ângulo de 90°, a função não terá nenhuma melhoria. Pior ainda, para o intervalo de $0 \leq \theta \leq 90°$, a função irá aumentar, exatamente o oposto da exigência da minimização. Portanto, a Eq. 34 precisa ser verdadeira para garantir a minimização.

O Método do Gradiente é um método derivativo que usa apenas a primeira derivada da função objetivo. Sendo o gradiente o maior aumento, ou diminuição, da função *f(x)* no ponto *x*, para a maximização de uma OF a direção de busca é simplesmente o gradiente dessa função (isso é chamado de subida mais íngreme). Dessa forma, para a minimização a direção de busca é justamente o negativo do gradiente da OF (chamada de descida mais íngreme).

Consequentemente, o próximo passo do método para um problema de minimização vai obedecer o que é definido pela Eq. 36:

$$x^{k+1} = x^k - \lambda^k \nabla f(x^k) \quad (36)$$

Onde x^k e x^{k+1} são pontos sucessivos em *x* para a *k-ésima* e a *k-ésima+1* iterações, λ^k é um escalar que determina o comprimento do passo na direção de busca e $\nabla f(x^k)$ é o gradiente da função no *k-ésimo* ponto.

O procedimento do método do gradiente é repetido até que a função objetivo não mude mais, sob uma determinada tolerância.

Outro método derivativo muito comum é o Método de Newton. O Método de Newton usa ambas as derivadas da função, de primeira e de segunda ordem, portanto este método pode levar em consideração a curvatura de *f(x)* no passo *k* e encontrar uma direção de busca melhor. O mínimo da função objetivo em x^k pode ser obtido pela soma do gradiente da função no passo *k* mais a derivada de segunda ordem, representada pela matriz Hessiana, no passo *k*, multiplicada pela distância *x* de *k-1* até *k*, conforme Eq. 37.

$$\nabla f(x^k) + \boldsymbol{H}(x^k)\Delta x^k = 0 \quad (37)$$

Para uma função quadrática, apenas um passo é necessário para chegar ao mínimo. No entanto, para uma função não-linear mais geral, mais passos devem ser necessários. Dessa forma, pode ser útil reescrever a Eq. 37 da seguinte maneira:

$$s^k = -[\boldsymbol{H}(x^k)]^{-1} \nabla f(x^k) \quad (38)$$

Onde:

$$s^k = \frac{x^{k+1} - x^k}{\lambda^k} = \frac{\Delta x^k}{\lambda^k} \quad (39)$$

Uma vez que para o Método de Newton o escalar do comprimento do passo λ é igual a um, a Equação 38 pode ser reescrita para evitar uma operação de inversão de matriz conforme a Equação 40, e então resolver para Δx^k:

$$\boldsymbol{H}(x^k)\Delta x^k = -\nabla f(x^k) \quad (40)$$

Um dos métodos derivativos mais eficientes é o algoritmo de Levenberg-Marquardt (LMA). O algoritmo de Levenberg-Marquardt é um método utilizado para resolver problemas de mínimos quadrados não-lineares, principalmente problemas de minimização, como a minimização de uma função objetivo para a estimação de parâmetros. O LMA foi proposto como uma melhoria do Método de Newton, que fica limitado em situações em que a matriz Hessiana não é positiva definida, para garantir que a matriz Hessiana seja bem condicionada e positiva definida.

Levenberg (1944) propôs uma forma nova de tratar o problema de minimização de mínimos quadrados adicionando um procedimento com pesos. Esse novo procedimento amortece a distância das variáveis de seus pontos iniciais. Esse amortecimento permite que o usuário solucione problemas de mínimos quadrados não-lineares mais complexos que os aplicados normalmente.

Mais tarde, Marquardt (1963) propôs um algoritmo para ser usado para a estimação de parâmetros não lineares através de mínimos quadrados utilizando um método de máxima verossimilhança. Neste método, é feita uma interpolação entre o método de séries de Taylor e o método do gradiente baseada no método da máxima verossimilhança.

O LMA está disponível no pacote de diversos softwares. Esse algoritmo foi baseado no trabalho de Moré (1977), o qual propôs uma implementação com variáveis implicitamente escalonadas e a escolha de um parâmetro de Levenberg-Marquardt através de um esquema baseado no esquema de Hebden (Hebden, 1973).

2.3.1.3 Métodos de otimização heurística

Apesar dos métodos de otimização já mencionados serem eficientes e robustos, podem conter problemas numéricos. Para que se possa superar problemas tais como o número elevado

de parâmetros em um modelo, a alta correlação entre os parâmetros e a natureza multimodal da função objetivo, métodos de otimização heurística (HOM) podem ser utilizados (Schwaab *et al*., 2008a).

Além disso, HOM podem ser úteis para evitar mínimos locais, um problema recorrente quando se utiliza métodos diretos ou indiretos.

Os Algoritmos Genéticos (GAs) são HOM muito importantes. Como descrito por Goldberg (1989), GAs são algoritmos de busca baseados nos mecanismos de seleção natural e genética natural. A cada passo de busca uma nova série de cordas (como pontos no espaço de busca) é criada para ser ajustada usando informações dos ajustes das séries anteriores. Ademais, ocasionalmente novas partes dessas cordas podem ser criadas, randomicamente, para serem ajustas para assegurar boas medidas.

É possível dizer que os GAs procuram informações históricas para explorar novos pontos de busca com melhoria do desempenho esperado. Portanto, os GAs são considerados randômicos, mas não inteiramente randômicos, pois utilizam informações obtidas das avaliações anteriores da OF. Também, deve-se notar que as avaliações são da própria função objetivo, uma vez que o método não utiliza as derivadas.

Também, com um pouco menos de caráter randômico, uma técnica chamada Arrefecimento Simulado (SA) foi proposta. A técnica foi apresentada por Kirkpatrick *et al*. (1983) como um método de otimização cujo o objetivo é evitar os mínimos locais em um problema de OF multivariável, a qual pode conter muitos mínimos locais. O SA foi baseado em uma analogia de que um meio de muitos corpos pode ser levado a um estado de baixa energia pelo arrefecimento (Donnelly, 1986).

Como reportado por Sofianopoulou (1992), o SA é um método estocástico de otimização que é baseado em melhorias iterativas com deteriorações da OF controladas para evitar mínimos locais.

Como o GA, ainda que seja estocástico, o SA usa informações de passos anteriores do problema com uma distribuição de probabilidade de Gibbs para fazer o próximo movimento, portanto, este é suportado por informações estatísticas. Com essa ideia de utilizar informações anteriores, o método de Otimização por Enxame de Partículas (PSO) foi então proposto.

O Enxame de Partículas, proposto por Kennedy e Eberhart em 1995, é um interessante e importante aprimoramento do AS. Na introdução do trabalho de Kennedy e Eberhart é deixado claro que o PSO é composto por duas ideias principais: a relação entre a vida artificial e a teoria de enxames e a computação evolucionária (EVOC).

Como os autores relacionam o método com a EVOC, é plausível assumir, conforme retratado no trabalho, que o PSO é conectado de alguma forma com o GA e a programação evolucionária (EVOP).

O PSO começou do desejo de simular a coreografia de uma revoada de pássaros. Os pássaros, ou agentes, deveriam estar sob duas ideias: coincidir velocidade com o vizinho mais próximo e aleatoriedade (Kennedy e Eberhart chamaram isto de 'loucura') de movimento.

Então, uma população de agentes era inicializada randomicamente com uma velocidade. O programa escolhia um agente foco e determinava a velocidade coincidente do agente mais próximo a cada iteração, o que criava um movimento síncrono. Para evitar um movimento unânime e inalterado da população, uma variável estocástica foi introduzida a cada iteração para simular com mais similitude o comportamento real dos pássaros.

Mais tarde, uma nova variável foi introduzida. Ao invés da variável da 'loucura', uma variável de 'pousar' foi adicionada e mais adiante a metáfora foi estendida para a capacidade dos pássaros de encontrarem comida. De alguma forma os pássaros sabem como encontrar um campo de milho, então estabeleceu-se que a informação do melhor lugar deve ser repassada pelo agente que encontrou o milho para os outros. Esse conhecimento direciona o movimento dos agentes para uma determinada direção.

Quando essas considerações foram aplicadas no algoritmo, percebeu-se que seria melhor retirar a 'loucura' e a velocidade igual dos vizinhos: a revoada se tornou um enxame.

Schwaab *et al*. (2008) demonstram uma representação numérica disso. O movimento das partículas pelo espaço de busca e sua troca de informações são expressas pelas Eqs. 41-42.

$$v_{p,d}^{k+1} = wv_{p,d}^{k} + c_1 r_1\left(x_{p,d}^{ind} - x_{p,d}^{k}\right) + c_2 r_2\left(x_{d}^{glo} - x_{p,d}^{k}\right) \tag{41}$$

$$x_{p,d}^{k+1} = x_{p,d}^{k} + v_{p,d}^{k+1} \tag{42}$$

Onde p representa a partícula, d representa a direção de busca, k é o número de iterações, v representa a velocidade das partículas, x é a posição da partícula e os subscritos ind and glo são a melhor posição encontrada por uma partícula e a melhor posição encontrada pelo enxame, respectivamente. r_1 e r_2 são números randômicos com distribuição uniforme entre zero e um. c_1 e c_2 são os parâmetros de cognição e social, respectivamente, e w é um parâmetro inercial adicionado para assegurar a convergência das partículas (Shi e Eberhart, 1998).

Portanto, o algoritmo PSO vai abandonando seu caráter randômico ao decorrer das iterações, sendo direcionado, a cada passo, para um ponto mais próximo do ótimo da função objetivo.

3 ENZYMATIC HYDROLYSIS MODELS SELECTION USING PARTICLE SWARM OPTIMIZATION

Christian L. da Silveira, Renata M. Collares, Lisiane M. Terra, Marcio A. Mazutti, Nina P. G. Salau*

Chemical Engineering Department, Federal University of Santa Maria.

Abstract

This work compares several enzymatic hydrolysis models composed by a single equation, most of them already presented in literature, and a new empirical model proposed in this work, to predict the kinetic of enzymatic hydrolysis of cassava starch. Several experimental data of the hydrolysis kinetics were analyzed order to evaluate the influence of the total enzymatic mass of pectinase, α-amylase, and amyloglucosidase as well as solid to liquid ratio on the profile of fermentable sugar released. Statistical tests were executed to compare models to the experimental data and to determine the best models for the enzymatic hydrolysis of cassava starch. Parameter estimation was performed to obtain the model parameters through the Particle Swarm Optimization with a Nonlinear Least Squares Objective Function minimization. Comparing all the models residuals, the difference between the predicted and the experimental data, the proposed empirical model was found to be the one with lowest residual. Also, applying the Corrected Akaike Informantion Criterion, which takes into account the number of parameters to be estimated, the same model was found to be the most suitable for fitting the experimental data.
Keywords: bioethanol, cassava starch, hydrolysis process modeling, particle swarm optimization.

* To whom all correspondence should be addressed. E-mail: ninasalau@ufsm.br
Address: Chemical Engineering Department, UFSM – Av. Roraima, 1000, Cidade Universitária - Bairro Camobi. 97105-900 Santa Maria, RS – Brazil.
Phone: +55-55-3220-8448- Fax: +55-55-3220-8030.

1. Introduction

The increase on energetic demand and atmospheric contamination by combustion gases have opened searching for new, safe, effective, and more accessible energy sources. For this aim, biofuels production from different agricultural sources, such as agro industrial by-products, vegetable materials as woods, among others, has recently received great attention [1]. In such scenario, biofuels have been an attractive alternative to current petroleum based fuels, as they can be used as transportation fuels with little change to current technologies, and have significant potential to improve sustainability and reduce green house gases emissions [2]. In developed countries there is a growing trend towards employing modern technologies and efficient bioenergy conversion using a range of biofuels, which are becoming economically competitive with fossil fuels [3]. Among the biofuels, the bioethanol has been reported as a biofuel that could reduce both CO_2 emissions and dependence on foreign oil [4]. Being Brazil´s experience with sugarcane-based biomass ethanol frequently cited as a success story, biofuels can be produced from different biomass sources. For example, an alternative to sugar-based feedstock is producing ethanol from starch of sources such as corn, barley, wheat, rice, cassava, and potato [5].

In a previous work [6], cassava (*Manihot esculenta crantz*) starch was used to obtain fermentable sugars to a subsequent ethanol production, the experimental data used in this work was obtained by Collares *et al.* [6]. Also called yuca, manioc, and mandioca, cassava is a perennial woody shrub, which grows in tropical and subtropical areas of the world [7]. It is the world's fourth most important crop, and is grown in many countries in Africa, Asia, and Latin America. Its root crop contains starch and is one of the most important sources of calories in the tropics. It is also widely employed as a raw material for many industrial applications in the animal feed industry and starch industry, and more recently for production of fuel ethanol [8]. Its pulp contains about 50%–70% starch on a dry weight basis and 20%–30% fibers, which are composed mainly of cellulose and other non-starch polysaccharides [9].

The exploration of novel efficient bioprocesses for underused biomasses is, thus, at the forefront of biotechnological research and industrial application [9,10]. In recent years, several studies were accomplished aiming to improve the yield of the enzymatic hydrolysis of cassava for ethanol production [11]. For example, Lin et al. [8] reports that at a concentration of 150 g/L of cassava flour the reaction have reached a yield of 53.3 g/L of ethanol, and 92% of efficiency was achieved after 48 hours of simultaneous saccharification and fermentation. Also, Nitayavardhana et al. [12] reported that "the reducing sugar release of sonicated samples improved as much as 181% with respect to control group" of his research. Leaes et al. [13] studied the effects of ultrasound on

hydrolysis of cassava to obtain fermentable sugar; in this work, Leaes et al. reported a yield increase of 30% of fermentable sugar production in comparison to the experiments without ultrasound, concluding that this procedure is useful for the hydrolysis process.

Taking into account that there is a considerable amount of studies focusing on experimental hydrolysis of cassava starch, the main objective of this work is to determine the best kinetic model to predict the enzymatic hydrolysis of cassava starch using pectinase, α-amylase, and amyloglucosidase with already established kinetic models. The effects of amyloglucosidase, pectinase, reaction time, and solid to liquid ratio were evaluated in a previous work [6] to obtain a consistent experimental data set. Afterwards, Michaelis-Menten [14], Tait et al. [15], Polakovič and Bryjak [16], Humphrey, Wald et al., Fan and Lee, Huang, Holtzapple et al. models [17], a Gompertz based model, a Sum of Hyperbolic Sins model, a Populational Logistic model, an Exponential Logistic model [18], and a purely empirical model, with derivative explicitly as a function of time, were evaluated and compared to verify which best describes the enzymatic hydrolysis of cassava starch.

2. Hydrolysis models

All the models presented in this work are models that comprises the whole enzymatic hydrolysis process in a single model equation; this approach was used in order to simplify the process simulations. In this manner, the enzymes used in the hydrolysis were not differentiated from each other.

The tested models are functions of time and total enzymatic mass, and usually accounting the substrate consumption and product obtained. As the experiments were performed in a batch bioreactor, all the kinetics equations will be represented as a rate of reaction, cf. Eq. (1).

$$\frac{d(TRS)}{dt} = r \tag{1}$$

Where r denotes the kinetic rate $[h^{-1}]$; TRS is the total released sugar (dimensionless); and t is the time of integration $[h]$.

In this work were used models already presented for enzymatic hydrolysis modeling, models for microorganism growth, mechanistic models, and purely empirical models, to predict the starch consumption and total released sugar (TRS). All the models were submitted to a parameter estimation procedure, using a Particle Swarm Optimization (PSO) algorithm, for further analysis and comparison. The models used can be seen in Table 1.

Table 1. Models used in this work.

Model (Nº)	Reference	Model equation	Number of Parameters
Michaelis-Menten (1)	[14]	$r = k_1\left(\frac{S}{S+k_2}\right)E$	2
Tait et al. (2)	[15]	$r = \frac{k_1 S}{k_2(1+G/k_3)+\frac{S^2}{k_4}+S}E$	4
Polakovič based (a) (3)	[16]	$r = \frac{k_1 SE}{k_2\left(1+\frac{G}{k_2}\right)+S}$	3
Polakovič based (b) (4)	[16]	$r = \frac{k_1 SE}{k_2\left(1+\frac{G}{k_3}\right)+\frac{S^2}{k_4}+S}$	4
Humphrey et al. based (5)	[17]	$r = k_1 S^{k_2}\left(\frac{E}{E+k_3}\right)$	3
Wald et al. (6)	[17]	$r = \frac{k_1 SE}{k_2+E}$	2
Fan and Lee (7)	[17]	$r = k_1 + \frac{k_2 SE}{k_3+S}$	3
Huang (8)	[17]	$r = \frac{k_1 SE}{k_2+S+k_3 E}$	3
Holtzapple et al. (9)	[17]	$r = \frac{k_1 SE}{k_2+\left(\frac{S-E-k_2+\sqrt{(k_2+E-S)^2+4k_2 S}}{2S}\right)S+k_3 E}$	3
Gompertz based (10)	[18]	$r = k_1 exp\left(-exp(k_2-k_3 S)\right)E$	3
Sum of hyperbolic sins (11)	[18]	$r = \left(\frac{k_1}{sinh(G^{k_2})}+\frac{k_3}{sinh(S^{k_4})}\right)E$	4
Populational Logistic Model (12)	[18]	$r = k_1 SG\left(1-\frac{G}{G_{max}}\right)E$	1
Exponential Logistic Model (13)	[18]	$r = \frac{k_1}{1+exp(k_2-k_3 S)}E$	3
Proposed model (14)		$r = \frac{-k_1 exp\left(-\frac{(-k_2-t)^2}{k_3^2}\right)(-2k_2-2t)}{k_3^2}E^{k_4}$	4

Where S represents the substrate, G represents the TRS, and E the enzymatic mass added to the hydrolysis process. The parameters were all referred as k_i, where i is the parameter number to characterize the total number of parameters to be estimated for each model.

All the state variables were adimensionalized and, consequently, normalized, to avoid stiffness of the equations, to maintain the same data magnitude for both starch consumption and total released sugar, and to avoid high parameters correlation, cf. Equation 2. Nevertheless, the usual parameters nomenclature was maintained.

$$var_{dimensionless} = \frac{var - var_{min}}{var_{max} - var_{min}} \quad (2)$$

Where var represents the state variables Starch, TRS, and total mass of enzymes used.

Many models proposed in the literature were tested; these models are presented in Table 1.

3. Parameter estimation problem

It consists in obtaining the kinetic parameters values from the experimental data set to validate a kinetic rate equation to be used in Eq. (1). This may be done by the minimization of an objective function (Eq. 3). The rate equations used in this work have a number of parameters to be estimated between 1 and 4, as showed in Table 1. The parameter estimation problem is solved here with Particle Swarm Optimization (PSO) [19,20], briefly described in the item 3.1.

The objective function should lead to the closest prediction of the experimental data. Therefore, the model proposed to be the objective function ought yield the minimization of the errors between the model predictions and the experimental data, *i.e.* zero, as can be seen in the Eq. (3).

$$\min_x f(x) = \left\| TRS_{\exp} - TRS_{sim} \right\|_2^2 = \left\| F(x) \right\|_2^2 = \sum_i F_i^2(x) \quad (3)$$

Also, the parameter estimation procedure, and the objective function minimization, accounted the specifics conditions of each experiment, and the parameter results are a global estimation, which means that one set of parameters was obtained to predict the whole behavior of the process, independently of the initial conditions.

3.1. Particle Swarm Optimization

The Particle Swarm Optimization (PSO), originally proposed by Kennedy (a social psychologist) and Eberhart (an electrical engineer) in 1995, is based on animals social behavior, more precisely with some analogies of birds

searching for corn [19]. It basically consists in the random positioning of a chosen number of particles in the space of the problem that shall be seek, and each one of these particles evaluates the objective function at its current location. Each particle determines its next movement through the search space, combining its own history, or its best fitness, with those of the other particles, with some random perturbations. The next iteration takes place only after all particles have been moved [20].

At the initial iterations, the random character is high, since the algorithm still do not know where must be the minimum, and the particles conduct to a global search over the seek region. As the iterations evolve, the particles concentrate around the more promising regions, consequently leaving its randomness behind [19].

3.2. Computational methods

A laptop (Intel® Core™ i7 2.20GHz processor and 8.00Gb of RAM memory) was used to perform the computations presented here. The PSO algorithm and the analyzed kinetic rate equations were implemented in Intel Fortran® Compiler 11. The PSO version was based in the Kennedy and Eberhart work [21], with 50 particles as default number, maximum number of iterations 50,000, and at least 3 successful minimizations for each model.

The Dormand-Prince method implemented in Matlab (*ode45*) was used to perform simulations after the parameters estimation.

4. Material and methods

Detailed information concerning to enzymes mass added, composition of raw material, and analytical methods can be obtained in the work of Collares et al. [6]. Briefly, the experimental procedure consisted in the use of jacketed vessel bioreactor produced by Metalquim LTDA, with 5 L volume and agitation, temperature and pressure controllers were used. The working volume used was 2 L. The raw material (cassava) was added to the bioreactor, considering a determined solid to liquid ratio ($R_{S/L}$). The pH was adjusted to 5.5 using 0.1M HCl, and the temperature to 45 ºC until reach the equilibrium. Then, pectinase was added, considering the levels defined in the experimental design, and the reactions were carried out for 60 minutes under agitation of 200 r/min. Afterwards, temperature was increased to 95 ºC until stabilization, then 0.8 µL/g starch of α-amylase was added and the reaction proceeded for more 60 minutes under the same agitation. Following, temperature was decreased to 60 ºC and pH of the solution adjusted to 4.0 using 0.1M HCl. Next, amyloglucosidase was added, according to the levels defined in the experimental design, and the reaction was carried out under the same agitation, until the specified time. At

the end, the samples were withdrawn from the bioreactor and the solids removed by filtration. The amount of total released sugar (TRS) released by the hydrolysis was measured in the supernatant, and the data collected were analyzed to estimate the parameters of the proposed kinetic reaction. The total sugar released was measured by 3,5-dinitrosalicylic acid method [22]. Table 2 presents the range of investigated variables in the experimental design to evaluate the hydrolysis of starch from cassava.

Table 2. Experimental independent variables.

Exp.	Reaction Time [h]	Total enzymatic mass [$\mu L.g^{-1}$]	$R_{S/L}$
1	21	3.99	0.70
2	21	8.37	0.70
3	55	3.67	0.70
4	55	3.99	0.70
5	55	8.69	0.70
6	21	3.67	0.38
7	21	8.37	0.38
8	21	8.69	0.38
9	55	3.67	0.38
10	55	8.69	0.38
11	55	8.37	0.38
12	55	8.69	0.38
13	38	5.86	0.50
14	38	10.88	0.50
15	38	6.18	0.33
16	38	6.18	0.50
17	38	6.18	0.50
18	38	6.18	0.50

5. Results and discussion

In this work the experimental data obtained by Collares et al. [6] are used to estimate the parameters of several models, presented in Table 1.

The parameters estimated for each model and their intervals with 95% of confidence can be seen in Table 3.

Table 3. Estimated parameters and their interval of confidence (95%).

Model n°	k_1	k_2	k_3	k_4
Michaelis-Menten (1)	0.1781E-1±0.0115	6.4360E-4±0.0076		
Tait et al. (2)	91.7664±0.8376	0.2301±0.0055	0.8996E-3±0.0006	0.2619E-2±0.0010
Polakovič based (a) (3)	0.5497±0.0221	0.1725E-5±0.0003	0.1038E-5±6E-7	
Polakovič based (b) (4)	58.8198±6.7626	0.2086E-1±0.0017	0.1269E-3±9E-5	0.4083E-2±0.0014
Humphrey et al. based (5)	0.2605±0.0781	8.5553E14±8.55E2	1.0636±0.2127	
Wald et al. (6)	0.2420±0.0401	0.5396E-12±0.0523		
Fan and Lee (7)	18.3994±4E-16	1.2055E5±1E-12	0.1031E-1±4E-17	
Huang (8)	7.113E9±5E12	0.6454±5E3	2.9348E10±2E13	
Holtzapple et al. (9)	48.9407±4.2957	0.3031E-3±0.0023	2.0314E2±17.2886	
Gompertz based (10)	7.7046E4±7E-15	0.2557E-2±1E-16	45.6441±4E-14	
Sum of hyperbolic sins (11)	0.5238E-4±5E-5	8.0369±0.0001	0.2974E-9±2E-10	1.5703±0.0015
Populational Logistic Model (12)	7.7233±0.0033			
Exponential Logistic Model (13)	0.1808±4E-17	0.1866±1E-16	43.6210±1E-15	
Proposed model (14)	2.0436E2±14.3056	61.1184±0.6218	26.3540±3.1625	0.0070±0.1470

Although the low residual values obtained for the models (Table 4), some of them resulted in parameters with confidence intervals bigger than their absolute values, as can be seen in Table 3. This big interval of confidence causes the parameters to cross the zero-line and, consequently, to have no meaning for the model at some point. Regardless, these intervals of confidence for the estimated parameters, the parameters were maintained in the models for further analysis. The confidence intervals were computed using the Student's inverse for 95% of confidence.

The models whose parameters crossed the zero line were the Michaelis-Menten (k_2), Polakovič based (a) (k_2), Wald et al. (k_2), Huang (k_2,k_3), Holtzapple et al. (k_2), and the empirical model (k_4).

Later, with the estimated parameters for each model, simulations were performed and the experimental data was compared to the model obtained data. Although the Least Squares Objective Function was used as a metric for the parameters estimation, for the comparison between the models it was preferred to measure the sum of residuals of the model, i.e., the linear difference between the experimental and the model data, cf. Eq. 4, because it would be used to the Akaike Criteria Information evaluation.

$$res = \sum_{j=1}^{18} \sum_{i=1}^{5} \left| TRS_{exp_{j,i}} - TRS_{mod_{j,i}} \right| \qquad (4)$$

Where the subscript j represents the number of experiments performed, and the subscript i represents the number of experimental data collected in each experiment.

The sums of residuals values for each model are presented in Table 4. It can be seen that almost all models are at the same magnitude of residuals, except the sum of hyperbolic sins models, which has the largest sum of residuals value, and the empirical model, which has the lowest residual.

Table 4. Sum of residuals and Corrected Akaike Information Criterion values for each model.

Model (Nº)	Number of parameters	Sum of residuals	AICc
Michaelis-Menten (1)	2	3.619077236	-121.4700882
Tait et al. (2)	4	3.473447961	-118.7427659
Polakovič based (a) (3)	3	3.516263511	-120.4554288
Polakovič based (b) (4)	4	3.473321908	-118.7441844
Humphrey et al. based (5)	3	3.541380297	-120.1772251
Wald et al. (6)	2	3.510522864	-122.6604323
Fan and Lee (7)	3	3.901964397	-116.3872589
Huang (8)	3	3.510557144	-120.5189119
Holtzapple et al. (9)	3	3.510455864	-120.5200395
Gompertz based (10)	3	3.557702738	-119.9974869
Sum of hyperbolic sins (11)	4	7.93112518	-86.4711052
Populational Logistic Model (12)	1	3.724985045	-122.4351633

Exponential Logistic Model (13)	3	2.793904324	-129.4437184
Proposed model (14)	4	1.163262	-161.5002561

As almost all the models presented the sum of residuals values very close to each other, and as some of the models have different number of parameters and different mathematical structures, some supported by mechanistic background, while others by an empirical approach, it is important to qualify which models is better.

The Akaike Information Criterion (AIC) was developed by Akaike [23] and latter a correction for this method (AICc) was proposed by Hurvich and Tsai [24] claiming to be superior to the traditional AIC.

The AICc was used in this work as a measure for chosen the best model to be used, as it accounts the errors (residuals) and the number of parameters of the model [25,26]. As the smaller the value obtained for the AICc, the better the model is [27,28].

As it can be seen from Table 4, the model with the lowest AICc value, the most negative value, was the Proposed model (14). Therefore, the model (14) was the model that best fitted to the experimental data for the hydrolysis of cassava starch.

In Figures 1-6, it is presented the comparison of the experimental data and the model (14) obtained values.

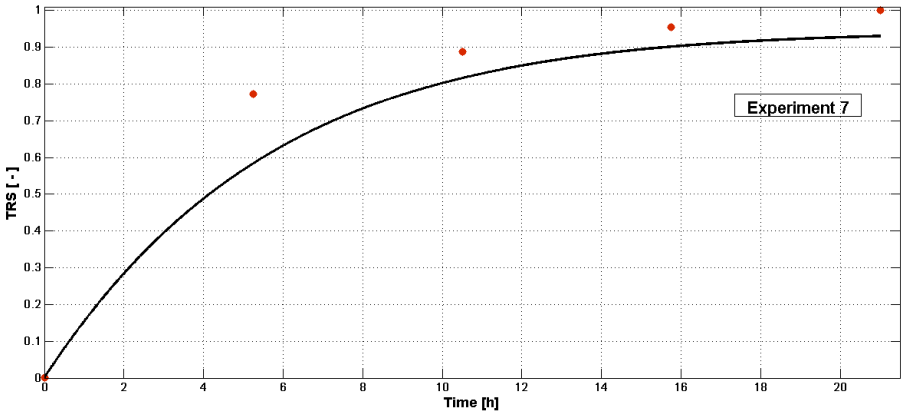

Figure 1 – TRS experimental data (●) vs. model data (-) for the Experiment 7.

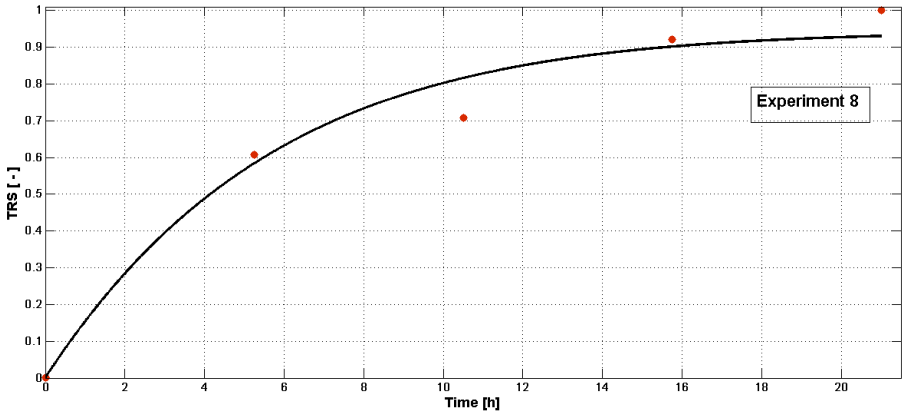

Figure 2 – TRS experimental data (●) vs. model data (-) for the Experiment 8.

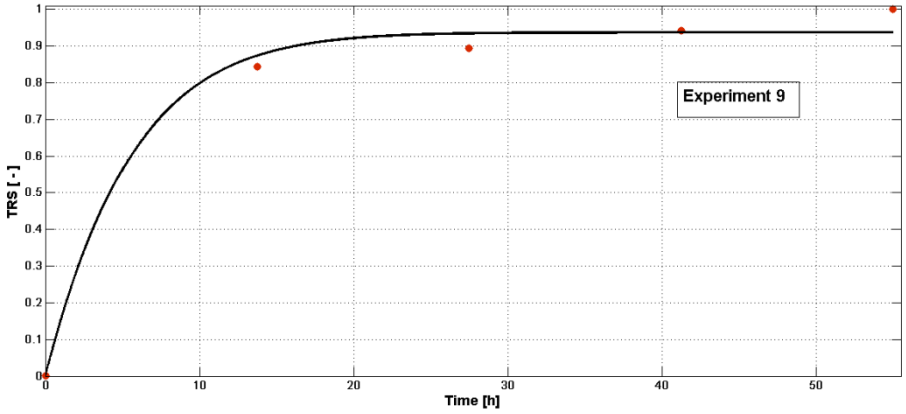

Figure 3 – TRS experimental data (●) vs. model data (-) for the Experiment 9.

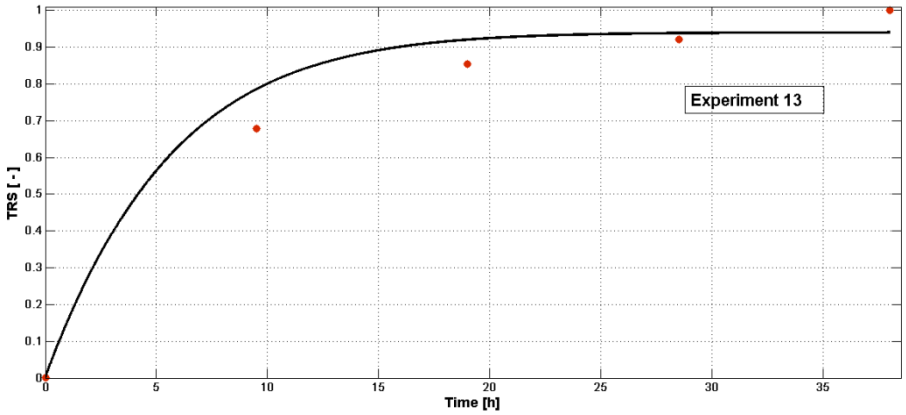

Figure 4 – TRS experimental data (●) vs. model data (-) for the Experiment 13.

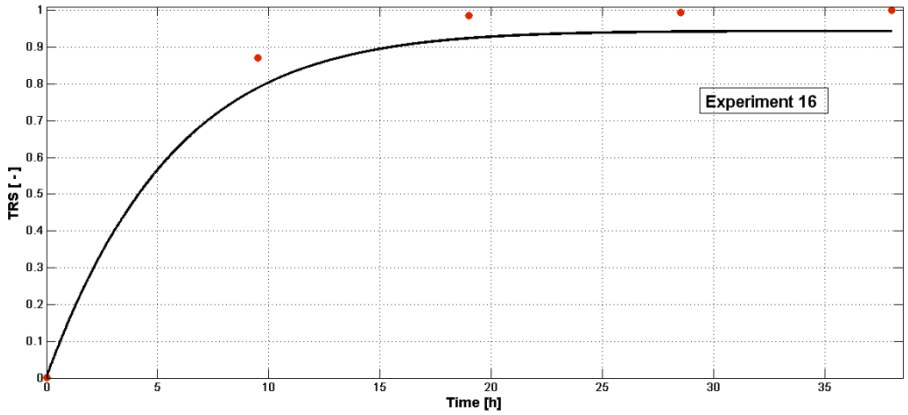

Figure 5 – TRS experimental data (●) vs. model data (-) for the Experiment 16.

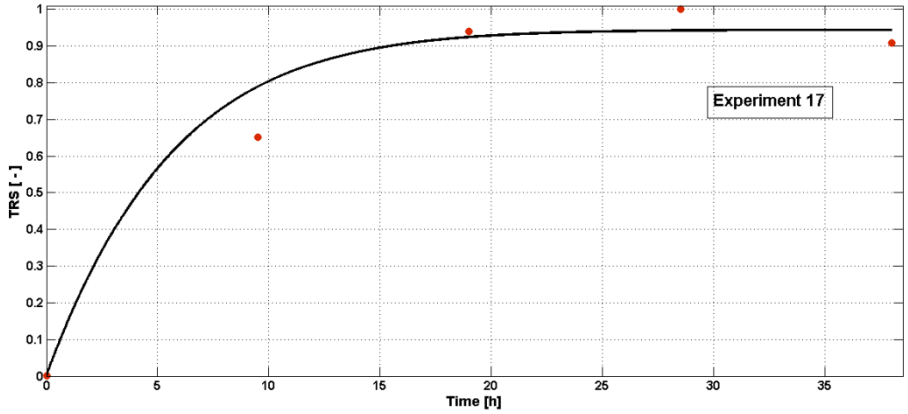

Figure 6 – TRS experimental data (●) vs. model data (-) for the Experiment 17.

The model has a behavior very similar to the hydrolysis kinetics obtained experimentally, as can be seen in Figures 1-6, showing to be reliable as a model to predict the process kinetics and the amount of TRS.

6. Conclusion

This work compared several enzymatic models for the hydrolysis of the cassava starch using α-amylase, pectinase, and amyloglucosidase as enzymes. Although, some models used to be compared were purely empiric, while others were theoretical supported models, the proposed model, which derivative was an explicit function of time, presented the lowest sum of residuals of all the fifteen tested models. Consequently, having the minor error between the experimental and the simulated data.

The model, however, had four parameters to be estimated, while other had one, two or three. In order to account the number of parameters of each model as a criteria of best model selection, it was used the Corrected Akaike Information Criterion. In this test, the proposed model (14) still had the best value, therefore it was chosen as the best among all the tested models.

Further, simulations were presented alongside the experimentally obtained data, to verify if the behavior of the data predicted by the models was close to the experimental ones. In this last analysis, the model behavior showed to be very similar to the experimental behavior of the enzymatic hydrolysis of cassava starch. In this manner, the models can be used as an empirical model to predict the kinetics of the referred process.

A further study can be performed using parameters identifiability and the sensitivity matrix of the parameters, to measure their importance in each one of the tested models.

Acknowledgment

The authors thank CNPq/Brazil, CAPES and FAPERGS for financial support and scholarships.

References

[1] Ruiz, M. I., Sanchez, C. I., Torres, R. G., Molina, D. R. (2011). Enzymatic hydrolysis of cassava starch for production of bioethanol with a Colombian wild yeast strain. *Journal of the Brazilian Chemistry Society*, 22 (12), 2337-2343.

[2] Carere, C. R., Sparling, R., Cicek, N., Levin, D.B. (2008). Third generation biofuels via direct cellulose fermentation. *International Journal of Molecular Science*, 9, 1342–1360.

[3] Puhan, S., Vedaraman, N., Rambrahaman, B.V., Nagarajan, G. (2005). Mahua (Madhuca indica) seed oil: a source of renewable energy in India. *Journal of Scientific and Industrial Research*, 64, 890–896.

[4] Roy, P., Orikasa, T., Tokuyasu, K., Nakamura, N., Shiina, T. (2012). Evaluation of the life cycle of bioethanol produced from rice straws. *Bioresource Technology*, 110, 239-244.

[5] Li, H., Wang, L., Shen, L. (2010). Potential CO_2 emission reduction by development of non-grain-based bioethanol in China. *Environmental Management*, 46, 555-564.

[6] Collares, Renata M., Miklasevicius, Luiza V. S., Bassaco, Mariana M., Salau, Nina P. G., Mazutti, Marcio A., Bisognin, Dilson A., Terra, Lisiane M. (2011). Optimization of enzymatic hydrolysis of cassava to obtain fermentable sugars. *Journal of Zhejiang University*, 13(7), 579–586.

[7] Dai, D., Hu, Z., Pu, G., Li, H., Wang, C. (2006). Energy efficiency and potentials of cassava fuel ethanol in Guangxi region of China. *Energy Conversion & Management*, 47, 1686-1699.

[8] Lin, H.J., Xian, L., Zhang, Q.J. , Luo, X.M., Xu, Q.S., Yang, Q., Duan, C.J., Liu, J.L., Tang, J.L., Feng, J.X. (2011). Production of raw cassava starch-degrading enzyme by Penicillium and its use in conversion of raw cassava flour to ethanol. *Journal of Industrial Microbiology & Biotechnology*, 38(6), 733-742.

[9] Rattanachomsri, U., Tanapongpipat, S., Eurwilaichitr, L., Champreda, V. (2009). Simultaneous non-thermal saccharification of cassava pulp by multi-enzyme activity and ethanol fermentation by Candida tropicalis. *Journal of Bioscience and Bioengineering*, 107(5), 488-493.

[10] Pan, S., Wu, S., Kim, J. (2011). Preparation of glucosamine by hydrolysis of chitosan with commercial α-amilase and glucoamylase. *Journal of Zhejiang University*, 12(11), 931-934.

[11] Chen, Y., Huang, S., Tang, Z., Chen, X., Zhang, Z. (2011). Structural changes of cassava starch granules hydrolyzed by a mixture of α-amylase and glucoamylase. *Carbohydrate Polymers*, 85 (1), 272-275.

[12] Nitayavardhana, S., Rakshit, S.K., Grewell, D., van Leeuwen, J., Khanal, S.K. (2008). Ultrasound pretreatment of cassava chip slurry to enhance sugar release for subsequent ethanol production. *Biotechnology and Bioengineering*, 101(3), 487-496.

[13] Leaes, E. X., Zimmermann, E., Souza, M., Ramon, A. P., Mezadri, E. T., Dal Prá, V., Terra, L. M., Mazutti, M. A. (2013). Ultrasound-assisted enzymatic hydrolysis of cassava waste to obtain fermentable sugars. *Biosystems Engineering*, 115 (1), 1-6.

[14] Mazutti, M. A., Corazza, M. L., Maugeri, F., Rodrigues, M. I., Corazza, F. C., Treichel, H. (2009). Inulinase production in a batch bioreactor using agroindustrial residues as the substrate: experimental data and modeling. *Bioprocess and Biosystems Engineering*, 32, 85-95.

[15] Tait, M. C. de A., Módenes, A. N., Kroumov, A. D. (2005). Application of system analysis to the process of simultaneous saccharification and fermentation of starch to ethanol utilizing genetically modified microorganisms. *Science & Engineering Journal*, 15(2), 7-12.

[16] Polakovič, M.; Bryjak, J (2004). Modelling of potato starch saccharification by an *Aspergillus niger* glucoamylase. *Biochemical Engineering Journal*, 18, 57-63.

[17] Brown, R. F.; Agbogbo, F. K.; Holtzapple, M. T. (2010). Comparison of mechanistic models in the initial rate enzymatic hydrolysis of AFEX-treated wheat straw. *Biotechnology for Biofuels*, 3-6.

[18] Zwietering, M. H.; Jongenburger, I.; Rombouts, F. M.; Riet, K van't (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, 56(6), 1875.

[19] Schwaab, M., Biscaia, E. C., Monteiro, J. L., Pinto, J. C. (2007). Nonlinear parameter estimation through particle swarm optimization. *Chemical Engineering Science*, 63, 1542 – 1552.

[20] Poli, R., Kennedy, J., Blackwell, T. (2007). Particle swarm optimization. *Swarm intelligence*, 1(1), 33-57.

[21] Kennedy, J., Eberhart, R. (1995) Particle Swarm Optimization. Proceedings of the IEEE International Conference on Neural Networks, 4, 1942-1948.

[22] Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry*, 31(3), 426-428.

[23] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.

[24] Hurvich, C. M.; Tsai, C. L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *Journal of Time Series Anal.*, 14, 271-279.

[25] Cavanaugh, J. E. (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*, 33, 201-208.

[26] Bezerra, R. M. F.; Dias, A. A.; Fraga, I.; Pereira, A. N. (2011). Cellulose hydrolysis by Cellobiohydrolase Cel7A shows mixed hyperbolic product inhibition. *Applied Biochemistry and Biotechnology*, 165(1), 178-189.

[27] Li, M. Y.; Sun, X. M.; Zhao, G. M., Huang, X. Q.; Zhang, J. W.; Tian, W.; Zhang, Q. H. (2013). Comparison of mathematical models of lactic acid bacteria growth in vacuum-packaged raw beef stored at different temperatures. *Journal of Food Science*, 78(4).

[28] López, S.; Prieto, M.; Dijkstra, J.; Dhanoa, M. S.; France, J. (2004). Statistical evaluation of mathematical models for microbial growth. *International Journal of Food Microbiology*, 96, 289-300.

4 MODELING THE MICROBIAL GROWTH AND TEMPERATURE PROFILE IN A FIXED-BED BIOREACTOR

Published in: Bioprocess and Biosystems Engineering, 37 (10): 1945-54, October, 2014.
DOI: 10.1007/s00449-014-1170-0

Christian L. da Silveira, Marcio A. Mazutti, Nina P. G. Salau*

Chemical Engineering Department, Universidade Federal de Santa Maria

Abstract

Aiming to scale up and apply control and optimization strategies, currently is required the development of accurate plant models to forecast the process nonlinear dynamics. In this work a mathematical model to predict the growth of the *Kluyveromyces marxianus* and temperature profile in a fixed-bed bioreactor for solid state fermentation using sugarcane bagasse as substrate was built up. A parameter estimation technique was performed to fit the mathematical model to the experimental data. The estimated parameters and the model fitness were evaluated with statistical analyses. The results have shown the estimated parameters significance, with 95% confidence intervals, and the good quality of process model to reproduce the experimental data.

Keywords: *Kluyveromyces marxianus*, solid-state fermentation, fixed-bed bioreactor, mathematical modeling, parameter estimation.

1. Introduction

Bifidobacterias are responsible for restoring the intestinal flora, after one submits himself to an antibiotic treatment, they also stimulate the immune system and produce vitamins, mainly of the B group. Hence, it is reasonable to assert that Bifidobacterias are benign microorganisms, benefiting the host [1].

* To whom all correspondence should be addressed. E-mail: ninasalau@ufsm.br

Address: Chemical Engineering Department, UFSM – Av. Roraima, 1000, Cidade Universitária - Bairro Camobi. 97105-900 Santa Maria, RS – Brazil.

Phone: +55-55-3220-8448- Fax: +55-55-3220-8030.

Boehm et al. [2] have discussed that a significant increase in the number of faecal bifidobacterias was seen in a group of preterm infants fed with bovine milk supplemented with oligosaccharides. Knol et al. [3] have also reported the number of bifidobacterias in a group fed with a supplemented formula which contained, in 1.0 g/dl, 90% of galacto-oligosaccharides and 10% of fructo-oligosaccharides, was significantly higher (34.47±28.08% supplemented vs. 5.16±9.37% non-supplemented).

The importance of inulinase production reflects on the bifidobacteria health matters, due its role on the production of oligosaccharides [4]. The authors have considered a maximum yield in the conversion of total oligosaccharides of 63%. On the other hand, Singh et al. [5] have demonstrated that *Kluyveromyces* yeast inulinase production reaches high yields.

In this context, solid state fermentation (SSF) is a possibility for the production of inulinase using *Kluyveromyces marxianus*. The method is usually defined as fermentation without free water, or almost none free water, where the substrate has just enough moisture to support the growth and the metabolism of the microorganisms. Likewise, SSF is referred as a process with low energy requirements, low wastewater generation, environmental-friendly, once it resolves some of the solid waste disposal, and lower operating costs, investment layouts and less downstream processing if compared to the typical submerged fermentation [6].

The SSF scale-up main problem is the removal of the heat generated by the metabolic activity of the cell growth [7], which is the largest contributor to the raise of temperature in the bioreactor. Empirical models are not ideal for process scale-up, hence, it is necessary a phenomenological model which accurately describes the process dynamics. However, once each yeast has its own specific growth velocity, depending not just of its genre but also of the moisture content of the medium [8], and each medium has its own density, specific heats, void fraction and others, it is also necessary to estimate such parameters to better fit the process model to the experimental data.

It shall be highlighted that control strategies for a complex bioprocess such as SFF require a detailed and representative mathematical model for capturing the nonlinear process behavior. Lekanda and Pérez-Correa [9] have shown that the control strategies are based on experimental measurements that may have noises due bed heterogeneity, so it is important to know the parameters along the reactor to build-up an accurate process model.

In the present work, the data referring to growth of *Kluyveromyces marxianus* NRRL Y-7571 and temperature profile in a fixed-bed bioreactor for solid-state fermentation presented by Mazutti et al. [13] were used for parameter estimation and model simulation.

2. Solid state fermentation modeling

The chosen model to describe the cell growth was the well-known Verhulst Logistic model, which takes into account a maximum growth velocity and a maximum of cells. According to Fanaei and Vaziri [10], in spite of its simplicity and accuracy, the logistic model cannot forecast a death phase representation. In their work, an additional parameter ϕ is used to represent the physiological state of the cells, increasing the number of parameters to be estimated. Here, the addition of the physiological factor to model equations has not improved the model accuracy, thus, the traditional logistic model was used, as given by Equation 1.

$$\frac{dX}{dt} = \mu X\left(1 - \frac{X}{X_m}\right) \tag{1}$$

The temperature model is closely related to the cells growth due to the metabolic heat generated by the growth of such cells. The chosen model to obtain the temperature profile was the modified distributed model, found in the work of Fanaei and Vaziri [10] and adapted here cf. Equation (2).

$$\rho_b C_{pb}\left(\frac{\partial T}{\partial t}\right) = \rho_s(1-\varepsilon)Y_Q\frac{dX}{dt} - \rho_a C_{pa}V_z\left(\frac{\partial T}{\partial z}\right) - \rho_a f\lambda V_z\left(\frac{\partial T}{\partial z}\right) \tag{2}$$

Besides, two algebraic equations (Eqs. 3, 4) are necessary to support the differential equations system.

$$\rho_b = \varepsilon\rho_a + (1-\varepsilon)\,\rho_s \tag{3}$$

$$C_{pb} = \frac{\varepsilon\rho_a\left(C_{pa}+f\lambda\right)+(1-\varepsilon)\rho_s C_{ps}}{\rho_b} \tag{4}$$

3. Model integration and parameter estimation

Parameter estimation can be seen as a key issue for bioreactors. This tool is the crucial step to obtaining predictions from a mathematical model. In general, only a fraction of the parameters can be experimentally measured, while the rest are often fitted by parameter estimation.

Here the parameters were estimated to fit the system model to the experimental data obtained by Mazutti et al. [13]. The estimated parameters were: μ (Eq. 1), ρ_s, C_{ps}, ε and Y_Q (Eqs. 2-4). A total number of 5 parameters (NP=5) were estimated, for 6 experiments (NE=6) measured at 25 different times (NY=25), and according to Equation (5), 120 degrees of freedom were available for the statistical tests.

$$DF = NE.NY - NP \tag{5}$$

where DF is the degrees of freedom of the system, NE the number of experiments, NY the number of measures for each experiment and NP the number of parameters.

The model fitting is based on experimental data, which are used to assign parameter values that minimize the objective function in the parameter estimation problem. The objective function of the non-linear least square method to be minimized is given by Equation 6. It is defined as the sum of the differences between experimental (subscript *exp*) and model data (subscript *mod*).

$$\begin{bmatrix} min\ f(X) \\ min\ f(T) \end{bmatrix} = \begin{bmatrix} \left\| X_{exp} - X_{mod} \right\|^2 \\ \left\| T_{exp} - T_{mod} \right\|^2 \end{bmatrix} = \sum F(x) \tag{6}$$

Firstly, the parameters were estimated using the six experimental dataset. Afterwards, the resulting parameters values were used as initial guess for a new parameter estimation step using separately each one of the six experimental dataset. In the next step, the mean and confidence intervals for each parameter were calculated. Schwaab and Pinto [16] suggests that usually the confidence interval is calculated using the Student's *t*-distribution, for the particular degrees of freedom, the standard deviation and the mean of the parameters estimated, cf. Equation 7.

$$\hat{\alpha} - t\sigma_\alpha < \alpha < \hat{\alpha} + t\sigma_\alpha \tag{7}$$

Where α is the real parameter, $\hat{\alpha}$ is the estimated parameter, σ_α is the standard deviation of the estimated parameter and the t is Student's t-distribution parameter based on the degrees of freedom of the problem.

The Dormand-Prince pair which is based on an explicit Runge-Kutta (4,5) [11] was chosen as numerical integration method to simulate the model composed by ordinary differential equations, since the Eq. 2 was

discretized through the bioreactor bed height. The parameter estimation problem was solved by the non-linear least squares method, using the Levenberg-Marquardt algorithm [12].

The Dormand-Prince pair solves non-stiff differential equations and the estimation algorithm of Levenberg-Marquardt would prioritize the fitting for the larger scale variable. Owing to the model state variables at different scales, the original system model was stiff, being not suited for the chosen numerical integration and parameter estimation methods. The stiffness has led to a high condition number of the system in the integration problem (4. 247146 $\times 10^{14}$), i.e., the ratio of the largest singular value of the Jacobian matrix to the smallest. To overcome such problem, the cells mass and the reactor temperature (states variables) were normalized between 0 and 1.

The Levenberg-Marquardt Algorithm (LMA) is an optimization method that was first develop by Kenneth Levenberg [14] and later improved by Donald Marquardt [15]. It consists in a damped Least Squares method for nonlinear functions minimization, it is iterative and may be subjected to local minimum if a bad initial guess is given. At each iteration the chosen parameter is replaced for a new estimation until to reach a minimum, which may not be the global one. Further information of the LMA can be found on the work of Moré [12].

The model was written Matlab® routine. The numerical integration function used for the model simulations was the Matlab® function *ode45*, which uses the Dormand-Prince pair, based on an explicit Runge-Kutta (4,5). The parameter estimation code was also written Matlab® code, being the estimations performed using the Matlab® function *lsqnonlin* from the Optimization Toolbox®, which uses non-linear least square method to be minimized alongside the Levenberg-Marquard algorithm. All the simulations were run in a desktop Intel® Core™ i7-3770 with 3.40GHz processor and 12.0 Gb of RAM memory.

4. Material and methods

Detailed information concerning to medium composition, substrate origins and treatments can be obtained in the work of Mazutti et al. [13]. Briefly, the sugarcane bagasse was supplemented with pre-treated cane molasses 10.0 wt%, corn steep liquor 30.0 wt% and soybean bran 20.0 wt%. The strain of *Kluyveromyces marxianus* NRRL Y-7571 was maintained on an YM agar medium for further use. Cell production for pre-inoculum was carried out in 50 mL test tubes with 10 mL of liquid YM medium. The cell mass contained in the pre-inoculum was determined by a direct weight before its use in the fermentation.

The packed-bed SSF bioreactor consists of a cylindrical stainless bed with an air supplier and the air humidity is between 95-100%. The saturated air enters in the bottom of the bioreactor and passes through the bed until it exits at the top, as can be seen on Figure 1. The bioreactor was filled with 3 kg of dry sugarcane bagasse, supplemented as defined above. The moisture content was corrected to 65% (w/w) and autoclaved at 121°C for 20 minutes. The fermentations runs were started with the inoculation of a volume of inoculum corresponding to the cell mass defined in the experimental design. The total weight (moisture basis) was about 10 kg. The microbial growth expressed in terms of mass of cells was calculated considering the results of oxygen uptake rate [13].

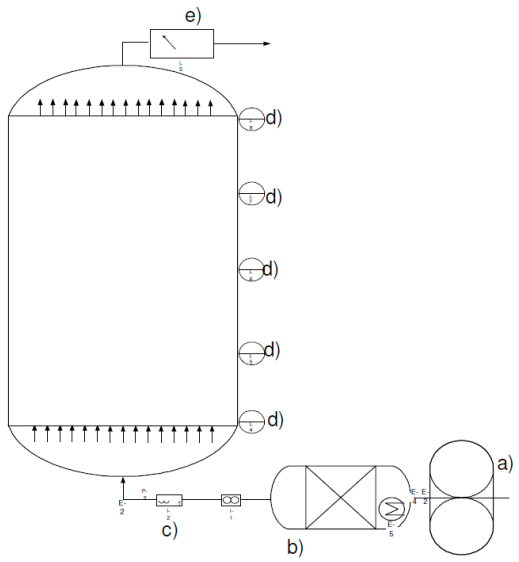

Figure 1: Schematic diagram of the packed bed bioreactor: a- compressor; b- humidifier; c- flowrate, temperature and humidity control; d- temperature sensors; e- CO_2, temperature and humidity measurer.

The experiments were carried out by 24 hours, the cell mass data acquisition was made hour by hour and the temperature measures were continuously taken at each height of the bed (inlet, 10, 20 and 30 cm and output - denoted by letter d in Figure 1) by a platinum resistance temperature detectors (RTD), called PT100 sensors. There was obtained 25 experimental points in each of the 6 experiments. The inlet air volumetric flow rate varied between 2.0 to 2.0 $[m^3/h]$, and its temperature between 27 to 33 $°C$.

5. Results and discussion

5.1 Parameters sensitivity and statistical analyses

The minimization of the objective function used in non-linear least squares method, cf. Equation 6, has resulted in 0.9843 and 1.0850 for the dimensionless values of cells mass and bioreactor temperature, respectively.

First of all, the parameters sensitivity analysis was performed in order to measure the influence of each parameter over the states. The parameters sensitivity matrix is computed over a steady-state of the process, and requires that the steady-state is extensible to other operation conditions and that the initial parameters values are reasonable. The equation for the parameters sensitivity matrix is exposed by Equation (8) [17].

$$0 = \frac{\partial f(\bar{x},\bar{u},\theta)}{\partial x}\frac{\partial \bar{x}}{\partial \theta} + \frac{\partial f(\bar{x},\bar{u},\theta)}{\partial \theta} \quad (8)$$

Where $f(\bar{x},\bar{u},\theta)$ represents the function for the cells and temperature at steady-state, x are the states (cells and temperatures) and θ are the parameters.

Table 1 presents the results of the analysis, which shows that the metabolic heat (Y_Q) has the major influence over the temperature of the bioreactor, and such influence its due the large value of the parameter and the lack of normalization - once there is no reference number which it may be submitted for normalization. Also, can be seen that for the cells growth profile only the specific growth rate (μ) has influence, which was expected since the observation of the cells growth differential equation.

Table 1. Parameters sensitivity analysis for cells and temperature over parameters.

	μ	ρ_S	C_{ps}	ε	Y_Q
X	6.5866×10^{-06}	-	-	-	-
T	-2.3106×10^{-05}	3.8503×10^{-04}	3.4438×10^{-06}	-1.6750×10^{-03}	$1.1841\times10^{+01}$

The 95% confidence interval of the estimated parameters can be seen in Table 2. Since all estimated parameters are significant (i.e., parameters significantly different from zero), all of them must be included in the model.

Table 2. Estimated parameters

μ	0.8641±0.1008
ρ_s	271.2540±0.2189
C_{ps}	2,836.0511±0.0003
ε	0.9995±0.00003
Y_Q	$8.3660\times10^{+6}$±101.4770

The Student's *t* test for 95% of confidence was performed to assure that the means are intersecting, thus it is not possible to consider the means different of each other [16]. The results presented in Table 3 show that the means for the experimental data and the model are equivalent, intersecting in an interval.

Table 3. Student's *t* test for cells mass and bioreactor temperature data

Cells mass			Temperature		
0.6293	$< \mu_{exp} = 0.7077 <$	0.7860	0.4606	$< \mu_{exp} = 0.4967 <$	0.5328
0.5974	$< \mu_{mod} = 0.6711 <$	0.7448	0.4159	$< \mu_{mod} = 0.4525 <$	0.4890

For the variances, the χ^2 and the *Fisher's exact test* for 95% of confidence were performed. Both tests have shown that the models cannot be distinguished, i.e., the variances intersect each other in the χ^2 test and the *Fisher's exact test* is into the lower and upper limits. These results can be seen in Tables 4 and 5, respectively.

Table 4. χ^2 test for cells mass and bioreactor temperature data

Cells mass			Bioreactor temperature		
0.1481	$< \sigma^2_{exp} = 0.1880 <$	0.2467	0.0314	$< \sigma^2_{exp} = 0.0398 <$	0.0523
0.1359	$< \sigma^2_{mod} = 0.1661 <$	0.2289	0.0333	$< \sigma^2_{mod} = 0.0407 <$	0.0519

Table 5. Fisher's exact test for cells mass and bioreactor temperature data

Cells mass			Bioreactor temperature		
lower limit	S^2_{exp}/S^2_{mod}	upper limit	lower limit	S^2_{exp}/S^2_{mod}	upper limit
0.6946	1.1322	1.4416	0.6946	0.9783	1.441

5.2 Model validation

In the previous work of Mazutti et al. [13], six experiments were carried out to observe the cells growth and temperature profile along the time. The cells growth kinetics phases can be distinguished among lag,

exponential and stationary phases. Further, the temperature profile can be analyzed as intimately connected to the growth, as the maximum temperature coincides with the exponential growth phase of the cells.

The cells mass experimental data used to estimate the parameters and fit the model for different inlet air velocity and temperature are shown in Figure 2. As aforementioned, the model used to estimate the cells mass was the Verhulst Logistic model. According to the results of Figure 2, for all six different operating conditions, the model has presented a good performance to predict the lag, the exponential growth and the stationary phases, which are intimately related to the temperature profile in the bioreactor. Analyzing the biomass profiles presented in Figure 2, it is possible to identify the microbial growth phases, where the lag phase occurs at 0–6 h, exponential phase at 6–11 h and stationary phase from 11–24 h.

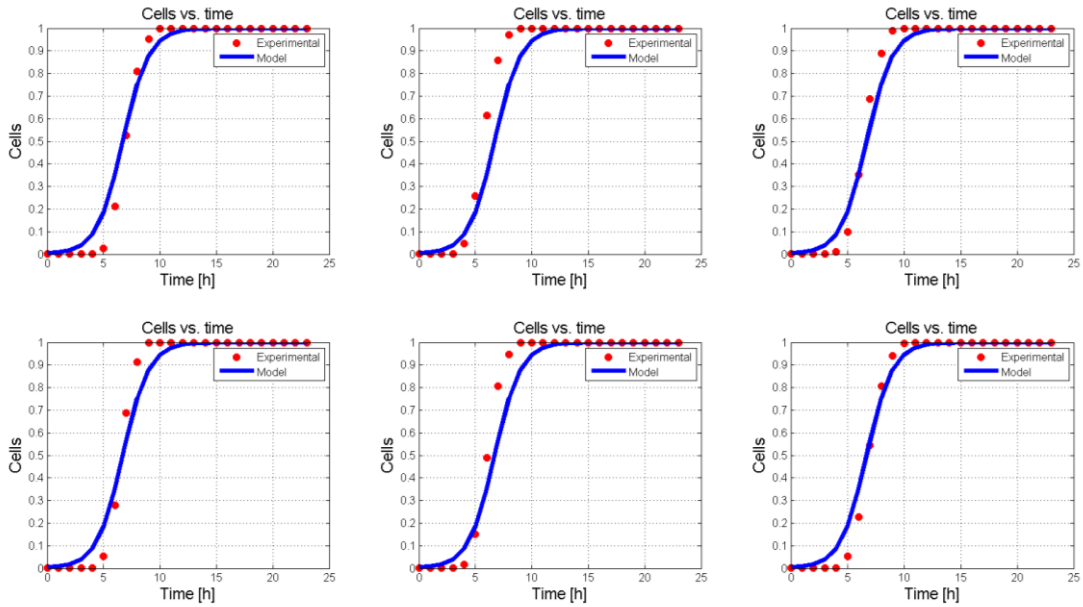

Figure 2: Comparison between the cells mass obtained by experimental data and by the process model using estimated parameters for different values of inlet air velocity $[m^3h^{-1}]$ and of temperature [°C], respectively: a) 2.0 and 27.0, b) 2.0 and 30.0, c) 2.0 and 33.0, d) 3.0 and 27.0, e) 3.0 and 30.0, f) 2.4 and 30.0.

In Figure 3 is shown the comparison between the bioreactor temperature obtained by experimental data and by the process model using estimated parameters for different inlet air velocity and temperature. As it can be observed, the phenomenological model could capture the bioreactor temperature dynamics for all six different operating conditions. The temperature gradients increase along the bioreactor height and, at the maximum gradient, the outlet air temperature reached values 20°C higher than the inlet air temperature. The increase in volumetric air flow rate did not significantly contribute to the metabolic heat removal, due to the low heat capacity of the air, as

already described in the literature, since the main contribution to metabolic heat removal is due to evaporation [13]. Increasing the inlet air temperature from 27 to 33°C caused a considerably decrease on the time where the highest temperature gradients were verified, whereas the volumetric air flow rate had little influence on the times of the maximum temperature gradients. In fact, the temperature is more influenced by evaporation than convective action of the inlet air.

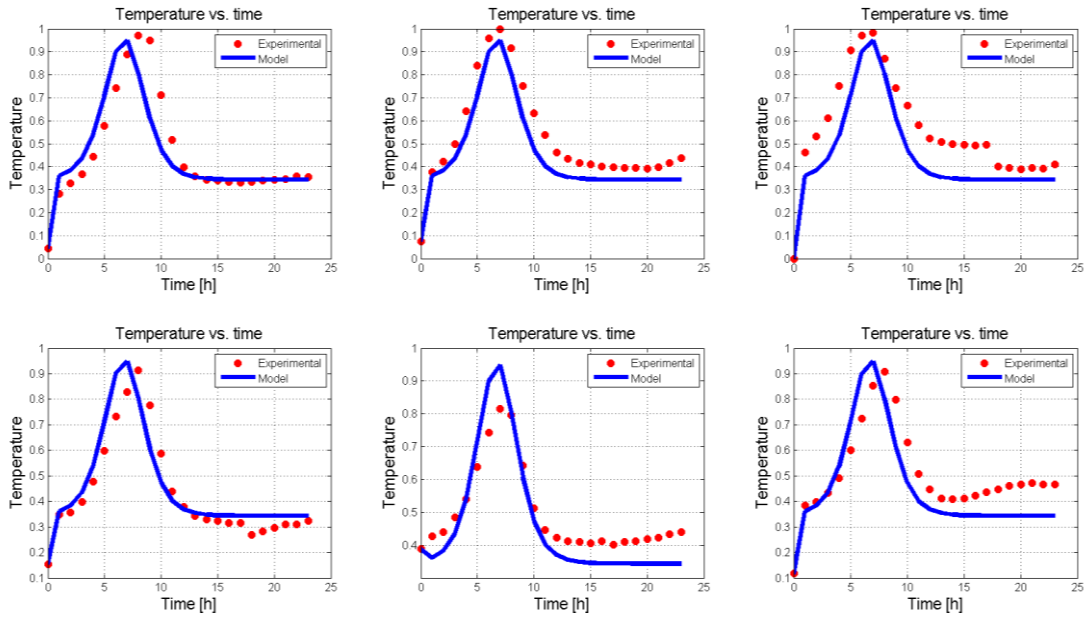

Figure 3: Comparison between the bioreactor temperature obtained by experimental data and by the process model, at the height *z* of 0.4 m, using estimated parameters for different values of inlet air velocity [m^3h^{-1}] and of temperature [°C], respectively: a) 2.0 and 27.0, b) 2.0 and 30.0, c) 2.0 and 33.0, d) 3.0 and 27.0, e) 3.0 and 30.0, f) 2.4 and 30.0.

The comparison between the observed (experimental) values and the predicted (model) values of cell mass and bioreactor temperature are shown in Figures 4 and 5, respectively. It is well-known, if predict values (circle) coincide perfectly with the observed ones (45º line), the model can be considered perfect fitted.

According to Figure 4, the differences between observed and predicted values of the cell mass are higher during the exponential growth phase. It is shown that such difference is negative in the first, fourth and sixth runs, and positive in the second and fifth ones. The best model fitting is shown by third run.

The residuals between observed and predicted values are due to the six experiments under different operating conditions. The Verhulst Logistic model fitted by a single set of estimated parameters, hence, fails in

predicting accurately all the six experiments. However, the satisfactory model forecast joint with the model simplicity become the Verhulst Logistic model suited to represent the cell growth.

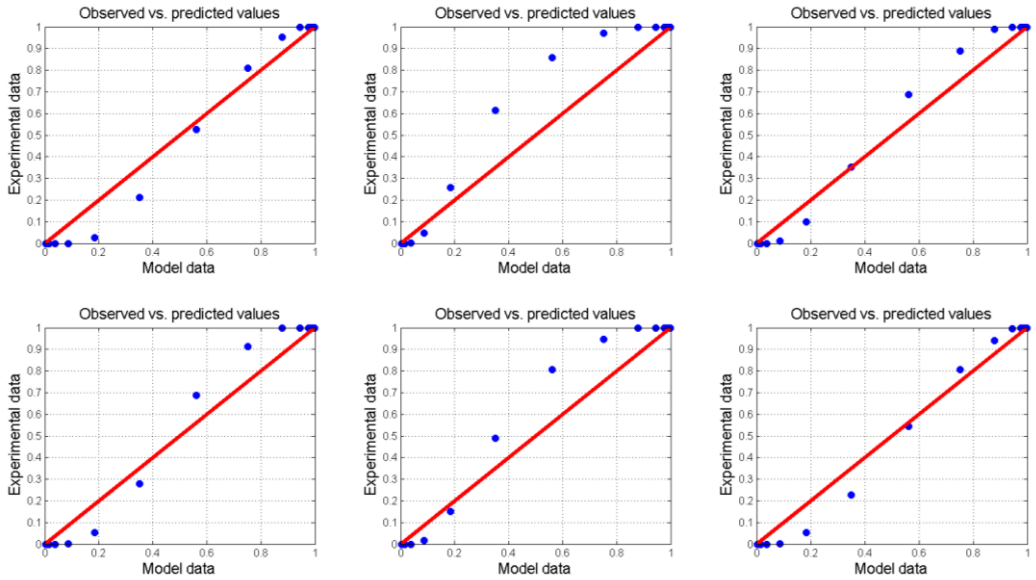

Figure 4: Observed (experimental) data vs. predicted (model) data of cells mass for the six runs.

In Figure 5 are illustrated the differences between observed and predicted values of bioreactor temperature. The major residuals are observed from the middle of the simulation time interval. From this point the cell growth is exponential and, due the derivative term in right-hand side of the Equation 2, the model shall be more sensitive to variations in cell population. Consequently, in this region the model cannot fit perfectly the experimental data. It is also verified by the larger distance between the predict and observed values in Figure 5.

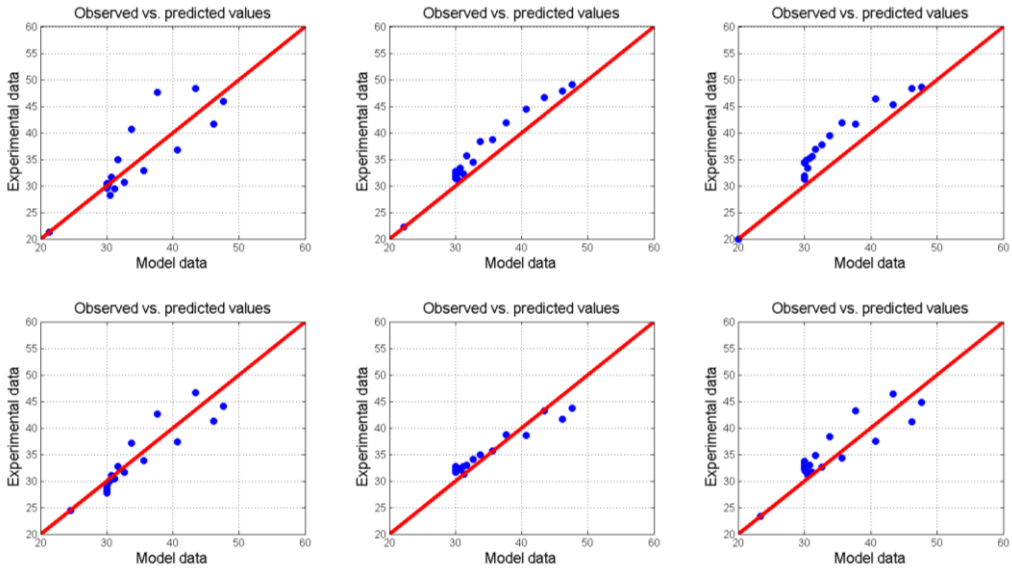

Figure 5: Observed (experimental) data vs. predicted (model) data of bioreactor temperature for the six runs.

The normalized numerical residuals, the difference between cells values obtained from experimental data and model simulation values, for each experiment (#1-6) at each measured time (0-24h) are given by Table 6. A similar comparison was made for temperature values in Table 7. The residual analyses have shown the biggest model prediction errors were observed in temperature. The total sum of the residuals for the cells and temperature are, respectively, 9.7800 and 11.1986. Also, it can be seen that the majority of the errors are amongst 8 to 10 hours, exactly when the cells starts to cease growing and the temperature falls fast, and due its behavior minor errors in the models results in major deviations from the experimental data (Figures 2-3).

Table 6. Normalized cells residuals (absolute errors) between experimental and simulated data for each time unit.

Time [h]	#1	#2	#3	#4	#5	#6
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0030	0.0030	0.0030	0.0030	0.0030	0.0030
2	0.0101	0.0101	0.0101	0.0101	0.0101	0.0101
3	0.0235	0.0235	0.0235	0.0235	0.0235	0.0235
4	0.0486	0.0455	0.0486	0.0486	0.0486	0.0486
5	0.0937	0.0478	0.0840	0.0913	0.0780	0.0912
6	0.1454	0.0880	0.0711	0.1174	0.0195	0.1153
7	0.0751	0.3275	0.0651	0.0087	0.2020	0.0594
8	0.0874	0.4173	0.2477	0.2487	0.3662	0.1049
9	0.2059	0.3675	0.2828	0.3087	0.3415	0.2012
10	0.2058	0.2513	0.2402	0.2500	0.2502	0.1919
11	0.1470	0.1477	0.1475	0.1477	0.1477	0.1426
12	0.0819	0.0819	0.0819	0.0819	0.0819	0.0818
13	0.0439	0.0439	0.0439	0.0439	0.0439	0.0438

14	0.0231	0.0231	0.0231	0.0231	0.0231	0.0230
15	0.0120	0.0120	0.0120	0.0120	0.0120	0.0119
16	0.0062	0.0062	0.0062	0.0062	0.0062	0.0061
17	0.0032	0.0032	0.0032	0.0032	0.0032	0.0031
18	0.0017	0.0017	0.0017	0.0017	0.0017	0.0015
19	0.0009	0.0009	0.0009	0.0009	0.0009	0.0008
20	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
21	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
22	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
23	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
24	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
Total	1.2193	1.9030	2.3974	1.4315	1.6641	1.1646

Table 7. Normalized temperature residuals (absolute errors) between experimental and simulated data for each time unit.

Time [h]	#1	#2	#3	#4	#5	#6
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0790	0.0173	0.0997	0.0137	0.0654	0.0241
3	0.0570	0.0392	0.1492	0.0261	0.0564	0.0152
4	0.0664	0.0642	0.1776	0.0354	0.0505	0.0011
5	0.0941	0.1052	0.2152	0.0597	0.0021	0.0460
6	0.1339	0.1273	0.1960	0.1133	0.0720	0.1098
7	0.1581	0.0584	0.0721	0.1684	0.1581	0.1753
8	0.0588	0.0511	0.0340	0.1207	0.1344	0.0966
9	0.1690	0.1141	0.0659	0.1106	0.0062	0.1037
10	0.3402	0.1443	0.1340	0.1684	0.0343	0.1890
11	0.2385	0.1595	0.1938	0.1148	0.0392	0.1595
12	0.1137	0.1378	0.1790	0.0381	0.0450	0.1069
13	0.0298	0.0917	0.1535	0.0092	0.0539	0.0779
14	0.0030	0.0786	0.1542	0.0107	0.0580	0.0580
15	0.0045	0.0676	0.1501	0.0183	0.0608	0.0608
16	0.0054	0.0634	0.1493	0.0225	0.0599	0.0668
17	0.0111	0.0576	0.1470	0.0283	0.0679	0.0782
18	0.0141	0.0546	0.1509	0.0278	0.0581	0.0924
19	0.0105	0.0514	0.0583	0.0757	0.0651	0.1029
20	0.0035	0.0515	0.0515	0.0619	0.0687	0.1168
21	0.0000	0.0481	0.0446	0.0481	0.0756	0.1237
22	0.0034	0.0550	0.0515	0.0344	0.0790	0.1271
23	0.0137	0.0722	0.0481	0.0344	0.0893	0.1237
24	0.0103	0.0928	0.0653	0.0206	0.0962	0.1237
Total	1.6181	1.8027	2.7409	1.3613	1.4962	2.1793

5.3. Distributed parameter model

A distributed parameter model has a major importance for a bioreactor because it can predict the process dynamics in two independent coordinates, e.g., time and bioreactor length. Here the mathematical model validated in the previous section was simulated not only using the time as independent variable but also the bioreactor length.

The time-varying measured temperature profiles in the bioreactor can be seen in Figure 6. This figure illustrates the experimental data obtained in five measured lengths of the bioreactor (inlet, outlet, 10, 20 and 30 cm).

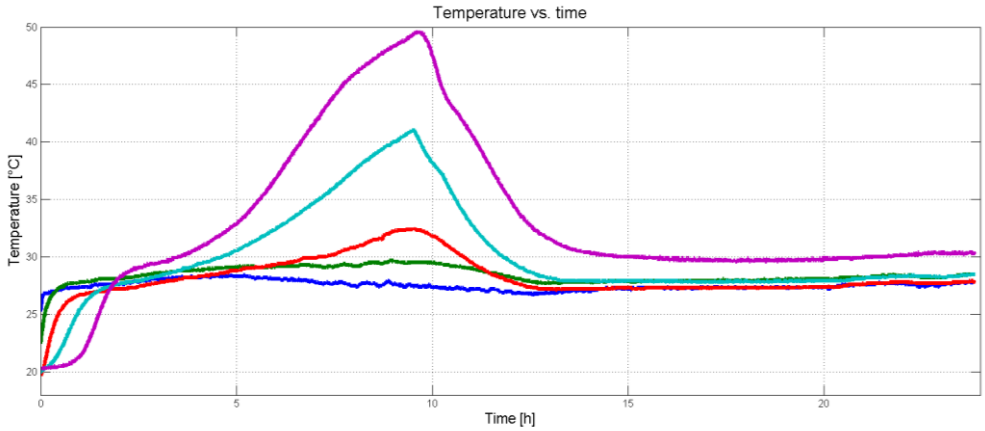

Figure 6: Experimental data of temperature profiles through time and bioreactor length: from inlet (*bottom*) to outlet (*upper*) for inlet air velocity of 2.0 m^3h^{-1} and inlet temperature of 27.0°C.

The results of temperature profiles obtained by the simulation of the model through time and bioreactor length are illustrated in Figure 7. If Figure 7 is compared to Figure 6, one can observe the similarity between the simulated and the experimental data of temperature profiles. It can be seen in Figures 6 and 7 that the most significant temperature variations can be observed in the bioreactor outlet (graph right-side). The simulated data of cells mass profile through time and bioreactor length are shown in Figure 8.

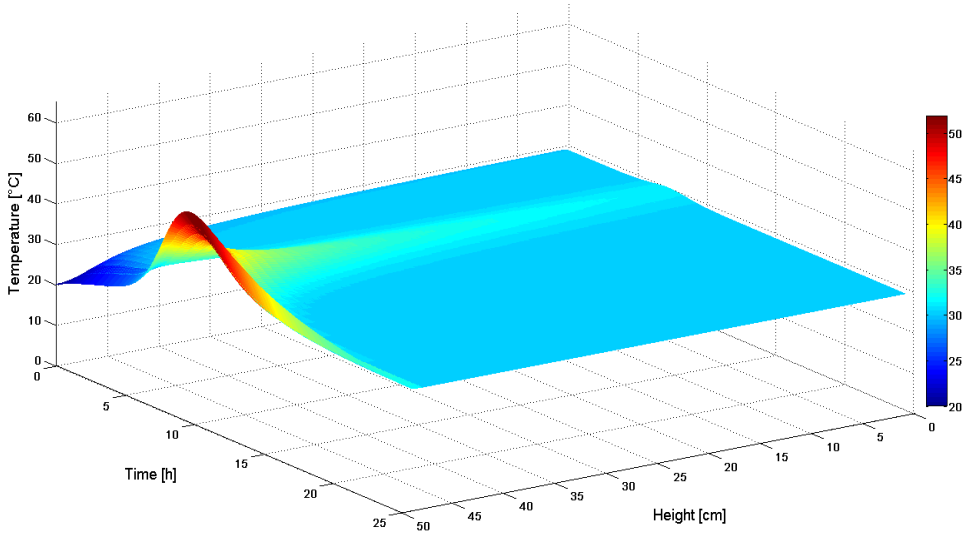

Figure 7: Simulated data of temperature profile through time and bioreactor length.

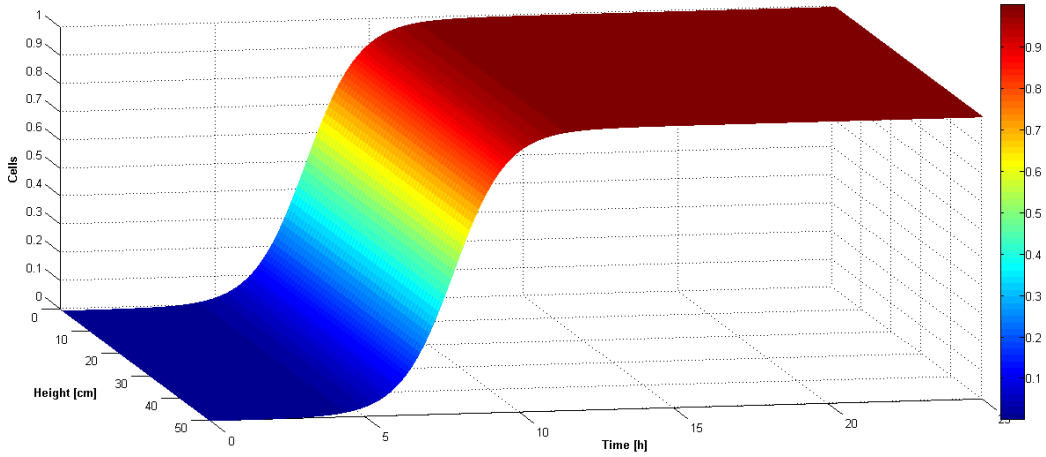

Figure 8: Simulated data of cells mass profile through time and bioreactor length.

According to Figures 7 and 8, it can be concluded that the validated model is able to predict the dynamics of both cells growth and bioreactor temperature profiles through time and space dimensions. Such model quality is essential for further development and application of control and optimization strategies.

Furthermore, according to the model (as it can be seen in Figure 7), the temperature has its dramatic increase almost exclusively in the middle of the bioreactor. The walls and a fairly height of the reactor remain almost unchanged, showing and important characteristic of a fixed-bed bioreactor for SSF that must be considered in making and scaling a bioreactor for this purpose. Else some problems may be faced during the process if the temperature increases too much at some areas, in this manner, a different configuration of the bioreactor cooling

system that allows a more homogeneous heat distribution would be interesting for the cells growth and process operation.

6. Conclusion

The main goal of this work was the building up and a posteriori validation of a mathematical model to predict the growth of the *Kluyveromyces marxianus* and temperature profile in a fixed-bed bioreactor for solid state fermentation using sugarcane bagasse as substrate was developed. The model chosen to fit to the cell mass experimental data was the Verhulst logistic. The bioreactor temperature profile was predicted with a model based on an energetic balance with time and space derivatives. Such model also takes into account the medium, the air heat exchange and the metabolic heat generated by the cells growth.

The mathematical model fitness to the experimental data was performed by a parameter estimation technique. For this task, it was considered six experimental datasets of cells mass and bioreactor temperature, obtained under different operating conditions. The parameter estimation problem was solved by the non-linear least squares method, using the Levenberg-Marquardt algorithm. The confidence intervals have demonstrated that the estimated parameters were significant. Further, the statistical analysis of Student's *t* test, χ^2 and Fisher's exact test have shown that the model is equivalent to the process experimental data – all tests were performed for 95% of confidence intervals. Low residues between the observed (experimental) and predicted (validated model) were found, endorsing the satisfactory model accuracy.

Finally, a distributed parameter model that explores the cells growth and bioreactor temperature profiles through time and space dimensions was simulated using the estimated parameters. The comparison between the distributed parameter model and the experimental data has shown that the model is suited not only to predict cells growth and bioreactor temperature profiles through the time but also trough the bioreactor length.

Currently a model-based predictive control strategy is being studied aiming to control the reactor temperature along the fermentation process. As further work, a more accurate model will be implemented in a Computational Fluid Dynamics (CFD) software may be composed, which will also consider axial variations into the bioreactor, giving a better tool to visualize the whole process behavior.

Appendix

Variable/Parameter	Description	Value	Units
μ	Specific growth rate	0.8641	h^{-1}
X	Cells mass	Time dependent	Dimensionless
X_m	Maximum possible cells concentration	1.000	Dimensionless
ρ_b	Bed density	2.4957	kg m^{-3}
ρ_a	Moist air density	1.1400	kg m^{-3}
ρ_s	Substrate density	271.2540	kg m^{-3}
C_{pb}	Heat capacity of the bed	3,389.8019	J $kg^{-1}°C^{-1}$
C_{pa}	Heat capacity of the moist air	1,180	J $kg^{-1}°C^{-1}$
C_{ps}	Heat capacity of the substrate	2,836.0511	J $kg^{-1}°C^{-1}$
ε	Void fraction	0.9995	Dimensionless
T	Reactor temperature	Time and cell mass dependent	°C
Y_Q	Metabolic heat coefficient	8.3660×10^{6}	J kg_{cells}^{-1}
V_z	Inlet air velocity	33.3300	m h^{-1}
ƒ	Water carrying capacity of air	0.00246	kg_{water} kg_{air}^{-1} $°C^{-1}$
λ	Water evaporation latent heat	2,414,300	J kg_{water}^{-1}

References

[1] GIBSON, G. R., WANG, X. Bifidogenic properties of different types of fructo-oligosaccharides. Food Microbiology. 11, 491-498 (1994).

[2] BOEHM, G., LIDESTRI, M. CASETTA, P. JELINEK, J. NEGRETTI, F., STAHL, B. MARINI, A. Supplementation of a bovine milk formula with an oligosaccharide mixture increases counts of faecal bifidobacteria in preterm infants. Arch Dis Child Fetal Neonatal. 86, 178-181 (2002).

[3] KNOL, J., BOEHM, G., LIDESTRI, M., NEGRETTI, F., JELINEK, J., AGOSTI, M., STAHL, B., MARINI, A., MOSCA, F. Increase of faecal bifidobacteria due to dietary oligosaccharides induces a reduction of clinically relevant pathogen germs in the faeces of formula-fed preterm infants. Acta Paediatrica. 94 (Suppl 449), 31-33 (2005).

[4] KIM, D. H., CHOI, Y. J., SONG, S. K., YUN, J. W. Production of inulo-oligosaccharides using endo-inulinase from a *Pseudomonas sp.*. Biotechnology Letters. Vol. 19, 369-371 (1997).

[5] SINGH, R. S., DHALIWAL, R., PURI, M. Production of inulinase from *Kluyveromyces marxianus* YS-1 using root extract of Aspargus racemosus. Process Biochemistry. 41, 1703-1707 (2006).

[6] PANDEY, A. Solid state fermentation. Biochemical Engineering Journal. 13, 81-84 (2003).

[7] NAGEL, F. J. I., TRAMPER, J., BAKKER, M. S. N., RINZEMA, A. Temperature control in a continuously mixed bioreactor for solid-state fermentation. Biotechnology and bioengineering, Vol. 72, 2 (2001).

[8] SAUCEDO-CASTANEDA, G., LONSANE, B. K., KRISHNAIAH, M. M., NAVARRO, J. M., ROUSSOS, S., RAIMBAULT, M. Maintenance of heat and water balances as a scale-up criterion for the production of ethanol by *Schwanniomyces castellii* in a solid state fermentation system. Process biochemistry. 27, 97-107 (1992).

[9] LEKANDA, J. S., PÉREZ-CORREA, J. R. Energy and water balances using kinetic modeling in a pilot-scale SSF bioreactor. Process biochemistry. 39, 1793-1802 (2004).

[10] FANAEI, M. A., VAZIRI, B. M. Modeling of temperature gradientes in packed-bed solid-state bioreactors. Chemical Engineering and Processing: Process Intensification. (2008).

[11] Dormand, J. R. and P. J. Prince. A family of embedded Runge-Kutta formulae. J. Comp. Appl. Math., 6, 19-26, 1980.

[12] MORÉ, J. J. The Levenberg-Marquardt Algorithm: implementation and theory. Numerical Analysis. Ed. G. A. Watson, Lectures Notes in Mathematics 630, Springer Verlag, 105-116 (1977).

[13] MAZUTTI, M. A., ZABOT, G., BONI, G., SKIVRONSKI, A., OLIVEIRA, D., LUCCIO, M., RODRIGUES, M. I., TREICHEL, H., MAUGERI, F. Kinetics of inulinase production by solid-state fermentation in a packed-bed bioreactor. Food Chemistry. 120, 163-173 (2010).

[14] LEVENBERG, K. A Method for the solution of certain non-linear problems in least squares. Quarterly applied mathematics. 2, 164-168 (1944).

[15] MARQUARDT, D. An Algorithm for least-squares estimation of nonlinear parameters. SIAM Journal on Applied Mathematics. 11 (2), 431-441 (1963).

[16] SCHWAAB, M.; PINTO, J. C. Experimental data analysis I. Fundamentals of statistics and parameter estimation. E-papers. Rio de Janeiro – RJ (2007).

[17] LI, R.; HENSON, A.; KURTZ, M. J. Selection of model parameters for off-line parameter estimation. IEEE Transactions on Control Systems Technology. 12(3), (2004).

5 IDENTIFIABILITY MEASURES TO SELECT THE PARAMETERS TO BE ESTIMATED IN A SOLID-STATE FERMENTATION MODEL

Christian L. da Silveira, Marcio A. Mazutti, Nina P. G. Salau*

Chemical Engineering Department, Universidade Federal de Santa Maria

Abstract

BACKGROUND: Process modeling can lead to of advantages such as helping in process control, reducing process costs and product quality improvement. This work proposes a solid-state fermentation model composed by seven differential equations to represent the process, and performs a parameters estimation with a parameters identifiability analysis (PIA), in order to build an accurate model with optimum parameters.

RESULTS: Statistical tests were performed to verify the model accuracy with the estimated parameters considering different assumptions. The results have shown that the model assuming substrate inhibition better represents the process. Also, it was shown that 8 model parameters were non-identifiable, and with the removal of these parameters from the estimation procedure the results obtained were better.

CONCLUSION: PIA can be useful to estimation procedure, since it may diminish the number of parameters that can be evaluated. Also, PIA improved the model results, showing to be an important procedure to be taken.

Keywords: solid-state fermentation, bioreactor, mathematical modeling, parameters identifiability analysis, optimization, parameter estimation.

1. Introduction

Process modeling can lead to varieties of advantages, Imamoglu and Sukan[1] have reported that it helps in process control, reduces process costs and can improve product quality. There are some examples in literature

* To whom all correspondence should be addressed. E-mail: ninasalau@ufsm.br

Address: Chemical Engineering Department, UFSM – Av. Roraima, 1000, Cidade Universitária - Bairro Camobi. 97105-900 Santa Maria, RS – Brazil.

Phone: +55-55-3220-8448- Fax: +55-55-3220-8030.

where the modeling and simulation were used in biological process. A kinetic model of ethanol produced from sugar beet raw juice, which can also be used for further bioethanol production process, is found in Dodic *et al.*[2]. In the work of Salau *et al.*[3] is discussed the benefits of modeling and process simulations to develop a controller for an industrial gas-phase polyethylene reactor.

Besides a reliable process model, good parameter estimation is required to keep the model adjusted to the process. In general, the process parameters cannot be measured and, hence, they may be estimated from available experimental data by minimization of some objective function[4] such as nonlinear least squares minimization[5]. In this context, Habibi *et al.*[6] have used the nonlinear regression for parameters estimation in a dynamic model for a biological process of formaldehyde degradation using bacteria.

Aiming the simulations reliability, the process model must be suitable to predict the observed phenomena or experimental data. Thus, Silveira *et al*.[7] have proposed a model, hereafter called short model, to represent a solid-state fermentation process, the case study of this work. The so-called short model is composed only by one ordinary differential equation (ODE) and one partial differential equation (PDE), which are used to predict the cells growth kinetics and bioreactor bed temperature profile, respectively. The model presented by Silveira *et al.*[7] is a simpler form of the model presented by Fanaei and Vaziri[8], which also has taken a physiological factor into account. Such factor is dependent on two Arrhenius-like algebraic equations with terms of synthesis and denaturation of the reactions. These terms were also considered in this work. All aforementioned models are based on a Logistic Equation develop by Verhulst in 1838[9] for the cells growth kinetics and in an energy balance for the bioreactor bed temperature profile. Further details are given in Section 3.

Solid-state fermentation (SSF) is defined as a fermentation process with no free water, that implies in the fact that the substrate must contain only enough moisture to support the metabolism microorganisms growth[10]. According to Mazutti *et al.*[11], SSF has received more interest because of its lower energy requirements and wastewater production, in comparison to submerged fermentation, and of its use in agro-industrial wastes treatment. In addition, the products obtained by solid-state fermentation are concentrated and their recovering process is easier if compared to the submerged fermentation[12]. Nevertheless, excessive heating in bioreactors for SSF are still a huge problem that must be controlled to scale-up the process[13].

The main goal of this work is to propose a model that represents the solid-state fermentation process in a packed-bed bioreactor using *Kluyveromyces marxianus*. The model proposed in this work is more complex than the one presented in the previous work[7], being composed by a set of six ODEs and one PDE used to predict the cells growth kinetics, the cells physiological state, the bioreactor bed temperature profile, the substrate

consumption, the ethanol production and both carbon dioxide and oxygen produced by the cells metabolism. Here, three different assumptions for the specific growth rate are compared to be used in the model. *Student's t-test* and the *Fisher's exact test* were performed in order to verify if the means and the variances of the full model using three different assumptions for the specific growth rate correspond to the experimental data. The sum of residuals between the experimental and model data was chosen as measure performance to choose the best model to represent the solid-state fermentation process studied in this work.

Furthermore, this work will also present a parameters identifiability analysis, in order to evaluate the most important parameters used in the solid-state fermentation model proposed.

2. SSF modeling

2.1. Previous model of solid-state fermentation

Hereafter, the equations that have been used to build the previous work model of the SSF process are presented. The solid-state fermentation modeling generally involves a S-curve, described by a logistic function developed by Verhulst[9] for the cells growth. Fanaei and Vaziri[8] have included a physiological factor (ϕ). Rajagopalan and Modak[14], Lagemaat and Pyle[15], among other researchers have considered the Verhulst equation (eq 1) as a suitable model for growth kinetics.

In a previous work of Silveira *et al.*[7], the model was proposed to represent a solid-state fermentation process in a less complex form. The model was able to represent only the yeast cells growth over time and the bioreactor bed temperature profile over both time and length of the bioreactor. Therefore, the model was composed by only one ordinary differential equation and one partial differential equation with five model parameters to be estimated. Further, no variation for the specific growth rate was used, i.e., it was considered as a constant, and no physiological factor was included to the model.

The previous work model is given by eqs 1-4.

$$\frac{dX}{dt} = \mu_{max} X \left(1 - \frac{X}{X_m}\right) \tag{1}$$

$$\rho_b C_{pb} \left(\frac{\partial T}{\partial t}\right) = \rho_s (1-\varepsilon) Y_Q \frac{dX}{dt} + \rho_a C_{pa} V_z \left(\frac{\partial T}{\partial z}\right) + \rho_a \lambda V_z \left(\frac{\partial T}{\partial z}\right) \tag{2}$$

$$\rho_b = \varepsilon\rho_a + (1-\varepsilon)\,\rho_s \quad (3)$$

$$C_{pb} = \frac{\varepsilon\rho_a(C_{pa}+f\lambda)+(1-\varepsilon)\rho_s C_{ps}}{\rho_b} \quad (4)$$

Where the eq 1 represents the cells growth based on the Logistic Equation proposed by Verhulst, as already mentioned; eq 2 results from the reactor energy balance, that takes into account the metabolic heat produced as a result of the cells growth (first term on the right-hand side), the energy exchange with the air that flows through the reactor (second term on the right-hand side), and the latent heat of moisture evaporation through the bed (third term on the right-hand side).

Eq 3 computes the bed density as the void fraction of the reactor being filled by air and the non-void fraction being the substrate density. Eq 4 depicts the specific heat capacity of the bed, again taking into account the air and substrate fractions.

2.2. Model of solid-state fermentation

Besides predicting the cells growth kinetics and bioreactor bed temperature profile, even as the previous model, the model presented in this work takes into account likewise: the cells physiological state, the substrate consumption, the ethanol production and both carbon dioxide and oxygen produced by the cells metabolism.

The use of a factor called physiological factor to represent the changes of the cells growth as a function of bioreactor bed temperature, water activity, and time was a priori reported by Baranyi *et al.*[16] and a posteriori other works have referred their work. Fanaei and Vaziri[17], for instance, have adapted the logistic equation (eq 1) to include the physiological factor of the yeast cells (ϕ), as given by eq 5.

$$\frac{dX}{dt} = \mu\phi X\left(1-\frac{X}{X_m}\right) \quad (5)$$

Mitchel *et al.*[17] have characterized the physiological factor as a function composed by two Arrhenius-type terms, with temperature dependence. Dalsenter *et al.*[18] extended this function to a so-called key component, describing the auto-synthesis and thermal denaturation of the component. Following the idea of both works, Fanaei and Vaziri[8] have proposed the eq 6 in which the physiological factor depends on the time. This eq 6 is supported by eq 7 and 8 that represent the synthesis and denaturation coefficients, respectively.

$$\frac{d\phi}{dt} = \gamma_s \phi(1 - \phi^{\alpha}) - \gamma_d \phi \quad (6)$$

$$\gamma_s = \gamma_{s0} e^{-\frac{E_S}{R(T+273)}} \quad (7)$$

$$\gamma_d = \gamma_{d0} e^{-\frac{E_d}{R(T+273)}} \quad (8)$$

The energy balance used in this model is the same presented in the previous work model, and it was already presented in eq 2.

Eqs 9-12 are intimately linked with the cells growth kinetics, since all of them – substrate consumption (total reducing sugar - TRS), *S*, and production of ethanol, *P*, carbon dioxide, CO_2 and oxygen, O_2, – are dependent on the eq 5, considering its respective conversion rate.

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}} \frac{dX}{dt} \quad (9)$$

$$\frac{dP}{dt} = Y_{P/X} \frac{dX}{dt} \quad (10)$$

$$\frac{dCO_2}{dt} = Y_{CO_2/X} \frac{dX}{dt} \quad (11)$$

$$\frac{dO_2}{dt} = Y_{O_2/X} \frac{dX}{dt} \quad (12)$$

Aiming to obtain an accurate model to represent the solid-state fermentation process, three different assumptions were compared: a) the system presents substrate inhibition and, thereby, the specific growth rate (μ) is a function of the substrate (eq 13); b) the Monod model is used to calculate the specific growth rate (μ) (eq 14) and c) the specific growth rate (μ) is a constant value, i.e., a model parameter.

$$\mu = \frac{\mu_{max} S}{K_S + S + k_1 S^2} \quad (13)$$

$$\mu = \frac{\mu_{max} S}{K_S + S} \quad (14)$$

All the experimental data of the state variables was normalized, keeping the results between 0 and 1. The normalization procedure is described by eq 15.

$$\Psi = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \quad (15)$$

3. Parameter estimation

The model parameters were estimated to fit them to the experimental data obtained by Mazutti *et al.*[11]. According to the model equations given in section 2.2, the estimated parameters were: μ_{max}, K_s, k_1, ε, ρ_s, C_{ps}, ρ_a, α, $Y_{x/s}$, $Y_{P/x}$, $Y_{x/O2}$, $Y_{x/CO2}$, Y_q, E_s, E_d, γ_{s0} and γ_{d0}. The parameters values, units and descriptions can be found in Table 1. A total number of seventeen parameters were estimated, for six experiments measured at twenty-five different time units. According to eq 16, 133 degrees of freedom were available for the statistical tests.

$$DF = NE.NY - NP \quad (16)$$

where DF is the degrees of freedom of the system, NE the number of experiments, NY the number of measures for each experiment and NP the number of parameters.

The model fitting is based on the experimental data. They are used to assign parameter values that minimize the objective function in the parameter estimation problem. The bioreactor bed temperature measurement chosen was the one from the sensor located at the bioreactor top because this variable suffers the greatest changes[7]. The objective function to be minimized is the non-linear least squares function, given by eq 17. It is defined as the sum of the differences between experimental (subscript *exp*) and model (subscript *mod*) data.

$$\begin{bmatrix} \min \begin{matrix} f(X) \\ f(T) \\ f(S) \\ f(P) \\ f(CO_2) \\ f(O_2) \end{matrix} \end{bmatrix} = \begin{bmatrix} \left\|X^i_{exp} - X^i_{mod}\right\|^2 \\ \left\|T^i_{exp} - T^i_{mod}\right\|^2 \\ \left\|S^i_{exp} - S^i_{mod}\right\|^2 \\ \left\|P^i_{exp} - P^i_{mod}\right\|^2 \\ \left\|CO^i_{2\,exp} - CO^i_{2\,mod}\right\|^2 \\ \left\|O^i_{2\,exp} - O^i_{2\,mod}\right\|^2 \end{bmatrix}_{\substack{i=0,\dots,NY-1 \\ j=1,\dots,NE}} =$$

$$\sum_{j=1}^{NE} \sum_{i=0}^{NY-1} F_{obj}[X(,i), T(j,i), S(j,i), P(j,i), CO_2(j,i), O_2(j,i)] \quad (17)$$

Parameters estimated are uncertain to some extent[4], as good the parameters are the smaller is the value of the objective function, represented by the least squares equation (eq 17). Thus, the estimation problem involves a data regression, an adjustment of the parameters to satisfy some constraint and process boundaries, and a minimization of an objective function[19].

The least squares objective function for minimization denotes the error sum of squares, i.e., the difference between the experimental and the model data. Since the experimentally observed data are fixed, the model data are obtained directly by the equations (eqs 2, 5, 9-12) and the objective function is a function of the parameters[5].

The Matlab® *lsqnonlin* solves a nonlinear least-squares problem as the one given by eq 17. The algorithm chosen for this problem was the Levenberg-Marquardt.

The Levenberg-Marquardt Algorithm (LMA) is a derivative-based method of optimization that was first develop by Kenneth Levenberg[20] and later improved by Donald Marquardt[21]. It consists in an iterative damped Least Squares method for minimization of nonlinear functions, however it may be subjected to local minimum if a bad initial guess is given. At each iteration, the chosen parameter is replaced by a new estimation until to reach a minimum, which may not be the global one[22].

3.1. Numerical integration

The numerical integrator used on this study was the Dormand-Prince pair, based on an explicit Runge-Kutta of fourth and fifth order, which solves non-stiff differential equations[23]. This algorithm was used through the Matlab® function *ode45*, which is attached to the software. This *ode45* function receives the low and upper limits for integration, in this case time from zero to twenty four ([0 24]), the initial conditions for each differential equation (the initial guesses were the first point of each experiment) and some options, in which were determinate Absolute Tolerance and Relative Tolerance to 1×10^{-6}.

The computer used to perform the integrations and estimation has an Intel® Core™ i7-3770 with 3.40 GHz processor and 12 Gb of RAM memory and it is running with the Windows 7 64 bits Operating System.

4. Material and methods

Detailed information concerning to medium composition, substrate origins and treatments can be obtained in the work of Mazutti *et al.*[11]. Concisely, the sugarcane bagasse was supplemented with pre-treated cane molasses

10.0 wt%, corn steep liquor 30.0 wt% and soybean bran 20.0 wt%. The strain of *Kluyveromyces marxianus* NRRL Y-7571 was maintained on an YM agar medium for further use. Cell production for pre-inoculum was carried out in 50 mL test tubes with 10 mL of liquid YM medium. The cell mass contained in the pre-inoculum was determined by a direct weight before its use in the fermentation[7].

The packed-bed SSF bioreactor, with 0.0363 m³, consists of a cylindrical stainless bed with a saturated, or at least 95% of water, air supplier. The saturated air flows through the bed from the bottom to the top, as can be seen on Figure 1. The inlet and outlet air temperatures were continuously monitored by a temperature probe (PT100, NOVUS, Brazil). The microbial growth expressed in terms of mass of cells was calculated considering the results of oxygen uptake rate[7,11].

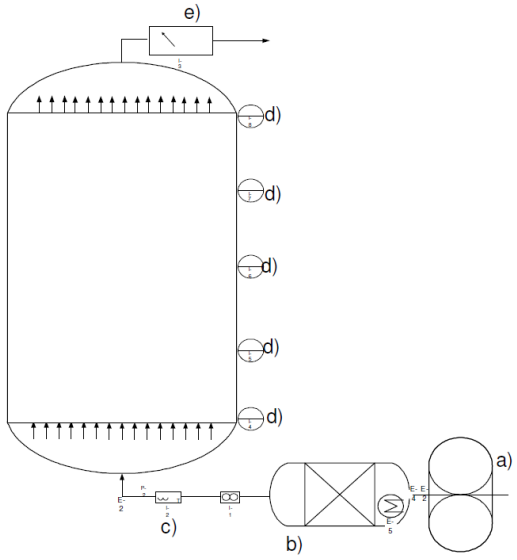

Figure 1: Schematic diagram of the packed bed bioreactor: a- compressor; b- humidifier; c- flowrate, temperature and humidity control; d- temperature sensors; e- CO_2, temperature and humidity measurer.

The experiments were carried out by 24 hours, the cell mass, total reducing sugar, ethanol, CO_2 and O_2 data acquisition were made hour by hour and the temperature measures were taken at each height of the bed (inlet, 10, 20 and 30 cm and output), denoted by letter d in Figure 1.

5. Modeling validation, results and discussion

5.1. Estimated parameters and statistical analyses

As referred in section 4, the parameters estimated for each one of the three different assumptions for the specific growth rate can be seen in Table 1. Some of the confidence intervals were too narrow and, consequently, they are not presented in the Table 1. Firstly it was performed a global estimation, using all the available data, then the estimated parameters were used as initial guesses for the parameter estimation using each set of experimental data individually. After each parameter estimation step, the residuals, i.e., the difference between the normalized experimental data and the model data obtained with the estimated parameter were computed and then a new parameter estimation step was performed considering the newest estimated parameters as initial guesses. The new residuals set were then compared to the one of the last parameter estimation step. This procedure was repeated many times using each experiment dataset until to achieve the smallest set of residuals.

Although the heat capacity of the substrate and the moist air density are physical properties, their values are unknown. Even though there might be values already used in literature, these values only give us some guidance in which region the parameters are, since we do not know the real values for these experiment conditions, and for this reason these parameters were also estimated.

The means of the parameters estimated using each experiment dataset and their confidence intervals were computed according to the works of Draper and Smith[5] and Schwaab and Pinto[24].

Table 1. Parameters estimated with intervals of 95% of confidence.

Parameter	Substrate inhibition	Monod	Constant specific growth rate	Units	Description
μ_{max}	0.8243±0.0438	0.8246*	0.8242±0.0775	h^{-1}	Maximum growth rate
K_s	0.00024±0.0149	4.1600E-06*	-----	g/L	Half-velocity constant
k_1	-0.3109±0.0787	-----	-----	L/g	Dissociation constant
ε	0.9991*	0.9989*	0.9990*	[ ]	Void fraction
ρ_s	269.9952*	269.9952*	269.9952*	$kg.m^{-3}$	Substrate density
C_{ps}	2499.9575*	2499.9575*	2499.9575*	$J.\ kg^{-1}.{}^{o}C^{-1}$	Heat capacity of substrate
ρ_a	0.9000*	0.8999*	0.9000*	$Kg.m^{-3}$	Moist air density

α	11.0000*	11.0000*	11.0000*	[ ]	Physiological exponent
$Y_{x/s}$	1.0241±0.0329	1.0241*	1.0241±0.0204	g/g	Substrate to cells yield coefficient
$Y_{P/x}$	1.0160±0.0097	1.0160*	1.0160±0.0072	g/g	Cells to product yield coefficient
$Y_{x/O2}$	1.0167±0.0091	0.9835*	0.9835±0.0028	g/g	Cells to [O_2] yield coefficient
$Y_{x/CO2}$	1.0647±0.0140	0.9392*	0.9392±0.0134	g/g	Cells to [CO_2] yield coefficient
Y_q	6.6928E+06*	6.2745E+06*	6.2745E+06*	$J.kg_{cells}^{-1}$	Metabolic heat coefficient
E_s	6.8137E+04*	6.8137E+04*	6.8137E+04*	$J.mol^{-1}$	Activation energy for the physiological factor synthesis
E_d	2.9451E+05*	2.9451E+05*	2.9451E+05*	$J.mol^{-1}$	Activation energy for the physiological factor denaturation
γ_{s0}	9.7603E+08*	9.7603E+08*	9.7603E+08*	h^{-1}	Frequency factor for the physiological synthesis
γ_{d0}	8.7400E+45*	8.7400E+45*	8.7400E+45*	h^{-1}	Frequency factor for the physiological denaturation

* Confidence interval too narrow (< E-05).

For the substrate inhibition assumption, it can be seen that the half-velocity constant, K_s, has a confidence interval larger than its absolute value, meaning that it could have no significance. Also, the dissociation coefficient has assumed negative values, showing a different response than the expected on the substrate inhibition.

In Figures 2-6 the residuals between the experimental and the model data for all the process variables for one experiment are illustrated. As closer the data is to the straight line the better is the fitness, meaning that the model value is very close to the experimental value. In all the four figures, it can be seen that the substrate inhibition assumption has significantly less residuals, as the Monod and constant specific growth rate assumptions have larger errors. Both Monod and constant specific growth rate have similar residuals plots, differing mostly in the bioreactor bed temperature profile.

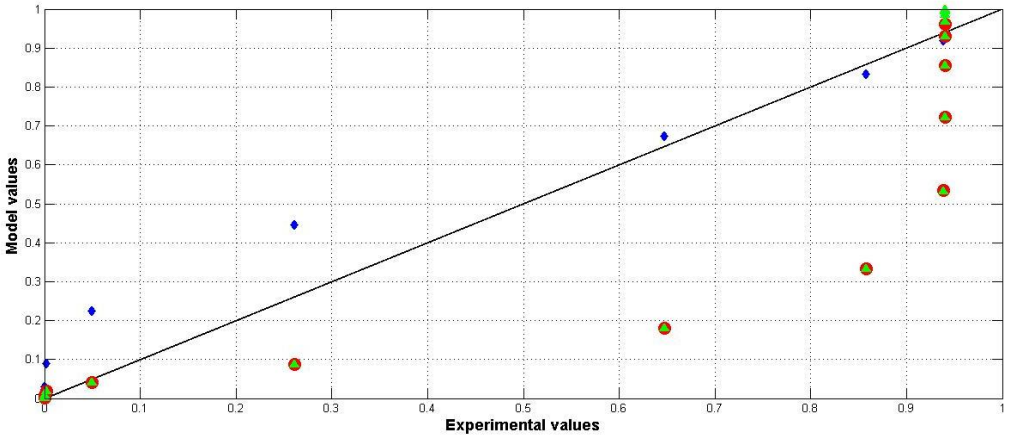

Figure 2. Model vs. Experimental values (residuals) for cells growth kinetics (substrate inhibition ⧫, Monod ●, constant specific growth rate▲).

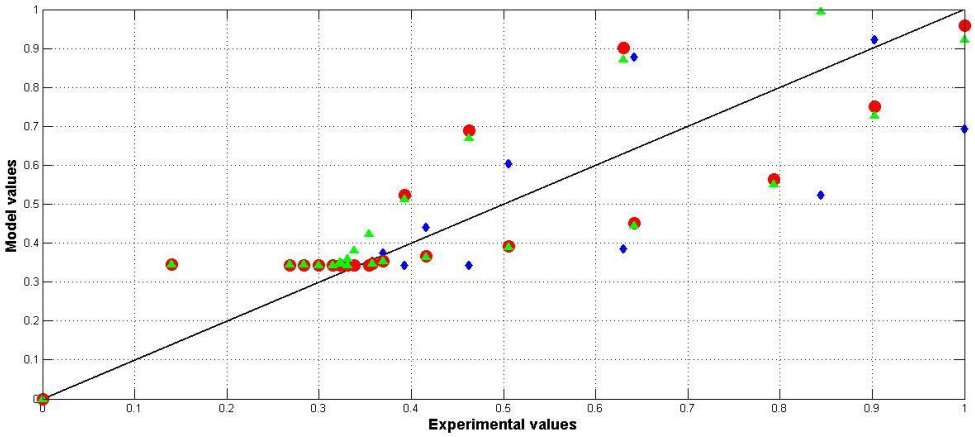

Figure 3. Model vs. Experimental values (residuals) for bioreactor bed temperature profile (substrate inhibition ⧫, Monod ●, constant specific growth rate▲).

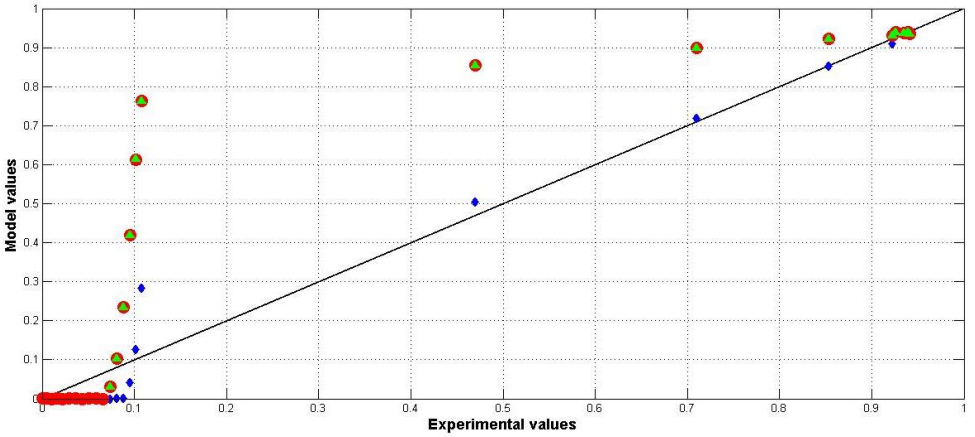

Figure 4. Model vs. Experimental values (residuals) for substrate consumption (substrate inhibition ⧫, Monod ●, constant specific growth rate▲).

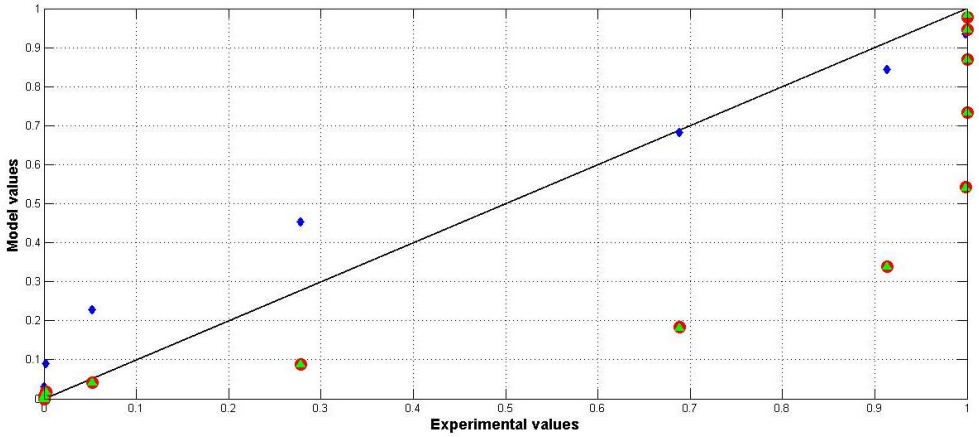

Figure 5. Model vs. Experimental values (residuals) for ethanol production (substrate inhibition ⧫, Monod ●, constant specific growth rate▲).

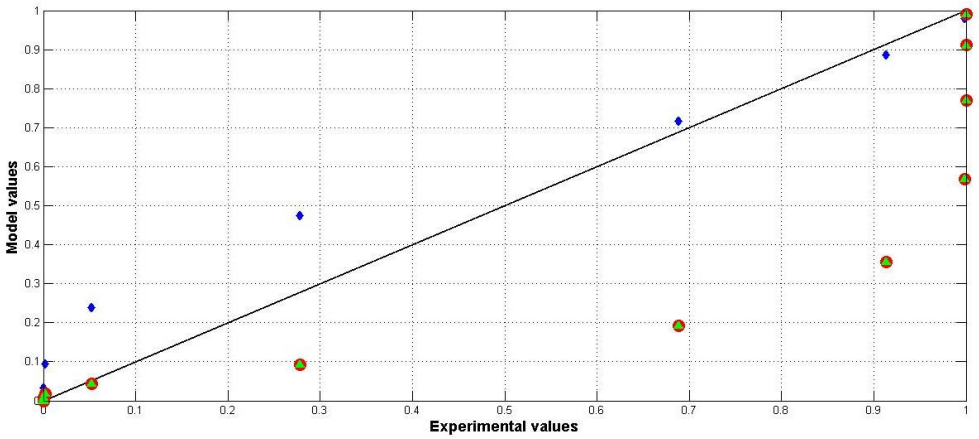

Figure 6. Model vs. Experimental values (residuals) for CO2 produced by the cells metabolism (substrate inhibition ⧫, Monod ●, constant specific growth rate▲).

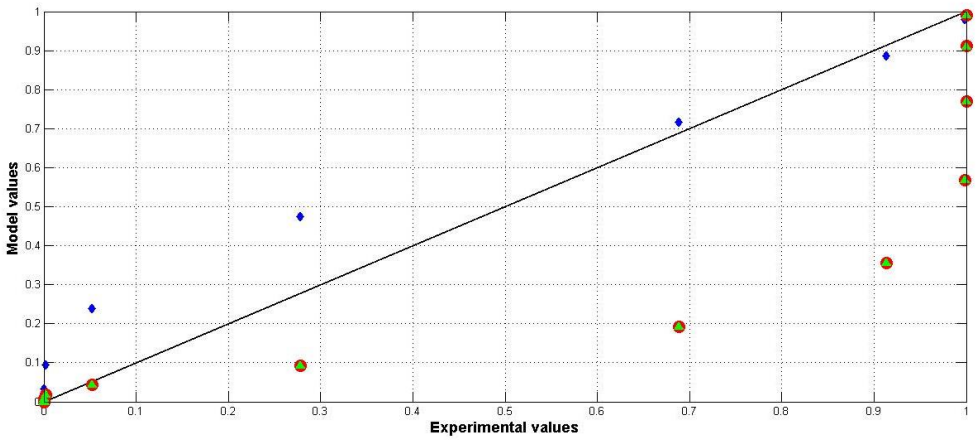

Figure 7. Model vs. Experimental values (residuals) for O2 produced by the cells metabolism (substrate inhibition ⧫, Monod ●, constant specific growth rate▲).

In addition, to quantify precisely the residuals of all the experiments for all the variables, concisely presented in Figure 2-7, Table 2 is shown to compare the objective function (eq 17) value for the three tested assumptions. It can be seen that the assumption of substrate inhibition has presented the smallest value for the objective function and, thereby, both parameters previously mentioned, K_s and k_1, were maintained in the model.

Table 2. Objective function values comparison for the three assumptions for the specific growth rate.

Model	Substrate inhibition	Monod	Constant specific growth rate
Objective function value	11.2542	48.3033	48.0561

Nevertheless, the models must be tested to verify whether the model fidelity to the process, i.e. whether the model is able to predict the process dynamics with the estimated parameters. *Student's t-test* and the *Fisher's exact test* were performed in order to verify if the means and the variances of the full model using three different assumptions for the specific growth rate correspond to the experimental data. The results can be seen in Tables 3 and 4, respectively.

Table 3. *Student's t-test* for all model equations for each one of the three different assumptions for the specific growth rate s ($\bar{x}$ is the mean of each state).

	Cells growth	Bed temperature	Total reducing sugar
Experimental	$0.6237 < \bar{x} < 0.7560$	$0.4702 < \bar{x} < 0.5399$	$0.1932 < \bar{x} < 0.2943$
Substrate inhibition	$0.5904 < \bar{x} < 0.7121$	$0.4346 < \bar{x} < 0.4991$	$0.2198 < \bar{x} < 0.3386$
Monod	$0.4966 < \bar{x} < 0.6252$	$0.4195 < \bar{x} < 0.4909$	$0.3047 < \bar{x} < 0.4303$
Constant specific growth rate	$0.5185 < \bar{x} < 0.6541$	$0.4233 < \bar{x} < 0.4902$	$0.2764 < \bar{x} < 0.4088$
	Ethanol	O_2	CO_2
Experimental	$0.6237 < \bar{x} < 0.7560$	$0.6237 < \bar{x} < 0.7560$	$0.6237 < \bar{x} < 0.7560$
Substrate inhibition	$0.5996 < \bar{x} < 0.7232$	$0.6000 < \bar{x} < 0.7237$	$0.6283 < \bar{x} < 0.7579$
Monod	$0.5042 < \bar{x} < 0.6349$	$0.5046 < \bar{x} < 0.6353$	$0.5284 < \bar{x} < 0.6653$
Constant specific growth rate	$0.5265 < \bar{x} < 0.6643$	$0.5269 < \bar{x} < 0.6647$	$0.5518 < \bar{x} < 0.6961$

Table 4. *Fisher's exact test* for all model equations for each one of the three different assumptions for the specific growth rate (lower limit < S_{exp}^2/S_{mod}^2< upper limit).

	Cells growth	Bed temperature	Total reducing sugar
Substrate inhibition	0.6669 < 1.2184 < 1.5728	0.6669 < 0.7223 < 1.5728	0.6669 < 0.7464 < 1.5728
Monod	0.6769 < 1.0848 < 1.5375	0.6769 < 0.9759 < 1.5375	0.6769 < 0.6645 < 1.5375
Constant specific growth rate	0.6854 < 0.9729 < 1.5089	0.6854 < 1.1076 < 1.5089	0.6854 < 0.5960 < 1.5089
	Ethanol	O_2	CO_2
Substrate inhibition	0.6669 < 1.1803 < 1.5728	0.6669 < 1.1787 < 1.5728	0.6669 < 1.0748 < 1.5728
Monod	0.6769 < 1.0508 < 1.5375	0.6769 < 1.0494 < 1.5375	0.6769 < 0.9569 < 1.5375
Constant specific growth rate	0.6854 < 0.9425 < 1.5089	0.6854 < 0.9412 < 1.5089	0.6854 < 0.8583 < 1.5089

In Table 3 can be observed that both assumptions substrate inhibition and the constant specific growth rate have intercepted the experimental intervals for the *t-student's* test for all the model equations. Differently, the Monod model assumption has not intercepted the experimental intervals of the total reducing sugar and, thus, it cannot be considered equivalent to the process data, showing that the assumption is not good enough when compared to the other two. The *Fisher's exact test* results of Table 4 have shown that both Monod and the Constant specific growth rate assumptions were inconsistent for the total reducing sugar data, as the variance ratio is even smaller than the lower limit of the analysis. Finally, according to the results of Tables 2-4, the substrate inhibition assumption can be considered to be the best among the three models, with a very small objective function value, to represent the case study of the solid-state fermentation.

It shall be highlighted that the model validation with the estimated parameters was made considering one experiment left out of the parameter estimation procedure.

5.2. Comparison between the previous model and this work model assuming substrate inhibition

The previous model comprehends only the cells growth kinetics and the bioreactor bed temperature profile. Although this work model comprehends more process variables, the comparisons between the two models were performed only with these two process variables, as they are common to both models.

According to the results of Figure 8, both models represent likewise the cells growth kinetics. However, the previous model better fits both lag and log (exponential) phases while the model presented in this work better

fits the stationary phase. It could be explained by the assumption of the specific growth rate as a function of the substrate concentration.

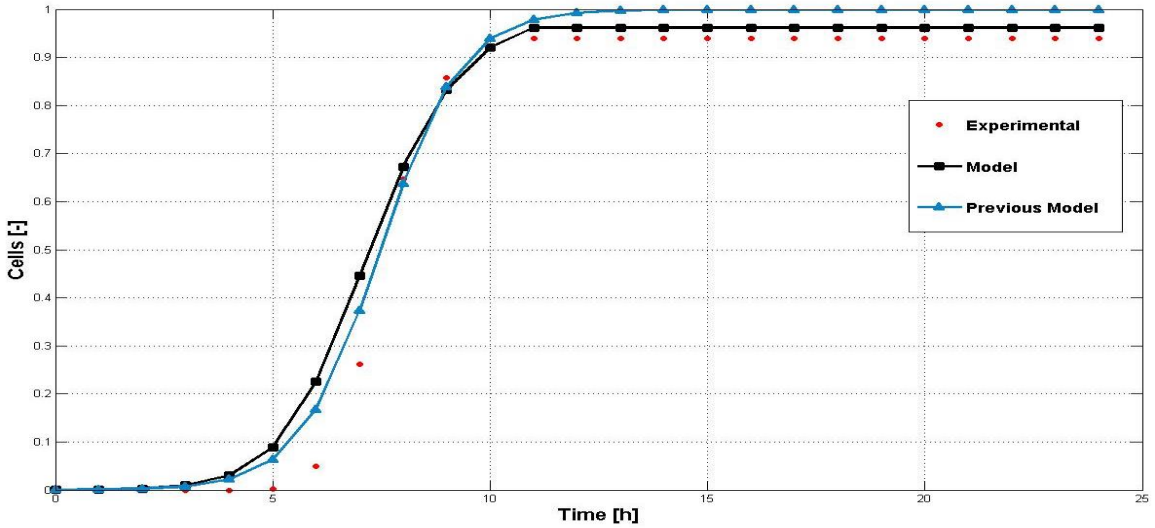

Figure 8. Comparison between the experimental and model data for the cells growth kinetics.

Both models present the same mismatch to predict the experimental data of bioreactor bed temperature profile. It can be observed in the results of Figure 9 that the bioreactor bed temperature profile of models data start to increase faster than the one of experimental data, exceeding a little the normalized boundary of 1. However, the model assuming substrate inhibition assumption has a dynamic quite more similar to the one presented by the experimental data. This fact probably results from the strong temperature dependence on the metabolic heat generated by the cells growth kinetics and due to lower value estimated to the metabolic heat coefficient using the this work model when compared to the one estimated using the previous work model.

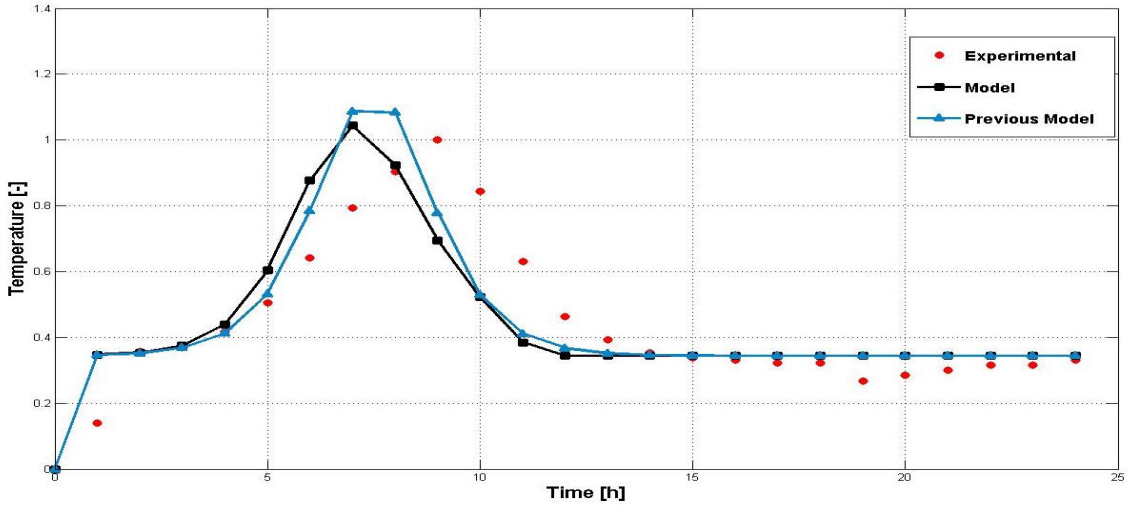

Figure 9. Comparison between the experimental and model data for the bioreactor bed temperature profile.

Although the Figures 8 and 9 have shown that both models can satisfactory represent the two considered process variables of the case study of this work, the model quality is better evaluated by using a performance measure. In this section, the sum of residuals between the experimental and model data was chosen as metric, so it can be easier to compare with Figures 8 and 9, and the results are illustrated in Table 5.

Table 5. Sum of the absolute residuals between the experimental and model data for both models.

	State	Cells growth		Bed temperature	
		Previous model	**Model**	**Previous model**	**Model**
Experiment	1	0.5879	0.9626	2.5104	2.5990
	2	0.8259	0.9609	1.8253	1.7861
	3	0.3574	0.8896	3.5750	3.5457
	4	0.5217	0.8497	2.0681	2.2027
	5	0.5817	0.5702	2.7749	2.8322
	6	1.5589	1.2392	3.7958	3.5375
	7	0.5082	0.8449	3.3976	3.4634
	Sum	4.9417	6.3172	19.9471	19.9670

According to the results of Table 5, the difference between the sum of the residuals between the experimental and model data of both models is quite small and could be neglected. However, the model presented in this work is more complex and, even with a bigger set of parameters, could represent the cells growth kinetics and bioreactor bed temperature profile as accurately as the previous work model. Although adding complexity to a model is not the ideal modeling practice, this model allows evaluating more important process variables such as the cells physiological state, the substrate consumption, the ethanol production and both carbon dioxide and oxygen produced by the cells metabolism (the previous model has not comprised these other variables). Therefore, the model assuming substrate inhibition will be used from here up to represent the solid-sate fermentation process studied in this work.

5.3. Remaining process variables simulation with the model assuming substrate inhibition

The following results are concerning to the remaining process variables, which were not present in the previous work. These simulations were also performed using the substrate inhibition assumption, which, as previously mentioned, was validated as the best assumption among the three ones compared to describe the process. The results of the model simulations for the remaining variables – substrate consumption, ethanol production, O_2 cells metabolism production and CO_2 cells metabolism production – can be seen in Figures 10-13, respectively.

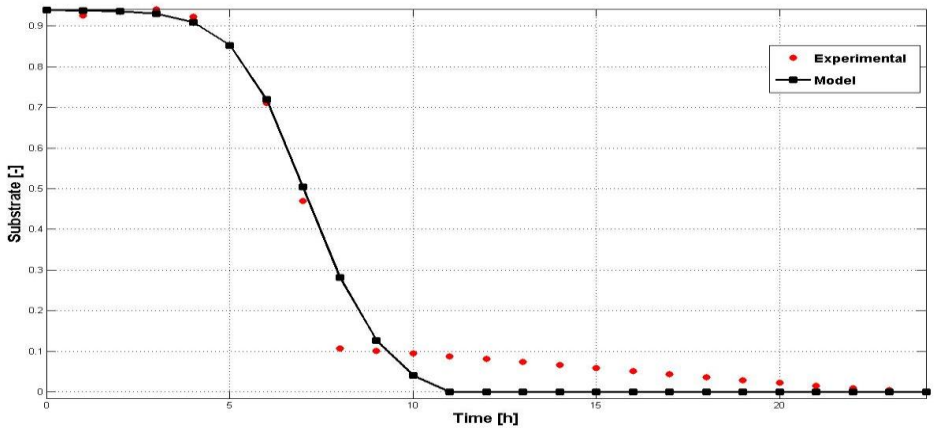

Figure 10. Comparison between the experimental and model data for the substrate consumption.

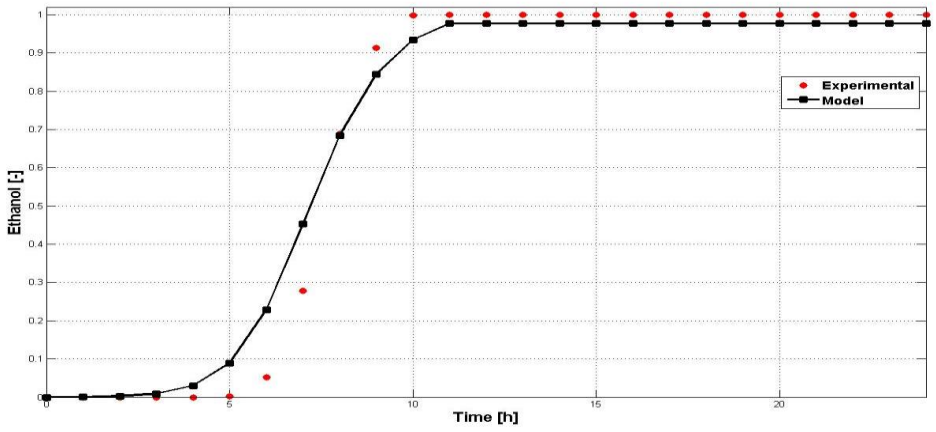

Figure 11. Comparison between the experimental and model data for the ethanol production.

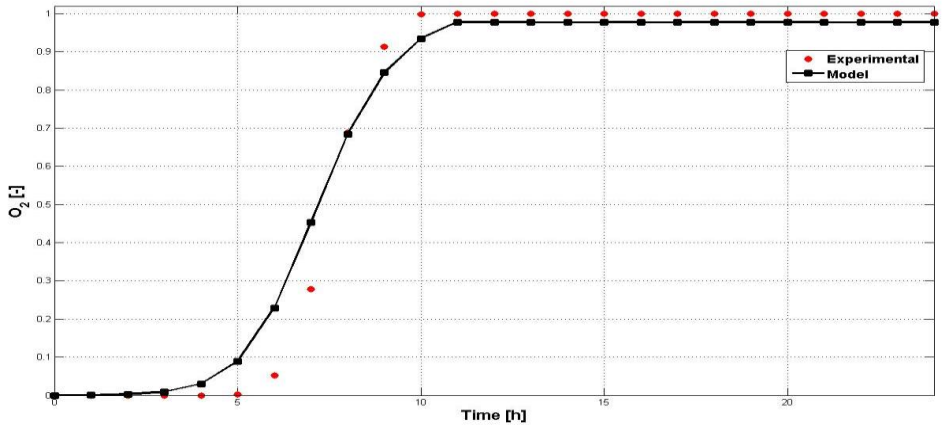

Figure 12. Comparison between the experimental and model data for the O_2 cells metabolism production.

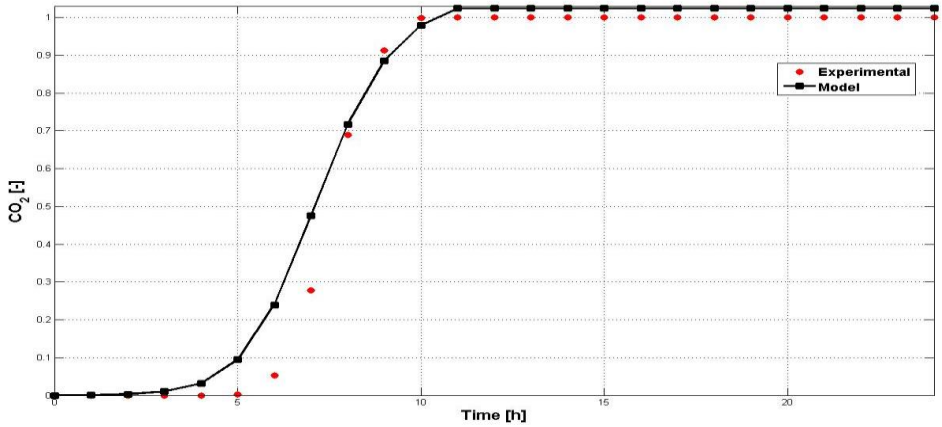

Figure 13. Comparison between the experimental and model data for the CO_2 cells metabolism production.

According to the results of the figures above, the model could satisfactorily predict the experimental data for the four remaining process variables: substrate consumption (Figure 10), ethanol production (Figure 11), O_2 cells metabolism production (Figure 12) and CO_2 cells metabolism production (Figure 13). A small model mismatch can be observed mainly in between 5 to 10 hours, period that corresponds exactly to the log (exponential) phase of cells growth kinetics.

6. Parameters Identifiability Analysis

After all the model evaluation and validation the Parameters Identifiability Analysis (PIA) was performed. Li *et al.*[25] suggests the removal of some parameters from the estimation procedure when they have weak effect on

the measured outputs or when the effects of the parameter on the measured output is linearly dependent. For problems of this nature, Secchi *et al.*[26] even proposed an algorithm (SELEST) to select and estimate only the identifiable parameters.

The presence of these weak effect parameters during the estimation procedure may cause degradation in the predictive capability of the model. In this manner, it is important to identify these parameters through a sensitivity matrix evaluation and reestimate the remaining model parameters, considering the substrate inhibition hypothesis, leaving the non-identifiable parameter outside the estimation procedure.

The magnitude of the parameter effect can be measured through eq 18.

$$\frac{dW_x(x,\theta)}{dt} = J_x(x,x)W_x(x,\theta) + J_p(x,\theta) \quad (18)$$

Where $W_x(x,\theta)$ represents the parameter-output sensitivity elements, $J_x(x,x)$ are the Jacobian derivatives of each state equation for each state variable, and $J_p(x,\theta)$ are the Jacobian derivatives of each state equation for each estimated parameter. The evaluation of this function results in the sensitivity matrix of the parameters, which will give their identifiability.

The resultant matrix, obtained from eq 18, is presented in Table 6. It can be seen that the evaluation of the parameters μ_{max}, K_s, k_1, $Y_{S/X}$, $Y_{P/X}$, $Y_{CO_2/X}$, $Y_{O_2/X}$, and Y_Q resulted in zero for all the state equations.

Table 6. Sensitivity matrix of all the parameters previously estimated for all the state equations.

	Cells	*Temperature*	*Substrate*	*Ethanol*	$\boldsymbol{\phi}$	O_2	CO_2
$\boldsymbol{\mu_{max}}$	0	0	0	0	0	0	0
$\boldsymbol{K_s}$	0	0	0	0	0	0	0
$\boldsymbol{k_1}$	0	0	0	0	0	0	0
$\boldsymbol{\varepsilon}$	0	-1.6600E-07	0	0	-4.394E-12	0	0
$\boldsymbol{\rho_s}$	0	4.5339E-08	0	0	1.200E-12	0	0
$\boldsymbol{C_{ps}}$	0	4.1981E-07	0	0	1.111E-11	0	0
$\boldsymbol{\rho_a}$	0	-1.5113E-10	0	0	-4.000E-15	0	0
$\boldsymbol{\alpha}$	0	0	0	0	7.343E-11	0	0
$Y_{x/s}$	0	0	0	0	0	0	0
$Y_{P/x}$	0	0	0	0	0	0	0
$Y_{x/O2}$	0	0	0	0	0	0	0
$Y_{x/CO2}$	0	0	0	0	0	0	0
Y_q	0	0	0	0	0	0	0
E_s	0	0	0	0	-1.3638E-5	0	0
E_d	0	0	0	0	0.02082	0	0

γ_{s0}	0	0	0	0	0.00652	0	0
γ_{d0}	0	0	0	0	-4.76E+36	0	0

In this manner, the referred parameters are non-identifiable for this process model, or at least for this amount of data. Their values will be held constant based on the values of the previous estimation, and all the other parameters whose sensitivity matrix elements were nonzero will be maintained for a new estimation. Thereby, the number of parameters to be estimated decreased from 17 to only 9.

The new parameters estimation results were very similar to the previous values obtained. Table 7 presents only the parameters that presented changes in their values.

Table 7. Identifiable parameters with changed values.

Parameter	Substrate inhibition	Units	Description
ε	0.9993	[ ]	Void fraction
ρ_s	269.9951	$kg.m^{-3}$	Substrate density
C_{ps}	2499.9572	$J.kg^{-1}.^oC^{-1}$	Heat capacity of substrate
ρ_a	0.9155	$Kg.m^{-3}$	Moist air density

The Least Squares Objective Function (OF), eq 17, was evaluated for both procedures – using all the parameters in the estimation and estimating only the identifiable parameters. The OF when all the parameters were used during estimation resulted in 11.2542, as already presented in Table 2; however, the OF when only the identifiable parameters were estimated has resulted in 10.3846.

This reduction in the OF, despite not being much significant, shows that those non-identifiable parameters were, indeed, depreciating the parameters estimation procedure. Also, their effect in the outputs were not so significant as well.

In order to represent the results more clearly, Figures 14-19 depicts both procedures, using all the parameters in the estimation and using only the identifiable parameters for cells growth and for the temperature profile.

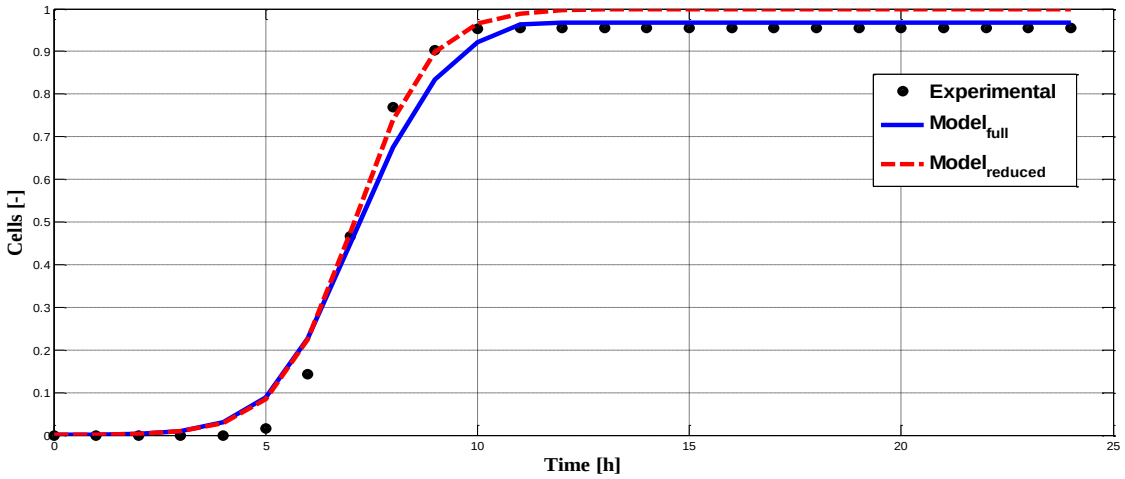

Figure 14. Cells growth comparison between the experimental data (●), all estimated parameters (solid line), and only identifiable parameters estimated (dashed line).

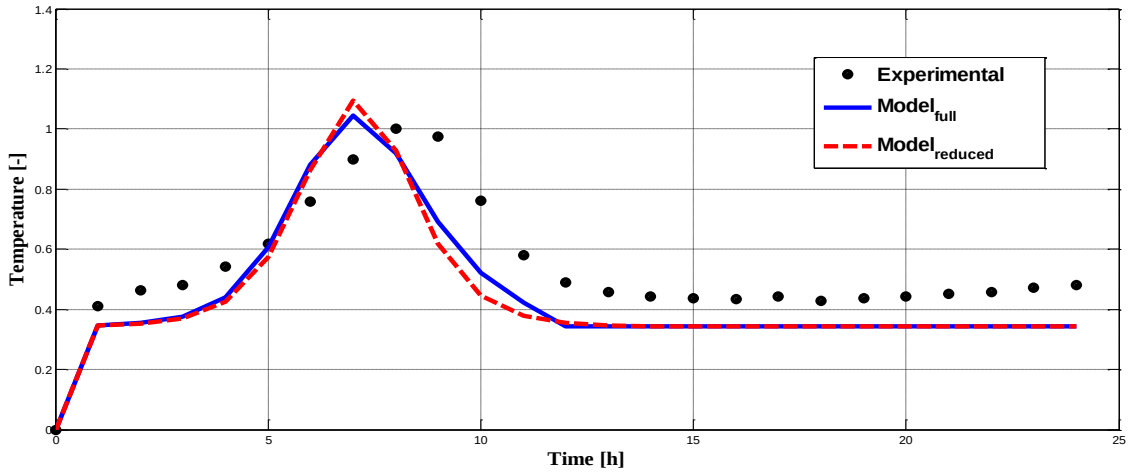

Figure 15. Temperature profile comparison between the experimental data (●), all estimated parameters (solid line), and only identifiable parameters estimated (dashed line).

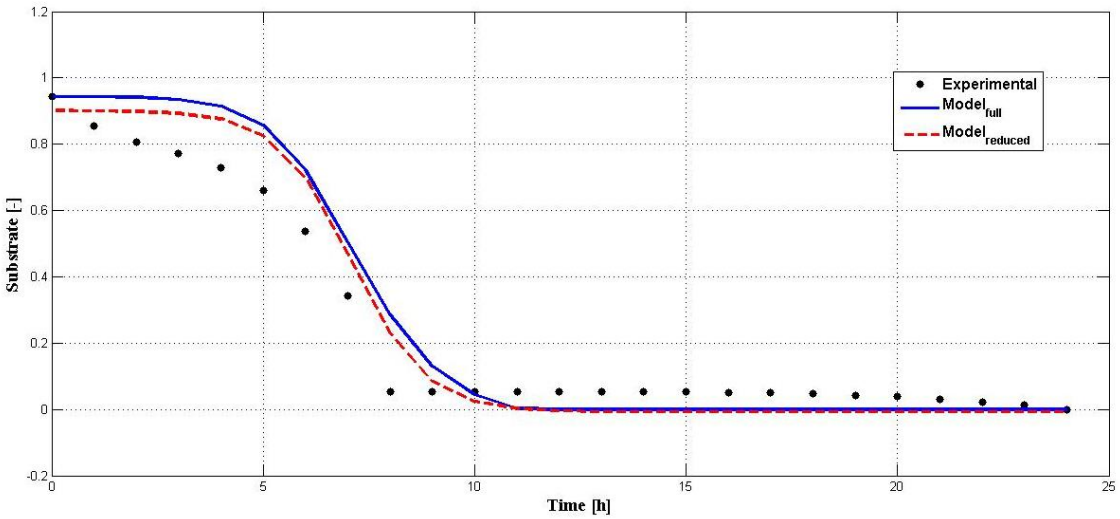

Figure 16. Substrate consumption comparison between the experimental data (●), all estimated parameters (solid line), and only identifiable parameters estimated (dashed line).

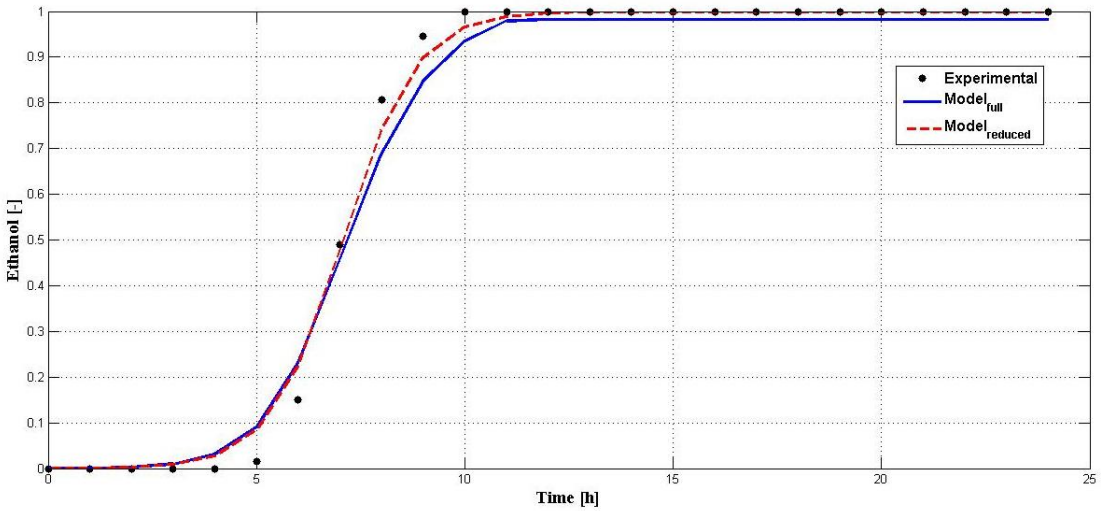

Figure 17. Ethanol production comparison between the experimental data (●), all estimated parameters (solid line), and only identifiable parameters estimated (dashed line).

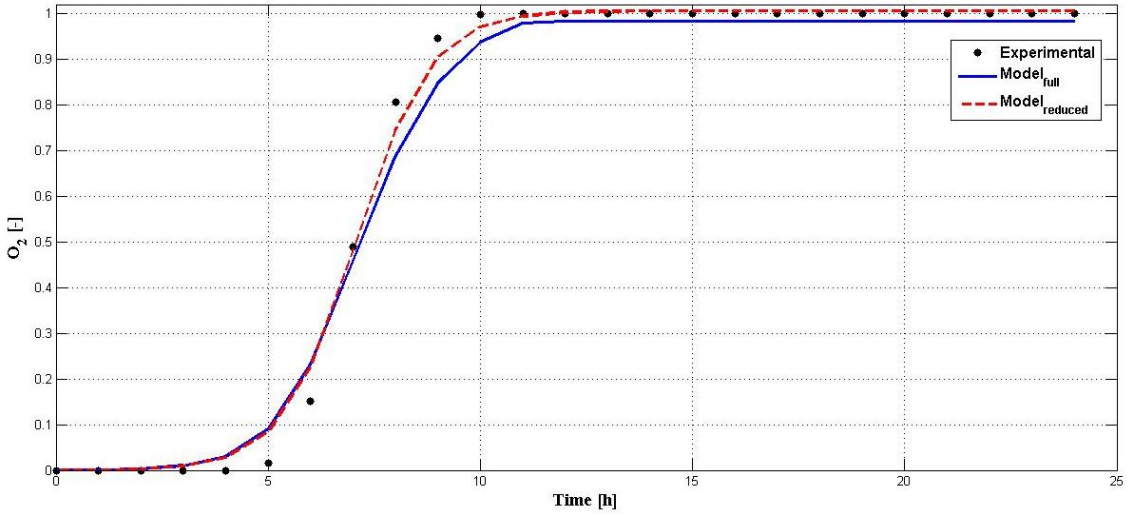

Figure 18. Oxygen yield comparison between the experimental data (●), all estimated parameters (solid line), and only identifiable parameters estimated (dashed line).

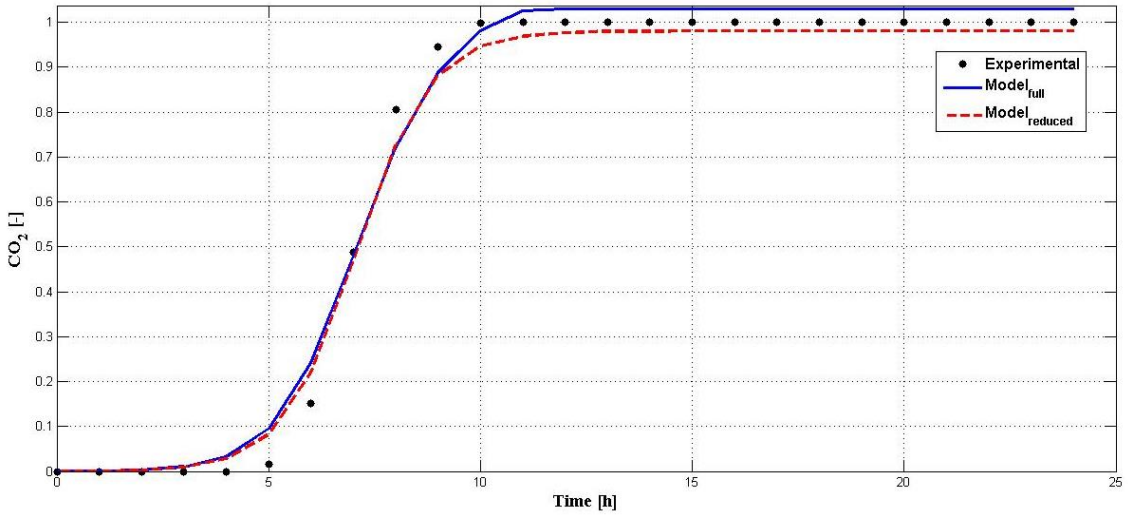

Figure 19. Carbon dioxide yield comparison between the experimental data (●), all estimated parameters (solid line), and only identifiable parameters estimated (dashed line).

7. Conclusions

In this work a model for a solid-state fermentation process was proposed. The model is composed by seven process variables. Among them, six process variables were experimentally measured (the cells growth kinetics, the bioreactor bed temperature profile, the substrate consumption, the ethanol production and both carbon dioxide and oxygen produced by the cells metabolism) and one process variable has accounted the physiological state of the cells. The model is composed by a set of six ordinary differential equations and one partial differential

equation to predict the mentioned process variables. This model is a more complete and complex form of a previous work model.

Firstly, three assumptions were compared for the specific growth rate in the model proposed in this work – substrate inhibition, Monod and constant value – and they were statistically analyzed. The *Student's t-test* for the Monod assumption has shown that the model failed to predict the total reducing sugar. According to the *Fisher's exact test*, Monod and constant specific growth rate assumptions lead the model to fail in the total reducing sugar prediction. Although the constant specific growth assumption has an equivalent mean to the experimental data, its variance is too large, and it fails the test, leading to conclude that the model using this assumption cannot represent the process. Among the three assumptions, the substrate inhibition assumption was considered the best one to describe the solid-state fermentation process, since it has succeeded in the *Student's t-test* and *Fisher's exact test* for all the process variables. In addition, the objective function value of parameter estimation has shown that the substrate inhibition assumption has presented much smaller residuals errors than the other two ones.

Afterwards, the model assuming substrate inhibition assumption was compared to the model presented in the previous work for both cells growth kinetics and bioreactor bed temperature profile, the only two process variables common to both models. The results have shown that this work model can satisfactory represent the dynamics of the two considered process variables of the case study of this work. Such results were strengthened by the small difference between the sum of the residuals between the experimental and model data, performance measure chosen to quantify the models accuracy, for both models. However, the model presented in this work has been preferred to represent the solid-state fermentation process studied because it is more complete, and even with a bigger set of parameters could represent the cells growth kinetics and bioreactor bed temperature profile as accurately as the previous model. Besides of allowing to evaluate more important process variables as aforementioned.

Later, comparisons between the experimental data and the model assuming substrate inhabitation data were performed for the four remaining measured process variables: substrate consumption, ethanol production, O_2 cells metabolism production and CO_2 cells metabolism production. The comparison results have emphasized that the proposed model can predict satisfactorily the experimental data even for the four remaining process variables, being the best model to represent the case study of the solid-state fermentation process.

Finally, a parameters identifiability analysis was performed in order to verify which were the most important parameters of the model and which were the weak effect parameters. Eight parameters were found to be non-indentifiable, therefore they were held constant while a new estimation was proceeded for the remaining

parameters. The model with this new estimation result was compared to both experimental data and substrate inhibition model. It was shown, either through the objective function results as through graphical analysis that the model with only identifiable parameters estimated is as good as the other, however, with less computational effort and less complexity.

APPENDIX A

Parameter	Constant value or Equation number	Unit	Description
X	Equation 1	g/L	Cells concentration
X_m	1.0	g/L	Maximum cells concentration
T	Equation 2	$^\circ$ C	Temperature
T_a	30.0	$^\circ$ C	Air entrance temperature
ρ_b	Equation 3	kg.m^{-3}	Bed density
C_{pb}	Equation 4	$\frac{\text{J}}{\text{kg}^\circ\text{C}}$	Heat capacity of bed
ϕ	Equation 5	Dimensionless	Physiological factor
γ_s	Equation 7	h^{-1}	Rate coefficient of the physiological factor synthesis
γ_d	Equation 8	h^{-1}	Rate coefficient of the physiological factor denaturation
S	Equation 9	g/L	Substrate concentration
P	Equation 10	g/L	Product concentration
CO_2	Equation 11	g/L	CO_2 concentration
O_2	Equation 12	g/L	O_2 concentration
C_{pa}	1180.0	$\frac{\text{J}}{\text{kg}^\circ\text{C}}$	Heat capacity of moist air [8]
Z	0.4	m	Evaluated bed height
λ	2414300.0	$\frac{\text{J}}{\text{kg}_{\text{H}_2\text{O}}}$	Latent heat of evaporation of water [8]
R	8.3144	$\frac{\text{J}}{\text{K mol}}$	Universal gas constant
V_Z	33.33	m/h	Moist air velocity
ƒ	0.00246	$\frac{\text{kg}_{\text{H}_2\text{O}}}{\text{kg}_{\text{air}}{}^\circ\text{C}}$	Water carrying capacity of air [8]
μ	Equations 13-14	h^{-1}	Growth rate
Ψ	Equation 15	Dimensionless	Symbolic normalized state
Y	Equation 15	---	Symbolic state

References

[1] IMAMOGLU, E., SUKAN, F. V. **Scale-up and kinetic modeling for bioethanol production**. Bioresource Technology, 144, 311-320 (2013).

[2] DODIĆ, J. M., VUČUROVIĆ, D. G., DODIĆ, S. N., GRAHOVAC, J. A., POPOV, S. D., NEDELJKOVIĆ, N. M. **Kinetic modelling of batch ethanol production from sugar beet raw juice**. Applied Energy, 99, 192-197 (2012).

[3] SALAU, N. P. G., NEUMANN, G. A., TRIERWEILER, J. O., SECCHI, A. R. **Dynamic behavior and control in an industrial fluidized-bed polymerization reactor**. Industrial & Engineering Chemistry Research, 47 (16), 6058-6069 (2008).

[4] SCHWAAB, M., BISCAIA, E. C., MONTEIRO, J. L. Jr., PINTO, J. C. **Nonlinear parameter estimation through particle swarm optimization**. Chemical Engineering Science, 63, 1542-1552 (2008).

[5] DRAPER, N. R., SMITH, H. **Applied regression analysis**. Wiley, New Work, 1998.

[6] HABIBI, A., VAHABZADEH, F., ZAIAT, M. **Dynamic mathematical models for biodegradation of formaldehyde by *Ralstonia eutropha* in a batch bioreactor**. Journal of Environmental Management, 129, 548-554 (2013).

[7] SILVEIRA, C. L. da, MAZUTTI, M. A., SALAU, N. P. G. **Modeling the microbial growth and temperature profile in a fixed-bed bioreactor**. Bioprocess and Biosystems Engineering, Published online (2014).

[8] FANAEI, M. A., VAZIRI, B. M. **Modeling of temperature gradients in packed-bed solid-state bioreactors**. Chemical Engineering and Processing: Process Intensification, 48(1), 446–451 (2009).

[9] VERHULST, P. F. **Notice sur la loi que la population poursuit dans son accroissement**. Correspondance mathématique et physique, 10,113-121 (1838).

[10] PANDEY, A. **Solid-state fermentation**. Biochemical Engineering journal, 13, 81-84 (2003).

[11] MAZUTTI, M. A., ZABOT, G., BONI, G., SKOVRONSKI, A., OLIVEIRA, D. de, LUCCIO, M. Di, RODRIGUES, M. I., TREICHEL, H., MAUGERI, F. **Kinetics of inulinase production by solid-state fermentation in a packed-bed bioreactor.** Food Chemistry, 120, 163-173 (2010).

[12] RAHARDJO, Y. S. P., TRAMPER, J., RINZEMA, A. **Modeling conversion and transport phenomena in solid-state fermentation: a review and perspectives**. Biotechnology Advances, 24, 161-179 (2006).

[13] HÖLKER, U., LENZ, J. **Solid-state fermentation – are there any biotechnological advantages?**. Current Opinion in Microbiology, 8 (3), 301-306 (2005).

[14] RAJAGOPALAN, S., MODAK, J. M. **Heat and mass transfer simulation studies for solid-state fermentation processes**. Chemical Engineering Science, 49(13), 2187-2193 (1994).

[15] LAGEMAAT, J. Van de, PYLE, D. L. **Modelling the uptake and growth kinetics of *Penicillinum glabrum* in a tannic acid-containing solid-state fermentation for tannase production**. Biotechnology and Biochemical engineering, 40, 1773-1782 (2005).

[16] BARANYI, J., ROBINSON, T. P., KALOTI, A., MACKEY, B. M. **Predicting growth of *Brochothrix thermosphacta* at changing temperature**. Food Microbiology, 27, 61-75 (1995).

[17] MITCHELL, D. A., von MEIEN, O. F., KRIEGER, N., DALSENTER, F. D. H. **A review of recent developments in modeling of microbial growth kinetics and intraparticle phenomena in solid-state fermentation**. Biochemical Engineering Journal, 17, 15-26 (2004).

[18] DALSENTER, F. D. H., VICCINI, G., BARGA, M. C., MITCHELL, D. A., KRIEGER, N. **A mathematical model describing the effect of temperature variations on the kinetics of microbial growth in solid-state culture**. Process Biochemistry, 40, 801-807 (2005).

[19] PRATA, D. M., SCHWAAB, M., LIMA, E. L. L., PINTO, J. C. **Nonlinear dynamics data reconciliation and parameter estimation through particle swarm optimization: Application for na industrial polypropylene reactor**. Chemical Engineering Science, 64(18), 3953-3967 (2009).

[20] LEVENBERG, K. **A Method for the solution of certain non-linear problems in least squares**. Quarterly applied mathematics, 2, 164-168 (1944).

[21] MARQUARDT, D. **An Algorithm for least-squares estimation of nonlinear parameters**. SIAM Journal on Applied Mathematics, 11 (2), 431-441 (1963).

[22] MORÉ, J. J. **The Levenberg-Marquardt Algorithm: implementation and theory**. Numerical Analysis, Ed. G. A. Watson, Lectures Notes in Mathematics 630, Springer Verlag, 105-116 (1977).

[23] DORMAND, J. R., PRINCE, P. J. **A family of embedded Runge-Kutta formulae**. Journal of Computational and Applied Mathematics, 6(1), 19-26 (1980).

[24] SCHWAAB, M.; PINTO, J. C. **Experimental data analysis I. Fundamentals of statistics and parameter estimation**. E-papers, Rio de Janeiro – RJ (2007).

[25] LI, R. J.; HENSON, M. A.; KURTZ, M. J. **Selection of model parameters for offline parameter estimation**. IEEE Transactions on Control Systems Technology, (12), 402–412 (2004).

[26] SECCHI, A. R.; CARDOZO, N. S. M.; NETO, E. A.; FINKLER, T. F. **An algorithm for automatic selection and estimation of model parameters**. International Symposium on Advanced Control of Chemical Processes (ADCHEM 2006), Gramado, Brazil (2006).

6 SOLID-STATE FERMENTATION PROCESS MODEL REPARAMETRIZATION PROCEDURE FOR PARAMETERS ESTIMATION USING PARTICLE SWARM OPTIMIZATION

Accepted for publication in: Journal of Chemical Technology & Biotechnology, 2015.

Christian L. da Silveira, Marcio A. Mazutti, Nina P. G. Salau*

Chemical Engineering Department, Universidade Federal de Santa Maria

Abstract

BACKGROUND: Solid-state fermentation is a well-known bioprocess. The development of a solid-state fermentation reactor faces difficulties in control temperature gradients inside the bed. To understand this behavior, a good phenomenological model must be used. Furthermore, the model parameters must be obtained by a reliable and robust parameters estimation procedure, since often the model parameters are not available in literature.

RESULTS: The heuristic Particle Swarm Optimization (PSO) method was used to estimate the parameters of the reparametrized model through the minimization of a nonlinear-square objective function, since it is less sensitive to initial guesses than derivative based methods. Statistical analyses were made to evaluate the parameter estimation quality. Also, simulations with the reparametrized model were performed. All the analyses have revealed the great accuracy of the proposed model, predicting with very small residuals (errors) several experimental data set of a solid-state fermentation process.

CONCLUSION: Parameter models reparametrization procedure can be a very useful mathematical tool to obtain accurate and reliable models with fewer parameters to be estimated, providing less parameters correlation and increasing the degrees of freedom to perform the statistical analysis. In addition, the Particle-Swarm Optimization was found to be a very robust and efficient parameter estimation method.

Keywords: Solid-State Fermentation, Bioreactor, Process Modeling, Parameter Estimation, Particle-Swarm Optimization.

* To whom all correspondence should be addressed. E-mail: ninasalau@ufsm.br

Address: Chemical Engineering Department, UFSM – Av. Roraima, 1000, Cidade Universitária - Bairro Camobi. 97105-900 Santa Maria, RS – Brazil.

Phone: +55-55-3220-8448- Fax: +55-55-3220-8030.

1. Introduction

Solid-state fermentation (SSF) is defined by Pandey[1] as a type of fermentation that involves just solids with no free water, or at least a very few quantity of free water, and substrate must have only enough moisture to enable microorganisms to grow; in this case the microorganism used was a yeast breed *Kluyveromyces marxianus*.

SSF processes offer some opportunities, once they can be taken with agricultural wastes as substrate and their costs with medium tend to be lower. Also, the products obtained by SSF are concentrated, making easier the products recovering process in comparison with the submerged fermentation[2]. Farinas *et al*.[3] report that enzymes produced by SSF have greater stability against changes in temperature and pH, and their susceptibility to substrate inhibition is decreased.

However, SSF is still facing some issues in scaling-up. Problems like excessive heating of bioreactor are still hindering the conduction of the process in large-scale. The microbial growth produces huge amount of heat, creating a process control problem hard to be overcome[4]. Lenz *et al*. report that the process simulations and computer models will enable the SSF scale-up, with parameters optimization and advanced process control[5].

Good assumptions are necessary to produce a model that predicts the experimental data, or process behavior, with great fidelity[6]. Also, optimum parameters are essential to maintain the model accuracy in predicting experimental data. In this manner, parameters are to be estimated through an objective function minimization, which will decrease errors between experimental and simulated data[7-10].

As objective function to be minimized in the parameter estimation technique, the nonlinear least squares was chosen. The minimization of this objective function was made using the Particle Swarm Optimization (PSO) method. The PSO consists in a heuristic optimization algorithm, which aims to reduce an objective function value by the parameters optimization. The PSO method is particularly advantageous when the parameters to be estimated are unknown, since it is not very sensitive to initial guesses as the derivative based methods[8].

However, high correlations (a strong dependence between the parameters) among the parameters to be estimated may cause numerical problems, difficult the estimation procedure, and lower the computational efficiency[11]. The reparametrization, leading to a more linear model, can reduce the computational effort and improve the robustness of the least-squares objective function computing[12]. Schwaab and Pinto[13] proposed a model reparametrization to minimize the correlation among parameter estimates for power function models, and exemplified with applications in problems involving kinetic constants[14], and problems involving multiple

reparametrizations[15]. In this work, to overcome any high correlation problem, a reparametrization approach is used as well.

In this manner, this work proposes a reparametrization of the SSF model presented by Silveira *et al*.[10] to compare with the non-reparametrized model and verify if the reparametrization procedure can improve the accuracy of the predictions. The parameter estimation procedure presented in this work is based on a nonlinear least squares objective function minimization using a PSO algorithm implemented in Fortran® 95. The reparametrized model is validated with statistical and residual analysis.

2. Solid-state fermentation modeling

Microbial growth on solid-state fermentation process has been descripted by the empiric model developed by Verhulst in 1838[16]. By that time, the Belgian mathematician Verhulst developed the logistic equation to describe his idea that population exponential growth could not last forever, and it would face some difficulties, like food and space limitations, reaching a so-called population steady state[17]. Verhulst model has been used and adjusted with success to microorganisms growth experimental data in solid-state fermentations[18-20], and it is represented by eq 1.

$$\frac{dX}{dt} = \mu X\left(1 - \frac{X}{X_m}\right) \tag{1}$$

Where μ represents the cells specific growth rate and it is described with the assumption of substrate inhibition, as can be seen on eq 2. The substrate inhibition assumption was adopted based on our previous work[10], in which three different assumptions for the specific growth rate were tested: a) substrate inhibition, b) Monod and c) constant value. The results of the work based on statistical and residual analysis have shown that the substrate inhibition assumption was the one that best fitted to the experimental data.

$$\mu = \frac{\mu_{max} S}{K_S + S + k_1 S^2} \tag{2}$$

As used in our previous work[7], the energy balance that gives the equation for the bioreactor bed temperature is represented by eq 3. The model for bioreactor bed temperature was based on the one presented by Fanaei and Vaziri[21], which takes into account specific heats and densities of the bioreactor bed, of the substrate,

and of the air moisture, void fraction of substrate, metabolic heat due microorganism growth, latent heat of water evaporation, and water carrying capacity of air, through time.

$$\frac{\partial T}{\partial t} = \frac{\rho_s(1-\varepsilon)Y_Q\frac{dX}{dt}+\rho_a C_{pa}V_z\left(\frac{\partial T}{\partial z}\right)+\rho_a f\lambda V_z\left(\frac{\partial T}{\partial z}\right)}{\rho_b C_{pb}} \quad (3)$$

For the equation of the bioreactor bed temperature, two algebraic equations are used to obtain the bed density and specific heat:

$$\rho_b = \varepsilon\rho_a + (1-\varepsilon)\,\rho_s \quad (4)$$

$$C_{pb} = \frac{\varepsilon\rho_a\left(C_{pa}+f\lambda\right)+(1-\varepsilon)\rho_s C_{ps}}{\rho_b} \quad (5)$$

Also, substrate consumption (total reducing sugar - TRS), *S*, and production of ethanol, *P*, carbon dioxide, CO_2, and oxygen, O_2, are descripted by eqs 6-9. The carbon dioxide and oxygen production referred by eqs 8-9 are concerning to the carbon dioxide and oxygen produced by cells metabolism.

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}}\frac{dX}{dt} \quad (6)$$

$$\frac{dP}{dt} = Y_{P/X}\frac{dX}{dt} \quad (7)$$

$$\frac{dCO_2}{dt} = \frac{1}{Y_{CO_2/X}}\frac{dX}{dt} \quad (8)$$

$$\frac{dO_2}{dt} = \frac{1}{Y_{O_2/X}}\frac{dX}{dt} \quad (9)$$

As can be seen from eqs 6-9, all these variables (substrate consumption, ethanol production, and CO_2 and O_2 produced by cells metabolism) are strongly dependent on the cells growth kinetics. Thus, the cells growth kinetics must be as accurate as possible to guarantee the substrate, ethanol, CO2, and O2 goodness of fit.

The proposed model has a relatively large amount of parameters to be estimated. However, one of the greatest problem is that some of these parameters are of the order, for instance, of 1×10^{-1} while others are of 1×10^{6}. Magnitude differences can cause problems during estimation, interfering in the precision of the method or, even worse, preventing the parameter estimation. Also, the high correlation of parameters can difficult the estimation, by making its computational time become too large or even by avoiding the solution to be found.

For this reason the reparametrization procedure was used, in order to keep the high parameters magnitude normalized with the others used in this work, to avoid any high correlation between the parameters, and to prevent numerical problems that the PSO algorithm may encounter. In addition, the procedure took in account old parameters values estimated in the previous work as initial guesses[7,10].

Thus, the greatest problem with parameters magnitude relied on the bioreactor bed temperature equation, which had the widest band of parameters scale in the model. To reduce these magnitudes differences, some steps were taken in order to normalize those parameters.

Firstly, the specific heat was substituted into eq 3 and simplified with the bed density, as it can be seen from eqs 10 to 11.

$$\frac{\partial T}{\partial t} = \frac{\rho_s(1-\varepsilon)Y_Q\frac{dX}{dt}+\rho_a C_{pa}V_z\left(\frac{\partial T}{\partial z}\right)+\rho_a f\lambda V_z\left(\frac{\partial T}{\partial z}\right)}{\rho_b\frac{\varepsilon\rho_a(C_{pa}+f\lambda)+(1-\varepsilon)\rho_s C_{ps}}{\rho_b}} \tag{10}$$

$$\frac{\partial T}{\partial t} = \frac{\rho_s(1-\varepsilon)Y_Q\frac{dX}{dt}+\rho_a C_{pa}V_z\left(\frac{\partial T}{\partial z}\right)+\rho_a f\lambda V_z\left(\frac{\partial T}{\partial z}\right)}{\varepsilon\rho_a(C_{pa}+f\lambda)+(1-\varepsilon)\rho_s C_{ps}} \tag{11}$$

Eq 11 can be rewritten as the following:

$$\frac{\partial T}{\partial t} = \frac{\rho_s Y_Q\frac{dX}{dt}-\varepsilon\rho_s Y_Q\frac{dX}{dt}+\rho_a C_{pa}V_z\left(\frac{\partial T}{\partial z}\right)+\rho_a f\lambda V_z\left(\frac{\partial T}{\partial z}\right)}{\varepsilon\rho_a C_{pa}+\varepsilon\rho_a f\lambda+\rho_s C_{ps}-\varepsilon\rho_s C_{ps}} \tag{12}$$

In addition, rearranging eq 12 so that all parameters become normalized:

$$\frac{\partial T}{\partial t} = \frac{\frac{dX}{dt} - \varepsilon\frac{dX}{dt} + \frac{\rho_a C_{pa}}{\rho_s Y_Q} V_z\left(\frac{\partial T}{\partial z}\right) + \frac{\rho_a f\lambda}{\rho_s Y_Q} V_z\left(\frac{\partial T}{\partial z}\right)}{\frac{\varepsilon\rho_a C_{pa}}{\rho_s Y_Q} + \frac{\varepsilon\rho_a f\lambda}{\rho_s Y_Q} + \frac{\rho_s C_{ps}}{\rho_s Y_Q} - \frac{\varepsilon\rho_s C_{ps}}{\rho_s Y_Q}} \quad (14)$$

From eq 14 two choices can be made. The first one considers the previous work estimations[7,10], which leads to the use of the void fraction (ε) parameter as 0.9995, that is almost 1.0. If this approach is chosen, the terms that differ from another only by the presence of ε can be considered almost equal, leading to the eq 15.

$$\frac{\partial T}{\partial t} = \frac{\varphi_1 V_z\left(\frac{\partial T}{\partial z}\right) + \varphi_2 V_z\left(\frac{\partial T}{\partial z}\right)}{\varphi_1 + \varphi_2} \quad (15)$$

Despite of the simplicity, eq 15 should not be used since it cannot account the heat generated by the cells growth. Hence, the second option is to gather all the parameters, which are of the same dimension:

$$\frac{\partial T}{\partial t} = \frac{P_1\frac{dX}{dt} + P_2 V_z\left(\frac{\partial T}{\partial z}\right) + P_3 V_z\left(\frac{\partial T}{\partial z}\right)}{P_4 + P_5 + P_6 - P_7} \quad (16)$$

$$\frac{\partial T}{\partial t} = \frac{\theta_1\frac{dX}{dt} + \theta_2 V_z\left(\frac{\partial T}{\partial z}\right)}{\theta_3} \quad (17)$$

This sum of the parameters with the same dimension reduces the number of parameter to be estimated, increasing the degrees of freedom, and requiring less computational effort. Thus, the new set of parameters referred in eq 17 are:

$$\theta_1 = (1 - \varepsilon) \quad (18)$$

$$\theta_2 = P_2 + P_3 = \left[\frac{\rho_a C_{pa}}{\rho_s Y_Q}\right] + \left[\frac{\rho_a f\lambda}{\rho_s Y_Q}\right] \quad (19)$$

$$\theta_3 = P_4 + P_5 + P_6 - P_7 = \frac{\varepsilon\rho_a C_{pa}}{\rho_s Y_Q} + \frac{\varepsilon\rho_a f\lambda}{\rho_s Y_Q} + \frac{\rho_s C_{ps}}{\rho_s Y_Q} - \frac{\varepsilon\rho_s C_{ps}}{\rho_s Y_Q} \quad (20)$$

By the reparametrization the number of parameters to be estimated for the differential equation of the temperature profile decrease from 5 to 3. Such simplification avoids an overparametrization of the problem and increases the degrees of freedom for statistical tests that must be performed to validate the model. Also, this

reparametrization prevents the great magnitude differences and decreases the high correlations between the model parameters.

All the experimental data of the process variables was normalized, keeping the results between 0 and 1. The normalization procedure is described by eq 21.

$$\Psi = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \tag{21}$$

3. Parameter Estimation Problem

Parameters estimated are uncertain to some degree[8], and they reach their optimum value when the objective function is minimized, since this objective function computes the differences between experimental and simulated data. Prata et al.[22] have reported that the estimation problem includes data regression and parameters adjustment to that data, satisfying some constraint and boundaries, to reach the minimization of a certain objective function.

The objective function used in this work is the nonlinear least squares function, which assembles the squared differences between the simulated data and the experimental data for each state variable. The objective function is represented by eq 22, where the subscripts *exp* and *mod* represents experimental and model data, respectively.

$$\begin{bmatrix} f(X) \\ f(T) \\ \min\ f(S) \\ f(P) \\ f(CO_2) \\ f(O_2) \end{bmatrix} = \begin{bmatrix} \left\| X^i_{exp} - X^i_{mod} \right\|^2 \\ \left\| T^i_{exp} - T^i_{mod} \right\|^2 \\ \left\| S^i_{exp} - S^i_{mod} \right\|^2 \\ \left\| P^i_{exp} - P^i_{mod} \right\|^2 \\ \left\| CO^i_{2\,exp} - CO^i_{2\,mod} \right\|^2 \\ \left\| O^i_{2\,exp} - O^i_{2\,mod} \right\|^2 \end{bmatrix}_{\substack{i=0,\dots,NY-1 \\ j=1,\dots,NE}} =$$

$$\sum_{j=1}^{NE} \sum_{i=0}^{NY-1} F_{obj}[X(j,i), T(j,i), S(j,i), P(j,i), CO_2(j,i), O_2(j,i)] \tag{22}$$

Experimental data for this solid-state fermentation process was obtained by Mazutti *et al.*[23]. A list of model parameters with their units and values can be found in Table 1.

Degree of freedom is an important value for the statistical validation of the model, and its value can be obtained by the use of eq 23, where *NP* is the number of parameters, *NE* the number of experiments, *NY* the number of state variables measurements, and *DF* the degree of freedom.

$$DF = NE.NY - NE.NP \quad (23)$$

Before the reparametrization procedure, there were 12 (*NP*=12) parameters to be estimated, and considering 6 experiments (*NE*=6) measured at 25 different times (*NY*=25), the degree of freedom value would be 78. However, with the number of parameters reduction to 10 (*NP*=10), the new degree of freedom is 90, which can favor the validation tests.

The algorithm chosen for this parameter estimation problem was the Particle Swarm Optimization (PSO), whose routine is implemented in Fortran® 95. An algorithm of the PSO implementation can be found in Appendix B. The method is relatively simple to be implemented, and its computational time and effort is usually low.

The PSO was firstly proposed in 1995 by Kennedy and Eberhart[24] to be used in animals behavior prediction. The PSO algorithm randomly positions a determined number of particles within a defined space (boundaries) in which the optimum parameter value must be contained. Using the value at those randomly positioned particles, the objective function is evaluated (exploration stage). In the next iterations, particles move towards a new position, taking in account their own and the others particles history, the previous iterations best fitness are evaluated, and added with a random term. Each iteration only takes place after all particles have been moved and evaluated, and at each iteration more information about the region is gathered, diminishing the randomness of the particles towards the minimum of the objective function[8,25]. As iterations proceed and this random character diminish, the particles concentrate around the more promising regions (exploitation stage), performing a local search which will improve the solution. According to Schwaab *et al.*[8], a proper balance between the exploration and exploitation stages is fundamental. If the exploration stage is emphasized, convergence may not be reached; however, if the exploitation stage is emphasized, the convergence may lead to a region far from the global minimum.

Although the high number of objective function evaluations in PSO, it becomes advantageous since it is not as sensitive to the initial guesses as derivative based methods, since it is assumed that the parameters are unknown.

4. Material and Methods

More details about substrate origins and composition, as well as the treatment of the medium and analysis can be found in the work of Mazutti *et al*.[23]. Briefly, for medium composition a sugarcane bagasse supplemented with 10.0 wt% of pretreated cane molasses, 30.0wt% of corn steep liquor and 20.0 wt% of soybean bran were used. The strain of yeast used was the *Kluyveromyces marxianus* NRRL Y-7571, which was maintained in an YM agar medium. Cell production for pre-inoculum was carried out in test tubes of 50 ml with 10 ml of liquid YM medium. Cell mass contained in pre-inoculum was determined by direct weighing before its use in fermentation[7].

The bioreactor used for the solid-state fermentation was a packed-bed cylinder, with 0.0363 m³, of stainless steel with sugarcane compost medium. An air supplier was attached to the bottom of the bioreactor, and the air supplied was 95-100% saturated, flowing through the height of the bed, as can be seen on Figure 1. Temperatures at inlet and outlet of the bioreactor were continuously monitored by a PT100 (Novus, Brazil). Considering the oxygen uptake rate, the microbial growth was calculated[7,23].

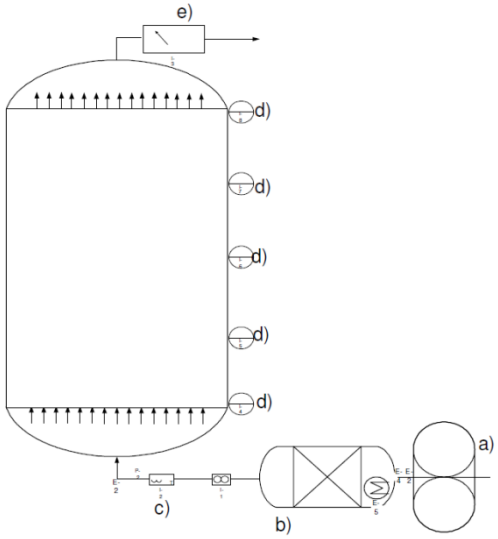

Figure 1. Diagram of the packed bed bioreactor used: a- compressor; b- humidifier; c- flowrate, temperature and humidity control; d- temperature sensors; e- CO_2, temperature and humidity measurer.

All experiments were carried out by 24 hours. Data acquisition of cells mass growth, substrate consumption, ethanol production, and CO_2 and O_2 produced by the cells metabolism was made hour by hour. Temperatures along the bed height were measured at the inlet, 10, 20, 30 cm and at the outlet (letter d at Figure 1) of the bioreactor.

The computer used for the parameters estimation has an Intel® Core™ i7-3770 with 3.40 GHz processor and 12 Gb of RAM memory, running with the Windows 7 64 bits Operating System.

5. Model validation, results and discussion

Eqs 1, 2, 6-9 and 17 describe the reparametrized model used in this paper. The estimated parameters can be seen in Table 1, with their descriptions and confidence intervals. The estimation was performed using a Particle Swarm Optimization algorithm, implemented in Fortran® 95. In this routine, the SSF experimental data was used, and the code is programmed to read, for each experiment, its different initial condition.

The space in which the estimator searched for parameters values to minimize the objective function was, initially, very wide. At each minimization, an analysis of parameters regions was made, so the intervals boundaries could be straightened, decreasing the time taken for the estimation.

No significant difference in results was obtained with the increasing of too many particles in the Particle Swarm Optimization algorithm; on the other hand, more time was required for the method to converge. Therefore, the number of particles was fixed as 50 particles and the iterations number fixed to 10,000 iterations. For each estimation, three minimizations were performed to assure more confidence that the global minimum was reached.

Table 1. Parameters estimated with intervals of 95% of confidence.

Parameter	Value	Unit	Description
μ_{max}	1.1434 ∓ 0.1021	h^{-1}	Maximum growth rate
K_s	$5 \times 10^{-5} \mp 0.0020$	g/L	Half-velocity constant
k_1	$-6.84 \times 10^{-4} \mp 0.0052$	L/g	Dissociation constant
θ_1	0.8738 ∓ 0.0384	Dimensionless	Normalized parameter
θ_2	$0.0032 \mp 6.1 \times 10^{-5}$	$°C^{-1}$	Normalized parameter
θ_3	0.3548 ∓ 0.0323	$°C^{-1}$	Normalized parameter
$Y_{x/s}$	1.0877 ∓ 0.0196	g/g	Substrate to cells yield coefficient
$Y_{P/x}$	1.0048 ∓ 0.0032	g/g	Cells to product yield coefficient
$Y_{x/O2}$	0.9953 ∓ 0.0084	g/g	Cells to [O_2] yield coefficient
$Y_{x/CO2}$	0.9951 ∓ 0.0083	g/g	Cells to [CO_2] yield coefficient

Confidence intervals presented in Table 1 shows, however, that parameters K_s and k_1 crosses the zero. This may imply in none or little significance of these parameters, consequently, there would be no substrate inhibition and cells growth kinetics would depend only upon a constant specific growth rate.

After the parameters estimation procedure, the obtained parameters were used for model simulation, and the modeling data was graphically compared to the experimental data in Matlab®. The simulation, using the parameters obtained from the PSO algorithm programmed in Fortran® 95, was made using the *ode45* Matlab integrator, which is based on an adaptive explicit Runge-Kutta of fourth and fifth orders, using a Dormand-Prince pair[7,26].

Several simulations were performed and compared to each set of experimental data to verify its goodness. In the Supplementary Data section are placed some of these simulations and residual graphical analysis. In order to overcome the graphical analysis, the *Student's t-test* was performed to verify if the means of the process variables predicted by the model were correspondent to the experimental means; likewise, the *Fisher's exact test* was performed to verify if variances of the process variables were also correspondent to experimental variances.

Table 2 presents *Student's t-test* results, where $\bar{x}$ is the mean value of the data. It can be seen that intervals of experimental and model data, for all process variables, intercept each other, meaning that the means of the experimental data for each process variable for all experiments are equivalent to those found through the model simulated data. In addition, intervals are very alike, leading to conclusion that the experimental and model data are very similar.

Table 2. *Student's t-test* for all states ($\bar{x}$ is the mean of each state).

	Cells	Temperature	Substrate
Experimental	$0.6237 < \bar{x} < 0.7560$	$0.4702 < \bar{x} < 0.5399$	$0.1932 < \bar{x} < 0.2943$
Model	$0.6320 < \bar{x} < 0.7606$	$0.4107 < \bar{x} < 0.4838$	$0.2019 < \bar{x} < 0.3201$
	Product	O_2	CO_2
Experimental	$0.6237 < \bar{x} < 0.7560$	$0.6237 < \bar{x} < 0.7560$	$0.6237 < \bar{x} < 0.7560$
Model	$0.6348 < \bar{x} < 0.7639$	$0.6347 < \bar{x} < 0.7639$	$0.6348 < \bar{x} < 0.7640$

Fisher's exact test was performed to verify if the model variances are also similar to experimental data variances. Table 3 presents the test results, where all variances ratios are contained into the intervals, meaning that model variances are also equivalent to experimental variances.

Table 3. *Fisher's exact test* for all states (lower limit < S^2_{exp}/S^2_{mod}< upper limit).

Cells	Temperature	Substrate
0.7144 < 1.0697 < 1.4207	0.7144 < 0.9161 < 1.4207	0.7144 < 0.7391 < 1.4207
Product	O_2	CO_2
0.7144 < 1.0594 < 1.4207	0.7144 < 1.0597 < 1.4207	0.7144 < 1.0593 < 1.4207

Therefore, the statistical analysis supports that the reparametrized model is reliable for predictions of the solid-state fermentation process. Some of the residuals plots can be found in the Supplementary Data section, where as closest the residuals are from the straight line the better the model.

Also, nonlinear least squares objective function values were computed for each process variable for each experiment in order to visualize where the greatest errors were located. Results can be seen on Table 4.

Table 4. Nonlinear Least Squares objective function value for each state at each experiment.

Experiment	**Cells**	**Temperature**	**Substrate**	**Product**	**[O_2]**	**[CO_2]**	**Experiment sum**
1	0.1688	0.1723	0.2154	0.1726	0.1725	0.1726	1.0742
2	0.0438	0.1241	0.2791	0.0417	0.0417	0.0417	0.5722
3	0.0401	0.5527	0.3103	0.0412	0.0412	0.0412	1.0268
4	0.0789	0.0966	0.0874	0.0801	0.0800	0.0801	0.5030
5	0.0190	0.3046	0.1958	0.0181	0.0182	0.0181	0.5738
6	0.3238	0.8885	0.3436	0.3192	0.3193	0.3191	2.5135
7	0.1436	0.2808	0.7391	0.1473	0.1472	0.1474	1.6055
State sum	0.8180	2.4196	2.1707	0.8202	0.8201	0.8203	

The results have shown that most of the errors comes from the temperature profile and substrate consumption data. In order to verify whether the reparametrization procedure have given or not any improvement in the model fidelity, the PSO method was also applied to the model in its normal form, as expresses by eq 3. A total number of 12 parameters were now estimated. The computational effort to estimate the 12 non-normalized parameters has become considerably larger, taking much more time to finish and, sometimes, not even converging.

The total sum of objective function of all process variables and all experiments for the reparametrized model provides a value of 7.8689. This objective function value is lower than the value reached in the non-reparametrized model, also using PSO method, which was 8.5596.

6. Conclusion

This work proposed a reparametrized model of a solid-state fermentation based on a previous work model[10] to apply a Particle Swarm Optimization (PSO) method to estimate the parameters. The reparametrization was done aiming to reduce the parameters correlation and increase the efficiency of the parameter estimation procedure, as the PSO was required to evaluate less times the objective functions in the estimation procedure. Also, as the number of parameters to be estimated is decreased, the degrees of freedom increase, and this is important for the model validation through statistical analysis.

The heuristic PSO method was used to estimate the parameters of the reparametrized model through the minimization of the objective function, since it is less sensitive to initial guesses than derivative based methods. The PSO routine was implemented in Fortran® 95. No significant changes in the results were obtained with increasing the particles number of the algorithm. However, a high number of particles may slow the computational time of the optimization and difficult the problem convergence.

Aiming to assure the accuracy of the model, statistical analyses were performed. The *Student's t-test* has shown that the process variables means are all equivalent to the experimental data means, since their intervals are all intercepting each other. Furthermore, *Fisher's exact test* has shown that the model variances are also corresponding to the experimental variances, which also ratifies that the model as valid. Thus, the statistical analysis of the model lead to conclude that it presents a satisfactory accuracy to be used to simulate the solid-state fermentation process studied here.

Moreover, the residual values of the nonlinear least squares objective function have shown that the most of the model errors are contained in the bioreactor temperature bed and in the substrate consumption. However, the value of the sum of objective function residuals of the reparametrized model is smaller than the value reached with the non-reparametrized model using the same PSO estimation procedure, showing that the model validated with the estimation of reparametrized parameters presented in this work is even better.

The reparametrized solid-state fermentation model can be considered accurate to simulate this fermentation process using the *Kluyveromyces marxianus* NRRL Y-7571 for the cells growth kinetics, the bioreactor bed temperature, the substrate consumption, the ethanol production, and the CO_2 and O_2 produced by the cells metabolism. Also, the PSO heuristic method successfully optimized the parameters of the referred model.

APPENDIX A

Parameter	Constant value or Equation number	Unit	Description
X	Equation 1	g/L	Cells concentration
X_m	1.0	g/L	Maximum cells concentration
T	Equation 2	°C	Temperature
T_a	30.0	°C	Air entrance temperature
ρ_b	Equation 3	$kg.m^{-3}$	Bed density
C_{pb}	Equation 4	$\frac{J}{kg\,°C}$	Heat capacity of bed
ϕ	Equation 5	Dimensionless	Physiological factor
γ_s	Equation 7	h^{-1}	Rate coefficient of the physiological factor synthesis
γ_d	Equation 8	h^{-1}	Rate coefficient of the physiological factor denaturation
S	Equation 9	g/L	Substrate concentration
P	Equation 10	g/L	Product concentration
CO_2	Equation 11	g/L	CO_2 concentration
O_2	Equation 12	g/L	O_2 concentration
C_{pa}	1180.0	$\frac{J}{kg\,°C}$	Heat capacity of moist air
Z	0.4	m	Evaluated bed height
λ	2414300.0	$\frac{J}{kg_{H_2O}}$	Latent heat of evaporation of water
R	8.3144	$\frac{J}{K\,mol}$	Universal gas constant
V_Z	33.33	m/h	Moist air velocity
ƒ	0.00246	$\frac{kg_{H_2O}}{kg_{air}\,°C}$	Water carrying capacity of air
μ	Equations 13-14	h^{-1}	Growth rate
Ψ	Equation 15	Dimensionless	Symbolic normalized state
Y	Equation 15	---	Symbolic state
$\bar{x}$	Table 2	---	Data mean

APPENDIX B

Step 1. Initialize the search parameters:

Define the number of iterations (N_{iter});

Define the number of search particles (N_{pt});

Define the number of searched dimensions (N_d);

Define the search boundaries (this will be a vector with length N_d);

Define the cognition and social parameters (c_1 and c_2);

Define the initial and final inertial weights (w_0 and w_f);

Initialize the iteration counter ($k = 0$);

Step 2. Compute the maximum particle velocities along each direction d:

$$v_d^{max} = \frac{\left(x_d^{max} - x_d^{min}\right)}{2}$$

Step 3. Compute initial particle positions and velocities:

$$x_{p,d}^k = x_d^{min} + r\left(x_d^{max} - x_d^{min}\right)$$

$$v_{p,d}^k = v_d^{max}(2r - 1)$$

Where r is a random number with uniform distribution between [0,1].

Step 4. Evaluate the objective function for each particle

Step 5. Store the particles position and objective function value to be used further in the construction of the confidence region

Step 6. Compute or update the best position found by the whole swarm (x^{glo}):

$$x_{k+1}^{glo} = x_k^{glo}$$

Step 7. Stop criteria 1

If the number of iterations exceeds the maximum

$$k = N_{iter}$$

Then stop the search.

Step 8. Update the objective function value for each particle with the best objective function value found for that particle:

$$x_p^{ind} = min\left|OF(x_p^{ind})\right|, p = 1..N_{pt}$$

Step 9. Compute the inertial weight value:

$$w = w_0 + (w_f - w_0)\frac{k}{N_{iter}}$$

Step 10. Update particle velocities:

$$v_{p,d}^{k+1} = wv_{p,d}^{k} + c_1 r_1\left(x_{p,d}^{ind} - x_{p,d}^{k}\right) + c_2 r_2\left(x_{d}^{glo} - x_{p,d}^{k}\right), \begin{matrix} p = 1..N_{pt} \\ d = 1..N_d \end{matrix}$$

Step 11. If the computed particle absolute velocity exceeds v_d^{max}:

$$v_{p,d}^{k+1} = v_d^{max} sign\left(v_{p,d}^{k+1}\right)$$

Step 12. Update particle positions:

$$x_{p,d}^{k+1} = x_{p,d}^{k} + v_{p,d}^{k}$$

Step 13. If $x_{p,d}^{k+1}$ exceeds the search boundaries, it is placed at the limit

Step 14. The close of the loop:

$$k = k + 1$$

Return to Step 4.

Supplementary Data

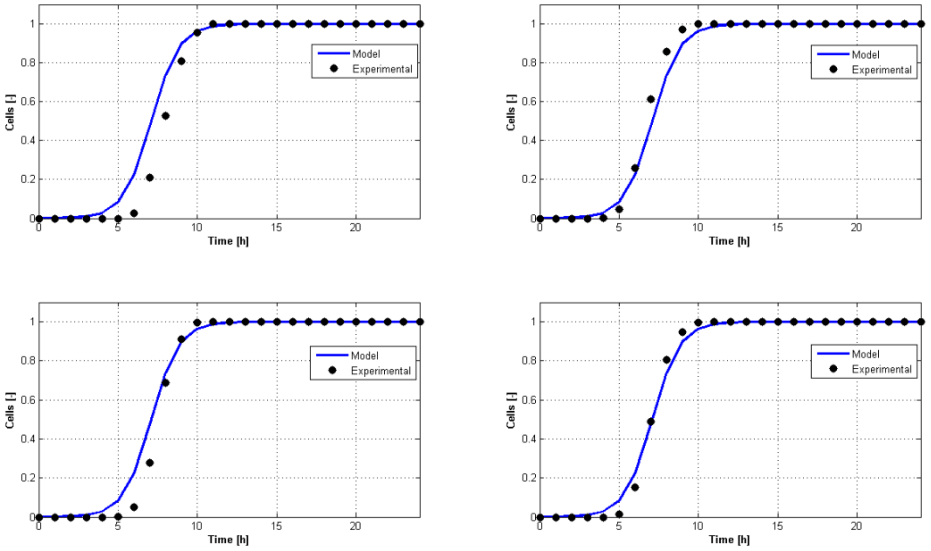

Figure 2. Cells growth vs time for experimental data (•) and model (-).

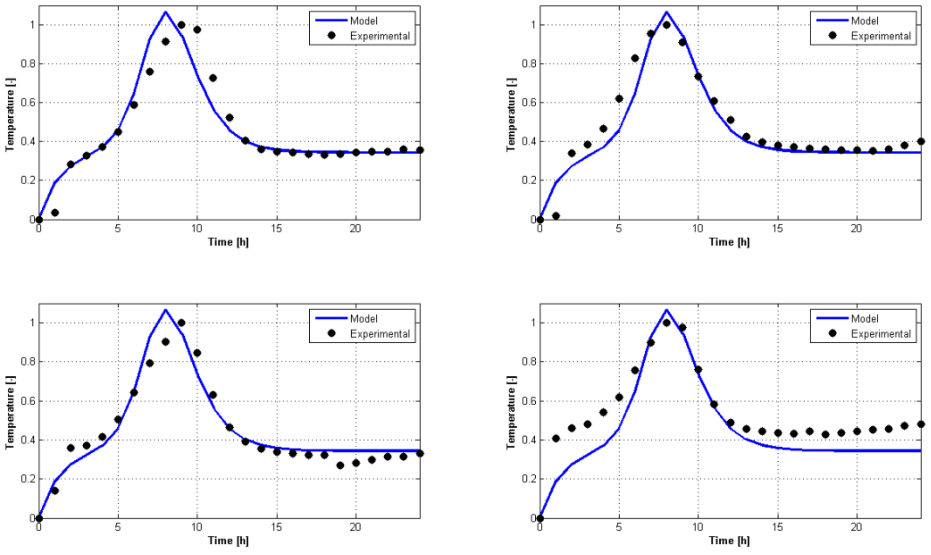

Figure 3. Temperature profile vs time for experimental data (•) and model (-).

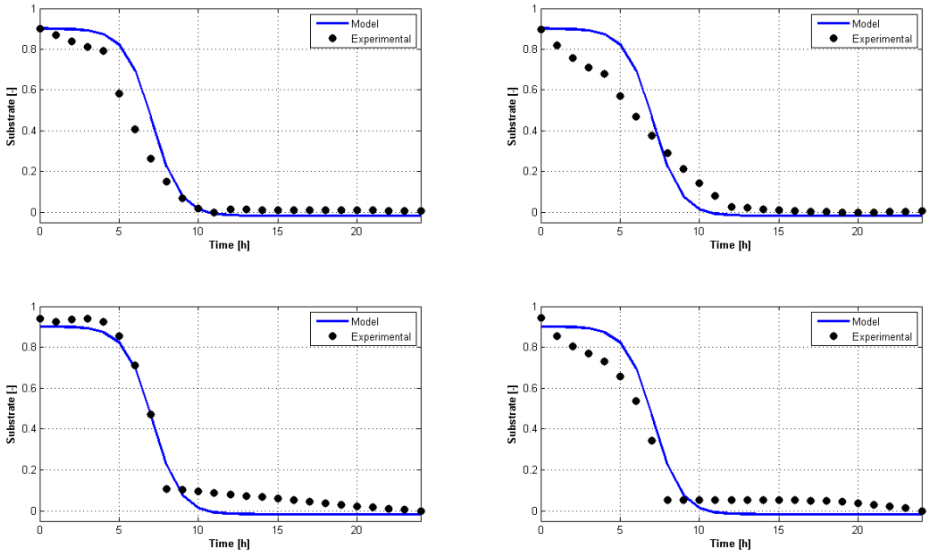

Figure 4. Substrate consumption vs time for experimental data (•) and model (-).

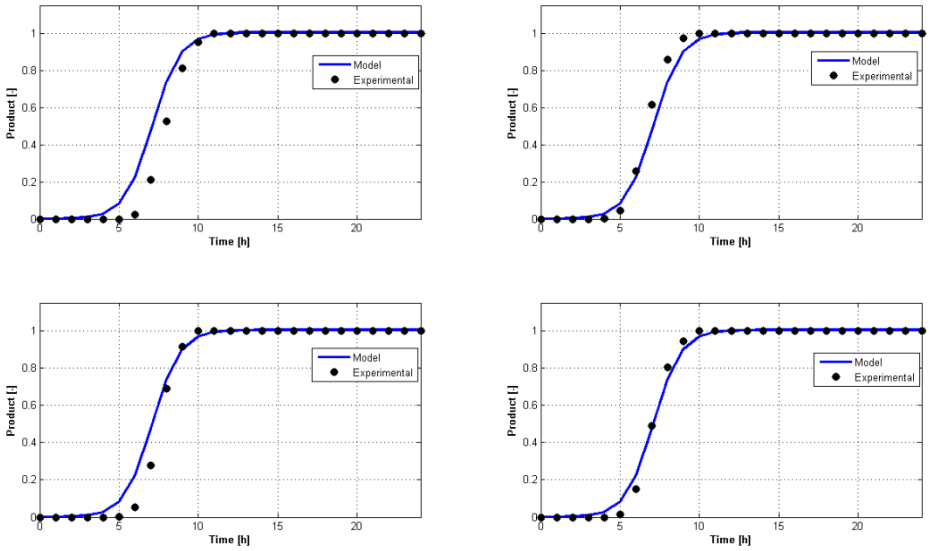

Figure 5. Ethanol production vs time for experimental data (•) and model (-).

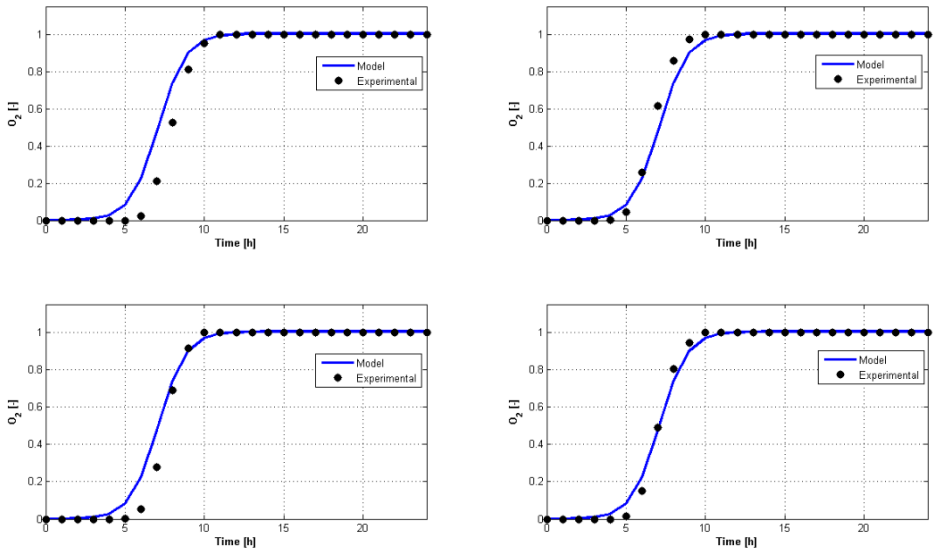

Figure 6. O_2 production vs time for experimental data (•) and model (-).

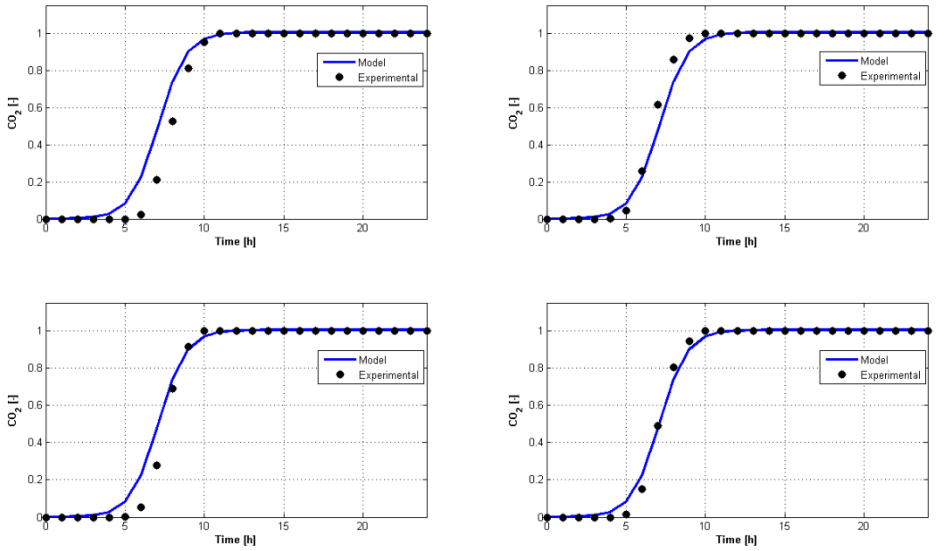

Figure 7. CO_2 production vs time for experimental data (•) and model (-).

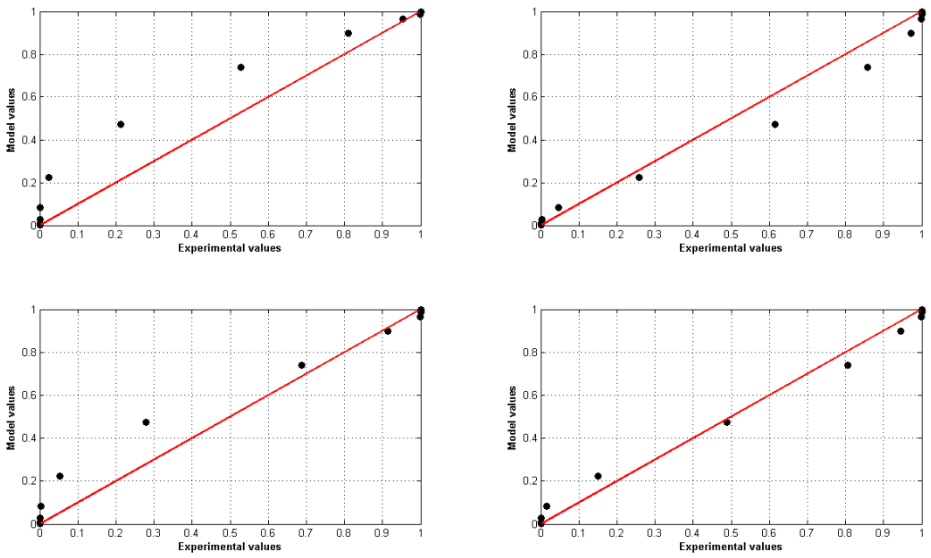

Figure 8. Model vs. experimental values (residuals) for cells growth kinetics.

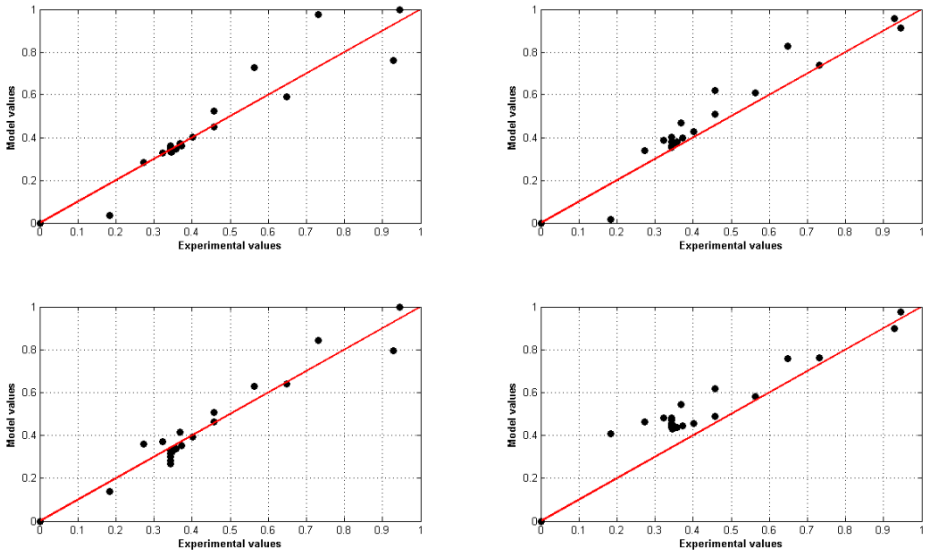

Figure 9. Model vs. experimental values (residuals) for bioreactor bed temperature.

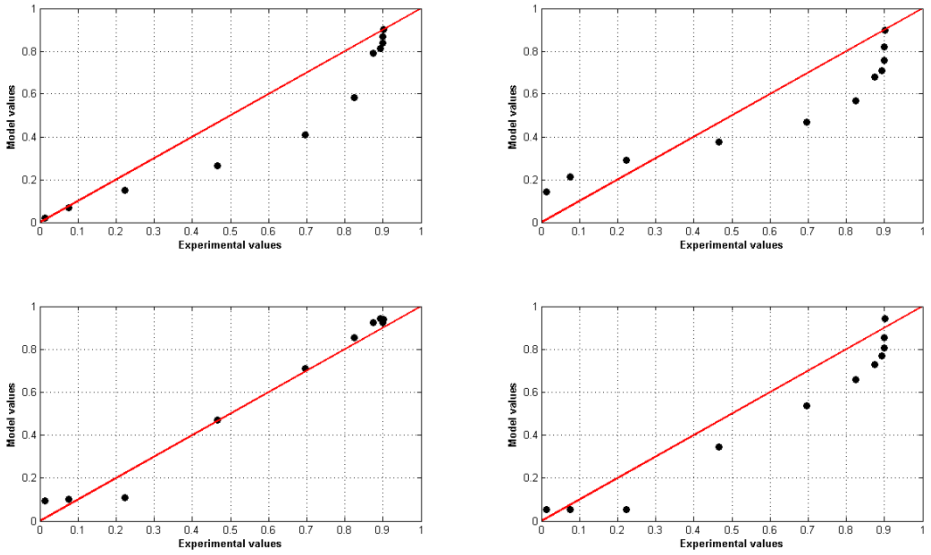

Figure 10. Model vs. experimental values (residuals) for substrate consumption.

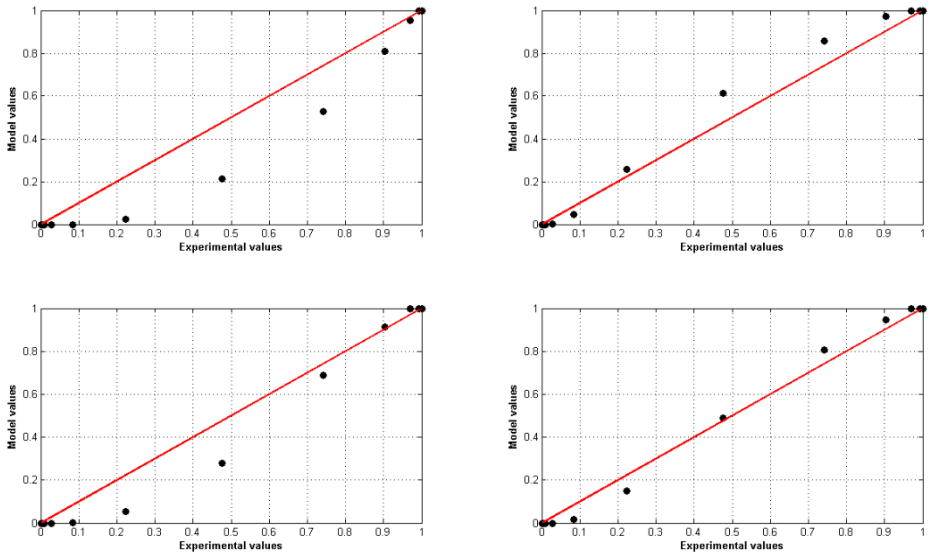

Figure 11. Model vs. experimental values (residuals) for product (ethanol).

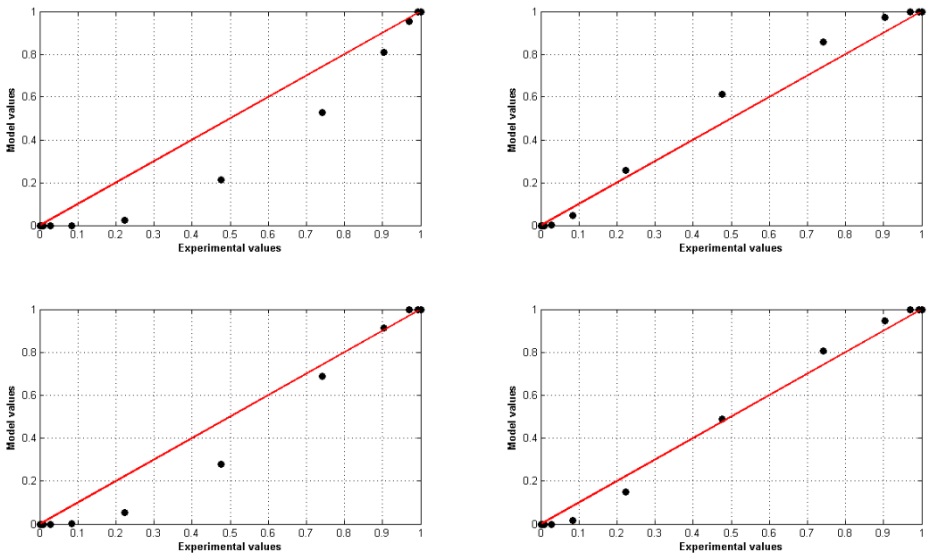

Figure 12. Model vs. experimental values (residuals) for O_2 produced by cells metabolism.

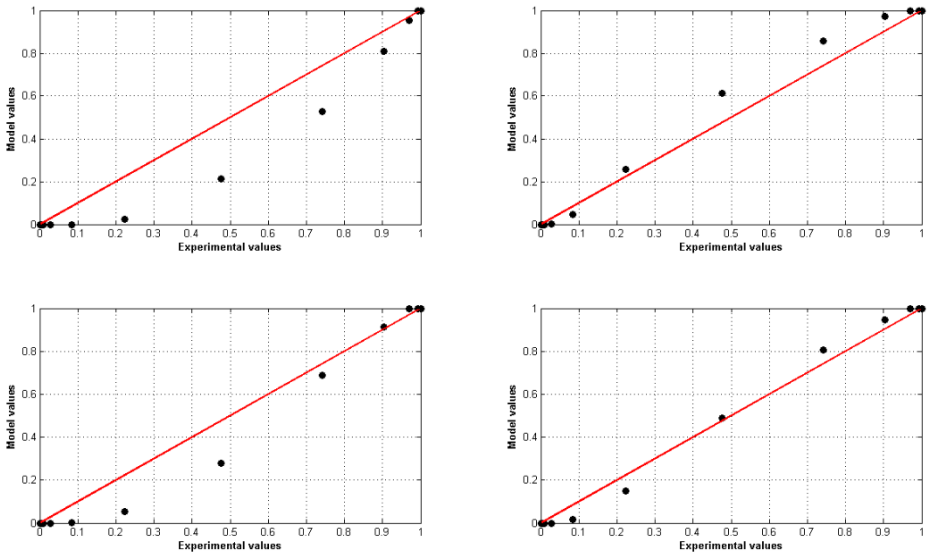

Figure 13. Model vs. experimental values (residuals) for CO_2 produced by cells metabolism.

References

[1] PANDEY, A. **Solid state fermentation**. Biochemical Engineering Journal, 13, 81-84 (2003).

[2] RAHARDJO, Y. S. P., TRAMPER, J., RINZEMA, A. **Modeling conversion and transport phenomena in solid-state fermentation: a review and perspectives**. Biotechnology Advances, 24, 161-179 (2006).

[3] FARINAS, C. S., VITCOSQUE, G. L., FONSECA, R. F., NETO, V. B., COURI, S. **Modeling the effects of solid state fermentation operating conditions on endoglucanase production using and instrumented bioreactor**. Industrial Crops and Products, 34, 1186-1192 (2011).

[4] HÖLKER, U., LENZ, J. **Solid-state fermentation – are there any biotechnological advantages?**. Current Opinion in Microbiology, 8 (3), 301-306 (2005).

[5] LENZ, J., HÖFER, M., KRASENBRINK, J. B., HÖLKER, U. **A survey of computational and physical methods applied to solid-state fermentation**. Applied Microbiology and Biotechnology, 65(1), 9-17 (2004).

[6] SCHWAAB, M.; PINTO, J. C. **Experimental data analysis I. Fundamentals of statistics and parameter estimation**. E-papers, Rio de Janeiro – RJ (2007).

[7] SILVEIRA, C. L., MAZUTTI, M. A., SALAU, N. P. G. **Modeling the microbial growth and temperature profile in a fixed bed bioreactor**. Bioprocess and Biosystems Engineering, 37, 1945-1954 (2014).

[8] SCHWAAB, M., BISCAIA, E. C., MONTEIRO, J. L. Jr., PINTO, J. C. **Nonlinear parameter estimation through particle swarm optimization**. Chemical Engineering Science, 63, 1542-1552 (2008).

[9] DRAPER, N. R., SMITH, H. **Applied regression analysis**. Wiley, New Work, 1998.

[10] SILVEIRA, C. L. da; MAZUTTI, M. A.; SALAU, N. P. G. **Identifiability measures to select the parameters to be estimated in a solid-state fermentation model**. Submitted to Biochemical Engineering Journal (2014).

[11] SCHWAAB, M.; LEMOS, L. P.; PINTO, J. C. **Optimum reference temperature for reparametrization of the Arrhenius equation. Part 2: problems involving multiple reparametrizations.** Chemical Engineering Science, 63, 2895-2906 (2008).

[12] ESPIE, D. M.; MACCHIETTO, S. **Nonlinear transformations for parameter estimation**. Industrial and Engineering Chemistry Research, 27 (11), 2175-2179 (1988).

[13] SCHWAAB, M.; PINTO, J. C. **Optimum reparametrization of power function models**. Chemical Engineering Science, 63, 4631-4635 (2008).

[14] SCHWAAB, M.; PINTO, J. C. **Optimum reference temperature for reparametrization of the Arrhenius equation. Part 1: Problems involving one kinetic constant**. Chemical Engineering Science, 62, 2750-2764 (2007).

[15] SCHWAAB, M.; PINTO, J. C. **Optimum reference temperature for reparametrization of the Arrhenius equation. Part 2: Problems involving multiple reparametrizations**. Chemical Engineering Science, 63, 2895-2906 (2008).

[16] VERHULST, P. F. **Notice sur la loi que la population poursuit dans son accroissement**. Correspondance mathématique et physique, 10, 113-121 (1838).

[17] BACAËR, N. **A short history of mathematical population dynamics**. Springer (2011).

[18] HAMIDI-ESFAHANI, Z. SHOJAOSADATI, S. A., RINZEMA, A. **Modeling of simultaneous effect of moisture and temperature on *A. niger* growth in solid-state fermentation**. Biochemical Engineering Journal, 21, 265-272 (2004).

[19] MITCHELL, D. A., GREENFIELD, P. F., DOELLE, H. W. **An empirical model of growth of *Rhizopus oligosporus* in solid-state fermentation**. Journal of Fermentation and Bioengineering, 72(3), 224-226 (1991).

[20] LAGEMAAT, J. van de; PYLE, D. L. **Modeling the uptake and growth kinetics of *Penicillium glabrum* in a tannic acid-containing solid-state fermentation for tannase production**. Process Biochemistry, 40, 1773-1782 (2005).

[21] FANAEI, M. A., VAZIRI, B. M. **Modeling of temperature gradients in packed-bed solid-state bioreactors**. Chemical Engineering and Processing: Process Intensification, 48(1):446–451 (2009).

[22] PRATA, D. M., SCHWAAB, M., LIMA, E. L. L., PINTO, J. C. **Nonlinear dynamics data reconciliation and parameter estimation through particle swarm optimization: Application for an industrial polypropylene reactor**. Chemical Engineering Science, 64(18), 3953-3967 (2009).

[23] MAZUTTI, M. A., ZABOT, G., BONI, G., SKOVRONSKI, A., OLIVEIRA, D. de, LUCCIO, M. Di, RODRIGUES, M. I., TREICHEL, H., MAUGERI, F. **Kinetics of inulinase production by solid-state fermentation in a packed-bed bioreactor.** Food Chemistry, 120, 163-173 (2010).

[24] KENNEDY, J.; EBERHART, R. **Particle Swarm Optimization**. Proceedings of the IEEE International Conference on Neural Networks, 4, 1942-1948 (1995).

[25] POLI, R., KENEDY, J., BLACKWELL, T. **Particle swarm optimization**. Swarm intelligence, 1(1), 33-57 (2007).

[26] DORMAND, J. R., PRINCE, P. J. **A family of embedded Runge-Kutta formulae.** Journal of computational and Applied Mathematics, 6(1), 19-26 (1980).

7 DISCUSSÃO

Esta dissertação foi estruturada com artigos científicos, dessa forma se faz importante discutir os principais tópicos que foram tratados em todo o trabalho e concatená-los em uma discussão.

O primeiro artigo trata do uso de diferentes tipos de modelos, mecanicísticos e empíricos, utilizados para representar a cinética de uma hidrólise enzimática. Cada modelo testado tinha um número determinado de parâmetros a serem estimados e, para que se pudesse avaliar o quão bom foi o ajuste de cada modelo, além da minimização da função objetivo, utilizou-se o Critério de Informação Akaike, o qual compreende o número de parâmetros estimados na análise. O método de otimização por Enxame de Partículas foi utilizado para minimizar a função objetivo. Este foi o primeiro problema de otimização de processos biotecnológicos tratado neste trabalho.

Do segundo ao quarto artigos, foram tratados problemas referentes ao processo de fermentação em estado sólido. Para o primeiro destes trabalhos, utilizou-se uma abordagem de um modelo simples, com apenas duas variáveis de estado – concentração celular e temperatura, utilizando um método derivativo (Levenberg-Marquardt) para a otimização dos parâmetros. Além disso, o método de integração numérica Dormand-Prince, conforme função *ode45* contida por padrão no Matlab®, foi introduzido para a simulação do processo.

No segundo artigo que trata do processo de fermentação em estado sólido, percebeu-se possível adicionar mais variáveis de estado ao modelo, ainda que aumentando sua complexidade. Dessa forma, pode-se acrescentar uma variável de estado fisiológico das células para relacionar seu crescimento e morte no reator. Também, adicionou-se o consumo de substrato e a obtenção de produto final, de CO_2 e de O_2 ao modelo. Além disso, uma matriz de sensibilidade dos parâmetros foi avaliada para que se pudesse obter a identifiabilidade dos parâmetros utilizados no modelo.

No último trabalho, uma nova abordagem utilizando uma reparametrização do modelo foi utilizada. Essa reparametrização diminuiu o número de parâmetros a serem estimados, a rigidez do modelo e também a correlação dos parâmetros. Além disso, o método do Enxame de Partículas foi utilizado novamente em conjunto com o método de Levenberg-Marquardt. Neste trabalho também utilizou-se o método de integração numérica Dormand-Prince para a simulação do processo.

Percebeu-se, durante o procedimento de estimação de parâmetros realizado em todos os trabalhos, que o número de parâmetros a serem estimados e, com maior importância, a diferença de magnitude desses parâmetros têm forte influência sobre a estimação. Quando os parâmetros a serem estimados possuíam uma diferença de magnitude muito grande, o procedimento de estimação demorava mais tempo para ser realizado do que quando eles não eram tão diferentes assim, até mesmo em situações onde haviam menos parâmetros serem estimados. Por tal motivo, a reparametrização foi utilizada neste último trabalho.

A otimização por Enxame de Partículas, como um método heurístico, tem grande funcionalidade para estimações quando a região de busca ainda é um tanto desconhecida; contudo, para o refinamento da estimação ainda é útil lançar mão do uso de métodos derivativos, tais como o Levenberg-Marquardt.

Ambos os processos biotecnológicos tratados neste trabalho, hidrólise enzimática e fermentação em estado sólido, foram modelados e validados, uma vez que os dados experimentais ficaram próximos aos modelos ajustados propostos com bastante precisão, apresentando baixos resíduos e perfis cinéticos bastante semelhantes. Isso apenas foi possível com cuidadosos procedimentos de estimação de parâmetros e, antes disso, por terem sido selecionados modelos adequados.

Considerando os testes estatísticos utilizados, além da minimização da função objetivo e das análises gráficas, os modelos utilizados com os parâmetros estimados são precisos e confiáveis para descrever os processos, principalmente no caso do modelo de fermentação em estado sólido, o qual é fenomenológico e, por esta razão, pode ser extrapolado para diferentes condições operacionais deste processo. Mostrando, também, que das três hipóteses testadas, a hipótese da inibição por substrato descreve com maior precisão os dados experimentais do processo.

8 CONCLUSÕES

O processo de hidrólise enzimática foi avaliado por diversos modelos. Alguns destes modelos tiveram seus resíduos, ou seja, seus erros comparados aos dados experimentais, mais baixos do que outros; porém, de uma maneira geral, todos os resíduos foram semelhantes entre si, dificultando a distinção sobre qual seria o melhor modelo a ser usado. Desta maneira, o Critério de Informação Akaike Corrigido (AICc) foi utilizado para avaliar, concomitantemente com os resíduos, o número de parâmetros do modelo. Entre todos os modelos que foram testados, o modelo que teve menor resíduo e melhor AICc foi o modelo empírico não-autônomo proposto. As simulações conduzidas confirmaram o quão bom foi o referido modelo.

O primeiro e mais simples modelo de fermentação em estado sólido que foi apresentado continha apenas duas equações diferenciais para descrever a cinética do crescimento celular e o perfil de temperatura no reator. O modelo utilizou uma equação logística para o crescimento celular e um balanço de energia para a temperatura. Os parâmetros foram ajustados utilizando mínimos quadrados não-lineares e o algoritmo de Levenberg-Marquardt. O modelo obteve um excelente ajuste, e para validá-lo foram conduzidos testes estatísticos. Os testes de *t de Student*, de χ^2 e de Fisher foram realizados para uma confiança de 95%, e os dados imulados mostraram que o modelo era confiável. Além disso, os resíduos e os gráficos gerados foram analisados. Os resíduos foram, de fato, muito baixos; e os gráficos mostraram que o comportamento obtido com os modelos era muito semelhante ao dos dados experimentais.

Como o modelo simples para predizer o comportamento do processo de fermentação em estado sólido foi bem sucedido, uma nova abordagem foi feita para que se pudesse contabilizar o estado fisiológico das células. Também, além da taxa de crescimento constante, foram testadas as hipóteses de Monod e de inibição por substrato, tornando-se então a taxa de crescimento específico como uma função do substrato disponível. Além disso, o consumo de substrato e a obtenção de produto, de CO_2 e de O_2 foram adicionados ao modelo. Essa nova proposta de modelo mostrou-se ser tão boa quanto a anterior, com a vantagem de poder predizer mais variáveis de estado do processo. Outrossim, percebeu-se que a hipótese da inibição por substrato melhorou o desempenho do modelo, diminuindo os resíduos. Neste trabalho, os testes de *t de Student*, de χ^2 e de Fisher foram realizados novamente, validando o modelo proposto conforme comparação com os dados experimentais.

No último trabalho, uma reparametrização do modelo do processo de fermentação em estado sólido foi proposta, de forma que se pudesse diminuir a correlação dos parâmetros e a

grande diferença de magnitude entre eles. Essa reparametrização também diminuiu o número de parâmetros a serem estimados, aumentando os graus de liberdade disponíveis para conduzir os testes estatísticos. Neste trabalho, o algoritmo de Levenberg-Marquardt foi utilizado para refinar a estimação de parâmetros feita pelo Enxame de Partículas, ainda que o Enxame de Partículas já tivesse fornecido uma minimização da função objetivo dos mínimos quadrados não-lineares bastante satisfatória. O modelo reparametrizado apresentou resíduos ainda menores do que os encontrados no modelo não-reparametrizado, levando-nos a concluir que, além das vantagens para o procedimento de estimação de parâmetros obtidas pela reparametrização, o modelo é ainda mais eficiente na predição do comportamento do processo de fermentação em estado sólido.

Dessa forma, mostrou-se que a otimização por Enxame de Partículas aliada ao algoritmo de Levenberg-Marquardt culminou em resultados melhores. Também, algumas manipulações numéricas são importantes para reduzir a correlação entre parâmetros e a diferença de magnitude entre eles, reduzindo, dessa maneira, o custo computacional e prevenindo problemas numéricos no procedimento.

Finalmente, concluiu-se que diferenças de magnitude entre as variáveis de estado pode causar dificuldade para o método Dormand-Prince durante o procedimento de integração. Dessa forma, novamente, percebe-se a importância da reparametrização do modelo.

Ambos os processos utilizados neste trabalho foram modelados e validados com dados experimentais e análises estatísticas e gráfica. Os seus resultados mostraram que eles são confiáveis e precisos para predizer os dados experimentais e, portanto, para descrever o comportamento geral do processo.

9 SUGESTÕES

Novos trabalhos podem ser feitos usando dados experimentais para o crescimento microbiano em diferentes alturas do leito do reator, de forma que se possa desenvolver um modelo de crescimento celular com gradientes ao longo do comprimento do reator.

Também, pode-se realizar um trabalho na busca de multiplicidades de estados estacionários para estes tipos de processo, principalmente para o processo de fermentação em estado sólido.

10 REFERÊNCIAS

AQUARONE, E.; BORZANI, W.; LIMA, U. de A.; SCHMIDELL, W. **Biotecnologia Industrial: Vol. 4 – Biotecnologia na Produção de Alimentos**. São Paulo: Ed. Blücher, v. 1, 2001.

ASCHER, U. M.; PETZOLD, L. R. **Computer methods for ordinary differential equations and differential-algebraic equations**. SIAM: Society for Industrial and Applied Mathematics, 1998.

AVERILL, B. A.; ELDREDGE, P. **Principles of general chemistry**. Disponível em: < http://2012books.lardbucket.org/books/principles-of-general-chemistry-v1.0/index.html>. Accessado em: 4 de Agosto, 2014.

BACAËR, N. **A short history of mathematical population dynamics**. Springer (2011).

BEQUETTE, B. W. **Process Dynamics: modeling, analysis, and simulation**. New Jersey: Prentice Hall PTR, 1998.

BEZERRA, R. M. F.; DIAS, A. A.; FRAGA, I.; PEREIRA, A. N. **Cellulose hydrolysis by Cellobiohydrolase Cel7A shows mixed hyperbolic product inhibition**. Applied Biochemistry and Biotechnology, v. 165 (1), 178-189, 2011.

BROWN, R. F.; AGBOGBO, F. K.; HOLTZAPPLE, M. T. **Comparison of mechanistic models in the initial rate enzymatic hydrolysis of AFEX-treated wheat straw**. Biotechnology for Biofuels, v. 3, 1754-6834, 2010.

CHAPRA, S.; CANALE, R. **Numerical methods for engineers**. McGraw-Hill, 6 ed., 2009.

COLLARES, R. M.; MIKLASEVICIUS, L. V. S.; BASSACO, M. M.; SALAU, N. P. G.; MAZUTTI, M. A.; BISOGNIN, D. A.; TERRA, L. M. **Optimization of enzymatic hydrolysis of cassava to obtain fermentable sugars**. Journal of Zhejiang University – Science B (Biomedicine & Biotechnology), v. 13 (7), 579-586, 2012.

COUTO, S. R.; SANROMÁN, M. Á. **Application of solid-state fermentation to food industry – A review**. Journal of Food Engineering, v. 76, 291-302, 2006.

DAVIS, M. E. **Numerical methods and modeling for chemical engineers**. New York: Wiley, 1984.

DOBRE, T. G.; MARCANO, J. G. S. **Chemical Engineering:** Modeling, Simulation and Similitude. Germany, Weinheim: WILEY-VCH, 2007.

DONNELLY, R. A. **Geometry optimization by simulated annealing**. Chemical Physics Letters, v. 136 (3, 4), 274-278.

DORMAND, J. R.; PRINCE, P. J. **A family of embedded Runge-Kutta formulae**. Journal of Computational and Applied Mathematics, 6 (1), 19-26, 1980.

PRINCE, P. J.; DORMAND, J. R. **High order embedded Runge-Kutta formulae**. Journal of Computational and Applied Mathematics, 7 (1), 67-75, 1981.

EDGARD, T. F.; HIMMELBLAU, D. M. **Optimization of Chemical Processes**. New York – NY, USA: McGraw-Hill, 1988.

ELQOTBI, M.; VLAEV, S. D.; MONTASTRUC, L.; NIKOV, I. **CFD modeling of two-phase stirred bioreaction systems by segregated solution of the Euler-Euler model**. Computers and Chemical Engineering, v. 48, 113-120, 2013.

ESPIE, D. M.; MACCHIETTO, S. **Nonlinear transformations for parameter estimation**. Industrial and Engineering Chemistry Research, v. 27 (11), 2175-2179, 1988.

EULER, L. **Introduction to analysis of the infinite**. New York: Springer, 2013.

FANAEI, M. A.; VAZIRI, B. M. **Modeling of temperature gradientes in packed-bed solid-state bioreactors**. Chemical engineering and Processing: Process Intensification, v. 48 (1), 446-451, 2009.

FEHLBERG, E. **Klassische Runge-Kutta-Formeln vierter und niedrigerer ordnung mit Schrittweiten-Kontrolle und ihre Anwendung auf Wärmeleitungsprobleme**. Computing, 6, 61-71, 1970.

GIANFRANCESCO, A.; TURCHIULI, C.; FLICK, D.; DUMOULIN, E. **CFD modeling and simulation of maltodextrin solutions spray drying to control stickiness**. Food and Bioprocess Technology, v. 3, 946-955, 2010.

GOLDBERG, D. E. **Genetic algorithms in search, optimization, and machine learning**. Addison Wesley Longman Inc., 1989.

HAMIDI-ESFAHANI, Z.; SHOJAOSADATI, S. A.; RINZEMA, A. **Modeling of simultaneous effect of moisture and temperature on *A. niger* growth in solid-state fermentation**. Biochemical Engineering Journal, v. 21, 265-272, 2004.

HAN, B.; ROMBOUTS, F. M.; NOUT, M. J. R. **A Chinese fermented soybean food**. International Journal of Food Microbiology, v. 65, 1-10, 2001.

HARDER, D. W. **4th-order Runge-Kutta and the Dormand-Prince Methods**. Slides from Waterloo Engineering, Department of Electrical and Computer Engineering, 2012. Available in: <https://ece.uwaterloo.ca/~ne216/Laboratories/04/4.DormandPrince.pptx>. Accessed in November 20, 2014.

HASHEMI, M.; MOUSAVI, S. M.; RAZAVI, S. H.; SHOJAOSADATI, S. A. **Mathematical modeling of biomass and α-amylase production kinetics by *Bacillus sp.* in solid-state fermentation based on solid dry weight variation**. Biochemical Engineering Journal, v. 53, 159-164, 2011.

HEBDEN, M. D. **An algorithm for minimization using exact second derivatives**. Relatório TP515 da Fundação de Pesquisas em Energia Atômica, Harwell, Inglaterra.

HÖLKER, U.; HÖFER, M.; LENZ, J. **Biotechnological advantages of laboratory-scale solid-state fermentation with fungi**. Applied Microbiology and biotechnology, v. 64 (2), 175-186, 2004.

HÖLKER, U.; LENZ, J. **Solid-state fermentation – are there any biotechnological advantages?**. Current Opinion in Microbiology, v. 8 (3), 301-306, 2005.

KENNEDY, J.; EBERHART, R. **Particle swarm optimization**. Proceedings of the IEEE International Conference on Neural Networks, v. 4, 1942-1948, 1995.

KIRKPATRICK, S.; GELATT, C. D. Jr.; VECCHI, M. O. **Optimization by simulated annealing**. Science, 220 (4598), 671-680, 1983.

KUTTA, W. **Beitrag zur näherungsweisen Integration totaler Differentialgleichungen**. Zeitschrift für angewandte Mathematik und Physik, 46, 435-453, 1901.

LAGEMAAT, J. van de; PYLE, D. L. **Modeling the uptake and growth kinetics of *Penicillium glabrum* in a tannic acid-containing solid-state fermentation for tanase production**. Process Biochemistry, v. 40, 1773-1782, 2005.

LENZ, J.; HÖFER, M.; KRASENBRINK, J. B.; HÖLKER, U. **A survey of computational and physical methods applied to solid-state fermentation**. Applied Microbiology and Biotechnology, v. 65 (1), 9-17, 2004.

LEVENBERG, K. **A method for the solution of certain non-linear problems in least squares**. Quarterly Applied Mathematics, 2 (2), 164-168, 1944.

MARQUARDT, D. **An algorithm for least-squares estimation of nonlinear parameters**. Journal of the society for Industrial and Applied Mathematics, 11 (2), 431-441, 1963.

MENG, X.; JIA, M.; WANG, T. **Neural network prediction of biodiesel kinematic viscosity at 313 K**. Fuel, v. 121, 133-140, 2014.

MITCHELL, D. A.; GREENFIELD, P. F.; DOELLE, H. W. **An empirical model of growth of *Rhizopus oligosporus* in solid-state fermentation**. Journal of Fermentation and Bioengineering, v. 72 (3), 224-226, 1991.

MITCHELL, D. A.; Meien, O. F. von; Krieger, N.; Dalsenter, F. D. H. **A review of recent developments in modeling of microbial growth kinetics and intraparticle phenomena in solid-state fermentation**. Biochemical Engineering Journal, v. 17, 15-26, 2004.

MORÉ, J. J. **The Levenberg-Marquardt Algorithm: implementation and theory**. Numerical Analysis. Ed. G. A. Watson, Lectures Notes in Mathematics 630, Springer Verlag, 105-116, 1977.

PANDEY, A. **Recent process developments in Solid-State Fermentation**. Process Biochemistry, v. 27, 109-117, 1992.

PANDEY, A. **Solid state fermentation**. Biochemical Engineering Journal, v. 13, 81-84, 2003.

PAOLUCCI-JEANJEAN, D.; BELLEVILLE, M.; ZAKHIA, N.; RIOS, G. M. **Kinetics of cassava starch hydrolysis with Termamyl® enzyme**. Biotechnology and Bioengineering, v. 68 (1), 71-77, 2000.

PARK, E. Y.; IKEDA, Y.; OKUDA, N. **Empirical evaluation of cellulose on enzymatic hydrolysis of waste office paper**. Biotechnology and Bioprocess Engineering, v. 7 (5), 268-274, 2002.

PINTO, J. C.; LAGE, P. L. C. **Métodos numéricos em problemas de Engenharia Química.** Rio de Janeiro – RJ: E-papers, 2001.

POLAKOVIC, M.; BRYJAK, J. **Modeling of potato starch saccharification by and *Aspergillus niger* glucoamylase**. Biochemical Engineering Journal, v. 18, 57-63, 2004.
RAMERIZ, W. F. **Computational methods for process simulation**. Boston: Butterworths, 1989.

RATTANACHOMSRI, U.; TANAPONGPIPAT, L. E.; CHAMPREDA, V. **Simultaneous non-thermal saccharification of cassava pulp by multi-enzyme activity and ethanol fermentation by *Candida tropicalis***. Journal of Bioscience and Bioengineering, v. 107, n. 5, 488-493, 2009.

REID, I. D. **Solid-state fermentations for biological delignification**. Enzyme and Microbial Technology, v. 11, 786-803, 1989.

RISSO, R. dos S. **Etanol de Batata-Doce: Otimização do pré-processamento da matéria-prima e da hidrólise enzimática**. Dissertation (Masters), Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Departamento de Engenharia Química, Programa de Pós-Graduação em Engenharia Química, Porto Alegre, RS, 2014.

RODRÍGUEZ-FERNÁNDEZ, D. E.; RODRÍGUEZ-LEÓN, J. A.; CARVALHO, J. C. de; STURM, W.; SOCCOL, C. R. **The behavior of kinetic parameters in production of pectinase and xylanase by solid-state fermentation**. Bioresource Technology, v. 102, 10657-10662, 2011.

RUFFELL, J.; LEVIE, B.; HELLE, S.; DUFF, S. **Pretreatment and enzymatic hydrolysis of recovered fiber for ethanol production**. Bioresource Technology, v. 101, 2267-2272, 2010.
RUNGE, C. **Ueber die numerische Auflösung von Differentialgleichungen**. Mathematische Annalen, 46 (2), 167-178, 1895.

RUSSEL, J. B. **Química Geral**. São Paulo – SP: McGraw-Hill do Brasil, 1981.

SCHWAAB, M.; BISCAIA, E. C.; MONTEIRO, J. L. Jr.; PINTO, J. C. **Nonlinear parameter estimation through particle swarm optimization**. Chemical Engineering Science, v. 63, 1542-1552, 2008a.

SCHWAAB, M.; LEMOS, L. P.; PINTO, J. C. **Optimum reference temperature for reparametrization of the Arrhenius equation**. **Part 2: problems involving multiple reparametrizations**. Chemical Engineering Science, v. 63, 2895-2906, 2008b.

SCHWAAB, M.; PINTO, J. C. **Análise de dados experimentais I -** Fundamentos de estatística e estimação de parâmetros. 1. ed. Rio de Janeiro – RJ: E-papers, 2007.

SHARMA, N.; RATHORE, M.; SHARMA, M. **Microbial pectinase: sources, characterizations and applications**. Reviews in Environmental Science and Biotechnology, v. 12, 45-60, 2013.

SHI, Y.; EBERHART, R. **A modified particle swarm optimizer**. Proceedings of the IEEE International Conference on Evolutionary Computation, Anchorage, Alaska, 69-73, 1998.

SILVEIRA, C. L.; MAZUTTI, M. A.; SALAU, N. P. G. **Modeling the microbial growth and temperature profile in a fixed bed bioreactor**. Bioprocess and Biosystems Engineering, 37, 1945-1954, 2014.

SOFIANOPOULOU, S. **Simulated annealing applied to the process allocation problem**. European Journal of Operational Research, v. 60, 327-334, 1992.

SOUZA, P. M. de; MAGALHÃES, P. de O. **Application of microbial α-amylase in industry – A review.** Brazilian Journal of Microbiology, v. 41, 850-861, 2010.

VATS, S.; NEGI, S. **Use of artificial neural network (ANN) for the development of bioprocess using *Pinus roxburghii* fallen foliages for the release of polyphenols and reducing sugar**. Bioresource Technology, v. 140, 392-398, 2013.

VÁZQUEZ, G.; ANTORRENA, G.; GONZÁLEZ, J.; FREIRE, S.; CRESPO, I. **The influence of acetosolv pulping conditions on the enzymatic hydrolysis of Eucalyptus pulps**. Wood Science and Technology, v. 34 (4), 345-354, 2000.

VERHULST, P. F. **Notice sur la loi que la population poursuit dans son accroissement**. Correspondance mathématique et physique, v. 10, 113-121, 1838.

VINIEGRA-GONZÁLEZ, G.; FAVELA-TORRES, E.; AGUILAR, C. N.; RÓMERO-GOMEZ, S. de J.; DÍAZ-GODÍNEZ, G.; AUGUR, C. **Advantages of fungal enzyme production in solid state over liquid fermentation systems**. Biochemical Engineering Journal, v. 13, 157-167, 2003.

APÊNDICE

Tableau de Butcher: Método de Dormand-Prince

0							
1/5	1/5						
3/10	3/40	9/40					
4/5	44/45	-56/15	32/9				
8/9	19372/6561	-25360/2187	64448/6561	-212/729			
1	9017/3168	-355/33	46732/5247	49/176	-5103/18656		
1	35/384	0	500/1113	125/192	-2187/6784	11/84	
	35/384	0	500/1113	125/192	-2187/6784	11/84	0
	5179/57600	0	7571/16695	393/640	-92097/339200	187/2100	1/40

Onde cada linha da primeira coluna é a soma de elementos à direita da mesma linha; os elementos na direita superior da tabela correspondem aos coeficientes apresentados na Equação 22; e os elementos na parte inferior são os coeficientes de pesos $b_{i,j}$ apresentados na Eq. 23, a primeira linha representa os pesos para a aproximação de 5ª ordem enquanto os da segunda para a de 4ª ordem.

ANEXO

```
function [tout,xout,Nsteps_acc,Nsteps_rej] =
ode45(FUN,tspan,x0,pair,ode_fcn_format,tol,trace,count,hmax,N_est_acc_steps)

% Copyright (C) 2001, 2000 Marc Compere
% This file is intended for use with Octave.
% ode45.m is free software; you can redistribute it and/or modify it
% under the terms of the GNU General Public License as published by
% the Free Software Foundation; either version 2, or (at your option)
% any later version.
%
% ode45.m is distributed in the hope that it will be useful, but
% WITHOUT ANY WARRANTY; without even the implied warranty of
% MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the GNU
% General Public License for more details at www.gnu.org/copyleft/gpl.html.
%
% ------------------------------------------------------------------
%
% ode45 (v1.15) integrates a system of ordinary differential equations using
% 4th & 5th order embedded formulas from Dormand & Prince or Fehlberg.
%
% The Fehlberg 4(5) pair is established and works well, however, the
% Dormand-Prince 4(5) pair minimizes the local truncation error in the
% 5th-order estimate which is what is used to step forward (local extrapolation.)
% Generally it produces more accurate results and costs roughly the same
% computationally.  The Dormand-Prince pair is the default.
%
% This is a 4th-order accurate integrator therefore the local error normally
% expected is O(h^5).  However, because this particular implementation
% uses the 5th-order estimate for xout (i.e. local extrapolation) moving
% forward with the 5th-order estimate should yield local error of O(h^6).
%
% The order of the RK method is the order of the local *truncation* error, d,
% which is the principle error term in the portion of the Taylor series
% expansion that gets dropped, or intentionally truncated.  This is different
% from the local error which is the difference between the estimated solution
% and the actual, or true solution.  The local error is used in stepsize
% selection and may be approximated by the difference between two estimates of
% different order, l(h) = x_(O(h+1)) - x_(O(h)).  With this definition, the
% local error will be as large as the error in the lower order method.
% The local truncation error is within the group of terms that gets multipled
% by h when solving for a solution from the general RK method.  Therefore, the
% order-p solution created by the RK method will be roughly accurate to O(h^(p+1))
% since the local truncation error shows up in the solution as h*d, which is
% h times an O(h^(p)) term, or rather O(h^(p+1)).
% Summary:   For an order-p accurate RK method,
%            - the local truncation error is O(h^p)
%            - the local error used for stepsize adjustment and that
%              is actually realized in a solution is O(h^(p+1))
%
% This requires 6 function evaluations per integration step.
%
% Both the Dormand-Prince and Fehlberg 4(5) coefficients are from a tableu in
% U.M. Ascher, L.R. Petzold, Computer Methods for  Ordinary Differential Equations
% and Differential-Agebraic Equations, Society for Industrial and Applied
Mathematics
% (SIAM), Philadelphia, 1998
%
% The error estimate formula and slopes are from
% Numerical Methods for Engineers, 2nd Ed., Chappra & Cannle, McGraw-Hill, 1985
%
% Usage:
%         [tout, xout] =
ode45(FUN,tspan,x0,pair,ode_fcn_format,tol,trace,count,hmax)
%
% INPUTS:
```

```
%    FUN   - String containing name of user-supplied problem description.
%            Call: xprime = fun(t,x) where FUN = 'fun'.
%            t      - Time (scalar).
%            x      - Solution column-vector.
%            xprime - Returned derivative COLUMN-vector; xprime(i) = dx(i)/dt.
%    tspan - [ tstart, tfinal ]
%    x0    - Initial value COLUMN-vector.
%    pair  - flag specifying which integrator coefficients to use:
%               0 --> use Dormand-Prince 4(5) pair (default)
%               1 --> use Fehlberg pair 4(5) pair
%    ode_fcn_format - this specifies if the user-defined ode function is in
%            the form:     xprime = fun(t,x)   (ode_fcn_format=0, default)
%            or:           xprime = fun(x,t)   (ode_fcn_format=1)
%            Matlab's solvers comply with ode_fcn_format=0 while
%            Octave's lsode() and sdirk4() solvers comply with ode_fcn_format=1.
%    tol   - The desired accuracy. (optional, default: tol = 1.e-6).
%    trace - If nonzero, each step is printed. (optional, default: trace = 0).
%    count - if nonzero, variable 'rhs_counter' is initalized, made global
%            and counts the number of state-dot function evaluations
%            'rhs_counter' is incremented in here, not in the state-dot file
%            simply make 'rhs_counter' global in the file that calls ode45
%    hmax  - limit the maximum stepsize to be less than or equal to hmax
%    N_est_acc_steps - an estimate of how many accepted steps there may be;
%                      used to preallocate memory for the [tout,xout] solutions
%
% OUTPUTS:
%    tout  - array of discretized times points (an Nsteps_acc by 1 column-vector).
%    xout  - solution, one solution column-vector per tout-value (an Nsteps_acc by
Nstates matrix)
%    Nsteps_acc - total number of accepted integration steps + 1
%    Nsteps_rej - total number of rejected integration steps
%
% The result can be displayed by: plot(tout, xout).
%
% The following relationship should hold after a completed run:
%    rhs_counter == (Nsteps_acc-1+Nsteps_rej)*6+1
%
% Marc Compere
% CompereM@asme.org
% created : 06 October 1999
% modified: 03 July 2001

if nargin < 10, N_est_acc_steps = (tspan(2)-tspan(1))*1e3; end
if nargin <  9, hmax = (tspan(2) - tspan(1))/2.5; end
if nargin <  8, count = 0; end
if nargin <  7, trace = 0; end
if nargin <  6, tol = 1.e-6; end
if nargin <  5, ode_fcn_format = 0; end
if nargin <  4, pair = 0; end

pow = 1/6; % see p.91 in the Ascher & Petzold reference for more infomation.

if (pair==0),
   % Use the Dormand-Prince 4(5) coefficients:
    a_(1,1)=0;
    a_(2,1)=1/5;
    a_(3,1)=3/40; a_(3,2)=9/40;
    a_(4,1)=44/45; a_(4,2)=-56/15; a_(4,3)=32/9;
    a_(5,1)=19372/6561; a_(5,2)=-25360/2187; a_(5,3)=64448/6561; a_(5,4)=-212/729;
    a_(6,1)=9017/3168; a_(6,2)=-355/33; a_(6,3)=46732/5247; a_(6,4)=49/176;
a_(6,5)=-5103/18656;
    a_(7,1)=35/384; a_(7,2)=0; a_(7,3)=500/1113; a_(7,4)=125/192; a_(7,5)=-
2187/6784; a_(7,6)=11/84;
    % 4th order b-coefficients
    b4_(1,1)=5179/57600; b4_(2,1)=0; b4_(3,1)=7571/16695; b4_(4,1)=393/640;
b4_(5,1)=-92097/339200; b4_(6,1)=187/2100; b4_(7,1)=1/40;
    % 5th order b-coefficients
```

```
    b5_(1,1)=35/384; b5_(2,1)=0; b5_(3,1)=500/1113; b5_(4,1)=125/192; b5_(5,1)=-
2187/6784; b5_(6,1)=11/84; b5_(7,1)=0;
    for i=1:7,
     c_(i)=sum(a_(i,:));
    end
else, % pair==1 so use the Fehlberg 4(5) coefficients:
    a_(1,1)=0;
    a_(2,1)=1/4;
    a_(3,1)=3/32; a_(3,2)=9/32;
    a_(4,1)=1932/2197; a_(4,2)=-7200/2197; a_(4,3)=7296/2197;
    a_(5,1)=439/216; a_(5,2)=-8; a_(5,3)=3680/513; a_(5,4)=-845/4104;
    a_(6,1)=-8/27; a_(6,2)=2; a_(6,3)=-3544/2565; a_(6,4)=1859/4104; a_(6,5)=-
11/40;
    % 4th order b-coefficients (guaranteed to be a column vector)
    b4_(1,1)=25/216; b4_(2,1)=0; b4_(3,1)=1408/2565; b4_(4,1)=2197/4104; b4_(5,1)=-
1/5;
    % 5th order b-coefficients (also guaranteed to be a column vector)
    b5_(1,1)=16/135; b5_(2,1)=0; b5_(3,1)=6656/12825; b5_(4,1)=28561/56430;
b5_(5,1)=-9/50; b5_(6,1)=2/55;
    for i=1:6,
     c_(i)=sum(a_(i,:));
    end
end % if (pair==0)


% Initialization
t0 = tspan(1);
tfinal = tspan(2);
t = t0;
hmin = (tfinal - t)/1e20;
h = (tfinal - t)/100; % initial step size guess
x = x0(:);            % ensure x is a column vector

Nstates = size(x,1);
tout = zeros(N_est_acc_steps,1);       % preallocating memory is not only faster
but, for long
xout = zeros(N_est_acc_steps,Nstates); % simulations, prevents disappointing
surprises later

Nsteps_rej = 0;
Nsteps_acc = 1;
tout(Nsteps_acc) = t;     % first output time
xout(Nsteps_acc,:) = x.'; % first output solution = IC's

if count==1,
 global rhs_counter
 if ~exist('rhs_counter'),rhs_counter=0; end
end % if count

if trace
 clc, t, h %, x
end

if (pair==0),

   k_ = x*zeros(1,7);  % k_ needs to be initialized as an Nx7 matrix where N=number
of rows in x
                       % (just for speed so octave doesn't need to allocate more
memory at each stage value)

   % Compute the first stage prior to the main loop.  This is part of the Dormand-
Prince pair caveat.
   % Normally, during the main loop the last stage for x(k) is the first stage for
computing x(k+1).
   % So, the very first integration step requires 7 function evaluations, then each
subsequent step
```

```
   % 6 function evaluations because the first stage is simply assigned from the
last step's last stage.
   % note: you can see this by the last element in c_ is 1.0, thus t+c_(7)*h = t+h,
ergo, the next step.
   if (ode_fcn_format==0), % (default)
      k_(:,1)=feval(FUN,t,x); % first stage
   else, % ode_fcn_format==1
      k_(:,1)=feval(FUN,x,t);
   end % if (ode_fcn_format==1)

   % increment rhs_counter
   if (count==1), rhs_counter = rhs_counter + 1; end

   % The main loop using Dormand-Prince 4(5) pair
   while (t < tfinal) & (h >= hmin),
      if t + h > tfinal, h = tfinal - t; end

      % Compute the slopes by computing the k(:,j+1)'th column based on the
previous k(:,1:j) columns
      % notes: k_ needs to end up as an Nxs, a_ is 7x6, which is s by (s-1),
      %        s is the number of intermediate RK stages on [t (t+h)] (Dormand-
Prince has s=7 stages)
      if (ode_fcn_format==0), % (default)
         for j = 1:6,
            k_(:,j+1) = feval(FUN, t+c_(j+1)*h, x+h*k_(:,1:j)*a_(j+1,1:j)');
         end
      else, % ode_fcn_format==1
         for j = 1:6,
            k_(:,j+1) = feval(FUN, x+h*k_(:,1:j)*a_(j+1,1:j)', t+c_(j+1)*h);
         end
      end % if (ode_fcn_format==1)

      % increment rhs_counter
      if (count==1), rhs_counter = rhs_counter + 6; end

      % compute the 4th order estimate
      x4=x + h* (k_*b4_); % k_ is Nxs (or Nx7) and b4_ is a 7x1

      % compute the 5th order estimate
      x5=x + h*(k_*b5_); % k_ is Nxs (or Nx7) and b5_ is a 7x1

      % estimate the local truncation error
      gamma1 = x5 - x4;

      % Estimate the error and the acceptable error
      delta = norm(gamma1,'inf');       % actual error
      tau = tol*max(norm(x,'inf'),1.0); % allowable error

      % Update the solution only if the error is acceptable
      if (delta<=tau),

         t = t + h;
         x = x5;    % <-- using the higher order estimate is called 'local
extrapolation'
         Nsteps_acc = Nsteps_acc + 1;
         tout(Nsteps_acc) = t;
         xout(Nsteps_acc,:) = x.';

         % Assign the last stage for x(k) as the first stage for computing x(k+1).
         % This is part of the Dormand-Prince pair caveat.
         % k_(:,7) has already been computed, so use it instead of recomputing it
         % again as k_(:,1) during the next step.
         k_(:,1)=k_(:,7);

      else, % unacceptable integration step
```

```
         Nsteps_rej = Nsteps_rej + 1;

      end % if (delta<=tau)

      if trace
         home, t, h, Nsteps_acc, Nsteps_rej
      end % if trace

      % Update the step size
      if (delta==0.0),
       delta = 1e-16;
      end % if (delta==0.0)
      h = min(hmax, 0.8*h*(tau/delta)^pow);

   end % while (t < tfinal) & (h >= hmin)

else, % pair==1 --> use RK-Fehlberg 4(5) pair

   k_ = x*zeros(1,6);  % k_ needs to be initialized as an Nx6 matrix where N=number
of rows in x
                       % (just for speed so octave doesn't need to allocate more
memory at each stage value)

   % The main loop using Dormand-Prince 4(5) pair
   while (t < tfinal) & (h >= hmin),
      if t + h > tfinal, h = tfinal - t; end

      % Compute the slopes by computing the k(:,j+1)'th column based on the
previous k(:,1:j) columns
      % notes: k_ needs to end up as an Nx6, a_ is 6x5, which is s by (s-1),  (RK-
Fehlberg has s=6 stages)
      %        s is the number of intermediate RK stages on [t (t+h)]
      if (ode_fcn_format==0), % (default)
         k_(:,1)=feval(FUN,t,x); % first stage
      else, % ode_fcn_format==1
         k_(:,1)=feval(FUN,x,t);
      end % if (ode_fcn_format==1)

      if (ode_fcn_format==0), % (default)
         for j = 1:5,
            k_(:,j+1) = feval(FUN, t+c_(j+1)*h, x+h*k_(:,1:j)*a_(j+1,1:j)');
         end
      else, % ode_fcn_format==1
         for j = 1:5,
            k_(:,j+1) = feval(FUN, x+h*k_(:,1:j)*a_(j+1,1:j)', t+c_(j+1)*h);
         end
      end % if (ode_fcn_format==1)

      % increment rhs_counter
      if (count==1), rhs_counter = rhs_counter + 6; end

      % compute the 4th order estimate
      x4=x + h* (k_(:,1:5)*b4_); % k_(:,1:5) is an Nx5 and b4_ is a 5x1

      % compute the 5th order estimate
      x5=x + h*(k_*b5_); % k_ is the same as k_(:,1:6) and is an Nx6 and b5_ is a
6x1

      % estimate the local truncation error
      gamma1 = x5 - x4;

      % Estimate the error and the acceptable error
      delta = norm(gamma1,'inf');       % actual error
      tau = tol*max(norm(x,'inf'),1.0); % allowable error
```

```
      % Update the solution only if the error is acceptable
      if (delta<=tau),

         t = t + h;
         x = x5;    % <-- using the higher order estimate is called 'local
extrapolation'
         Nsteps_acc = Nsteps_acc + 1;
         tout(Nsteps_acc) = t;
         xout(Nsteps_acc,:) = x.';

      else, % unacceptable integration step

         Nsteps_rej = Nsteps_rej + 1;

      end % if (delta<=tau)

      if trace
         home, t, h, Nsteps_acc, Nsteps_rej
      end % if trace

      % Update the step size
      if (delta==0.0),
       delta = 1e-16;
      end % if (delta==0.0)
      h = min(hmax, 0.8*h*(tau/delta)^pow);

   end % while (t < tfinal) & (h >= hmin)

end % if (pair==0),


% if necessary, trim outputs
if (Nsteps_acc<N_est_acc_steps),
   tout = tout(1:Nsteps_acc);
   xout = xout(1:Nsteps_acc,:);
end % if (Nsteps_acc<N_est_acc_steps)


if (t < tfinal),
   disp('Step size grew too small.')
   t, h, Nsteps_acc, Nsteps_rej
end
```