

# IMPLEMENTAÇÃO E AVALIAÇÃO DE TÉCNICAS DE IDENTIFICAÇÃO DE SISTEMAS LINEARES E NÃO LINEARES USANDO O CÓDIGO LIVRE SCILAB

Rosiane Rocha Ribeiro

**Resumo:** Neste trabalho são aplicados algoritmos de identificação em um sistema que representa vias bioquímicas, conforme modelo proposto por Mendes et al (1998). O modelo descreve as variações de 8 concentrações de espécies envolvidas. A caracterização e a avaliação da dinâmica do processo foram feitas aplicando-se vários tipos de perturbações na entrada, concluindo-se sobre a não linearidade do modelo. As representações nos espaços de estados foram obtidas através de um algoritmo implementado usando o código livre Scilab. A seleção do modelo foi feita através de índices e critérios estruturais apropriados, como MSE, AIC, FPE e BIC. A sintonia do modelo foi executada através da evolução diferencial. Os parâmetros do modelo estimados representaram adequadamente o processo, demonstrando a adequação do procedimento de estimação.

**Palavras-chave:** Identificação de sistema, Representação em espaço de estados, sintonia de modelo, evolução diferencial.

---

<sup>1</sup> Bolsista da FAPEMIG – Faculdade de Engenharia Química, Universidade Federal de Uberlândia. Avenida João Naves de Ávila, 2121, Bloco K, Campus Santa Mônica.  
e-mail: rosi.ribeiro.rocha@hotmail.com

<sup>2</sup> Professora Orientadora – Faculdade de Engenharia Química, Universidade Federal de Uberlândia. Avenida João Naves de Ávila, 2121, Bloco K, Campus Santa Mônica.  
Laboratório de Modelagem, Controle e Otimização de Processos.  
CP: 593, Uberlândia-MG, CEP: 38408-100.

# IMPLEMENTATION AND EVALUATION OF TECHNIQUES OF IDENTIFICATION OF LINEAR AND NON-LINEAR SYSTEMS USING FREE SOFTWARE

Autor 1<sup>3</sup>, Autor 2<sup>4</sup>

**Abstract:** In this work, algorithms of identification are applied in a system that represents biochemistry routes, according to the model presented by Mendes et al (1998). The model describes the variation of 8 concentrations of the species involved. The characterization and evaluation of the process dynamics had been done applying many kinds of perturbation at inlets, which means that the model is nonlinear. The representations of states' spaces were obtained through an algorithm developed with the free software Scilab. The model was chosen by indexes and appropriated structural criteria, like MSE, AIC, FPE and BIC. The model's line was executed by differential evolution. The estimated parameters of the model represent the process properly, showing the fitness of the estimation procedure.

**Keywords:** systems identification, data representation of space, model tuning, differential evolution.

---

<sup>3</sup> Scholarship holder from FAPEMIG – School of Chemical Engineering, Federal University of Uberlândia. João Naves de Ávila Avenue, 2121, Building K, Campus Santa Mônica. E-mail: rosi.ribeiro.rocha@hotmail.com

<sup>4</sup> Professor Guide – School of Chemical Engineering, Federal University of Uberlândia. João Naves de Ávila Avenue, 2121, Building K, Campus Santa Mônica. Laboratory of Modeling, Control and Process Optimization. C.P: 593, Uberlândia-MG, CEP: 38408-100.

# 1 – INTRODUÇÃO

## *1.1- Identificação de um sistema*

Os modelos matemáticos têm adquirido cada vez mais importância em biotecnologia, sendo já uma ferramenta indispensável no estudo de processos bioquímicos. A simulação de processos bioquímicos tem suscitado um grande interesse, quer no meio acadêmico quer industrial, apresentando-se como um instrumento efetivo no projeto e desenvolvimento de algoritmos robustos, na otimização, monitorização, controle e caracterização desses processos.

O objetivo da modelagem matemática consiste em desenvolver um modelo com boa capacidade de predição e cujo procedimento de identificação e validação não se torne mais complexo do que o necessário.

A formulação do modelo matemático permite simular e otimizar condições operacionais de um equipamento ou de um processo com menor esforço experimental, especialmente em processos biotecnológicos cujos comportamentos dinâmicos são complexos e cujos experimentos são Estes modelos não prescindem, entretanto, de dados experimentais de boa qualidade que permitam a sua validação e que permitam a obtenção de equações constitutivas e parâmetros necessários para o fechamento do número de graus de liberdade do modelo (Velo, 2006).

A capacidade de representar um objetivo, uma ação ou um processo em termos matemáticos é uma das aspirações mais contundentes dos engenheiros, de modo a melhorar a monitorização e o controle de processos químicos em geral e bioquímicos em particular. Tem-se verificado um crescente interesse na utilização de estratégias baseadas em modelos, tanto no meio acadêmico quanto no meio industrial, para o projeto, o monitoramento, o controle e a otimização de processos, com vistas a ganhar vantagens competitivas. É necessário estabelecer-se a estrutura mais adequada do modelo e identificar o melhor valor para os seus parâmetros, de modo a obter-se a melhor aproximação aos dados experimentais. Assim, os dados experimentais têm um papel de destaque na modelação, uma vez que permitem identificar áreas de deficiência do modelo, levando à formulação de novas hipóteses e contribuindo para a obtenção de um modelo mais robusto e para a sua validação (Velo, 2006).

Há várias formas de classificar as técnicas de modelagem. Uma delas agrupa os métodos em três categorias: modelagem caixa branca, modelagem caixa preta e modelagem

caixa cinza. Na modelagem caixa branca é necessário conhecer bem o sistema em estudo, bem como as leis físicas que descrevem o sistema a ser modelado. Por essa razão esse tipo de modelagem é também conhecido como modelagem conceitual. A Identificação de Sistemas é uma área do conhecimento que estuda técnicas alternativas à modelagem caixa branca. Uma das características dessas técnicas é que pouco ou nenhum conhecimento prévio do sistema é necessário e, conseqüentemente, tais métodos são também referidos como modelagem (ou identificação) caixa preta ou modelagem empírica.

Em muitos casos será preferível usar técnicas de identificação para se obter modelos que descrevem o comportamento de um sistema. O que se pretende descrever com tais modelos são as relações de causa e efeito entre as variáveis de entrada e de saída. Nesse caso, o tipo de modelos, as técnicas usadas, e os requisitos necessários são bastante distintos dos correspondentes na modelagem pela natureza do processo.

A motivação para o estudo de técnicas de Identificação de Sistemas surge do fato que frequentemente não se conhecem as equações envolvidas no funcionamento de um determinado sistema ou elas são conhecidas, mas seria impraticável, por limitações de tempo e recursos para levantar tais equações e estimar seus respectivos parâmetros.

Uma categoria de técnicas que pode ser colocada entre a modelagem pela física ou natureza do processo e a identificação caixa preta é chamada identificação caixa cinza. As técnicas desse grupo caracterizam-se por usar informação auxiliar que não se encontra no conjunto de dados utilizados durante a identificação. O tipo de informação auxiliar e a forma com que ela é usada variam muito entre as diversas técnicas disponíveis. Sendo assim, existem métodos de identificação caixa cinza mais “claros”, que usam mais informação auxiliar e mais “escuros” (Aguirre, 2007).

## *1.2 – Desenvolvimento do modelo*

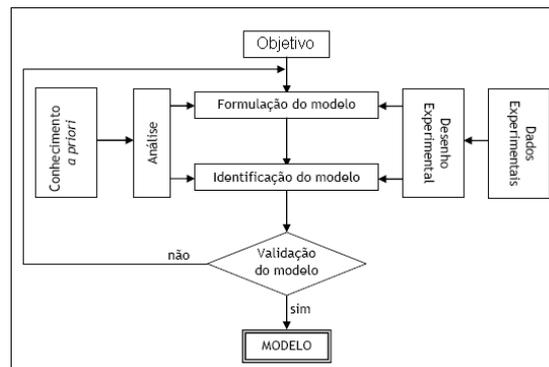
O desenvolvimento de um modelo matemático capaz de simular o comportamento de um dado processo é uma tarefa iterativa, que decorre em várias etapas fortemente dependentes umas das outras: (1) formulação do modelo; (2) identificação do modelo e (3) validação do modelo. Na Figura 1 são representadas, de uma forma esquemática, as etapas do desenvolvimento do modelo matemático.

A etapa de formulação do modelo consiste na definição e caracterização da sua estrutura. Nesta fase são definidos os limites do sistema, as variáveis de entrada e saída assim

como o tipo e o nível de complexidade do modelo a utilizar. São também definidas as várias relações entre as variáveis.

Durante a etapa de identificação pretende-se estimar o valor numérico dos parâmetros envolvidos nas funções que relacionam as variáveis do modelo. Esta é normalmente a etapa crítica do desenvolvimento do modelo matemático, devido essencialmente às dificuldades experimentais, dificuldades numéricas decorrentes de estabilidade, dificuldades de convergência, problemas com robustez dos métodos usados e ao tempo que exige.

A etapa final do desenvolvimento do modelo matemático consiste na sua validação, sendo feito nesta etapa o teste de sua capacidade de previsão. As previsões do modelo são comparadas com os resultados experimentais que não foram utilizados na identificação dos seus parâmetros. Se o modelo representar satisfatoriamente os resultados experimentais conclui-se que o modelo está validado, caso contrário, se as previsões discordarem significativamente desses resultados, não conseguindo também descrever as suas tendências, a estrutura do modelo deverá ser alterada e as etapas anteriores devem ser novamente consideradas.



**Figura 1:** Etapas no desenvolvimento de um modelo matemático.

Na modelação matemática de processos bioquímicos são exploradas essencialmente três fontes de informações: o objetivo do modelo, ou seja, o fim a que se destina deverá ser definido claramente no começo; o conhecimento *a priori*, que pode ser baseado no conhecimento empírico, físico, bioquímico e/ou microbiológico do sistema; e os dados experimentais obtidos durante operações normais ou durante experiências planejadas para a identificação do modelo (Rao e Iyengar, 1984; Vanrolleghem e Dochain, 1998; Hodgson *et al.*, 2004).

A modelação de processos bioquímicos normalmente envolve um maior número de iterações do que demais processos químicos. Na maior parte das aplicações físicas e químicas o conhecimento já adquirido previamente do sistema possui informação de elevada qualidade,

permitindo que grande parte da estrutura do modelo seja deduzida através desse conhecimento. A validação experimental permite a identificação dos seus parâmetros e ultrapassar algumas incertezas que ainda possam existir na estrutura do modelo, recorrendo eventualmente a algumas iterações do procedimento de modelação.

Contrariamente, as características inerentes dos processos bioquímicos, como a sua não linearidade, conjugadas com técnicas analíticas pouco adequadas e robustas, fazem com que o procedimento de modelação matemática não possa ser aplicado sem alteração, implicando muitas vezes em um número de iterações relativamente elevado. Por outro lado, os dados experimentais pouco precisos, afetados de algumas incertezas, são também um problema na definição da estrutura e identificação dos parâmetros do modelo, pois aumentam as possibilidades quer de se obter um modelo inválido, quer de não se conseguir estimar um valor único para os seus parâmetros. Este problema é ainda mais evidente na identificação em linha, onde se utilizam técnicas que têm associadas mais incertezas para se obter a informação em tempo real. Quando a calibração do modelo é feita, permite o uso de varias técnicas analíticas complementares, que apesar de exigirem um maior tempo de análise, conduzem a dados experimentais mais precisos (Vanrolleghem e Dochain, 1998).

Neste trabalho, o objetivo é aplicar técnicas de Identificação de Sistemas em um modelo de vias bioquímicas descrito por Mendes e Kell (1998), para encontrar o modelo que descreve o processo. Na seção 1 é apresentada uma breve introdução das características e das etapas de uma Identificação. Na Seção 2 será apresentado os materiais e métodos utilizados. Na Seção 3 a descrição do processo e dinâmica do mesmo. Na Seção 4 será apresentado a aplicação de métodos de Identificação de Sistemas não lineares, utilizando modelos em Espaço de Estados. Na Seção 5 será apresentado a sintonia do modelo por Evolução Diferencial para o conhecimento de parâmetros do modelo. Na Seção 6 será apresentado os resultados. E por fim, na Seção 7, as conclusões.

## **2 – MATERIAL E MÉTODOS**

### *2.1 – O código Scilab*

Todo o processo de identificação do modelo que será apresentado a seguir foi desenvolvido utilizando como ferramenta, o *código* livre Scilab (disponível em <http://www.scilab.org>). O Scilab (**Scientific Laboratory**) é uma plataforma gráfica usada na resolução de problemas numéricos. Disponível gratuitamente (com o código fonte), o Scilab

foi desenvolvido para ser um sistema aberto onde o usuário pode definir novos tipos de dados e operações; possui centenas de funções matemáticas com a possibilidade de interação com programas em várias linguagens como o C e Fortran; tem uma sofisticada estrutura de dados que inclui objetos como funções racionais, polinômios, listas, sistemas lineares etc., possui um interpretador e uma linguagem de programação (estruturada) própria.

## 2.2 – Evolução diferencial

O processo de sintonia do modelo para estimação de seus respectivos parâmetros foi realizado através da evolução diferencial.

A Evolução Diferencial foi desenvolvida por Storn e Price em meados da década de noventa e surgiu de tentativas de resolver o problema de ajuste polinomial de Chebychev (Storn and Price, 1995). O algoritmo é iniciado criando uma população inicial escolhida aleatoriamente devendo cobrir todo o espaço de busca. Geralmente, é criada por uma distribuição de probabilidade uniforme, quando não há nenhum conhecimento sobre o problema.

A idéia principal da evolução diferencial é gerar novos indivíduos, denotados vetores modificados ou doadores, pela adição da diferença ponderada entre dois indivíduos aleatórios da população a um terceiro indivíduo. Esta operação é chamada mutação.

As componentes do indivíduo doador são misturadas com as componentes de um indivíduo escolhido aleatoriamente (denotado vetor alvo), para resultar o chamado vetor tentativa, ou vetor experimental. O processo de misturar os parâmetros é referido frequentemente como "cruzamento" na comunidade dos algoritmos evolutivos.

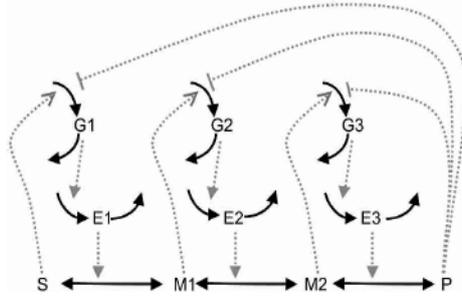
Se o vetor experimental resultar um valor da função objetivo menor que o vetor alvo, então o vetor experimental substitui o vetor alvo na geração seguinte. Esta última operação é chamada seleção. O procedimento é finalizado através de algum critério de parada (Arantes *et al.*, 2006 ).

## 3 – O MODELO MATEMÁTICO

### 3.1 – Descrição do processo

Nesse trabalho será utilizado um modelo de vias bioquímicas descrito por Mendes e Kell (1998) e ilustrado na Figura 2. As setas contínuas representem o fluxo de massa, as setas

tracejadas representam o regulamento cinética. S e P são respectivamente o substrato e o produto e são mantidos em concentrações constantes.  $M_1$  e  $M_2$  são os metabolismos intermediários da via.  $E_1$ ,  $E_2$  e  $E_3$  são as enzimas.  $G_1$ ,  $G_2$  e  $G_3$  são as espécies RNAm para as enzimas (Mendes *et al.*, 1998).



**Figura 2:** Esquema do modelo de vias bioquímicas.

A formulação matemática que descreve o modelo dinâmico não linear é descrita através das equações 1 a 8.

$$\frac{dG_1}{dt} = \frac{V_1}{1 + \left(\frac{P}{Ki_1}\right)^{ni_1} + \left(\frac{Ka_1}{S}\right)^{na_1}} - k_1 G_1 \quad (1) \quad \frac{dG_2}{dt} = \frac{V_2}{1 + \left(\frac{P}{Ki_2}\right)^{ni_2} + \left(\frac{Ka_2}{M_1}\right)^{na_2}} - k_2 G_2 \quad (2)$$

$$\frac{dG_3}{dt} = \frac{V_3}{1 + \left(\frac{P}{Ki_3}\right)^{ni_3} + \left(\frac{Ka_3}{M_2}\right)^{na_3}} - k_3 G_3 \quad (3) \quad \frac{dE_1}{dt} = \frac{V_4 G_1}{K_4 + G_1} - k_4 E_1 \quad (4)$$

$$\frac{dE_2}{dt} = \frac{V_5 G_2}{K_5 + G_2} - k_5 E_2 \quad (5) \quad \frac{dE_3}{dt} = \frac{V_6 G_3}{K_6 + G_3} - k_6 E_3 \quad (6)$$

$$\frac{dM_1}{dt} = \frac{kcat_1 E_1 \left(\frac{1}{km_1}\right) (S - M_1)}{1 + \frac{S}{Km_1} + \frac{M_1}{Km_2}} - \frac{kcat_2 E_2 \left(\frac{1}{Km_3}\right) (M_1 - M_2)}{1 + \frac{M_1}{Km_3} + \frac{M_2}{Km_4}} \quad (7)$$

$$\frac{dM_2}{dt} = \frac{kcat_2 E_2 \left(\frac{1}{km_3}\right) (M_1 - M_2)}{1 + \frac{M_1}{Km_3} + \frac{M_2}{Km_4}} - \frac{kcat_3 E_3 \left(\frac{1}{Km_5}\right) (M_2 - P)}{1 + \frac{M_2}{Km_5} + \frac{P}{Km_6}} \quad (8)$$

sendo que  $M_1$ ,  $M_2$ ,  $E_1$ ,  $E_2$ ,  $E_3$ ,  $G_1$ ,  $G_2$  e  $G_3$  representam as concentrações das espécies envolvidas nas diferentes reações bioquímicas, e S e P são as concentrações controláveis. Um

total de 16 simulações foram realizadas, alterando apenas as variáveis controláveis S e P conforme apresentado na Tabela 1 e mantendo as demais condições iniciais das variáveis de estados do processo constantes em cada experimento, conforme apresentado na Tabela 2. (Rodriguez-Fernandez *et al.*, 2006).

**Tabela 1:** Concentrações de Substrato e Produto nos experimentos

Experimento	Substrato	Produto	Experimento	Substrato	Produto
1	0,10000	0,05000	9	2,1544	0,05000
2	0,10000	0,13572	10	2,1544	0,13572
3	0,10000	0,36840	11	2,1544	0,36840
4	0,10000	1,00000	12	2,1544	1,00000
5	0,46416	0,05000	13	10,0000	0,05000
6	0,46416	0,13572	14	10,0000	0,13572
7	0,46416	0,36840	15	10,0000	0,36840
8	0,46416	1,00000	16	10,0000	1,00000

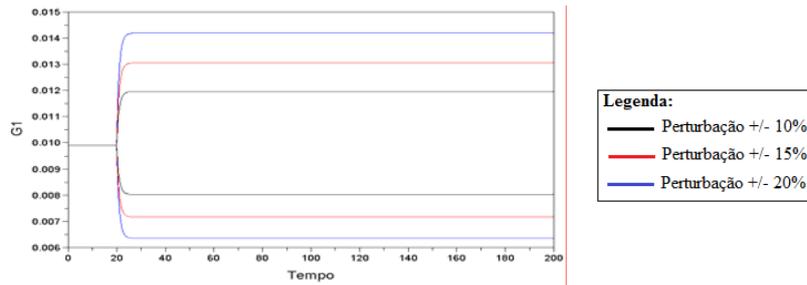
A Tabela 2 apresenta os valores exatos dos 36 parâmetros presentes no modelo matemático descrito pelas equações 1 a 8.

**Tabela 2:** Valores exatos dos 36 parâmetros do modelo.

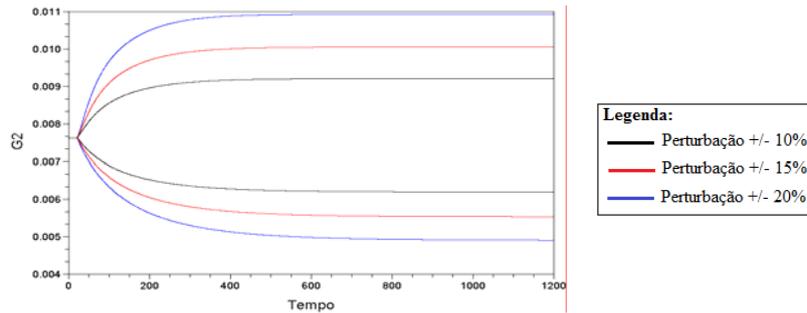
Parâmetro	Valor	Parâmetro	Valor	Parâmetro	Valor
$V_1$	1,00	$V_3$	1,00	$V_6$	0,10
$K_{i1}$	1,00	$K_{i3}$	1,00	$K_6$	1,00
$n_{i1}$	2,00	$n_{i3}$	2,00	$k_6$	0,10
$K_{a1}$	1,00	$K_{a3}$	1,00	$k_{cat1}$	1,00
$n_{a1}$	2,00	$n_{a3}$	2,00	$K_{m1}$	1,00
$k_1$	1,00	$k_3$	1,00	$K_{m2}$	1,00
$V_2$	1,00	$V_4$	0,10	$k_{cat2}$	1,00
$K_{i2}$	1,00	$K_4$	1,00	$K_{m3}$	1,00
$n_{i2}$	2,00	$k_4$	0,10	$K_{m4}$	1,00
$K_{a2}$	1,00	$V_5$	0,10	$k_{cat3}$	1,00
$n_{a2}$	2,00	$K_5$	1,00	$K_{m5}$	1,00
$k_2$	1,00	$k_5$	0,10	$K_{m6}$	1,00

### 3.2 - Análise de linearidade do modelo

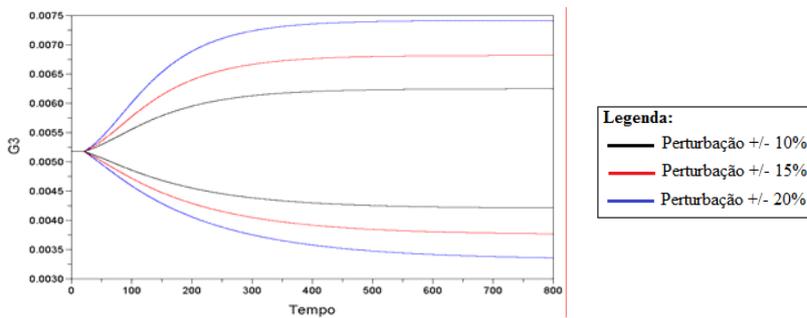
Com o objetivo de avaliar a linearidade do modelo foram realizadas perturbações degrau e perturbações randômicas a partir do estado estacionário. Para as perturbações degrau foram considerados os níveis de 10, 15 e 20% a mais e a menos do estado estacionário. Os resultados obtidos são apresentados nas Figuras 3 a 10.



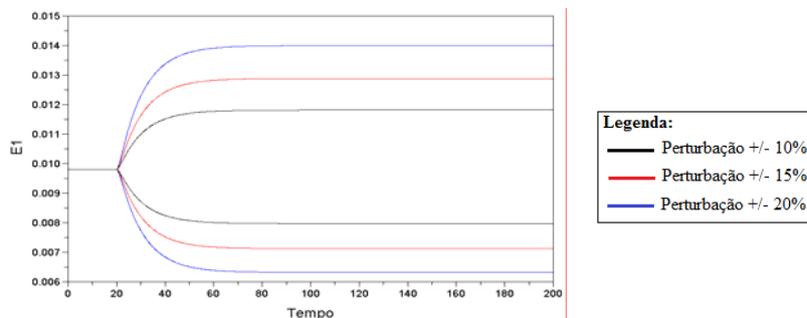
**Figura 3:** Respostas da variável  $G_1$  frente às perturbações de degraus aplicadas a partir do estado estacionário.



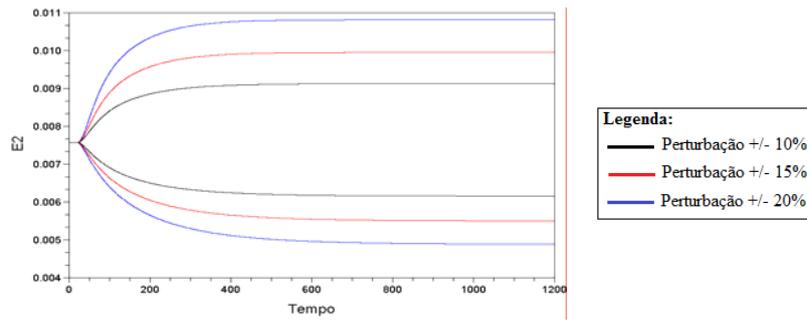
**Figura 4:** Respostas da variável  $G_2$  frente às perturbações de degraus aplicadas a partir do estado estacionário.



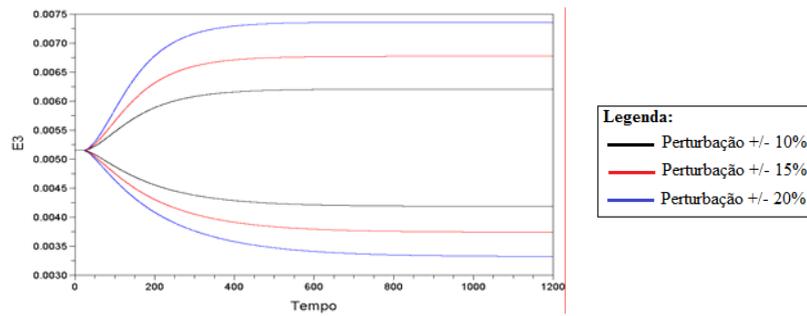
**Figura 5:** Respostas da variável  $G_3$  frente às perturbações de degraus aplicadas a partir do estado estacionário.



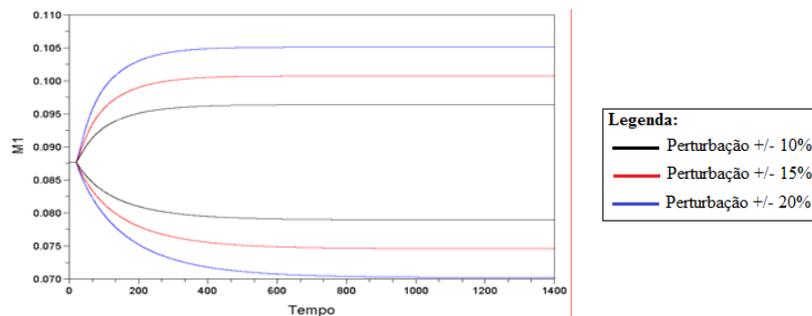
**Figura 6:** Respostas da variável  $E_1$  frente às perturbações de degraus aplicadas a partir do estado estacionário.



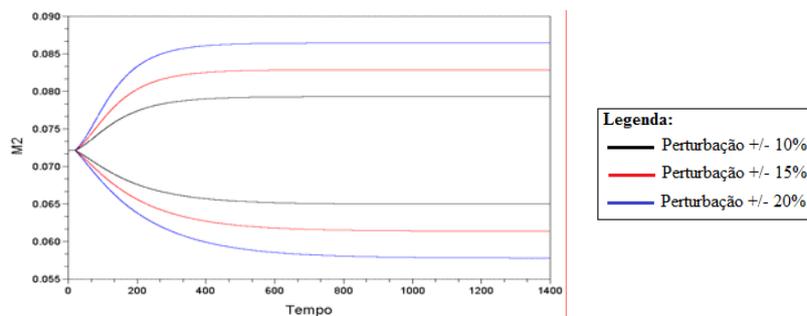
**Figura 7:** Respostas da variável  $E_2$  frente às perturbações de degraus aplicadas a partir do estado estacionário.



**Figura 8:** Respostas da variável  $E_3$  frente às perturbações de degraus aplicadas a partir do estado estacionário.



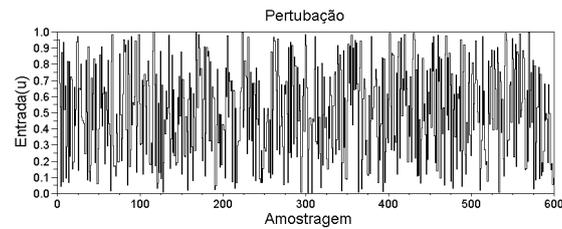
**Figura 9:** Respostas da variável  $M_1$  frente às perturbações de degraus aplicadas a partir do estado estacionário.



**Figura 10:** Respostas da variável  $M_2$  frente às perturbações de degraus aplicadas a partir do estado estacionário.

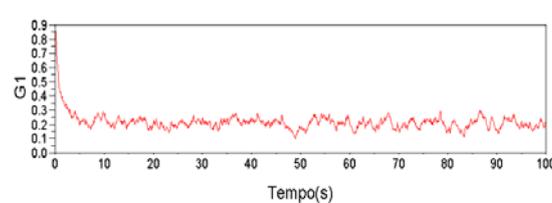
A análise destas figuras mostra que em pontos próximos do estado estacionário o sistema pode ser considerado linear, pois perturbações simétricas geraram respostas simétricas das saídas.

As perturbações randômicas aplicadas foram mais intensas e aplicada sobre a variável de controle S, no intervalo de [0.0;1.0], conforme ilustra a Figura 11.

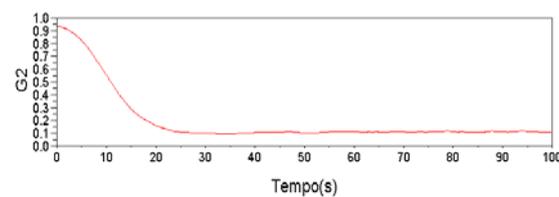


**Figura 11:** Perturbação randômica aplicada na variável S.

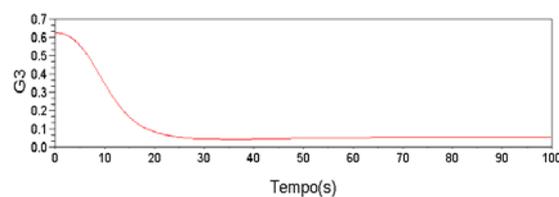
As respostas nas variáveis de saída são apresentadas nas Figuras 12 a 19.



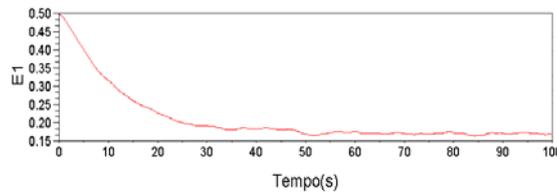
**Figura 12:** Gráfico da resposta da variável  $G_1$  á perturbação randômica aplicada.



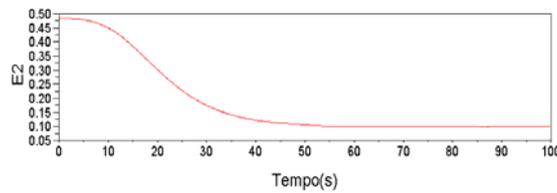
**Figura 13:** Gráfico da resposta da variável  $G_2$  á perturbação randômica aplicada.



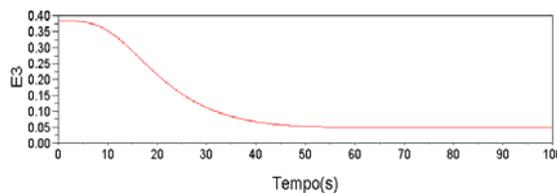
**Figura 14:** Gráfico da resposta da variável  $G_3$  á perturbação randômica aplicada.



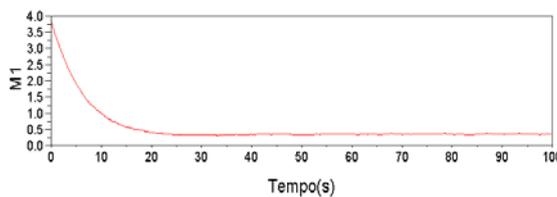
**Figura 15:** Gráfico da resposta da variável  $E_1$  à perturbação randômica aplicada.



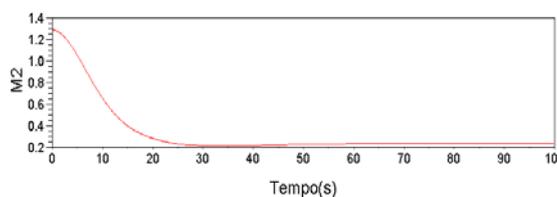
**Figura 16:** Gráfico da resposta da variável  $E_2$  à perturbação randômica aplicada.



**Figura 17:** Gráfico da resposta da variável  $E_3$  à perturbação randômica aplicada.



**Figura 18:** Gráfico da resposta da variável  $M_1$  à perturbação randômica aplicada.



**Figura 19:** Gráfico da resposta da variável  $M_2$  à perturbação randômica aplicada.

A perturbação randômica possibilita excitar a planta em vários pontos, permitindo obter um modelo muito mais condizente com a realidade, se a identificação for realizada utilizando os dados desse tipo de perturbação.

Pela análise dos gráficos gerados por esse tipo de perturbação percebemos que apenas a variável de estado  $G_1$  responde de forma mais intensa a perturbação randômica aplicada e as

outras variáveis respondem de forma mais suave. Concluindo que a variável  $G_1$  é a mais sensível à variação do substrato.

A partir dos dados gerados pela perturbação randômica, a identificação do modelo por representações em Espaço de Estados será realizada, como pode ser visto na próxima seção.

#### 4 - IDENTIFICAÇÃO DO SISTEMA

A Identificação do sistema já descrito será efetuada com o intuito de encontrar a melhor representação no Espaço de Estados do modelo, já que tal representação se mostra adequada em representações de sistemas não lineares.

##### 4.1 - Representação no Espaço de Estados

Uma das características básicas dessa representação é que ela descreve a dinâmica entre duas (ou mais de duas, no caso multivariável) variáveis, denominadas entrada e saída. Esse tipo de representação é válida no domínio do tempo e é mais conveniente para representar sistemas não-lineares e multivariáveis do que a função ou matriz de transferência. Um modelo linear típico em espaço de estados tem a seguinte forma:

$$\begin{aligned}\dot{x} &= Ax + Bu \\ y &= Cx + Du,\end{aligned}\tag{9}$$

sendo que  $x \in \mathbb{R}^n$  é o vetor de estado  $n$ -dimensional, o ponto indica a derivada temporal;  $u(t) \in \mathbb{R}^r$  é o vetor de entradas formado por  $r$  funções temporais;  $y(t) \in \mathbb{R}^p$  é o vetor  $p$ -dimensional de saídas medidas e  $A$ ,  $B$ ,  $C$  e  $D$  são matrizes constantes. O modelo será multivariável se  $r > 1$  e/ ou  $p > 1$ . Se houver apenas uma entrada  $r = 1$  e uma saída  $m = 1$ , o modelo é dito monovariável.

Duas observações importantes sobre a representação em espaço de estados são: *i*) o conhecimento do vetor de estado em qualquer instante  $t_0$  especifica o estado ou condição do sistema nesse instante; *ii*) a representação em espaço de estados não é única (Aguirre, 2007).

As possíveis representações no espaço de estados do modelo em questão foram obtidas pelo código Scilab e avaliados através de índices e critérios adequados. Dos 1000 dados

gerados pela perturbação randômica, 500 foram usados para a identificação e 500 para a validação do modelo.

A avaliação de um modelo pode ser executado através de índices e critérios estruturais, como MSE, AIC, FPE e BIC. O índice MSE (erro quadrático médio) é dado pela Eq. (10):

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (10)$$

sendo que  $n$  representa o número de amostras,  $y_i$  é o valor de referência e  $\hat{y}_i$  é o valor previsto pelo modelo para a  $i$ -ésima amostra.

O critério AIC é obtido através da função de Máxima Verossimilhança, a partir dos parâmetros ajustados para os modelos conforme os métodos. Este critério foi desenvolvido a partir da distância ou informação de Kulback e Leibler (1951). Esta distância é uma medida de discrepância entre as linhas do modelo verdadeiro e o modelo aproximado. Akaike (1983) relacionou a distância de Kulback e Leibler com a Máxima Verossimilhança, surgindo o AIC. O AIC é dado pela seguinte Eq.(11):

$$AIC = -2 \log L + 2K \quad (11)$$

sendo  $L$  a Verossimilhança Maximizada do modelo candidato e  $K$ , o número de parâmetros deste modelo.

Já o critério do erro de predição final (FPE), proposto por Akaike (1969) é dado pela Eq.(12):

$$FPE(p) = \hat{\rho}_p \left( \frac{N + (p+1)}{N - (p+1)} \right) \quad (12)$$

na qual  $N$  é o número de amostras,  $p$  é a ordem do modelo e  $\hat{\rho}_p$  é a estimativa da variância do ruído branco de entrada para o modelo de ordem  $p$ .

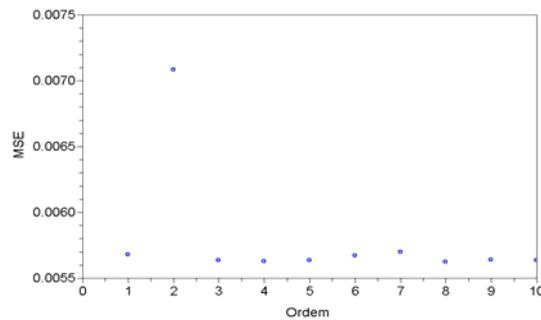
E finalmente, o índice BIC é o critério B de informação, que usa conceitos estatísticos de Bayes (Akaike, 1977; Akaike, 1978; Schwartz, 1978). A função a ser minimizada pela ordem ótima é dada pela Eq.(13):

$$BIC(p) = N \cdot \ln(\hat{\rho}_p) - (N - p) \cdot \ln\left(1 - \frac{p}{N}\right) + p \cdot \ln(N) + p \cdot \ln\left\{p^{-1} \cdot \left(\frac{\hat{\rho}_x}{\hat{\rho}_p} - 1\right)\right\} \quad (13)$$

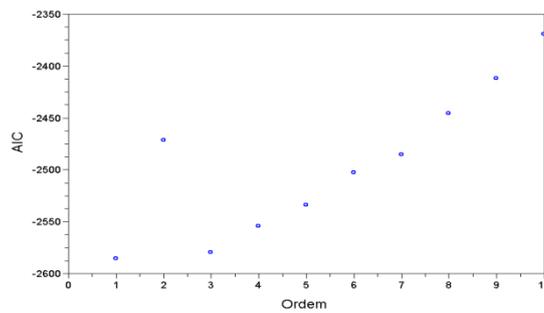
na qual  $\hat{\rho}_x$  é a variância da saída do modelo AR (Shirotai *et al.*, 2002).

A avaliação desses índices e critérios estruturais em modelos do espaço de estados com ordens variando de 1 a 10 possibilita escolher qual desses modelos melhor representa cada variável do sistema de vias bioquímicas já apresentado.

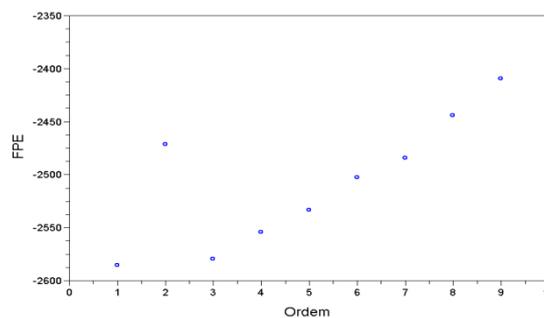
Para a variável  $G_1$ , o modelo de ordem 1 é o que melhor representa o processo, por apresentar menores índices e critérios e menor esforço computacional. As Figuras 20 a 23 descrevem as avaliações executadas para essa variável.



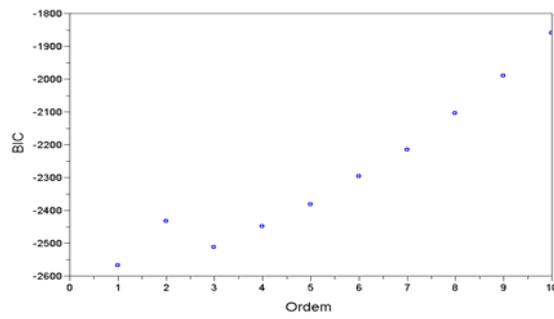
**Figura 20:** Avaliação do índice MSE para a variável  $G_1$ .



**Figura 21:** Avaliação do critério AIC para a variável  $G_1$ .



**Figura 22:** Avaliação do critério FPE para a variável  $G_1$ .



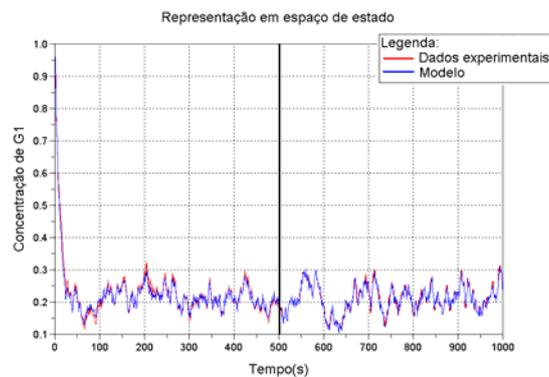
**Figura 23:** Avaliação do índice BIC para a variável  $G_1$ .

Essas mesmas avaliações foram realizadas em todos os modelos de cada uma das outras variáveis. Os menores índices e melhores modelos da representação de espaço de estados que descreve o sistema em questão, estão apresentados na tabelas 3.

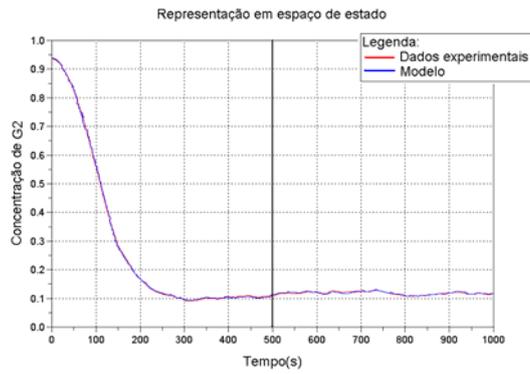
**Tabela 3:** Menores valores de MSE, AIC, FPE, BIC e melhor modelo para cada variável.

Variável	Ordem	MSE	AIC	FPE	BIC
$G_1$	1	0,00559	-2596,69	-2596,6962	-2579,8379
$G_2$	1	0,09946	-1334,69	-1334,6882	-1317,8300
$G_3$	1	0,04858	-1683,92	-1683,9282	-1667,0699
$E_1$	2	0,01316	-2392,57	-2392,5643	-2354,6347
$E_2$	1	0,04049	-2118,24	-2118,2376	-2101,3794
$E_3$	1	0,03156	-2237,73	-2237,7318	-2220,8736
$M_1$	1	0,76255	-235,419	-235,41966	-218,56140
$M_2$	1	0,14051	-1094,17	-1094,1446	-1077,3163

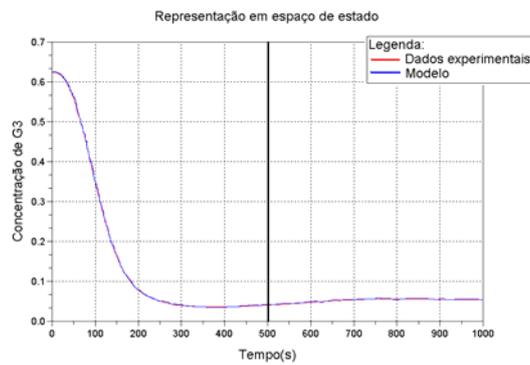
Nas próximas Figuras 24 a 31 são apresentadas as identificações e as validações no espaço de estados.



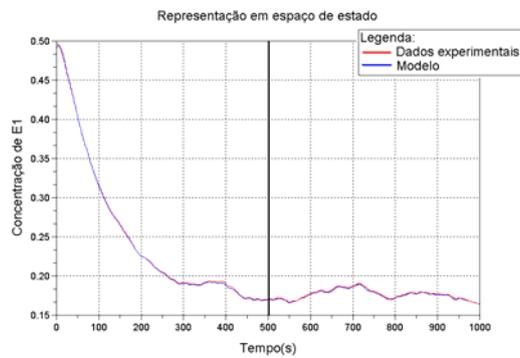
**Figura 24:** Identificação e validação da representação em Espaço de Estados da variável  $G_1$ .



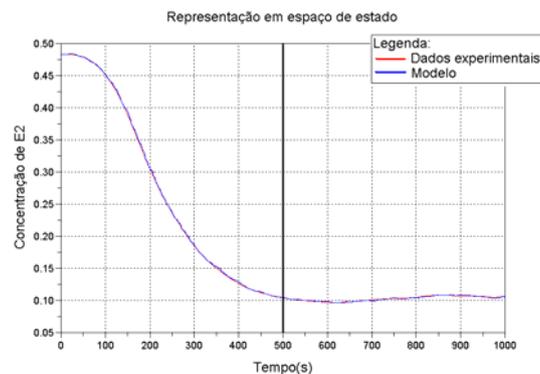
**Figura 25:** Identificação e validação da representação em Espaço de Estados da variável  $G_2$ .



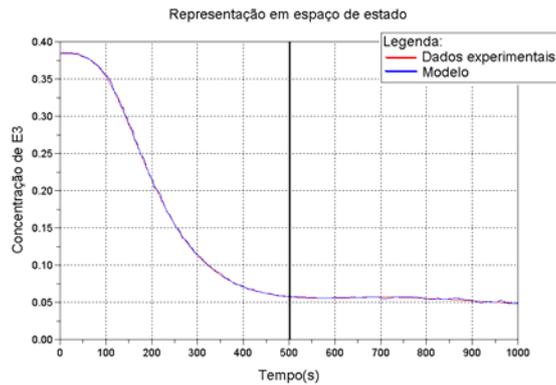
**Figura 26:** Identificação e validação da representação em Espaço de Estados da variável  $G_3$ .



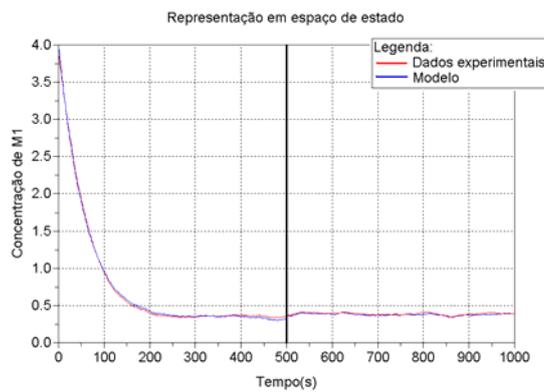
**Figura 27:** Identificação e validação da representação em Espaço de Estados da variável  $E_1$ .



**Figura 28:** Identificação e validação da representação em Espaço de Estados da variável  $E_2$ .



**Figura 29:** Identificação e validação da representação em Espaço de Estados da variável  $E_3$ .

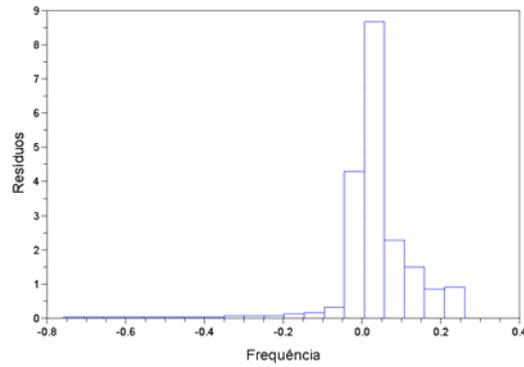


**Figura 30:** Identificação e validação da representação em Espaço de Estados da variável  $M_1$ .

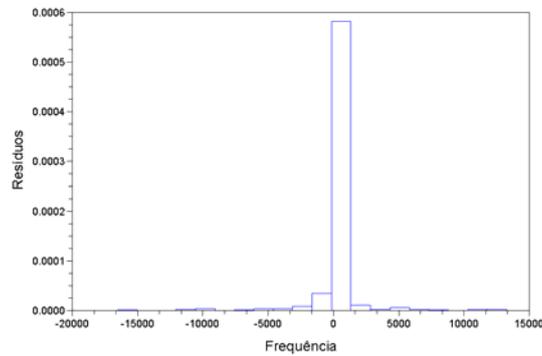


**Figura 31:** Identificação e validação da representação em Espaço de Estados da variável  $M_2$ .

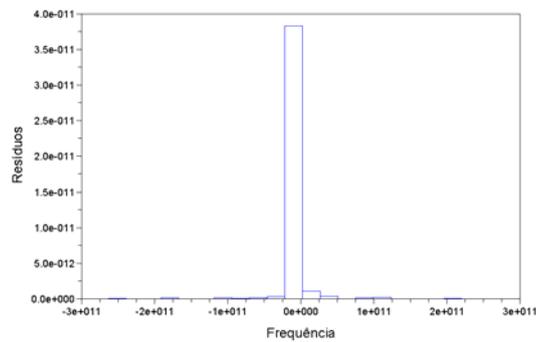
Uma análise de resíduos foi realizada para conferir os erros cometidos na identificação do modelo. Os histogramas dos resíduos na modelagem das equações que descrevem cada uma das variáveis são apresentados nas Figuras 32 a 39.



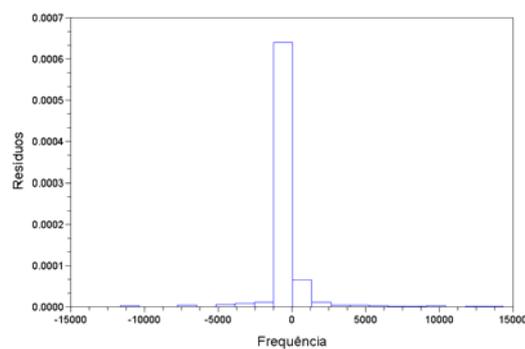
**Figura 32:** Histograma dos resíduos da identificação do modelo em espaço de estados da variável  $G_1$ .



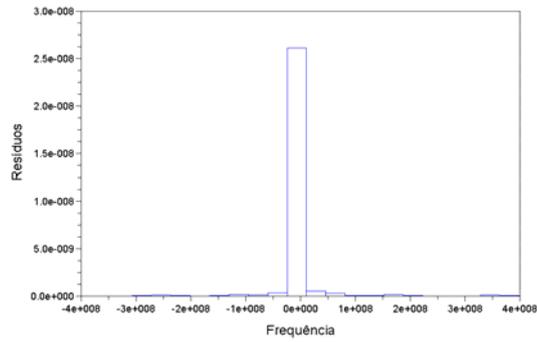
**Figura 33:** Histograma dos resíduos da identificação do modelo em espaço de estados da variável  $G_2$ .



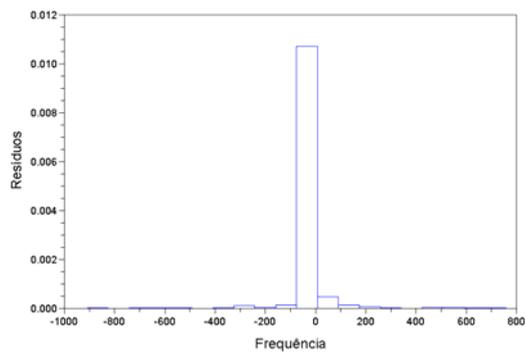
**Figura 34:** Histograma dos resíduos da identificação do modelo em espaço de estados da variável  $G_3$ .



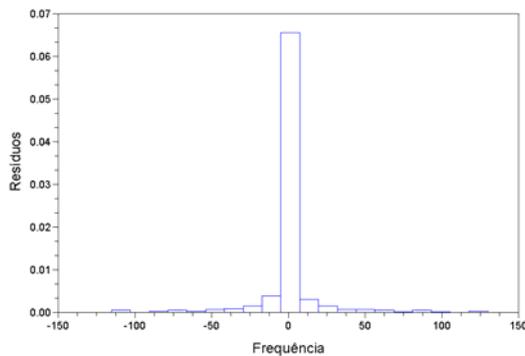
**Figura 35:** Histograma dos resíduos da identificação do modelo em espaço de estados da variável  $E_1$ .



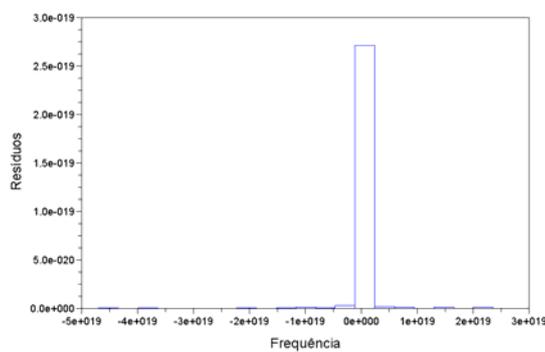
**Figura 36:** Histograma dos resíduos da identificação do modelo em espaço de estados da variável  $E_2$ .



**Figura 37:** Histograma dos resíduos da identificação do modelo em espaço de estados da variável  $E_3$ .



**Figura 38:** Histograma dos resíduos da identificação do modelo em espaço de estados da variável  $M_1$ .



**Figura 39:** Histograma dos resíduos da identificação do modelo em espaço de estados da variável  $M_2$ .

Realizando uma análise de todos os histogramas, percebe-se que a modelagem apresentou poucos erros. O modelo em espaço de estados, na forma matricial, fornecido pelo Scilab, é dado pela Eq. (9), já apresentada.

As matrizes para a variável de estado  $G_1$  são as seguintes:  $A=(0,8900701)$ ,  $B=(-0.0903895)$ ,  $C=(-0.4610149)$  e  $D=(0.0473016)$ . Para a variável  $G_2$  são:  $A=(0.9929008)$ ,  $B=(-0.0010272)$ ,  $C=(-0.2706209)$  e  $D=(-0.0018396)$ . Para a variável  $G_3$  são:  $A=(0.9927234)$ ,  $B=(0.0006186)$ ,  $C=(-0.2709079)$  e  $D=(-0.0008190)$ . Para a variável  $E_1$  são:  $A=\begin{pmatrix} 0,9912469 & -0,2262625 \\ 0,0004377 & 0,8904758 \end{pmatrix}$ ,  $B=\begin{pmatrix} -0,0069941 \\ 0,0027140 \end{pmatrix}$ ,  $C=(-0.2737446 \quad -0.5205832)$  e  $D=(0.0001644)$ . Já para a variável  $E_2$ , temos que:  $A=(0.9981969)$ ,  $B=(0.0023839)$ ,  $C=(-0.2612702)$  e  $D=(-0.0005780)$ . As matrizes para a variável  $E_3$  são:  $A=(0.9981965)$ ,  $B=(0.0026950)$ ,  $C=(-0.2612808)$  e  $D=(-0.0002574)$ . Para a variável  $M_1$  são:  $A=(0.9831151)$ ,  $B=(-0.0368776)$ ,  $C=(-0.2891654)$  e  $D=(0.0112919)$ . Finalmente, para a variável  $M_2$ , as matrizes do modelo em espaço de estados são:  $A=(0.9908611)$ ,  $B=(-0.0093538)$ ,  $C=(-0.2743558)$  e  $D=(0.0016358)$ .

## 5 – SINTONIA DO MODELO

A sintonia do modelo em questão será efetuada através do algoritmo da Evolução Diferencial implementado no código livre Scilab. Tal método é utilizado para a estimação dos parâmetros presentes no modelo.

### 5.1 - Identificação dos parâmetros utilizando evolução diferencial

Utilizando a Evolução Diferencial, foram efetuadas 10 execuções para encontrar os 36 parâmetros do modelo descrito nas Equações de 1 a 8. Com o objetivo de tornar a identificação mais realista, foram considerados um novo conjunto de dados com erro relativo de medição normalmente distribuído de 5%. Estes dados foram gerados usando a equação (14):

$$y^*(i) = y(i) \pm \sigma \quad (14)$$

sendo  $v$  uma variável aleatória com distribuição normal com média zero e desvio padrão unitário e  $\sigma$  é o desvio padrão de erros experimentais adicionados ao  $y^*$ . O erro relativo  $r$  é usado para definir esse desvio padrão, isto é,  $\sigma = r * y(i)$ .

Os melhores valores encontrados pela Evolução, juntamente com os parâmetros teóricos estão apresentados na Tabela 4, a seguir. Para comparação entres esses valores, a Tabela também apresenta os erros relativos entres eles.

**Tabela 4:** Valores identificados e teóricos dos parâmetros e os erros relativos entre eles.

Parâmetro	Valor teórico	Valor identificado	Erro relativo	Parâmetro	Valor teórico	Valor identificado	Erro relativo
$V_1$	1,00	1,001000	0,10%	$V_4$	0,10	0,100100	0,01%
$Ki_1$	1,00	1,001000	0,10%	$K_4$	1,00	0,999162	0,03%
$ni_1$	2,00	2,000100	0,01%	$k_4$	0,10	0,100011	0,00%
$Ka_1$	1,00	0,999238	0,08%	$V_5$	0,10	0,099968	0,03%
$na_1$	2,00	1,999257	0,04%	$K_5$	1,00	1,001000	0,10%
$k_1$	1,00	0,999145	0,09%	$k_5$	0,10	0,099968	0,03%
$V_2$	1,00	1,001000	0,10%	$V_6$	0,10	0,100004	0,00%
$Ki_2$	1,00	1,001000	0,10%	$K_6$	1,00	1,000704	0,01%
$ni_2$	2,00	1,999373	0,04%	$k_6$	0,10	0,100100	0,00%
$Ka_2$	1,00	0,999000	0,08%	$kcat_1$	1,00	0,999138	0,08%
$na_2$	2,00	1,999000	0,04%	$Km_1$	1,00	1,000604	0,05%
$k_2$	1,00	0,999678	0,09%	$Km_2$	1,00	1,000462	0,05%
$V_3$	1,00	0,999618	0,09%	$kcat_2$	1,00	0,999000	0,08%
$Ki_3$	1,00	1,001000	0,10%	$Km_3$	1,00	1,000338	0,05%
$ni_3$	2,00	1,999572	0,04%	$Km_4$	1,00	0,999409	0,08%
$Ka_3$	1,00	1,001000	0,10%	$kcat_3$	1,00	0,999000	0,08%
$na_3$	2,00	2,001000	0,10%	$Km_5$	1,00	1,000462	0,05%
$k_3$	1,00	1,000231	0,01%	$Km_6$	1,00	0,999459	0,08%

A Figura 40 apresenta o experimento numérico que apresentou o melhor resultado.

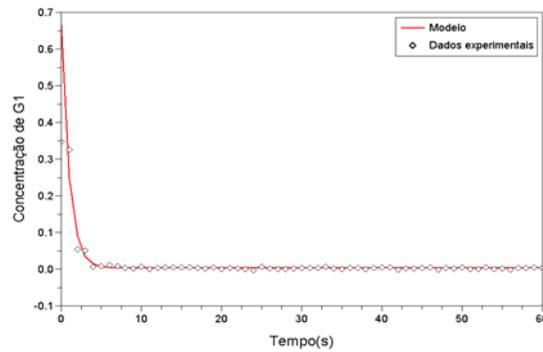
```

*** C A L C U L A N D O ...
Iteration: 10, Best: 359.223396, F: 0.800000, CR: 0.800000, NP: 72
best(1) = 1.001000
best(2) = 1.001000
best(3) = 2.000100
best(4) = 0.999238
best(5) = 1.999257
best(6) = 0.999145
best(7) = 1.001000
best(8) = 1.001000
best(9) = 1.999373
best(10) = 0.999000
best(11) = 1.999000
best(12) = 0.999678
best(13) = 0.999618
best(14) = 1.001000
best(15) = 1.999572
best(16) = 1.001000
best(17) = 2.001000
best(18) = 1.000231
best(19) = 0.100100
best(20) = 0.999162
best(21) = 0.100011
best(22) = 0.099968
best(23) = 1.001000
best(24) = 0.099968
best(25) = 0.100004
best(26) = 1.000704
best(27) = 0.100100
best(28) = 0.999138
best(29) = 1.000604
best(30) = 1.000462
best(31) = 0.999000
best(32) = 1.000338
best(33) = 0.999409
best(34) = 0.999000
best(35) = 1.000462
best(36) = 0.999459

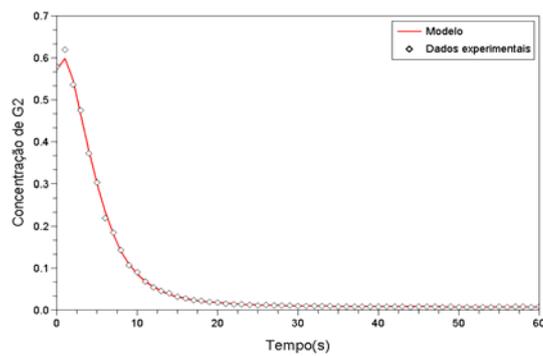
```

**Figura 40:** Arquivo de saída do experimento numérico que resultou na melhor estimativa dos parâmetros do modelo.

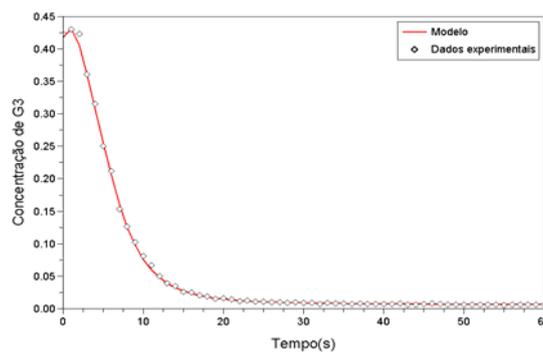
Os parâmetros identificados pela evolução diferencial foram bem próximos dos teóricos, enfatizando que o modelo com os parâmetros encontrados descreve de forma satisfatória o processo. As Figuras 41 a 48 representam os dados experimentais com erro e o modelo utilizando os parâmetros identificados na evolução diferencial.



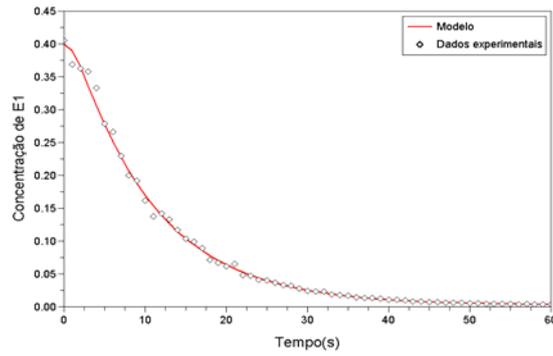
**Figura 41:** Validação do modelo com os parâmetros identificados por comparação com os dados experimentais para a variável  $G_1$ .



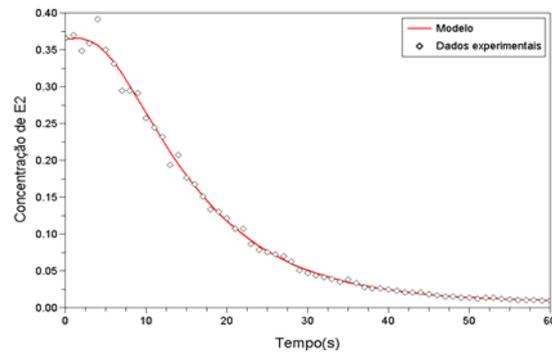
**Figura 42:** Validação do modelo com os parâmetros identificados por comparação com os dados experimentais para a variável  $G_2$ .



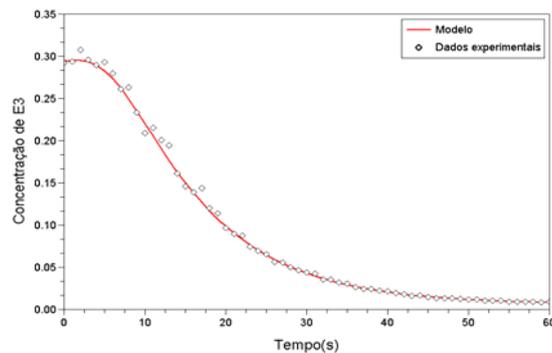
**Figura 43:** Validação do modelo com os parâmetros identificados por comparação com os dados experimentais para a variável  $G_3$ .



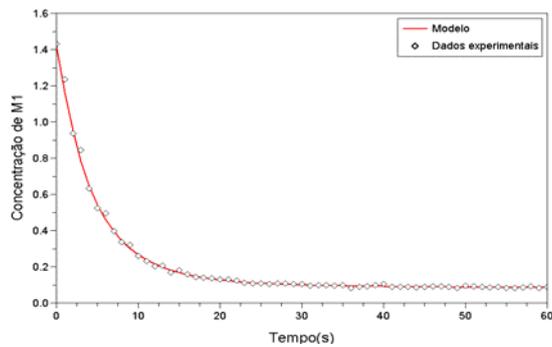
**Figura 44:** Validação do modelo com os parâmetros identificados por comparação com os dados experimentais para a variável  $E_1$ .



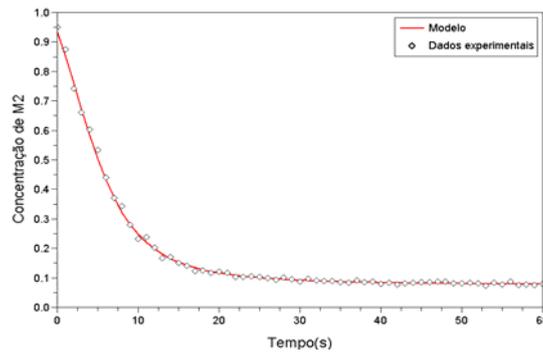
**Figura 45:** Validação do modelo com os parâmetros identificados por comparação com os dados experimentais para a variável  $E_2$ .



**Figura 46:** Validação do modelo com os parâmetros identificados por comparação com os dados experimentais para a variável  $E_3$ .



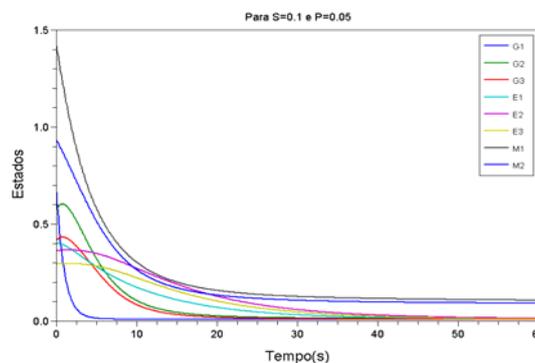
**Figura 47:** Validação do modelo com os parâmetros identificados por comparação com os dados experimentais para a variável  $M_1$ .



**Figura 48:** Validação do modelo com os parâmetros identificados por comparação com os dados experimentais para a variável  $M_2$ .

## 6 - RESUSTADOS

O perfil do modelo pode ser obtido tendo os valores dos seus parâmetros e as condições iniciais de cada variável de estado. Considerando o experimento 1 descritos na Tabela 1 e as condições iniciais da Tabela 2, a dinâmica do processo foi desenvolvida no código livre Scilab com os parâmetros identificados na Evolução Diferencial. A Figura 49 representa essa dinâmica.



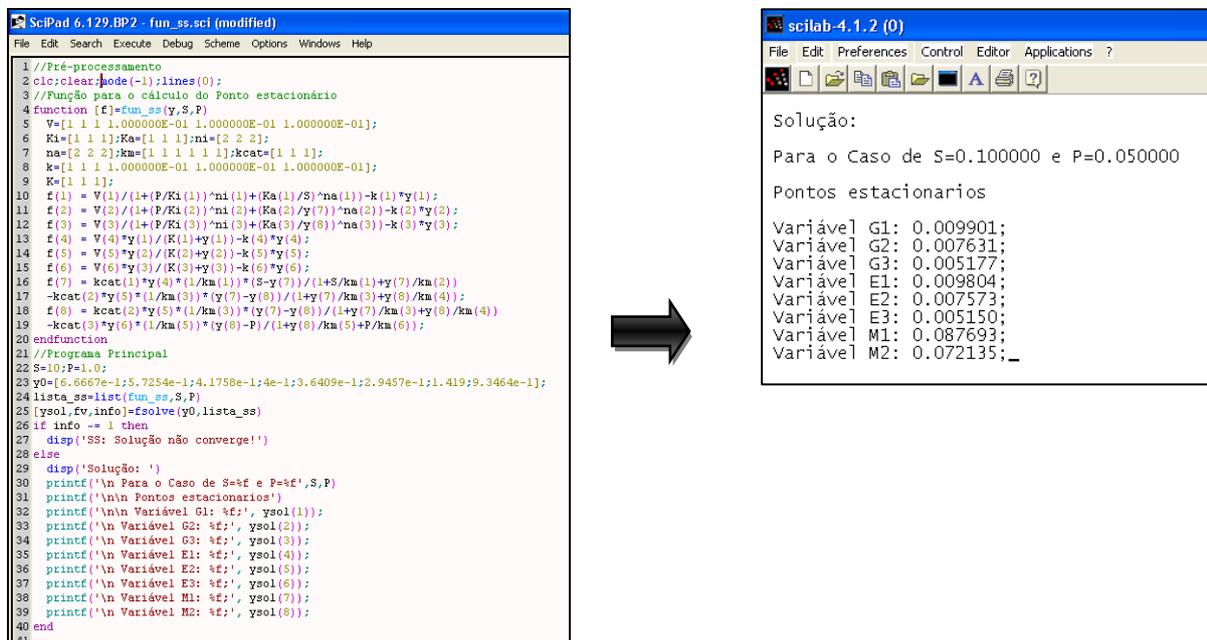
**Figura 49:** Variação dinâmica das concentrações no experimento 1.

Os respectivos estados estacionários das variáveis de estado no Experimento 1 são apresentados na Tabela 5. Esses valores foram obtidos através da utilização da função **fsolve** do Scilab.

**Tabela 5:** Valores de estado estacionário das variáveis de estados no Experimento 1

Variável de estado	Estado Estacionário
$G_1$	0.009901
$G_2$	0.007631
$G_3$	0.005177
$E_1$	0.009804
$E_2$	0.007573
$E_3$	0.005150
$M_1$	0.087693
$M_2$	0.072135

O algoritmo desenvolvido com seu respectivo resultado é apresentado na Figura 50.



**Figura 50:** O algoritmo desenvolvido no Scilab para o experimento 1, com seu respectivo resultado.

## 7 – CONCLUSÃO

Considerando que o modelo apresenta 36 parâmetros a serem identificados e pela análise dos erros relativos entre os valores teóricos e os valores identificados, pode-se concluir que a sintonia foi excelente e os parâmetros foram identificados de forma satisfatória. O código livre Scilab mostrou-se uma excelente alternativa como ferramenta de identificação de sistemas lineares e não lineares.

## 8 – AGRADECIMENTOS

À FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) e ao PIBIC (Programa Institucional de Bolsa de Iniciação Científica) pelo apoio financeiro.

## 9 -REFERÊNCIAS BIBLIOGRÁFICAS

AGUIRRE, L. A. (2007). Introdução à Identificação de Sistemas: Técnicas Lineares e Não Lineares Aplicadas a Sistemas Reais, 3 ed., Editora UFMG, Belo Horizonte, BH.

ARANTES, M. B., OLIVEIRA, G.T.S. e SARAMAGO, S.F.P., *Evolução Diferencial aplicada à solução de alguns problemas de Engenharia de Produção*, FAMAT em revista, nº 6, Maio de 2006, Uberlândia, MG.

HODGSON, B.J., TAYLOR, C.N., USHIO, M., LEIGH, J.R., KALGANOVA, T. E BAGANZ, F., (2004). Intelligent modelling of bioprocesses: a comparison of structured and unstructured approaches. *Bioprocess and Biosystems Engineering*, 26 (6), 353-359.

MENDES, P., KELL, D.B., 1998. *Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation*. *Bioinformatics* 14 (10), 869–883.

RAO, M.S. E IYENGAR, S.S., (1984). Application of Statistical Techniques in Modelling of Complex Systems. Em: Iyengar, S.S. (ed), *Computer Modeling of complex Biological Systems*. pp. 29-53, CRC Press, Boca Raton, Florida (USA).

RODRIGUEZ-FERNANDEZ, M., MENDES, P., BANGA, J.R., 2006. *A hybrid approach for efficient and robust parameter estimation in biochemical pathways*. *BioSystems* 83, 248-265.

SHIROTAI, C. e ITIKI, C., 2002, *Determinação da ordem na modelagem autoregressiva de sinais eletromiográficos*, Produção em Iniciação Científica da EPUSP, São Paulo, SP.

VANROLLEGHEM, P.A. E DOCHAIN, D., (1998). Bioprocess model identification. Em: Van Impe, J.F.M., Vanrollegheem, P.A. e Iserentant, D.M. (eds), *Advanced*

*Instrumentation Data Interpretation, and Control of Biotechnological Processes.* pp. 251-318, Kluwer Academic Publishers, Holanda.

VELOSO, A.C.A. *Optimização de Estratégias de Alimentação para a Identificação de Parâmetros de um Modelo de Utilização do Modelo em Monitorização e Controlo.* Portugal. Tese (Doutorado em Engenharia Química e Biológica) Universidade do Minho, 2006.