

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA QUÍMICA
ÁREA DE CONCENTRAÇÃO: SISTEMAS DE PROCESSOS
QUÍMICOS E INFORMÁTICA

**PROCESSO DE PRODUÇÃO DE BIOEMULSIFICANTE
POR *Candida lipolytica*:
OTIMIZAÇÃO, AMPLIAÇÃO DE ESCALA E
DESENVOLVIMENTO DE SOFTSENSOR BASEADO EM
REDES NEURAIS ARTIFICIAIS.**

Tese apresentada à Faculdade de Engenharia Química, como
parte dos requisitos exigidos para obtenção do grau de Doutor
em Engenharia Química.

Autora: Clarissa Daisy da Costa Albuquerque ✓

Orientadora: Prof^a. Dr^a. Ana Maria Frattini Fileti

Co-Orientadora: Prof^a. Dr^a. Galba Maria de Campos Takaki ✓

Campinas
São Paulo
2006

BIBLIOTECA CENTRAL
DESENVOLVIMENTO
COLEÇÃO
UNICAMP

UNIDADE BC
Nº CHAMADA TI UNICAMP
AL15p
V EX
TOMBO BC/ 69947
PROC 16.123.06
C D X
PREÇO 11,00
DATA 11/09/06
Nº CPD

BIB ID: 386341

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

AL15p

Albuquerque, Clarissa Daisy da Costa

Processo de produção de biomulsificante por *Candida lipolytica*: otimização, ampliação de escala e desenvolvimento de softsensor baseado em redes neurais artificiais. / Clarissa Daisy da Costa Albuquerque. -- Campinas, SP: [s.n.], 2006.

Orientadores: Ana Maria Frattini Fileti, Galba Maria de Campos Takaki
Tese (Doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Química.

1. Biosensores. 2. Detectores. 3. Redes neurais (Computação). 4. Inteligência artificial - Processamento de dados. 5. Agentes ativos de superfícies. 6. Surfactantes biodegradáveis. 7. *Candida lipolytica*. 8. Otimização matemática. 9. Planejamento experimental. 10. Superfícies de resposta (Estatística). I. Fileti, Ana Maria Frattini. II. Takaki, Galba Maria de Campos. III. Universidade Estadual de Campinas. Faculdade de Engenharia Química. IV. Título.

Titulo em Inglês: Palavras-chave em Inglês: Biomulsifier production process by *Candida lipolytica*: optimization, scale-up and development of artificial neural network based softsensor

Área de concentração: Sistemas de Processos Químicos e Informática

Titulação: Doutora em Engenharia Química

Banca examinadora: Ricardo Ribeiro Gudwin, Aline Carvalho da Costa, Flávio Vasconcelos da Silva, Maria da Graça Stuppiello Andrietta e Elias Basile Tambourgi

Data da defesa: 22/02/2006

66.9000

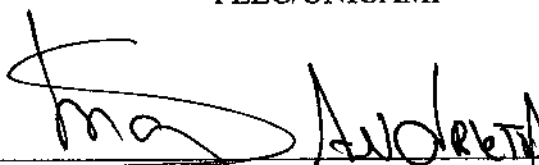
Tese de doutorado defendida por Clarissa Daisy da Costa Albuquerque e aprovada em 22 de fevereiro de 2006 pela Banca constituída pelos doutores:



Prof.^a Dr.^a Ana Maria Frattini Fileti – Orientadora
FEQ/UNICAMP



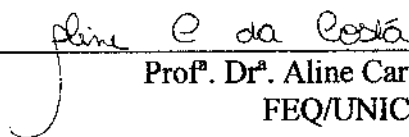
Prof. Dr. Ricardo Gudwin
FEEC/UNICAMP



Dr.^a Maria da Graça Stupiello Andrietta
CPOBA/UNICAMP



Prof. Dr. Flávio Vasconcelos da Silva
FEQ/UNICAMP

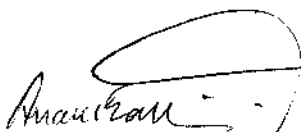


Prof.^a Dr.^a Aline Carvalho da Costa
FEQ/UNICAMP



Prof. Dr. Elias Basile Tambourgi
FEQ/UNICAMP

Esta redação corresponde à versão final da Tese de Doutorado defendida por Clarissa Daisy da Costa Albuquerque em 22 de fevereiro de 2006.

A handwritten signature in black ink, appearing to read 'Ana Maria', is written over a horizontal line. The signature is stylized with a large, looping flourish that extends to the right.

Profa. Dra. Ana Maria Frattini Fileti
Orientador

**A minha família,
minha estrela guia.**

Uma Águia Chamada "Circunstâncias"

A águia empurrou gentilmente seus filhotes para a beirada do ninho.
Seu coração se acelerou com emoções conflitantes,
ao mesmo tempo em que
sentiu a resistência dos
filhotes a seus insistentes cutucões.

Pensou ela:
Por que a emoção de voar tem que começar com o medo de cair?
O ninho estava colocado bem no alto de um pico rochoso.
Abaixo, somente o abismo e o ar para sustentar as asas dos filhotes.
E se justamente agora isto não funcionar? Ela pensou.

Apesar do medo, a águia sabia que aquele era o momento.
Sua missão estava prestes a se completar;
restava ainda uma tarefa final:
o empurrão.

A águia encheu-se de coragem.
Enquanto os filhotes não descobrirem suas asas,
não haverá propósito para as suas vidas.
Enquanto eles não aprenderem a voar,
não compreenderão o privilégio que é nascer águia.

O empurrão era o melhor presente que ela poderia oferecer-lhes.
Era seu supremo ato de amor.
Então, um a um, ela os precipitou para o abismo.
E eles voaram!

Às vezes, nas nossas vidas, as circunstâncias fazem o papel de águia.
São elas que nos empurram para o abismo.
E quem sabe não são elas, as próprias circunstâncias, que nos fazem
descobrir que temos asas para voar?

AGRADECIMENTOS

A Prof^ª. Dr^ª. Ana Maria Frattini Fileti pela orientação, apoio, incentivo, amizade e confiança depositada em todos os momentos e pelo compartilhamento de sua experiência e conhecimentos científicos.

A Prof^ª.Dr^ª. Galba Maria de Campos-Takaki pela co-orientação, apoio, incentivo, amizade e confiança depositadas em todos os momentos e pelos exemplos de fé, força, trabalho, equilíbrio, determinação, persistência, visão e capacidade científicas.

A Banca Examinadora formada pelos Profs. Drs. Aline Carvalho Costa, Elias Basile Tambourgi, Flávio Vasconcelos da Silva, Maria da Graça Stupiglio Andrietta e Ricardo Gudwin pelas sugestões fornecidas.

A Universidade Católica de Pernambuco pelos suportes científico, técnico e financeiro fornecidos durante o desenvolvimento desta tese.

Ao Magnífico Reitor da Universidade Católica de Pernambuco, Prof.Dr.Pe. Theodoro Peters pela minha liberação para realizar doutorado na Universidade Estadual de Campinas.

Ao Pró-Reitor de Pesquisa e Pós-Graduação da Universidade Católica de Pernambuco, Prof.Dr. Junnot Cornélio Matos pelo apoio e pareceres favoráveis às minhas licenças para realizar o doutorado na Universidade Estadual de Campinas.

Ao Chefe do Departamento de Estatística e Informática Prof. Jessé Gomes de Oliveira pela amizade, incentivo e apoio à realização deste trabalho.

À Prof^ª.Dr^ª. Maria do Carmo Rocha Marçal, à Prof^ª.Dr^ª. Leonie Asfora Sarubbo e à Dr^ª. Mabel Hannah Vance-Harrop pelos trabalhos precursores na área de biopolímeros com atividade de emulsificação, fundamentais para o desenvolvimento desta tese.

Ao Prof. Edson Gomes de Souza e ao Prof. Dr.Francisco Luiz dos Santos pela amizade, pelo apoio técnico e pelo uso de instalações e instrumentos do Laboratório de Polímeros da Universidade Católica de Pernambuco.

Ao Prof. Dr. Florival Rodrigues de Carvalho pela amizade, apoio e incentivo a realização deste trabalho e pelo uso de instalações e instrumentos do Laboratório de Combustíveis da Universidade Federal de Pernambuco.

Ao Prof. João Teotônio Manzi pela permissão de uso por tempo indeterminado de licença do software Matlab pertencente ao Laboratório de Controle e Automação de Processos da Universidade Católica de Pernambuco.

Ao Prof. Francisco Mauger (DEA/UNICAMP) pela possibilidade de visitar as instalações do Laboratório de Engenharia de Bioprocessos (LEB) e acompanhar a montagem e operação do biorreator BioFlo III.

Ao Prof. José Luis (LIKA/UFPE) pela possibilidade de visitar as instalações do Laboratório de Imunopatologia Keizo Asami (LIKA) e acompanhar e participar da montagem, desmontagem e operação do Bioflo 2000.

Ao Prof.Dr.Daniel Atala, ao Prof.Dr.Luciano Armiliato, à Profª.Drª Luciana Franco e à Profª.Drª.Daniella Bruneska pela alegria e desprendimento científico ao transmitir experiências e fundamentos necessários para a operação do biorreator BioFlo 2000.

A Profª Drª. Alexandra Salgueiro pela amizade e generosidade científica demonstradas no convívio diário no Núcleo de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco.

A Sônia, Sr.Humberto e Salatiel pelo apoio técnico durante a realização da parte experimental deste trabalho no NPCIAMB/UNICAP.

A Helena, Léa, Gilberto, Almir, Robson Lins, Sílvia Bandeira, Robson Pequeno, Francisco Madeiro, Ana Elisa, Mozart, Karina e Múcio Gomes pela amizade e apoio em momentos difíceis deste trabalho.

Aos demais colegas e amigos do Núcleo de Ciências em Pesquisas Ambientais e da Universidade Católica de Pernambuco, e em especial aos do Departamento de Estatística e Informática - César, Cristina, Fernando Bertino, Fernando César, Elianete, Gilka, Ilzes e Sílvia - pelo incentivo e apoio durante os anos do doutorado.

A Terezinha Ponce de Leon pela amizade, apoio e incentivo para que realizasse o doutorado.

A Sylvinha e Nicole pelo desprendimento, solidariedade, companheirismo, carinho e amizade demonstrados na minha chegada em Campinas.

A Sueli, Cíntia e Marlei pela amizade e bons momentos compartilhados em Campinas, na UNICAMP e em “nossa república”.

A D.Conceição, Vanda, Janaina, Sione, Nilza, Letícia, Lili, Marzué, Bárbara, Danielle, Tina, Eliana, Azize, Maruska e Renita, pela amizade e alegria e por me fazerem rir e me sentir em casa em momentos tensos deste trabalho.

Aos colegas Arnaldo Antunes, Fábio Certo, Gaston Buonous, Rafael, Paulo Pereira, Tatiana Pacianotto, Fábio Delattore, Luis Bertevello, Regina Santos, Marcele Carulo, César Fraia, Thaís, Valéria, Priscila, Márcio, Ana, Júnior, Adriana, Ângela, Juliana, Kleber, Sérgio, Jones, Roberto, Christian, Rafael, Renilson, Giovana, Débora, Richard, Martin, Patrícia e Rose, pela agradável convivência e horas alegres compartilhadas no Laboratório de Controle e Automação de Processos (LACAP/FEQ/DESQ/UNICAMP), na FEQ, na UNICAMP e em Campinas.

Aos colegas Pervane, Gustavo e Juliano pela amizade e pela eficiência na administração dos computadores, periféricos e problemas dos “Bichinhos Inteligentes Operando os Sistemas (BIOS)” do Departamento de Engenharia de Sistemas Químicos da Universidade Estadual de Campinas.

Aos colegas Ricardo Kenji, Mabel Kalina, Norma Sueli, Patrícia, Thaysa, Mariluce, Marcos Moraes, Marcos Lima, Raquel, Petrusk, Juliana, Charles, Thayse e Fabíola pela convivência sadia e enriquecedora nos laboratórios do NPCIAMB/UNICAP, durante o tempo em que trabalhamos juntos.

A Luís Otávio Aquino (AMBRIEX/Milipore) pela excelência do atendimento e pela rapidez na obtenção de novos manuais sobre a instalação e a operação do biorreator BioFlo 2000® (New Brunswick Scientific Co. Inc., New Jersey).

Aos funcionários e engenheiros, especialmente a Mário Buarque, da Divisão de Serviços Gerais da Universidade Católica de Pernambuco e ao Sr. André Monteiro da JATOMAQ pela agilidade e eficiência na assistência e conserto do compressor de ar do biorreator Bioflo 2000.

A Universidade Estadual de Campinas pela oportunidade de realização do doutorado.

Ao CNPq, CnPq/CT-PETRO, FINEP/CT-PETRO pelo apoio financeiro.

E em especial, a quem com pequenos e silenciosos grandes gestos, sempre me apoiou, quando mais precisei.

RESUMO

Entre as técnicas convencionais usadas para desenvolvimento de *software sensores*, redes neurais artificiais têm mostrado ser um instrumento poderoso em modelagem e controle de bioprocessos complexos. O objetivo geral do presente trabalho é o desenvolvimento de *software sensores* baseados em redes neurais para estimação e previsão em tempo real da concentração de biomassa e da atividade de emulsificação no processo de produção de bioemulsificante por *Candida lipolytica*. Bioemulsificantes normalmente apresentam vantagens como biodegradabilidade, baixa toxicidade e biocompatibilidade em relação a emulsificantes sintetizados quimicamente. Adicionalmente, eles apresentam potencial para serem sintetizados a partir de substratos de baixo custo e são normalmente efetivos em condições extremas de pH, temperatura e salinidade. Consequentemente, bioemulsificantes têm sido aplicados com sucesso em áreas como biorremediação e recuperação de óleos. Contudo, os bioemulsificantes não são ainda largamente empregados por conta do seu alto custo de produção, resultante primeiramente da baixa produtividade dos microrganismos empregados e do alto custo de recuperação. Entretanto, neste trabalho foi mostrado que o desenvolvimento de *softsensores* neurais juntamente com otimização de componentes de meio de produção e ampliação de escala do processo podem contribuir para tornar a produção de bioemulsificante mais eficiente e econômica. O modelo quadrático obtido na otimização dos componentes do meio usando planejamento composto central com três fatores e metodologia de superfície de resposta mostrou significância estatística e capacidade preditiva. A máxima atividade de emulsificação para emulsões água-em-hexadecano obtida foi de 4,415 UAE e as concentrações ótimas de uréia, sulfato de amônio e fosfato monobásico de potássio foram respectivamente iguais a 0,544 % (w/v), 2,131 % (w/v) and 2,628 % (w/v). O rendimento do processo foi otimizado em 272%. A ampliação do processo da escala de frascos para a escala de fermentador de bancada foi realizada com sucesso e os efeitos e interações da temperatura e velocidade de agitação sobre a atividade de emulsificação do bioemulsificante produzido por *Candida lipolytica* foram investigados. Os conjuntos de dados necessários para o treinamento, validação e teste dos *softsensores* foram obtidos de experimentos de produção de bioemulsificante realizados em biorreator de 5L, sob diferentes condições de temperatura e agitação. Os conjuntos de treinamento, validação e teste dos *softsensores* foram suavizados e expandidos usando interpolação com *spline cúbica*. Várias topologias de redes neurais com uma camada escondida foram testadas. As variáveis de entrada do processo incluíram pH, oxigênio dissolvido, densidade ótica e salinidade do líquido metabólico livre de células. O algoritmo de treinamento usado foi o algoritmo de retropropagação baseado em Levenberg-Marquardt em conjunção com regularização bayesiana. A raiz do erro quadrático médio (RMSE) e o coeficiente de determinação global (R_g^2) foram usados entre outros índices para comparar o desempenho dos modelos. Os resultados mostram que *softsensores* neurais fornecem estimação e previsão on-line de concentração de biomassa e de atividade de emulsificação dentro de uma variação aceitável de 5% dos valores experimentais. Coeficientes de determinação global superiores a 0,90 indicam o excelente ajuste dos modelos de redes neurais com os valores experimentais testados, obtidos para concentração de biomassa e atividade de emulsificação.

Key words: *software sensores*, *softsensores*, redes neurais artificiais, bioemulsificantes, *Candida lipolytica*, otimização, metodologia de superfície de resposta, scale-up.

ABSTRACT

Among conventional techniques used for development of 'software sensors', artificial neural networks have showed to be a powerful tool for modelling and control of complex bioprocess. The present work deals with the development of neural network based software sensors for real time estimation and prediction of biomass concentration and emulsification activity in a bioemulsifier production process by *Candida lipolytica*. Bioemulsifiers commonly have the advantages of biodegradation, low toxicity, and biocompatibility over chemically synthesized emulsifiers. In addition, they can potentially be synthesized from cheap substrates and are commonly effective at extremes of pH, temperature, and salinity. As a result, bioemulsifiers have found successful application in areas such as bioremediation and oil recovery. However, bioemulsifiers are not widely available because of their high production costs, which results primarily from low strain productivities and high recovery expenses. Therefore, in this work was showed that on-line neural softsensor development jointly with media optimization and scale up of the process can make bioemulsifier production more efficient and more economical. The second order model obtained in the optimization of the medium components using three-factor central composite design and response surface methodology showed statistical significance and predictive ability. It was found that the maximum emulsification activity to water-in-hexadecane emulsion produced was 4,415 UEA and the optimum levels of urea, ammonium sulfate and potassium dihydrogen orthophosphate were, respectively, 0,544 % (w/v), 2,131 % (w/v) and 2,628 % (w/v). The emulsifier production process yield was optimized in 272 %. Successful scale-up from flasks to laboratory scale bioreactor was attained and the effects and interactions of the temperature and agitation rate on the emulsification activity of the bioemulsifier produced by *Candida lipolytica* were investigated. The data sets required to training, validation and test the neural software sensors were obtained from bioemulsifier production experiments carried out using corn oil and sea water based media in a 5L bioreactor, under different temperature and agitation conditions. The training, validation and test sets were smoothed and expanded by interpolation using a piecewise smoothing cubic spline. Several neural network topologies with one hidden layer were tested. The input process variables included pH, dissolved oxygen, optic density and free cell metabolic liquid salinity. The training algorithm used was the Levenberg-Marquardt based backpropagation algorithm, in conjunction with Bayesian regularization. The root mean square error (RMSE) and the global determination coefficient (R_g^2) among others index were used to compare model performances. The results showed that neural 'software sensors' supplied for biomass concentration and emulsification activity on-line estimation and prediction within an acceptable variation of 5% of the experimental values. Global coefficients of determination higher than 0.90 indicated excellent agreement of the neural network models with experimental test values, obtained for biomass concentration and emulsification activity.

Key words: software sensors, softsensors, artificial neural network, bioemulsifiers, *Candida lipolytica*, optimization, surface response methodology, scale-up.

SUMÁRIO

AGRADECIMENTOS.....	V
RESUMO	VIII
ABSTRACT	IX
SUMÁRIO	X
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	7
2.1. INTRODUÇÃO	7
2.2. PROCESSO DE PRODUÇÃO DE BIOSSURFACTANTES POR <i>CANDIDA LIPOLYTICA</i> A PARTIR DE SUBSTRATOS DE BAIXO CUSTO.....	10
2.3. SOFTWARES SENSORES BASEADOS EM REDES NEURAIIS APLICADOS A BIOPROCESSOS	19
3. OTIMIZAÇÃO DE COMPONENTES DE MEIOS DE PRODUÇÃO DE BIOEMULSIFICANTES POR <i>Candida lipolytica</i>	37
3.1. INTRODUÇÃO	37
3.2. MATERIAIS E MÉTODOS.....	39
3.2.1. <i>Microrganismo e Meio de Preservação</i>	39
3.2.2. <i>Meios de Cultivo e Produção de Bioemulsificante</i>	39
3.2.3. <i>Preparação do inóculo</i>	41
3.2.4. <i>Determinação da salinidade</i>	41
3.2.5. <i>Determinação da concentração da biomassa</i>	42
3.2.6. <i>Determinação da atividade de emulsificação</i>	42
3.2.7. <i>Determinação da Tensão Superficial</i>	42
3.2.8. <i>Otimização de meios de produção de bioemulsificantes</i>	43
3.3. RESULTADOS E DISCUSSÃO	46
3.3.1. <i>Planejamento Fatorial Completo 2⁴</i>	49
3.3.2. <i>Planejamento Composto Central para Três Fatores</i>	68
3.4. CONCLUSÃO	89
4. AMPLIAÇÃO DO PROCESSO DE PRODUÇÃO DE BIOEMULSIFICANTE POR <i>Candida lipolytica</i> DA ESCALA DE FRASCOS PARA ESCALA DE BIORREATOR DE BANCADA.....	91
4.1. INTRODUÇÃO	91
4.2. MATERIAIS E MÉTODOS.....	91
4.2.1. <i>Descrição do Equipamento</i>	91
4.2.2. <i>Microrganismo e Meio de Preservação</i>	92
4.2.3. <i>Preparação do Inóculo</i>	93
4.2.4. <i>Preparação dos Meios de Crescimento e Produção</i>	95
4.2.5. <i>Produção em Batelada de Bioemulsificante</i>	96
4.2.6. <i>Métodos Analíticos</i>	98
4.3. RESULTADOS E DISCUSSÃO	100
4.3.1. <i>Ensaaios Preliminares</i>	100
4.3.2. <i>Ensaaios do Planejamento Experimental</i>	101
4.4. CONCLUSÃO	134

5. DESENVOLVIMENTO DE SOFTSENSOR BASEADO EM REDES NEURAIS PARA ESTIMAÇÃO E PREVISÃO DE BIOMASSA E DE ATIVIDADE DE EMULSIFICAÇÃO DE BIOEMULSIFICANTE PRODUZIDO POR <i>Candida lipolytica</i>	135
5.1. INTRODUÇÃO	135
5.2. MATERIAIS E MÉTODOS	136
5.2.1. Metodologia de desenvolvimento dos softsensores neurais	136
5.2.2. Pré-Processamento dos Dados	137
5.2.3. Arquitetura e topologia dos softsensores neurais	143
5.2.4. Ambiente de programação e treinamento dos protótipos	152
5.2.5. Algoritmo de treinamento dos protótipos	153
5.2.6. Procedimentos de treinamento, validação e teste	155
5.2.7. Medidas de avaliação do desempenho dos protótipos	156
5.2.8. Critério de seleção dos melhores softsensores neurais	161
5.3. RESULTADOS E DISCUSSÃO	162
5.3.1. Treinamentos e validações dos protótipos de softsensores neurais	164
5.3.2. Testes dos protótipos de softsensores neurais	244
5.3.3. Ratificação da suficiência das variáveis dos padrões de entrada	245
5.4. CONCLUSÃO	270
6. CONCLUSÃO GERAL E PERSPECTIVAS FUTURAS	271
7. REFERÊNCIAS BIBLIOGRÁFICAS	272
ANEXOS	286

1. INTRODUÇÃO

O uso de computadores para controle de processos de fermentação tem aumentado significativamente na última década devido à diminuição no custo, à melhoria na velocidade computacional, ao aumento de confiabilidade do *hardware* computacional e à crescente pressão econômica por produto, produtividade e controle de qualidade. O advento dos microprocessadores teve um enorme impacto sobre aplicações computacionais em tempo real para processos fermentativos. Microprocessadores fornecem meios confiáveis e baratos para integrar o controle computacional em processos de fermentação e realizar identificação *on-line*, controle adaptativo e otimização de processos de fermentação. A questão, hoje, não é mais se é economicamente viável ter um processo monitorado e controlado por computador, mas se qualquer indústria pode permanecer competitiva sem ele. Os benefícios da aquisição e análise automática de dados com proposta de manutenção de registros e supervisão gerencial são suficientes para justificar uma operação controlada por computador.

Até pouco tempo a indústria de fermentação tinha largado atrás de outros processos industriais na implementação de controle e otimização tecnológica, por várias razões. Processos fermentativos são menos entendidos por conta de sua complexidade, envolvendo um grande número de reações bioquímicas dinâmicas e processos de transferências. Medidas *on-line* de determinados parâmetros bioquímicos e fisiológicos são difíceis de serem realizadas e, conseqüentemente, a instrumentação necessária para medi-los é extremamente cara ou inexistente. Microrganismos têm seus próprios mecanismos de regulação celular, cabendo, portanto, aos sistemas de controle externos a importante tarefa de manipular o ambiente extracelular, visando afetar corretamente os mecanismos intracelulares.

Um pré-requisito para melhorar a eficiência de processos fermentativos é a habilidade para fornecer informações confiáveis em tempo real sobre as variáveis do processo. Atualmente, apenas algumas poucas variáveis de estado são medidas *on-line* com sensores viáveis técnica e economicamente. Medidas diretas *on-line* de variáveis primárias do processo, tais como biomassa, substrato e concentrações do produto, usualmente são

indisponíveis. O estado da fermentação, portanto, tem de ser inferido de medidas de variáveis secundárias e de conhecimento prévio do processo (PONS, 1992).

Numerosas técnicas têm sido desenvolvidas para estimação indireta de dados de sistemas biológicos. Tanto aproximações baseadas em modelo como aproximações não baseadas em modelo têm sido propostas por vários pesquisadores. Entretanto, muitas das técnicas desenvolvidas fundamentam-se sobre análise de sistemas lineares e, conseqüentemente, elas não representam o comportamento não linear dinâmico do processo precisamente. Nos casos em que estimadores não lineares têm sido propostos, medidas *off-line* intermitentes são freqüentemente requeridas para validar estados estimados, pois caso contrário estimativas enviesadas podem ser obtidas (PONS, 1992).

Portanto, no monitoramento e controle de bioprocessos, a falta de sensores confiáveis para variáveis biológicas, induziu o desenvolvimento de *softwares sensores*, que associam um sensor (*hardware*) e um algoritmo de estimação (*software*) visando fornecer estimativas *on-line* de parâmetros cinéticos e variáveis não mensuráveis. Estes *softsensores* são importantes ferramentas no monitoramento e controle de bioprocessos (CHÉRUUY, 1997) e várias aplicações bem sucedidas têm sido relatadas (STEPHANOPOULOS e SAN, 1990; BASTIN e DOCHAIN, 1990; GRAINDORGE *et al.*, 1994).

Várias técnicas de estimação têm sido propostas na literatura. Entre essas técnicas, quatro têm sido reconhecidas por terem grande potencial na estimação *on-line* de bioprocessos: (1) estimação através de balanços de elementos; (2) observador adaptativo, (3) técnicas de filtro (filtro de Kalman, filtro de Kalman estendido) e (4) redes neurais artificiais (ASSIS e MACIEL FILHO, 2000).

Nesse contexto, a aplicação de técnicas de redes neurais para estimação e previsão de variáveis do processo pode ser útil. Nos últimos anos, tem aumentado significativamente o número de aplicações da teoria das redes neurais para estimação e previsão de variáveis em processos biotecnológicos. Processos bioquímicos, em geral, exibem um alto grau de não linearidade e dinâmicas de tempo variante. Limitações no conhecimento do processo também contribuem para a complexidade (PONS, 1992).

Modelos de redes neurais artificiais têm sido desenvolvidos para previsão *on-line* de valores interamostrais de baixa frequência de medidas de biomassa, tanto quanto de outras variáveis, usando medidas secundárias de alta frequência. O potencial de redes neurais para modelar bioprocessos dinâmicos não lineares, visando fornecer um estimador *on-line*, tem sido demonstrado pela aplicação e, portanto parece apropriado considerá-las para desenvolvimento de *softsensores* (KARIM e RIVIERA, 1992ab; Di MASSIMO *et al.*, 1992ab; SYU e TSAO, 1993; YET-POLE, WU e LIU, 1996; GLASSEY *et al.*, 1997). A vantagem desta técnica é não ser necessário qualquer conhecimento antecipado sobre a taxa de crescimento cinético. A desvantagem é a falta de dados experimentais para treinar a rede (ASSIS e MACIEL FILHO, 2000).

Este trabalho teve como objetivo geral o desenvolvimento de *softsensor* baseado em redes neurais para estimação e previsão em tempo real da biomassa e da atividade de emulsificação de biossurfactante produzido por *Candida lipolytica*, usando óleo de milho como substrato de baixo custo e água do mar diluída e suplementada como meio basal. A otimização dos componentes do meio de produção de bioemulsificante por *Candida lipolytica* e a ampliação de escala do processo da escala de frascos para escala de biorreator de bancada foram objetivos específicos do presente trabalho, diretamente ligados ao desenvolvimento do *softsensor* neural, desde que o desempenho de uma rede neural artificial é influenciado pela qualidade e pela quantidade dos dados do conjunto de treinamento.

A criação, em escala mundial, de novas e rigorosas leis de proteção e controle do meio ambiente tem contribuído significativamente para o aumento da demanda industrial por biossurfactantes com atividade de emulsificação como alternativa a surfactantes de origem petroquímica.

Os surfactantes com atividade de emulsificação apresentam inúmeras vantagens sobre os emulsificantes químicos, por oferecerem baixa toxicidade e maior aceitabilidade ecológica. Para que possam competir economicamente com os emulsificantes químicos, os biossurfactantes necessitam de uma produção cuidadosa e planejada, com a utilização de substratos de baixo custo (ZHOU e KOSARIC, 1995).

Problemas ambientais, principalmente nas regiões costeiras e estuarinas, provocados pelo derramamento de petróleo e derivados, têm ocorrido com frequência, nas últimas décadas, chamando atenção pela contaminação por hidrocarbonetos aromáticos policíclicos, gases tóxicos e inflamáveis. Acidentes semelhantes podem ser observados no processo de abastecimento de embarcações nos portos, podendo também levar ao derramamento do petróleo e liberação de gases. Destaca-se no processo de derramamento, a persistência do petróleo e de seus derivados no ambiente, principalmente dos hidrocarbonetos aromáticos policíclicos os quais reconhecidamente são os maiores contaminantes aquáticos e terrestres, apresentando alto potencial carcinogênico e mutagênico. Quando colocados na água, onde ocorreu o acidente, os biossurfactantes se ligam às moléculas do petróleo, realizando sua biodegradação (CAMPOS-TAKAKI, 1999; VANCE-HARROP e CAMPOS-TAKAKI, 2001, CAMEOTRA e BOLLAG, 2003).

O processo de biodegradação representa um dos mecanismos primários pelos quais o petróleo e outros hidrocarbonetos poluentes podem ser eliminados do ambiente. Os microrganismos têm demonstrado a capacidade de oxidar hidrocarbonetos aromáticos policíclicos - derivados do petróleo, em especial naftaleno e fenantreno, em metabólitos oxidados, hidroxilados e posteriormente conjugados, com ausência de toxicidade, através de reações mediadas pelo citocromo P-450. Destacam-se, no processo de biotransformação dos xenobióticos, as enzimas envolvidas, cujo modelo de atuação tem sido estabelecido através de equações matemáticas (ATLAS, 1984; HARAYAMA, 1997; CARVALHO, 1999).

Geralmente, biossurfactantes são produzidos durante o crescimento de microrganismos sobre hidrocarbonetos, que normalmente são caros e, portanto aumentam o custo global do processo. Entretanto, outros substratos mais baratos solúveis em água, como glicose e etanol, são alguma vezes usados. Na pesquisa por matérias primas de baixo custo para produção de biossurfactantes, efluentes industriais têm demonstrado ser uma boa alternativa (DESAI e BANAT, 1997). Recentemente, demonstrou-se a produção de biopolímeros por *Candida lipolytica* com atividade emulsificante, utilizando óleos vegetais típicos da região Norte e Nordeste do Brasil (MARÇAL, 1991; ROCHA, 1996; SARUBBO, 1997; VANCE-HARROP, 2000; VANCE-HARROP, 2004). Os

biossurfactantes produzidos por estes microrganismos, quando cultivados em substratos insolúveis, apresentam grande importância na remoção de poluentes ambientais (FIECHTER, 1992; LIN, 1996).

A tecnologia utilizada para produção, em escala de frascos, de biossurfactantes por *Candida lipolytica*, utilizando fontes nutricionais de baixo custo, vem sendo utilizada com sucesso há vários anos por pesquisadores do Núcleo de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco (SARUBBO, 1997; SARUBBO, PORTO e CAMPOS-TAKAKI, 1999; VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2000; SARUBBO *et al.*, 2001; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2003, VANCE-HARROP, 2004). Com o presente trabalho iniciou-se a ampliação (*scale-up*) do processo para escala de biorreator de bancada, pretendendo-se migrar posteriormente para escalas maiores (piloto ou mesmo industrial).

Desde que, na ampliação de escala, muitos processos bioquímicos criam produtos impuros e diluídos, existe uma grande necessidade de aumentar a produtividade volumétrica e a concentração dos produtos. Para se obter bons resultados na ampliação de escala de bioprocessos, é preciso investir no desenvolvimento e otimização de tecnologias eficientes e econômicas. As formas normalmente empregadas para reduzir os custos de bioprocessos são através de: a) melhoramentos genéticos de microrganismos (engenharia genética); b) otimização de meios de cultivo e c) engenharia de sistemas de bioprocessos, ou seja, desenvolvimento de sistemas computacionais aplicados a processos biotecnológicos (SHIMIZU, 1996). A otimização do meio de produção de biossurfactante, em escala de frascos e o desenvolvimento de *softsensor* neural para fazer estimação e previsão da concentração de biomassa e da atividade de emulsificação do biossurfactante produzido por *Candida lipolytica* são dois objetivos, deste trabalho, diretamente ligados à ampliação de escala do processo e aos dois últimos itens da lista supracitada.

Estes fatos deixam evidente a importância da otimização, do monitoramento e do controle do processo e da necessidade de estimativas em linha da concentração da biomassa e da atividade enzimática do biossurfactante produzido por *Candida lipolytica*, no Núcleo

de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco (NPCIAMB/UNICAP). O desenvolvimento do *softsensor* proposto visa diminuir os custos e atrasos associados às medidas destas variáveis, além de contribuir para o monitoramento, controle avançado e otimização do processo de produção de biossurfactante por *Candida lipolytica*.

Esta tese está organizada em mais seis capítulos. O capítulo 2 apresenta uma revisão sobre o processo de produção de biossurfactantes por *Candida lipolytica* a partir de substratos de baixo custo e sobre *softsensores* baseados em redes neurais aplicados a bioprocessos. O capítulo 3 trata do processo de otimização de componentes de meios de produção de bioemulsificantes por *Candida lipolytica*. O capítulo 4 discute a ampliação do processo de produção de biossurfactante por *Candida lipolytica* da escala de frascos para a escala de biorreator de bancada. O capítulo 5 descreve o desenvolvimento de *softsensor* baseado em redes neurais para estimação e previsão de biomassa e de atividade de emulsificação de biossurfactante produzido por *Candida lipolytica*. As conclusões gerais e as perspectivas de futuros trabalhos são discutidas no capítulo 6. No capítulo 7 encontram-se as referências bibliográficas.

2. REVISÃO DE LITERATURA

2.1. Introdução

A capacidade de controlar, automática e precisamente, processos de fermentação em seu estado ótimo, reduzindo custos, aumentando a produção e mantendo a qualidade do produto metabólico, vem despertando cada vez mais o interesse das indústrias de biotecnologia. Entretanto, é preciso salientar as dificuldades em se projetar sistemas de controle de biorreatores, devido à natureza não linear e dinâmica dos mesmos, a incertezas nos modelos, à falta de sensores *on-line* confiáveis e à lentidão da resposta. Para solucionar o primeiro problema deve-se considerar uma estratégia de controle adaptativa. Para o segundo problema, o sistema de controle deve ser robusto às incertezas do modelo e capaz de rejeitar distúrbios. Para o terceiro problema, em muitos casos, os sensores *on-line* efetivos são de uso limitado e assim alguns estimadores de parâmetros e/ou observadores necessitam ser projetados a partir das variáveis mensuráveis disponíveis. Para o quarto problema, estratégia de controle do tipo preditivo deve ser considerada. Muitas estratégias de controle têm sido propostas e aplicadas na prática para resolver esses problemas e muitas outras ainda estão em progresso (SHIMIZU, 1996).

Esforços consideráveis de pesquisa têm sido direcionados para o desenvolvimento de biossensores para medidas de condições de fermentação crítica. Montague, Morris e Ward (1989) reviram o estado da arte da tecnologia de biossensor na regulação da fermentação. Atualmente, muito sucesso tem sido alcançado com aplicações em equipamentos em escala de laboratório. As consequências de falhas do sensor, conduzindo a contaminação da fermentação, são mais severas em sistemas em escala industrial, portanto, em processos onde a economia potencial é o mais importante, é comum restringir a instrumentação à mais padrão possível.

Uma alternativa para a tecnologia do biossensor é o desenvolvimento de algoritmos de estimação *on-line* para fornecer as medidas de desempenho necessárias. Tal aproximação é também conhecida como “soft-sensing” (estimação inferencial). O conceito

é relativamente simples: medidas disponíveis *on-line* (medidas secundárias) que estão relacionadas a variáveis chaves (primárias) do processo podem ser usadas juntas com informação de ensaio *off-line*, para fornecer uma estimativa *on-line* da variável primária. Essa filosofia, em sua maior simplicidade, pode conduzir a obtenção de uma correlação linear entre variáveis primárias e secundárias (THAM *et al.*, 1991).

Embora uma aproximação linear possa ser útil em condições normais, ela freqüentemente não suporta perturbações transientes. Di Massimo *et al.* (1992a) consideraram os méritos relativos das técnicas de estimação linear aplicadas a sistemas de fermentação e demonstraram a necessidade de se recorrer a uma aproximação adaptativa para superar a discordância modelo/processo. A aproximação mais conveniente seria o desenvolvimento de um modelo a partir do entendimento dos princípios básicos das relações existentes. Modelos de sistemas não lineares podem aumentar o intervalo de aplicabilidade do sistema, mas freqüentemente são muito difíceis e consomem tempo para serem desenvolvidos e podem ser impossíveis de serem especificados, especialmente quando os mecanismos básicos não são completamente entendidos. As dificuldades encontradas em se obter uma relação representativa razoável entre variáveis primárias e secundárias do processo fermentativo são, entretanto, contornáveis.

Um amplo espectro de técnicas que se situam entre aquelas baseadas sobre uma pura correlação caixa preta e aquelas usando um complexo modelo do sistema pode ser postulado. A literatura contém numerosas referências que dão suporte a esta afirmação, merecendo especial destaque às revisões realizadas por Jonhson (1987) e por Montague, Morris e Ward (1989). Mesmo que não seja proposta deste trabalho detalhar tais desenvolvimentos, um tema comum é a importância que um modelo representativo do processo joga na obtenção de um estimador com desempenho razoável.

Redes neurais têm sido usadas como uma alternativa aos tradicionais modelos matemáticos para simular padrões complexos. A simplicidade de implementação torna as redes neurais apropriadas para simulação das principais variáveis de estado (biomassa, substrato e concentrações de produto) de processos fermentativos, como uma função do progresso do processo (SYU e TSAO, 1993; GLASSEY, MONTAGUE, WARD e KARA,

1994a; ZHANG *et al.*, 1994). A descrição matemática de culturas de microrganismos é uma tarefa complexa que envolve grandes custos em desenvolvimento e computação. Embora modelos não estruturados de culturas de microrganismos - ou seja, aqueles em que o sistema biológico é descrito através de balanços de massa de um número reduzido de variáveis - sejam relativamente fáceis de formular, sua solução é complicada por causa da avaliação das relações cinéticas. Em muitos casos expressões matemáticas descrevendo estes termos como funções de variáveis envolvidas não são conhecidas. Por outro lado, se estas expressões estão disponíveis, elas freqüentemente envolvem um grande número de parâmetros que têm de ser ajustados usando técnicas de otimização não linear (BRAVO *et al.*, 1999). Pesquisas recentes têm sido dirigidas para o desenvolvimento de técnicas de estimação para variáveis do processo não mensuráveis. Entre estas, redes neurais artificiais vêm despertando interesse pela capacidade de modelarem processos não lineares, sem necessitarem de pré-especificação de relações dinâmicas detalhadas.

A seguir será apresentada uma visão geral sobre o processo de produção de biossurfactantes por *Candida lipolytica*, patentado pelos pesquisadores do Núcleo de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco (NPCLAMB/UNICAP) e sobre as técnicas mais usadas para estimação de estados e parâmetros não mensuráveis em bioprocessos, dando uma ênfase especial àquelas baseadas em redes neurais artificiais. Apesar de na literatura existirem diversos trabalhos referentes ao desenvolvimento de *softsensores* neurais usando medidas disponíveis para reconstruir a evolução de variáveis de estado e/ou para estimar parâmetros de bioprocessos, não existe nenhuma pesquisa sobre a aplicação dos mesmos em processo de produção de biossurfactante por *Candida lipolytica*.

2.2. Processo de produção de biossurfactantes por *Candida lipolytica* a partir de substratos de baixo custo

Com o advento da Biotecnologia, os limites de aplicação dos aditivos sintéticos expandiram-se e novos produtos foram desenvolvidos baseados na capacidade sintética dos microrganismos. Essas perspectivas, relacionadas com os produtos de elevado interesse industrial, têm conduzido nos últimos anos a investigação e desenvolvimento de modelos que constituem as bases das novas tecnologias na produção de agentes surfactantes por microrganismos (ROBERT *et al.*, 1991).

O mercado dos surfactantes gira em torno de \$ 9,4 bilhões por ano, e sua demanda aumentou cerca de 35% até o fim dos anos noventa. Praticamente, a maioria dos surfactantes usados é formada por compostos químicos derivados do petróleo; contudo, o interesse em surfactantes de origem microbiológica tem aumentado nos últimos anos. Os biossurfactantes possuem muitas vantagens sobre os surfactantes químicos, como baixa toxicidade, maior biodegradabilidade, maior compatibilidade ambiental, maior poder espumante, maior seletividade e atividade específica em pH extremos e em altas salinidades, além da vantagem de serem produzidos por fontes renováveis (DESAI e BANAT,1997).

O melhoramento do processo em muitos sistemas biotecnológicos é responsável por mais de 60% do custo total de produção. Devido a razões econômicas, deve-se levar em consideração não só o custo do cultivo como um todo, como também o de outras preparações brutas. Deve ser considerada no processo a atividade emulsificante que puder ser afetada por outros materiais presentes nestas operações. A recuperação dos biossurfactantes depende principalmente da sua carga iônica, solubilidade em água e localização, intra ou extracelular (DESAI e BANAT,1997).

O custo de produção de biossurfactantes é cerca de 3 a 10 vezes maior do que o custo dos surfactantes sintéticos (MULLIGAN e GIBBS,1993). Conseqüentemente, o futuro das aplicações dos biossurfactantes será governado principalmente pelo ganho econômico na produção versus o custo da aplicação.

O grande responsável pela redução de custos de produção de biossurfactantes é o processo de fermentação. Estima-se que o material bruto contabiliza 10 a 30% do custo total de produção (MULLIGAN e GIBBS, 1993). Apesar de melhoramento genético, otimização de meios de produção e ampliação de escala do processo reduzirem custos; os testes de toxicidade requeridos para um novo produto ser adotado na indústria de alimentos, cosméticos e de produtos de limpeza, elevam consideravelmente o orçamento. Certamente, a indústria que investir na redução dos custos de produção, obtendo a aprovação necessária para o produto, terá benefícios na introdução de um novo produto biotecnológico.

Diferentemente dos surfactantes sintéticos, os quais são classificados de acordo com a natureza dos seus grupos polares, os biossurfactantes são categorizados principalmente de acordo com sua composição química e origem microbiana. Em geral, suas estruturas incluem uma porção hidrofílica, constituída por aminoácidos ou peptídeos aniônicos ou catiônicos, mono, di ou polissacarídeos e uma porção hidrofóbica, constituída de ácidos graxos. Em sua maior parte os surfactantes são constituídos por glicolipídeos; que são carboidratos combinados com ácidos alifáticos de cadeia longa ou com ácidos hidroalifáticos. Além dos glicolipídeos existem ainda os lipopeptídeos, lipoproteínas, ácidos graxos, fosfolipídeos, lipídeos neutros e biossurfactantes poliméricos.

Os biossurfactantes são produzidos por uma grande variedade de microrganismos, excretados ou ligados às células, principalmente quando cultivados em substratos insolúveis em água. O contato direto da célula com as gotas de hidrocarbonetos e sua interação com as gotas emulsificadas têm sido descritos na literatura.

Compostos de superfície-ativa de origem biológica ou biossurfactantes foram descritos na literatura há poucas décadas. Devidos às vantagens de biodegradabilidade e de produção de fontes renováveis, os biossurfactantes vêm ganhando destaque e suas aplicações sendo largamente utilizadas (ROSENBERG, 1986; FIECHTER, 1992; BANAT, 1995).

Bioemulsificantes são moléculas anfipáticas, que podem ser classificadas em dois grupos: (i) compostos de baixo peso molecular, tais como glicolipídeos, ácidos gordurosos,

fosfolipídeos e lipopeptídeos, que baixam a tensão interfacial entre líquidos hidrofóbicos e a água, reduzindo assim a energia necessária para formar emulsões e (ii) bioemulsificantes poliméricos, que estabilizam emulsões (KAPLAN, ZOSIM e ROSENBERG, 1987; ABURUWAIDA *et al.*, 1991; NAVON-VENEZIA *et al.*, 1995).

Entre leveduras, o gênero *Candida* tem sido largamente empregado para fermentação de hidrocarbonetos, especialmente para produção de proteína de célula simples (SCP- Single Cell Protein), sendo relatado produção de agentes de superfície-ativa (SINGH *et al.*, 1992). *Candida lipolytica* tem mostrado crescimento sobre diversos substratos de carbono imiscíveis em água e produção de polímeros extracelulares com propriedades emulsificantes (CIRIGLIANO e CARMAN, 1985; SARUBBO, 1997; SARUBBO, PORTO e CAMPOS-TAKAKI, 1999; VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2000; SARUBBO *et al.*, 2001; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2003, VANCE-HARROP, 2004).

Compostos surfactantes (polímeros extracelulares) de *Candida lipolytica*, cultivada sobre n-alcano, foram isolados e caracterizados por Pareilleux (1979), apresentando carboidratos (450 g/L), proteína (250 g/L) e lipídeos (30 g/L) em suas composições. Entretanto, quando a levedura foi cultivada sobre glicose, não houve produção de surfactantes.

Um bioemulsificante chamado *liposan*, composto por 830 g/L de carboidratos, 170 g/L de proteína, foi isolado e caracterizado por Cirigliano e Carman (1984,1985) quando *Candida lipolytica* crescia sobre substrato de carbono imiscível em água (n-hexadecano). Lipídeos não foram detectados. Os autores também observaram a produção de uma atividade de emulsificação negligenciável, quando *Candida lipolytica* crescia no meio de cultura YNB (Yeast Nitrogen Base) suplementado com 1% de glicose.

A produção de biopolímeros por *Candida lipolytica*, com atividade emulsificante, utilizando-se como substratos óleos vegetais típicos das regiões Norte e Nordeste do Brasil, foi demonstrada por diversos pesquisadores (MARÇAL, 1991; ROCHA, 1996,

SARUBBO, 1997; SARUBBO, PORTO e CAMPOS-TAKAKI, 1999; VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2000; SARUBBO *et al.*, 2001; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2003, VANCE-HARROP, 2004).

A produção de biopolímeros extracelulares por *Candida lipolytica* em meios suplementados com óleo de babaçu, de coco e de dendê, como fonte de carbono, foi estudada por Marçal (1991), visando à aplicação futura destas substâncias como aditivos alimentares. O líquido metabólico livre de células, obtido de diferentes cultivos, demonstrou atividade de emulsificação a partir da fase de crescimento exponencial e durante toda a fase estacionária. O aumento do inóculo não foi proporcional ao aumento da produção de biopolímeros, sendo o maior rendimento evidenciado com inóculo de 10^2 U.F.C. (Unidade Formadora de Cultura) /mL. O aumento da concentração dos óleos babaçu e dendê, não estimularam a produção de biopolímeros extracelulares. Entretanto, nos meios suplementados com óleo de coco, foi evidenciada maior atividade de emulsificação a partir da concentração de 7,5% a 10,0% . Com relação à aeração, foi verificado que a agitação orbital favoreceu a produção de biopolímeros nos meios suplementados com os três óleos vegetais. Os biopolímeros isolados pela mistura de clorofórmio/metanol, 2:1(v/v), apresentaram solubilidade em água e atividade de emulsificação. Quimicamente estas substâncias, apresentaram-se constituídas por proteínas, carboidratos e lipídeos. A análise através de eletroforese revelou bandas de 18,4 Kd para os biopolímeros isolados do cultivo com óleo de babaçu e óleo de dendê; exceto para o produzido em óleo de coco, que apresentou banda de 24,0 Kd. Os produtos obtidos por hidrólise ácida evidenciaram a presença de D-glicose, D-Galactose e ácido D-galacturônico, nos três polímeros isolados. Os três produtos demonstraram baixa toxicidade, com DL₅₀ (1600 a 1700 mg/kg de peso corporal) e ausência de efeitos mutagênicos direto e indireto. Os biopolímeros produzidos por *Candida lipolytica* devido às suas características como bioemulsificantes, demonstraram favorecer estudos e perspectivas de utilização na indústria alimentícia.

A produção por *Candida lipolytica* de biopolímeros com atividade de emulsificação foi estudada por Sarubbo (1997), em processos de batelada e batelada alimentada, utilizando glicose como controle e óleo de babaçu como fonte de carbono.

Comparação entre as linhagens de *Candida lipolytica* IA 1055 e IA 1120 demonstrou ser a primeira melhor produtora de biopolímeros com atividade de emulsificação. As culturas suplementadas com óleo de babaçu e glicose (controle), em diferentes condições de batelada e batelada alimentada, apresentaram atividade de emulsificação no fim da fase exponencial e início da fase estacionária de crescimento, sendo observado maior produção de biopolímeros no cultivo em batelada. Os biopolímeros foram extraídos das fermentações em condições selecionadas: batelada com glicose a 1% (controle) e batelada com óleo de babaçu a 5% (condição 1). O maior rendimento foi observado em extração com solventes orgânicos e o maior índice de purificação foi obtido em sistema líquido-líquido com duas fases aquosas. A constituição química dos biopolímeros, com atividade de emulsificação, indicou a presença de carboidratos, proteínas e lipídeos, em menor quantidade.

A capacidade da levedura *Candida lipolytica* IA 1055 crescer sobre óleo de babaçu e produzir um emulsificante extracelular, sob várias condições, em cultura em batelada ou em cultura em batelada alimentada, a 27° C, foi demonstrada por Sarrubo, Porto e Campos-Takaki (1999). A atividade de emulsificação foi detectada depois que a levedura entrou na fase estacionária de crescimento, sugerindo que os bioemulsificantes fossem produzidos como metabólitos secundários. Observou-se também que o pH diminuiu de 5,5 para 2,3 durante a fermentação e que a atividade de emulsificação máxima foi alcançada depois que a cultura atingiu valores mínimos de pH. O biopolímero, obtido com o cultivo em batelada, foi isolado após 144h de fermentação e apresentou a seguinte composição: 600 g/L de carboidratos, 230 g/L de proteína e 11 g/L de lipídeos.

Fermentações em batelada, para produção de bioemulsificantes por *Candida lipolytica*, empregando glicose, nas concentrações de 0,5% 1,0 e 1,5%, como fonte de carbono, foram realizadas por Sarrubo *et al.* (2001). A levedura *Candida lipolytica* IA 1055 produziu atividade de emulsificação extracelular induzida, a 27° C, nas diferentes concentrações empregadas. Em todas as concentrações estudadas, a atividade de emulsificação máxima foi detectada na fase de crescimento, após o pH alcançar valores mínimos. Considerando o consumo de glicose e a produção de biomassa, a concentração da fonte de carbono não afetou o nível de crescimento máximo. Este comportamento foi provavelmente devido à utilização da fonte de carbono para produção de metabólito

secundário (bioemulsificante). O bioemulsificante isolado foi um biopolímero constituído por proteínas, carboidratos e lipídeos. O precipitado analisado do cultivo de *Candida lipolytica* em glicose a 1,5% apresentou em sua composição 47% de proteína, 45% de carboidrato, e 5% de lipídeos. Nenhum registro foi encontrado em literatura anterior, relatando a capacidade de *Candida lipolytica* produzir surfactante com atividade de emulsificação usando glicose como fonte de carbono. Os resultados obtidos no trabalho mostraram que a biossíntese de um bioemulsificante não é simplesmente um pré-requisito para a degradação de hidrocarboneto extracelular.

A introdução e utilização de meios de baixo custo, à base de água do mar, suplementados com óleo de milho, para a produção de biopolímeros extracelulares, com atividade de emulsificação, foi descrita pela primeira vez por Vance-Harrop *et al.* (1999) e Vance-Harrop (2000). Dois meios de baixo custo, suplementados com óleo de milho, foram formulados e avaliados visando o estudo das condições de produção de biosurfactante por *Candida lipolytica*. O meio denominado de YWS-C (*Yeast Salt Water-Corn*) foi formulado com a seguinte composição: 0,1% de sulfato de amônio; 0,1% de uréia; 5% de óleo de milho; 100ml de água do mar e água destilada, na proporção 1:1 V/V. O meio denominado YWS-C1 (*Yeast Salt Water-Corn* 1) foi formulado com a mesma composição do meio YWS-C, acrescido de 10mM de fosfato de potássio e pH ajustado para 5,3. Controles foram realizados com os dois meios, utilizando-se D-glicose como fonte de carbono. Os resultados indicaram que o meio de cultura YWS-C1, utilizado como base nutriente, suplementado com óleo de milho ou D-glicose como controle, possibilitou a produção de biosurfactante com atividade de emulsificação, quando o pH era controlado através de tampão fosfato. A atividade de emulsificação foi observada no meio YWS-C1, na fermentação em batelada, no período de 168 horas. Os resultados demonstram que a *Candida lipolytica* apresenta habilidade de crescimento nos meios com salinidade de 13% e os biopolímeros extracelularmente produzidos apresentam alta atividade de emulsificação, possibilitando a introdução de um novo meio de produção com baixo custo. Adicionalmente, a introdução de uréia como fonte de nitrogênio e de tampão fosfato induziu a produção de biosurfactante por *Candida lipolytica*, com propriedades de emulsificação.

Os resultados obtidos por Vance-Harrop *et al.* (1999) são corroborados pelas informações de Reisfield, Rosenberg e Gutnick (1972) que observaram a absoluta dependência da fonte de nitrogênio e do tampão fosfatos na produção de biossurfactante.

Estudos realizados por Vance-Harrop *et al.* (2000) sobre a influência da água do mar na produção de surfactante por *Candida lipolytica* demonstraram a capacidade de a levedura produzir biopolímeros com atividade de emulsificação após a fase estacionária de crescimento nos meios YSW-B2, YSW-B3 (*Yeast Salt Water-Bababassu 2 e 3*) e YSW-C1. A produção de biossurfactante realizada nos meios previamente descritos por Vance-Harrop *et al.* (1999), utilizando água do mar diluída e suplementada com óleo de babaçu e D-glicose como fonte de carbono, não foi afetada pela salinidade alta dos meios de cultura. Esses resultados são corroborados por Zinjarde *et al.* (1997), que obtiveram a produção de um emulsificante por *Yarrowia lipolytica* em um meio contendo água do mar artificial e n-hexadecano como fonte de carbono, visando sua utilização na dispersão da poluição dos derramamentos de óleos em ecossistemas marinhos.

Com o objetivo de estudar o crescimento de quatro espécies de *Candida* (*C. lipolytica*, *C. sphaerica*, *C. famata*, *C. glabrata*) isoladas em regiões contaminadas por petróleo e derivados, observar a produção de enzimas e avaliar a atividade de emulsificação em meios contendo óleo diesel, óleo da indústria têxtil, borra de óleo da indústria têxtil e glicose como substratos; Silva *et al.* (2001) realizaram fermentações em meio Sabourad líquido (50 ml), contendo 5% dos substratos e inoculados com 10^4 células/ml, sob agitação de 120 rpm, a 37°C. Para determinação da atividade de emulsificação, seguiu-se a metodologia descrita por Cooper e Goldenberg (1987) e para determinação da atividade lipásica foi utilizado o meio YMA (*Yeast Mold Agar*), modificado por Nascimento e Campos-Takaki (1994), utilizando trioleína como substrato. A trioleína foi substituída pelos substratos utilizados para avaliação do crescimento e produção enzimática a partir da inoculação de 10^7 células/ml em “cup plate”. As placas foram encubadas a 28° C durante 48 horas. O crescimento observado para as amostras de *Candida* estudadas, inclusive a *Candida lipolytica*, demonstrou o potencial biotecnológico em emulsificar fontes de carbono insolúveis e, principalmente, derivados de petróleo. As amostras mostraram-se também promissoras como agentes produtores de surfactantes e tensoativos

(ROSENBERG, 1986, SARUBBO, PORTO e CAMPOS-TAKAKI, 1999), apresentando potencialidade de aplicação industrial, especialmente na recuperação de óleos e limpeza de tanques.

A capacidade de *Candida lipolytica* produzir biossurfactantes extracelulares com atividade de emulsificação, através de fermentação utilizando óleo de babaçu e D-Glicose como fontes de carbono foi investigada por Vance-Harrop *et al.* (2003). A água do mar diluída a 50% suplementada com uréia, sulfato de amônio e fosfato foi utilizada como meio basal. Os melhores resultados foram atingidos com os meios YSW-B2 (contendo uréia, sulfato de amônio e óleo de babaçu) e YSW-B3 (contendo uréia, sulfato de amônio, fosfato e óleo de babaçu), através de fermentação em batelada alimentada com 5% de óleo de babaçu. A velocidade específica de crescimento foi de $0,02\text{ h}^{-1}$ e $0,04\text{ h}^{-1}$; tempo de geração de $34,6\text{ h}^{-1}$ e $17,3\text{ h}^{-1}$ e atividade de emulsificação igual a 0,666 e 0,158 unidades, respectivamente. As moléculas dos novos bioemulsificantes demonstraram ser constituídas por carboidratos, proteínas e lipídeos.

O potencial biotecnológico de *Candida lipolytica* na produção de biossurfactante com habilidade de emulsificação – utilizando meios de cultura de baixo custo à base de água do mar, adicionados de óleo de babaçu e glicose como controle da fonte de carbono – foi investigado por Vance-Harrop (2004). Os meios YSW-B2 e YSW-B3 apresentaram excelente capacidade de emulsificação sendo considerados como novos emulsificantes constituídos quimicamente por carboidratos, proteínas e lipídeos. Planejamento fatorial fracionário com dois níveis e cinco fatores foi realizado para investigar a influência do extrato de levedura, do sulfato de amônio, da uréia, do fosfato de potássio e do óleo de babaçu sobre a produção de biossurfactante. Os resultados obtidos sugeriram maior produção com os níveis utilizados de extrato de levedura, fosfato e uréia nos seus valores superiores, enquanto que o da amônia foi fixado em seu nível inferior. A fonte de carbono, óleo de babaçu, não exerceu efeito significativo sobre a produção de biossurfactante. Cinco condições de meios de culturas selecionados do planejamento foram utilizadas para realização de estudos com substratos mistos (solúvel e insolúvel) para os processos de remoção e bioissorção de pireno (derivado de petróleo). Os resultados das análises por cromatografia líquida de alta eficiência demonstraram uma alta taxa de remoção de pireno

nos meios I e IV com percentuais de 99,02% e 97,51%, respectivamente. Estudos subsequentes com a biossorção de pireno foram realizados utilizando a biomassa de *Candida lipolytica* imobilizada em álcool polivinil como matriz inerte. Os resultados obtidos sugeriram que a biomassa imobilizada de *Candida lipolytica* apresenta grande potencial nos processos de biossorção do pireno tendo em vista o baixo custo, da produção e da imobilização da biomassa.

A produção de biossurfactantes pela levedura *Candida lipolytica* em meio contendo resíduo de refinaria de óleos vegetais e ácido glutâmico como substratos foi avaliada por Rufino, Sarubbo e Campos-Takaki (2005). A tensão superficial no líquido metabólico livre de células foi medida após 72hs de cultivo, observando-se uma acentuada redução na tensão inicial do meio de produção de 50mN/m para 26mN/m. O índice de emulsificação foi determinado, obtendo-se 79% de emulsificação do óleo motor, embora o óleo de milho e o n-hexadecano não tenham sido efetivamente emulsificados. Os resultados obtidos sugerem que a utilização do resíduo industrial e do ácido glutâmico favorecem a produção de biossurfactantes promissores no controle da poluição ambiental causada por compostos oleosos.

2.3. Softwares sensores baseados em redes neurais aplicados a bioprocessos

Em processos de fermentação, nenhum controle ou otimização pode ser realizado sobre parâmetros chaves, a menos que estes parâmetros possam ser medidos *on-line*, para fornecer a informação requisitada pelo controlador ou otimizador. Entretanto, medidas *on-line* de parâmetros fisiológicos chaves são freqüentemente difíceis de serem realizadas. Apesar do esforço para desenvolver sensores que possam ser usados para detectá-las, a disponibilidade e confiabilidade destes instrumentos são ainda limitadas. Portanto, normalmente, para obter a informação necessária, recorre-se à estimação daqueles parâmetros chaves a partir de outros parâmetros que são relativamente simples e fáceis de medir (PONS,1992).

O sucesso da operação, controle e otimização de processos fermentativos depende fortemente da disponibilidade de uma avaliação rápida e precisa do desempenho do sistema. Isto, por sua vez requer informações confiáveis de variáveis do processo, em tempo real. A escassez de biossensores para medidas *on-line* diretas de variáveis primárias do processo - tais como biomassa e concentrações do substrato e do produto - torna o procedimento de identificação difícil. Muitas políticas de controle de fermentação industrial são baseadas sobre o uso de análises não freqüentes de amostras *off-line*, conduzindo para operação e controle pobre do processo. O estado da fermentação, portanto, tem de ser inferido de medidas de variáveis secundárias e de qualquer conhecimento prévio do processo dinâmico (KARIM e RIVERA, 1992a,b).

A falta de sensores confiáveis para monitorar e controlar parâmetros chaves em processos biotecnológicos, estimulou o desenvolvimento de “softwares sensores” que associam um sensor (*hardware*) e um algoritmo de estimação (*software*) para fornecer estimativas *on-line* de variáveis não mensuráveis e parâmetros cinéticos (Figura 1).

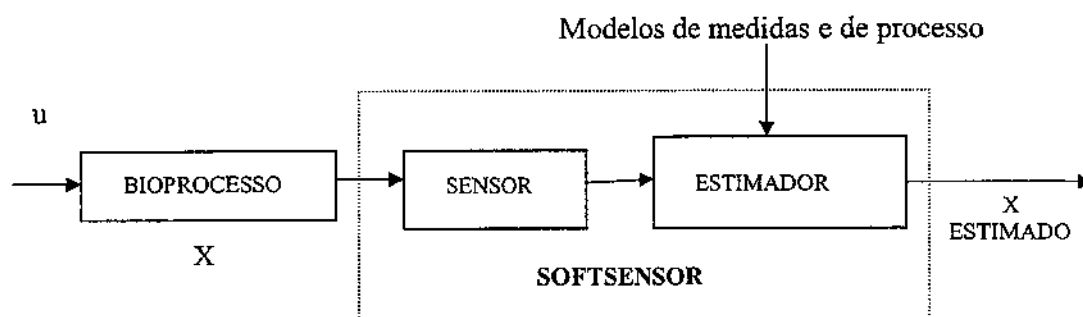


Figura 1. Princípio do *Software Sensor* (CHÉRUY,1997)

Em um *software sensor*, o (*hardware*) sensor tem que fornecer medidas confiáveis e suficientemente informativas para deduzir variáveis não mensuráveis do processo e parâmetros, e o algoritmo de estimação tem de ser projetado para produzir estimativas *on-line* que converjam para os valores reais das variáveis não mensuráveis, segura e rapidamente, qualquer que sejam os distúrbios do processo. A síntese de um algoritmo oferecendo tal garantia não é sempre possível, isto depende dos conhecimentos disponíveis sobre as medidas e o processo. Este conhecimento é geralmente expresso através de modelos matemáticos, cujas propriedades e validades influenciam fortemente as possibilidades do projeto e os desempenhos dos algoritmos de estimação (CHÉRUY, 1997). Apesar disto, numerosos *softwares* sensores têm sido desenvolvidos e aplicados com sucesso à engenharia de bioprocessos. Estes *softwares* sensores vêm demonstrando ser ferramentas úteis e com aplicações bem sucedidas no monitoramento e controle de bioprocessos (STEPHANOPOULOS e STEPHANOPOULOS,1986; STEPHANOPOULOS e SAN, 1990; BASTIN e DOCHAIN, 1990, GRAINDORGE *et al.*, 1994).

Nos últimos anos, técnicas de estimação propostas e revistas na literatura (PONS, 1992; CHÉRUY 1997, ASSIS e MACIEL FILHO, 2000) - baseadas em medidas indiretas, em balanço material macroscópico, em observador adaptativo, em filtragem estocástica, como o filtro de Kalman e o filtro de Kalman estendido e em redes neurais artificiais - vêm se destacando pelo sucesso de suas aplicações na estimação *on-line* de bioprocessos.

Redes neurais artificiais são ferramentas matemáticas poderosas capazes de aproximar qualquer relação entre variáveis dependentes e independentes. A teoria básica de redes neurais artificiais foi revista por Najjar, Basheer e Hajmeer (1997) e encontra-se disponível em qualquer livro texto de redes neurais, como Hagan, Demuth e Beale (1996). Em particular, esta revisão de literatura se deterá sobre aplicações em bioprocessos de *softwares* sensores baseados em redes neurais. As redes neurais, atualmente, aparecem como uma possível solução para superar as dificuldades de modelar a atividade microbiana em processos biotecnológicos, que são inerentemente não lineares e variantes no tempo. A tecnologia usada pela rede neural é inspirada no cérebro humano. Em redes neurais artificiais a informação é armazenada na forma de pesos entre neurônios artificiais, de forma similar ao que se supõe ser o armazenamento de conhecimento no cérebro. A estratégia de uso de redes neurais para resolver um determinado problema é escolher a arquitetura de rede que melhor se ajuste ao problema e um processo de aprendizagem. Exemplos representativos do conhecimento a ser adquirido pela rede são apresentados à rede que se auto-organiza para integrar a informação, dentro de sua estrutura, através da atualização dos pesos entre os neurônios.

A capacidade das redes neurais de modelar dinâmicas de bioprocessos foi investigada por Thibault, Van Breusegem e Chérury (1990), que mostraram que uma rede neural simples era capaz de generalizar o comportamento dinâmico de uma dada fermentação com alto grau de precisão e com certo grau de imunidade ao ruído. Uma rede neural treinada com o algoritmo de retropropagação dinâmico (com janela móvel) foi usada pelos autores para fazer previsões de concentrações de células e de substrato, no intervalo de amostragem seguinte, a partir da taxa de diluição e das concentrações de células e substrato no intervalo de amostragem atual. Usando resultados de simulações de um biorreator contínuo de tanque agitado (CSTR – *Continuous Stirred Tank Reactor*), os autores compararam as previsões das concentrações de células e de substrato fornecidas pelo modelo neural com as fornecidas por um modelo baseado no filtro de Kalman estendido (EKF – *Extended Kalman Filter*). Os resultados obtidos com redes neurais foram da mesma ordem de magnitude das correspondentes previsões realizadas pelo filtro de Kalman. As capacidades preditivas das redes neurais e do filtro de Kalman mostraram-se muito boas, a despeito das relativamente grandes e súbitas mudanças de fluxo afetando o

biorreator. Uma vantagem distintiva entre redes neurais e filtro de Kalman estendido, apontada pelos autores é o fato das redes neurais não necessitarem de conhecimento perfeito do modelo dinâmico do processo, o que as torna uma ferramenta promissora para modelagem, estimação e previsão de sistemas complexos e de grande porte, tais como bioprocessos, cujas dinâmicas são pobremente conhecidas.

Previsões *on-line* por técnicas de redes neurais - de concentrações de células e substrato no instante posterior a amostragem a partir de medidas de taxa de diluição e de concentrações de células e substrato no instante da amostragem - foram analisadas por Van Breusegem, Thibault e Chérui (1991). Foi mostrado que a precisão de previsões *on-line* baseadas em redes neurais, obtidas de uma sequência de aprendizagem inicial, diminui quando mudanças cinéticas ocorrem durante o curso da fermentação. As influências do comprimento da janela de aprendizagem, do número de iterações e do modelo neural inicial sobre a precisão preditiva de modelos neurais adaptativos foram investigadas. A grande contribuição do trabalho foi propor um algoritmo adaptativo, em que esquemas de aprendizagem com janelas deslizantes, inspirados por técnicas recursivas de estimação de parâmetros, foram usados para restaurar progressivamente o conhecimento integrado no modelo neural. Além disso, em vez do comumente usado algoritmo de retropropagação, um método quase Newton, que converge mais rapidamente para uma solução precisa, foi usado para adaptar os pesos da rede neural. O desempenho do algoritmo proposto foi avaliado sob várias condições de medidas, incluindo medidas com atraso e/ou pouco freqüentes, e demonstrou melhora significativa em relação ao algoritmo de retropropagação.

Uma rede neural com três camadas e com capacidade de aprendizagem suficiente para estimação de variáveis de bioprocessos foi desenvolvida por Linko e Zhu (1991). A rede foi capaz de aprender eficientemente uma sequência de dados de tempos passados de bioprocessos típicos através do algoritmo de retropropagação e fazer previsões um passo a frente de variáveis de estado em processos em batelada e em batelada alimentada. A entrada da rede consistiu de quatro neurônios (volume médio e concentrações de biomassa, substrato e produto no tempo presente) e a saída de três neurônios (concentrações de biomassa, substrato e produto no tempo seguinte). A camada escondida foi composta por quatro neurônios. Com respeito à capacidade preditiva, a rede neural apresentou respostas

rápidas e uma boa adaptabilidade, podendo ser incorporada como parte de um controlador de processo ou ser usada em processo de otimização.

Aplicação de rede neurais em fermentação de glicoamilase por *Aspergillus niger* foi realizada com sucesso por Linko e Zhu (1992). Para estimação, em tempo real, de atividade enzimática e biomassa, os autores empregaram redes neurais com seis neurônios na camada de entrada (pH, velocidade de agitação, taxa de consumo de oxigênio, taxa de evolução de dióxido de carbono, taxa de evolução de dióxido de carbono acumulado e taxa de utilização de nitrogênio acumulado), dez neurônios na camada escondida e dois neurônios na camada de saída (atividade enzimática e biomassa). Para previsão de atividade enzimática e biomassa, em tempo real, dois passos à frente, os autores empregaram redes neurais com as mesmas entradas e número de neurônios escondidos, mas com seis neurônios na camada de saída (atividade enzimática e biomassa nos instantes t , $t+1$ e $t+2$). A previsão, em tempo real, vários passos à frente foi excelente quando as condições operacionais foram mudadas. As redes foram treinadas usando algoritmo de retropropagação com termo de momento, em que a taxa de aprendizagem é aumentada sem conduzir a oscilação. Tanto na estimação quanto na previsão, os valores obtidos pelos modelos neurais foram muito próximos aos valores reais de atividade enzimática e biomassa obtidos em análises *off-line*.

Uma rede neural para estimar a concentração de biomassa, durante o curso da fermentação, a partir de variáveis do processo prontamente disponíveis, foi desenvolvida por Di Massimo *et al.* (1992). A rede foi usada para fornecer informação *on-line*, durante fermentação industrial para produção de penicilina G por *Penicillium chrysogenum*. A topologia especificada para rede foi de quatro camadas: uma camada de entrada, duas camadas escondidas e uma camada de saída. A rede foi treinada com os dados de fermentação selecionados, usando o algoritmo de *chemotaxis* (BREMERMAN e ANDERSON, 1989 *apud* Di MASSIMO *et al.*, 1992). A aplicação deste algoritmo para determinar a mais apropriada topologia resultou na especificação de seis nós por camadas, nas duas camadas escondidas. As medidas *on-line* selecionadas como entradas da rede foram: a alimentação do substrato, separada em duas componentes (devido à existência de duas alimentações de substrato primárias), o tempo e a taxa de evolução de dióxido de

carbono. A concentração da biomassa foi escolhida como saída. Estimativas da rede foram comparadas com dados de biomassa obtidos *off-line*. A qualidade dos resultados demonstrou que uma relação razoavelmente precisa foi identificada. Ficou demonstrada também a capacidade de uma rede neural, de escala relativamente modesta, capturar dinâmicas complexas de bioprocessos.

A aplicabilidade das redes neurais artificiais como importante ferramenta de apoio no desenvolvimento de estimadores inferenciais e de controladores preditivos não lineares foi discutida detalhadamente por Willis *et al.* (1992). Os autores aplicaram a técnica de modelagem através de redes neurais artificiais ao processo de produção de penicilina G por *Penicillium chrysogenum*. Uma rede neural artificial foi desenvolvida para fornecer estimativas da concentração de biomassa durante o curso da fermentação a partir de variáveis do processo disponíveis *on-line*. As entradas da rede foram a taxa de evolução de dióxido de carbono (CER – *Carbon Evolution Rate*), tempo e alimentação do substrato, dividida em duas componentes. A topologia empregada foi 4-6-6-1, quatro entradas, duas camadas escondidas com seis neurônios por camada e uma saída. As estimativas de biomassa feitas pela rede neural foram comparadas com medidas de biomassa obtidas de análises *off-line* e apresentaram boa precisão, sugerindo que a rede neural tinha capturado as características não lineares essenciais do processo e revelando o potencial exibido pelas redes neurais como *softsensores*. Adicionalmente, foi demonstrado que melhoras significativas na regulação do processo poderiam ser alcançadas se as estimativas produzidas pelo estimador neural fossem usadas como sinais de realimentação para controle. A possibilidade de uso de redes neurais diretamente dentro de uma estratégia de controle preditivo baseada em modelo também foi considerada, fazendo uso de uma rotina de otimização *on-line* para determinar as futuras entradas que minimizam os desvios entre as saídas previstas e desejadas. O melhor desempenho do controlador neural em relação a técnicas lineares convencionais foi demonstrado.

Dois *softwares* sensores para determinação *on-line* de biomassa, baseados em modelos genéricos – um baseado em modelo linear adaptativo aplicado a um processo industrial utilizando *Fusarium graminearum* e outro baseado em redes neurais artificiais, aplicado a produção industrial de penicilina G por *Penicillium chrysogenum* - foram

analisados por Montague, Morris e Tham (1992). No desenvolvimento do estimador adaptativo linear para determinação *on-line* de biomassa, os autores usaram como variável secundária medida, o dióxido de carbono na saída de gás do biorreator, como variável controlada, a biomassa e como variável manipulada a alimentação do biorreator. No desenvolvimento do estimador neural foram usadas como entrada da rede a alimentação do substrato separada em duas componentes, o tempo e o fluxo de ar deixando o reator. A topologia empregada foi de quatro entradas, duas camadas escondidas com 6 neurônios cada e uma saída. Os resultados mostraram que os *softsensors* foram capazes de fornecer boas estimativas da concentração de biomassa e ressaltaram melhoramentos no monitoramento *on-line* de bioprocessos que podem ser alcançados através do uso de *softwares* sensores.

A aplicação de redes neurais para estimação e previsão de variáveis de estado em processos biotecnológicos foi investigada por Karim e Rivera (1992a). Dois estudos de casos foram discutidos, em particular em operações batelada e com reciclo de células, na produção de etanol por *Zymomonas mobilis*. O primeiro caso consistiu de um cultivo em batelada de *Zymomonas mobilis*. Pesquisas anteriores em controle e otimização de produção de etanol por *Zymomonas mobilis* revelaram a importância das temperaturas de cultivo para o rendimento ótimo de etanol. Vários conjuntos de dados de cultivos foram obtidos em diferentes temperaturas, fornecendo uma fonte conveniente de informação sobre o comportamento do processo em várias condições ambientais. Neste primeiro caso, a rede neural empregada tinha quatro nós de entrada correspondendo a quatro medidas *on-line* do sistema (temperatura, potencial de redox, porcentagem de gás carbônico no gás exaurido e densidade ótica) e três nós de saída correspondendo a três variáveis de estado (concentrações de biomassa, glicose e etanol) para serem estimadas. Um *bias* foi incluído na camada de entrada para fornecer um limiar para ativação dos neurônios. Este caso de estudo ilustrou bem o uso de dados históricos experimentais no treinamento de redes neurais em bioprocessos. O segundo caso envolveu a simulação do cultivo de *Zymomonas mobilis* em sistemas com reciclo de células usando um modelo não estruturado. Neste caso, foram assumidas como entradas para a rede, a concentração de biomassa e a taxa de diluição e como saídas as concentrações de glicose e de etanol. As simulações foram realizadas em diferentes condições de tempo para fornecer dados de treinamento suficientes

para as redes neurais. Os treinamentos das redes neurais foram realizados com o algoritmo de otimização de gradiente conjugado. Estes estudos representaram problemas com sistemas altamente não lineares e variantes no tempo, cujos resultados foram apresentados e demonstraram a utilidade de estimadores baseados em redes neurais como ferramenta de identificação de sistemas e também de diagnóstico de contaminação de bioprocessos.

Dois estimadores baseados em redes neurais foram desenvolvidos por Karim e Rivera (1992b) para prever as principais variáveis do processo envolvidas na produção fermentativa de etanol pela bactéria anaeróbica *Zymomonas mobilis*. Redes neurais recorrentes e com alimentação progressiva (*feedforward*) foram investigadas e demonstraram ser apropriadas como estimadores de estado não mensuráveis. As redes neurais foram capazes de aproximar o comportamento não linear do processo fermentativo, usando dados de medidas brutos e sem requerer um modelo de processo simplificado derivado de princípios fundamentais. Os resultados da estimação foram comparados com dados experimentais obtidos de um biorreator em escala de laboratório rodando em modo batelada e mostraram que ambas as redes são capazes de estimar a biomassa e as concentrações de etanol relativamente bem. Entretanto, a estimativa da concentração de glicose sofreu, várias vezes, desvios do seu valor verdadeiro, provavelmente porque o modelo da rede foi incapaz de capturar e representar todos os fenômenos relevantes do processo com as quatro entradas fornecidas. De acordo com os autores, neste caso, a rede produziria uma melhor estimativa com a inclusão da taxa de produção de ácido como entrada adicional às quatro entradas fornecidas.

Redes neurais foram comparadas a experimentos fatoriais, por Kennedy, Prapulla e Thakur (1992), como técnicas de planejamento de meios de fermentação para produção de óleo intracelular por *Rhodotorula gracilis* (*Rhodosporium toruloides*). Para investigar três fatores - as concentrações de melaço, de nitrato de amônio e de extrato de levedura - em três níveis, vinte e sete experimentos foram requeridos para um planejamento fatorial completo. Em contraste, as redes neurais foram treinadas com 10 experimentos (uma redução de 63% em relação ao número de experimentos requeridos pelo planejamento fatorial) e previram os experimentos testes com razoável precisão. A porcentagem de erro da rede neural, baseada sobre um conjunto de dados de treinamento de 10 experimentos e

conjunto de dados de testes de 5 experimentos foi de 13,5% para biomassa, 8,4 % para o teor de lipídio e 2,8% para utilização de açúcar.

Uma pesquisa experimental sobre o uso de redes neurais para realizar o controle não linear de um biorreator de tanque agitado contínuo, com meio de crescimento contendo substrato rico em carbono (glicose) e cuja atividade biológica era sustentada visando converter o substrato em álcool foi realizada por Chtourou *et al.* (1993). A taxa de diluição na entrada e a concentração do substrato foram selecionadas como variáveis de controle. O algoritmo de retropropagação foi usado para identificação *on-line* e *off-line* do modelo inverso que forneceu a ação de controle. A arquitetura da rede usada para proposta de controle tinha cinco neurônios na camada de entrada e três neurônios na camada escondida. As entradas foram as concentrações do substrato $y^*(t+d)$, $y(t)$ e $y(t-1)$ e as taxas de diluição $u(t-1)$ e $u(t-2)$, com a primeira entrada (denotada por $y^*(t+d)$) representando ou a saída da planta na fase de aprendizagem ou a saída desejada na operação de controle. A saída da rede foi a concentração do substrato no tempo t . A saída foi medida no tempo t e a variável de controle foi avaliada e aplicada ao biorreator no tempo $t + \tau$, onde τ é o tempo de computação relacionado ao algoritmo de controle. Os resultados experimentais demonstraram o bom desempenho e a simplicidade de implementação do controlador neural adaptativo.

A possibilidade de aplicação de técnicas de redes neurais artificiais em controle de processo de fermentações microbianas foi investigada por Zhang *et al.* (1994). Os autores realizaram estudos sobre: (1) o efeito da estrutura da rede sobre a precisão da simulação, (2) os efeitos da preparação dos dados de treinamento sobre os resultados do treinamento, (3) o desenvolvimento e implementação de um protótipo de sistema de controle supervisorio baseado em redes neurais para fermentações de *Bacillus, thuringiensis*, em escala de laboratório. O objetivo de controle deste sistema foi regular o biorreator do laboratório para alcançar a máxima taxa de crescimento de microrganismos durante a fase de crescimento exponencial. O protótipo do sistema foi implementado tanto em simulações computacionais como em experimentos em escala de laboratório, com resultados promissores. O protótipo consistiu de uma unidade digital de medida e controle (DCU—*Digital measurement and Control Unit*) e de uma unidade de rede neural supervisora (NNSU—*Neural Network*

Supervisory Unit). A NNSU foi usada para estimar as condições do processo e supervisionar a operação da DCU de acordo com o estado do processo estimado. A unidade de rede neural supervisora funcionou como um sistema decisório que pesquisou os mais apropriados *set-points* para os controladores PID (*Proportional Integral and Derivative Controllers*) no DCU, durante o curso da operação. A componente chave no NNSU foi um estimador neural treinado com o algoritmo de retropropagação. Um conjunto de dados históricos do processo obtido de 18 fermentações foi usado no treinamento, para previsão da densidade ótica do meio de crescimento, usando medidas tomadas a cada hora, do tipo de inóculo, do tempo de processo acumulado, da densidade ótica do meio de crescimento, da temperatura do meio, do nível de pH e da concentração de oxigênio dissolvido. Os resultados obtidos comprovaram a eficiência do uso de redes neurais para controle automático de processos biológicos.

Uma estratégia geral usando redes neurais artificiais como alternativa para procedimento de planejamento fatorial de processos fermentativos foi desenvolvida por Glassey *et al.* (1994b), haja vista que apesar das técnicas convencionais de planejamento fatorial estarem disponíveis para apoiar a otimização de processos de fermentação, tais técnicas são limitadas em sua eficiência, pelas não linearidades características dos bioprocessos. Uma rede neural com topologia 5-2-3-1, isto é, cinco neurônios de entrada (taxa de alimentação, idade da batelada, concentração da batelada, tempo inicial da alimentação e temperatura da fermentação) duas camadas escondidas com dois e três neurônios em cada, respectivamente e um neurônio de saída (biomassa) foi utilizada para predizer os perfis de concentração biomassa em fermentações industriais com *Escherichia coli* K-12.. Uma rede neural com topologia 5-6-1, com as mesmas entradas usadas na previsão de biomassa (taxa de alimentação, idade da batelada, concentração da batelada, tempo inicial da alimentação e temperatura da fermentação), uma camada escondida com seis neurônios foi usada para fazer previsão da concentração de proteína recombinante. O esquema de otimização foi testado sob condições industriais e comparações das previsões do modelo com os desempenhos reais de subseqüentes fermentações confirmaram a precisão satisfatória da previsão de biomassa e a precisão aceitável de variáveis relacionadas a proteína na região mais próxima às condições operacionais.

Para controle de fermentações em lote de ácido láctico, a previsão do pH para um horizonte de tempo longo é uma informação muito útil. Dois modelos de redes neurais para previsão de pH sobre horizontes longos foram propostos por Latrille, Corrieu e Thibault (1994) para leites fermentados com culturas de bactérias de ácido láctico, puras (*Streptococcus thermophilus*) e mistas (*Streptococcus thermophilus* e *Lactobacillus bulgaricus*). Para fermentações isotérmicas (42°C) das culturas puras, um modelo de rede neural estático mostrou-se adequado para definir uma única curva de referência, que foi usada em conjunção com quatro métodos geométricos móveis, para prever o tempo final da fermentação com precisão menor que 20%. Para fermentações das culturas mistas conduzidas em várias temperaturas, um modelo dinâmico foi obtido com uma rede neural recorrente. Arquiteturas de pequena ordem foram testadas e procedimentos de aprendizagem diretos e semidiretos foram avaliados. As vantagens dos métodos semidiretos foram mostradas e a rede neural resultante foi capaz de prever o tempo final de fermentação com erro relativo médio de 7,7% na temperatura média de 40°-48° C. O bom desempenho alcançado pelas redes neurais - na previsão do tempo final de fermentação e dos valores futuros de pH, para lotes de culturas puras e misturadas de bactérias de ácido láctico - demonstrou que as mesmas possuíam capacidade preditiva e precisão suficiente para aplicação industrial.

Controle *on-line* de aminoácidos é complicado por incertezas típicas de processos biológicos e pelas dificuldades de monitoramento em tempo real de variáveis-chaves do processo. Lisina é um aminoácido essencial em nutrição humana e também é largamente usado em ração animal. Linko, Rajalahti e Zhu (1995) discutiram a construção e aplicação de redes neurais *feedforward* (alimentadas adiante ou com alimentação progressiva), treinadas com algoritmo de retropropagação, como *softwares sensores* em estimação de estado e previsão vários passos à frente de lisina produzida e açúcar consumido. Os resultados demonstraram que redes neurais bem treinadas podem ser usadas com sucesso em produção e controle de aminoácidos.

As características especiais de culturas de células plantas tornam difícil o uso convencional de técnicas analíticas para monitoração *on-line* de biomassa. Todavia, resultados promissores têm sido obtidos usando modelos matemáticos e algoritmos de

estimação recursivos. Entretanto, nesses casos, esforços experimentais adicionais são necessários para obter uma razoável descrição do processo. Recentemente, técnicas usando aproximações mais empíricas têm sido propostas para descrever processos complexos, minimizando o trabalho experimental necessário para sua aplicação. A evolução de biomassa em plantas de culturas de células foi monitorada por uma rede neural artificial *feedforward* desenvolvida por Albiol *et al.* (1995). A camada de entrada da rede foi formada pelo tempo atual, pela biomassa no tempo anterior e pelas concentrações de sucrose, glicose e frutose, nos dois tempos precedentes. A camada de saída foi constituída pela biomassa no tempo atual e pelas concentrações de sucrose, glicose e frutose no tempo anterior. O número de neurônios na camada escondida foi igual a três. Os resultados apresentados demonstraram que com o uso de apenas dois experimentos, a rede neural permitiu descrever precisamente a evolução da biomassa e do açúcar, superando a principal desvantagem da modelagem convencional.

Várias estratégias para produção de lipase por *Candida rugosa* ATCC 14830 em reator de tanque agitado foram investigadas por Wang *et al.* (1995) que também demonstraram a aplicação de modelos de redes neurais treinadas para estimação de biomassa e de atividade de lipase. Redes neurais relativamente simples, treinadas com o algoritmo de retropropagação, com topologia 4-7-1, com vetor de entrada formado pelo oxigênio dissolvido, dióxido de carbono acumulado produzido e dois valores passados da variável, concentração de biomassa ou atividade de lipase, a ser estimada e como saída da rede a concentração de biomassa ou a atividade enzimática no tempo atual. As redes neurais foram otimizadas para as entradas e o número de neurônios escondidos por tentativa e erro. As redes treinadas com dados experimentais de fermentações de lipase para estimação de biomassa e de atividade de lipase apresentaram excelentes ajustes ($R^2=0,999$ para biomassa e $R^2=0,997$ para atividade de lipase).

Uma rede neural empregando o algoritmo de retropropagação, com janela móvel, foi aplicada por Yet-Pole, Wu e Liu (1996) para modelagem dinâmica e estimação *on-line* de variáveis de estado não mensuráveis na fermentação de L-lisina. Os organismos foram mutantes auxotróficos para L-homoserina, resistentes a S-(2-aminoetil)-L-cisteína. Dados de 12 experimentos foram usados para treinar o estimador baseado em rede neural. A taxa

de absorção de oxigênio integrado, a taxa de produção de dióxido de carbono integrada, o quociente respiratório, a alimentação e a concentração inicial de açúcar foram escolhidas como entradas do estimador neural. A concentração de açúcar consumido, a concentração de L-lisina e a densidade ótica foram selecionadas como saídas do estimador neural. A topologia do estimador neural foi 13-11-3, ou seja, treze neurônios na camada de entrada, 11 neurônios na camada escondida e três neurônios na camada de saída. A concentração de açúcar reduzido foi calculada *on-line* através da concentração de açúcar consumido estimada e foi mantida em um valor determinado por meio de uma simples estratégia de alimentação compensatória. Com o estimador, a cultura de alimentação em lote pôde ser satisfatoriamente realizada. E mesmo com certo grau de variação do substrato, o estimador neural forneceu ainda estimativas satisfatórias das variáveis-chaves da fermentação.

Lisina, um aminoácido essencial para nutrição humana, normalmente é produzido por fermentação em grande escala, em biorreatores de tanque agitado. Zhu, Rajalahti e Linko (1996) investigaram a produção de lisina, em batelada alimentada, usando uma variedade industrial de *Brevibacterium flavum* crescida em um fermentador de 115 metros cúbicos, em um meio a base de melaço de beterraba. As dificuldades no monitoramento *on-line* de consumo de substrato e formação de produto complicam o controle em tempo real do processo. Os autores demonstraram que redes neurais multicamadas, bem treinadas com algoritmo de retropropagação podem ser empregadas para resolver tais problemas sem conhecimento prévio detalhado das relações das variáveis do processo. Várias topologias de redes neurais foram construídas e aplicadas em estimação de estado e previsão vários passos à frente, do açúcar consumido e da lisina produzida, usando variáveis medidas *on-line*, para proposta de controle de processo. Na estimação da lisina produzida, em todos os casos a precisão do ajuste sobre o período total de fermentação foi alta, com um coeficiente de determinação da mesma ordem de magnitude ($0,974 < R^2 < 0,977$). Um excelente ajuste foi obtido com uma rede neural simples de topologia 2-6-1, com o vetor de entrada consistindo apenas de oxigênio consumido e dióxido de carbono produzido. A estimativa simultânea de açúcar consumido e lisina produzida no tempo t , tendo como vetor de entrada o dióxido de carbono produzido e o quociente respiratório, também foi satisfatória do ponto de vista de fermentação industrial, com R^2 de 0,985 para açúcar consumido e R^2 de 0,987 para lisina produzida. Na previsão vários passos à frente, a melhor topologia entre as

testadas consistiu de dióxido de carbono acumulado e quociente respiratório, medidos *on-line* juntos com o açúcar consumido nos instantes $t-1$ e $t-2$. Os resultados também foram bastante satisfatórios, com valores de R^2 de 0,998, 0,997 e 0,994 (nos instantes t , $t+1$ e $t+2$ respectivamente) para açúcar e valores de R^2 de 0,959, 0,984 e 0,985 (nos instantes t , $t+1$ e $t+2$ respectivamente) para lisina. Os resultados obtidos na aplicação de redes neurais como *softsensores* no monitoramento de processos biológicos foram encorajadores e também mostraram que o sucesso do *softsensor* depende da boa representatividade dos dados usados no treinamento e na validação estatística.

Estimação *on-line* e previsão de atividade enzimática e concentração de biomassa, tanto na produção de glicoamilase a partir de fungos como na produção de lipase, a partir de levedura, foram realizadas satisfatoriamente por Linko, Luopa e Zhu (1997). Os experimentos para produção de glicoamilase foram realizados usando uma variedade industrial de *Aspergillus niger* e tiveram o pH controlado com amônia que também funcionou como fonte de nitrogênio para os fungos. Os experimentos para produção de lipase não tiveram o pH controlado e foram realizados com uma variedade de *Candida rugosa*. Várias topologias de redes neurais foram testadas, no processo de produção de glicoamilase, para estimação e previsão de atividade enzimática e concentração de biomassa através de medidas *on-line* de variáveis como pH, taxa de agitação, taxa de absorção de oxigênio, taxa de evolução de dióxido de carbono, total de oxigênio consumido, total de dióxido de carbono consumido. Procedimentos de testes com uma rede de topologia 4-10-6 apresentaram resultados muito satisfatórios, especialmente na previsão da atividade da glicoamilase. Uma rede neural com topologia 4-7-1 foi usada com sucesso para estimação da atividade enzimática da lipase. O uso de rede neural para previsão da atividade da lipase também possibilitou predizer o ponto final do processo de produção de enzima com base em dados *on-line* monitorados rotineiramente. Em ambos os casos, os resultados mostraram que redes neurais bem treinadas podem ser empregadas como *softwares* sensores para estimação *on-line* e previsão de parâmetros-chaves do processo em tempo real.

A produção de penicilina G por *Penicillium chrysogenum* IFO 8644 foi simulada por Cruz, Hokka e Giordano (1997), empregando uma rede neural *feedforward* com três camadas (uma camada de entrada, uma camada oculta com sete neurônios e uma camada de saída). O objetivo do algoritmo foi inferir variáveis não medidas *on-line*. A camada de entrada incluiu variáveis medidas *on-line*, tais como taxa de fluxo de ar, velocidade de agitação e concentração de oxigênio e as concentrações do produto e do substrato com seis horas de atraso (tempo suficiente para sua medida experimental). A camada de saída estimou as concentrações do microorganismo, do substrato e do produto. O treinamento da rede neural usou um algoritmo combinando dois procedimentos: pesquisa randômica e retropropagação. Os resultados foram promissores e observou-se que a rede neural foi capaz de descrever o comportamento não linear do processo. Além disso, os resultados mostraram que a técnica pode ser aplicada com sucesso a algoritmos de controle de processo devido a seu tempo de processamento e a sua flexibilidade na incorporação de novos dados.

As capacidades de diferentes redes neurais para prever as principais variáveis de estado (biomassa, concentrações de etanol e substrato, as variáveis de saída) em fermentações em lote de *Zymomonas mobilis* CP4 foram testadas por Shene, Diez e Bravo (1999). Dados experimentais registrados de fermentações em lote realizadas sob diferentes condições (meio, composição e temperatura) foram usados para treinar a rede e testar suas previsões. Primeiro, uma rede neural de caixa preta foi projetada para prever os três valores de saída fornecidos como valores de entrada da condição atual do sistema. A configuração da rede neural que forneceu melhores resultados possuía 10 neurônios na camada escondida (erro de treinamento total igual a $0,575 \times 10^{-3}$, erro de teste total igual a $0,719 \times 10^{-3}$). Melhores resultados foram obtidos usando redes neurais diferentes para cada uma das variáveis de saída (erro de teste total igual a $0,527 \times 10^{-3}$). Tentativas foram feitas para aumentar a capacidade de previsão dos modelos de rede neural, usando um modelo matemático do processo e uma rede neural para calcular os parâmetros cinéticos do modelo. Entretanto o erro foi superior ao dos casos anteriores. Neste caso, a melhor previsão foi obtida usando uma rede neural com 30 neurônios na camada escondida (erro total de treinamento igual a $5,535 \times 10^{-3}$, erro total de teste igual a $2,225 \times 10^{-3}$).

O estado da arte de redes neurais empregadas como *softwares* sensores em engenharia de processos, foi revisto por Linko, S.; Zhu e Linko, P. (1999). Vários fatores que afetam o desempenho de redes neurais (tais como arquitetura, topologia e função da rede, e seleção de variáveis de entrada em treinamentos e aplicações) são analisados e ilustrados com exemplos tirados de processos envolvendo produção de α -amilase, β -galactosidase, glicoamilase, lipase e xilase.

Um classificador de modelos de fermentação batelada capaz de selecionar entre vários modelos o que melhor se ajusta à dinâmica do processo foi desenvolvido por Vasilache *et al.* (2001). Um pré-tratamento dos dados denotado de auto-normalização foi proposto. Através de um estudo de parâmetro de sensibilidade demonstrou-se que a auto-normalização associa a uma família de modelos (mesma estrutura de modelo com diferentes parâmetros) uma representação aproximadamente única. Esta representação foi usada para treinamento de rede neural recorrente. A dimensão do conjunto de aprendizagem foi consideravelmente reduzida. A rede neural treinada foi usada para classificação de dados reais de fermentação láctica. O modelo que melhor se ajustou aos dados experimentais foi determinado e, deste o principal fenômeno governando o processo. A resposta do classificador neural representa apenas uma medida comparativa das características de cada um dos modelos considerados. Os resultados mostraram a boa capacidade da rede para reconhecer o melhor modelo. Esta técnica pode ser usada como uma ferramenta de apoio para modelagem de processos biotecnológicos.

Um sistema de controle e monitoramento *on-line* baseado em computador pessoal foi desenvolvido para fermentação de microrganismo por Liu, Wang e Lee (2001). Um sistema de HPLC *on-line* para análise de glicose e etanol do crescimento foi conectado ao fermentador via um equipamento de auto-amostragem, capaz de realizar pipetagem, filtração, diluição da amostra e injeção final dentro do HPLC através de automação baseada em um procedimento programável. As interfaces analógica/digital (A/D) e digital/analógica (D/A) foram equipadas a fim de processar os sinais dos eletrodos e do detector de HPLC, e para dirigir as bombas de alimentação, o motor do agitador e controlador da taxa de fluxo do gás. O programa que supervisiona o controle da velocidade de agitação, a taxa de fluxo de gás, o valor do pH, a taxa de alimentação do meio e as medidas *on-line* das

concentrações de etanol e glicose usando o Microsoft Visual Studio sob o sistema operacional Windows. Os sinais para os picos cromatográficos do HPLC *on-line* foram bem capturados e processados usando um filtro RC e um algoritmo de suavização. Este sistema de controle e monitoramento foi demonstrado ser efetivo na fermentação de etanol por *Zymomonas mobilis*, em ambos os modos de operação: batelada e batelada alimentada. Na adição do substrato e na determinação *on-line* por HPCL da concentração do produto, a concentração da biomassa na fermentação por *Zymomonas mobilis*, pode também ser estimada *on-line* através de controle de pH e de um *software* sensor implementado. O perfil da concentração do substrato na fermentação *feedback* seguiu bem o perfil do *set-point* devido à ação do *feedback* no controle da taxa de fluxo da alimentação.

A massa celular seca e o rendimento de glucan dependem principalmente da concentração dos componentes do meio. A relação entre as concentrações dos componentes do meio com o rendimento de glucan e com a massa celular seca é importante do ponto de vista de controle e otimização do processo, mas ainda não está bem estabelecida. Um modelo de rede neural artificial foi desenvolvido por Desai *et al.* (2004) para estimar a massa seca de levedura (*Saccharomyces cerevisiae* NCIM 3458) e seu teor em glucan, usando como entrada os seguintes componentes do meio: glicose, peptona, extrato de levedura, extrato de malte, manganês e magnésio. A capacidade de generalização da rede foi testada usando um conjunto de dados (conjunto de teste) diferente do conjunto de dados usado no treinamento. A rede previu o rendimento de glucan dentro de um intervalo de $\pm 3,5\%$ e a biomassa dentro de um intervalo de $\pm 5,5\%$ dos valores experimentais.

A presente revisão teve como propósito fornecer uma visão geral das técnicas e resultados obtidos em dois campos que envolvem e produzem novos resultados muito rapidamente: produção de biossurfactantes e desenvolvimento de *softsensors* neurais. Os bons resultados relatados na literatura sobre o desempenho de *softsensors* baseados em redes neurais como estimadores de variáveis-chaves de processos fermentativos e a não existência, na literatura analisada, de trabalhos envolvendo *softwares* sensores baseados em redes neurais para determinação de biomassa e atividade de emulsificação de biossurfactante produzido por *Candida lipolytica*, serviram de motivação para o desenvolvimento deste trabalho de doutorado. Portanto, o presente trabalho pretende

investigar e contribuir científica e tecnologicamente com as seguintes áreas, inexploradas e /ou pouco relatadas na literatura, no que diz respeito ao processo de produção de biossurfactante por *Candida lipolytica*:

- Otimização de meios de produção de biossurfactantes;
- Ampliação de escala de bioprocessos;
- Modelagem e controle de bioprocessos baseados em redes neurais artificiais e
- Biorremediação ambiental.

3. OTIMIZAÇÃO DE COMPONENTES DE MEIOS DE PRODUÇÃO DE BIOEMULSIFICANTES POR *Candida lipolytica*

3.1. Introdução

Processos biotecnológicos são muito complexos e freqüentemente pouco entendidos. Geralmente o conhecimento do fenômeno bioquímico em estudo é incompleto. Portanto é impossível otimizar bioprocessos apenas com base em suposições teóricas. Experimentos têm de ser realizados. Devido ao custo dos bioprocessos é necessário planejar experimentos biotecnológicos cuidadosamente com respeito aos seus objetivos. Para otimização de bioprocessos uma seqüência de experimentos tem de ser planejada a fim de alcançar a máxima produtividade com mínimo esforço experimental.

Muitos bioemulsificantes e seus processos de produção têm sido patenteados, mas poucos têm sido comercializados. O processo de fermentação possui a chave do melhoramento da economia global no processo de produção de bioemulsificantes. Desenvolvimentos recentes na área de otimização das condições de fermentação têm resultado em um significativo aumento nos rendimentos dos processos tornando eles mais atrativos comercialmente (DESAI e BANAT, 1997).

Para reduzir os custos de produção dos bioemulsificantes em relação aos emulsificantes sintéticos, melhoramentos dos meios de cultura e produção são realizados usualmente através de metodologia univariável (“um fator por vez”). A desvantagem desta técnica é que ela não inclui os efeitos das interações entre as variáveis empregadas e não descreve o efeito líquido dos vários constituintes do meio sobre a atividade de emulsificação. Por outro lado, planejamento fatorial completo incluindo todas as possíveis combinações para cada um dos fatores é uma ferramenta poderosa para o entendimento de processos complexos. Otimização através de planejamento fatorial e metodologia de superfície de resposta (RSM – *Response Surface Methodology*) é uma prática comum em biotecnologia para otimização de componentes de meios e condições de cultura (RAO,

JAYARAMAN e LAKSHMANA, 1993; CHEN, 1996; CROLA e KENNEDY, 2001; MURALIDHAR *et al.*, 2001; ADINARAYANA e ELLAIAH, 2002; PURI, BEG e GUPTA, 2002; RAO e SATYANARAYANA, 2003). Otimização estatística não apenas permite rápida varredura de um largo domínio experimental, mas também reflete o papel de cada um dos componentes. Basicamente a otimização do processo envolve três grandes passos: realização de experimentos estatisticamente planejados, estimação dos coeficientes de um modelo estatístico, e previsão da resposta e teste da adequação do modelo (Rao e Satyanarayana, 2003).

Planejamento fatorial e metodologia de superfície de resposta foram usados no presente estudo, para otimizar a atividade de emulsificação do bioemulsificante produzido por *Candida lipolytica* e também para paralelamente investigar a relação entre atividade de emulsificação e produção de biomassa, haja vista a não existência de consenso na literatura sobre a atividade de emulsificação ser um fenômeno dependente da densidade celular (GUERRA-SANTOS, KÄPELLI e FIECHTER, 1984; KITAMOTO *et al.*, 1992; NAVON- VENEZIA *et al.*, 1995; ARINO, MARCHAL e VANDECASTEELE, 1998; ROSENBERG e RON, 1999) e o uso crescente de biomassa como fonte alimentar.

3.2. Materiais e Métodos

3.2.1. Microrganismo e Meio de Preservação

A levedura *Candida lipolytica* UCP 988 foi obtida da coleção do banco de culturas do Núcleo de Pesquisas em Ciências Ambientais da Universidade Católica de Pernambuco. As colônias de leveduras foram mantidas em 4° C, em tubos de ensaio com o meio YMA (*Yeast Mold Agar*) inclinado, contendo 3 g/L de extrato de levedura, 3 g/L de extrato de malte, 10 g/L de D-glicose, 5 g/L de triptona e 15 g/L de agar. O pH foi ajustado para 5,0 com HCl. Transferências usando técnicas assépticas foram realizadas mensalmente para placas com agar fresco para assegurar a viabilidade celular.

3.2.2. Meios de Cultivo e Produção de Bioemulsificante

A produção de bioemulsificante por *Candida lipolytica* foi otimizada usando óleo vegetal como fonte de carbono e meio basal de baixo custo constituído de água do mar diluída em 50% suplementada com uréia, sulfato de amônio e fosfato monobásico de potássio. As concentrações dos componentes dos meios variaram de acordo com os valores e combinações especificados nas tabelas 2 ,3, 4 e 5, apresentadas a seguir.

3.2.2.1. Água do Mar

As amostras de água do mar, com salinidade de 35 ppt (*part per thousand*) e gravidade específica de 1025, usadas como meio basal, foram coletadas em garrafa de polietileno com capacidade de 2000 mL na praia do L, em Olinda, Pernambuco. Antes de serem usadas foram filtradas para retirar sólidos em suspensão e autoclavadas em vapor fluente.

3.2.2.2. Fonte de Carbono

Óleo de milho de padrão alimentar, originado da moagem do milho *Zea mays*, foi utilizado como fonte de carbono e suplemento dos meios de produção. Em função dos ácidos graxos principais em sua composição, o óleo de milho faz parte do grupo dos ácidos oléico/linoléico. Esse grupo compõe-se principalmente de óleos que contém em sua maioria ácidos graxos insaturados. Os ácidos graxos saturados estão presentes abaixo de 20%, sendo que os triacilgliceróis trissaturados estão quase que totalmente ausentes nos óleos deste grupo. Entre os ácidos graxos insaturados, contém 30,5% de ácido oléico (ω -9), 52,0% de ácido linoléico (ω -6) – um ácido graxo essencial que o organismo não é capaz de sintetizar – e 1 % de ácido linolênico (ω -3), em média. Sua composição média em ácidos graxos está apresentada na tabela 1, mas depende da variedade do milho e da localização da plantação, sendo que o milho é tolerante a uma variedade de condições ambientais e se adapta facilmente a diferentes climas, altitudes e estações do ano (HUI, 1996).

Tabela 1. Composição em ácidos graxos do óleo de milho (de RODRIGUES, 2002).

Nome	Nº de Carbonos	Insaturação	Composição(%)
	< 14	0	< 0,1
Ácido Mirístico	14	0	< 0,1
Ácido Palmítico	16	0	8,0 - 19,0
ÁcidoPalmitoléico	16	1	< 0,5
Ácido Esteárico	18	0	0,5 – 40,0
Ácido Oléico	18	1	19,0 – 50,0
Ácido Linoléico	18	2	34,0 – 62,0
Ácido Linolênico	18	3	< 2,0
Ácido Araquídico	20	0	< 1,0
Ácido Araquidênico	20	1	< 0,5
Ácido Behênico	22	0	< 0,5
Ácido Lignocérico	24	0	< 0,5

3.2.2.3. Fontes de Nitrogênio e Fósforo

Uréia, sulfato de amônio e fosfato monobásico de potássio foram selecionados como constituintes dos meios de baixo custo, a base de água do mar, empregados nos experimentos descritos a seguir, em função de informações existentes na literatura que sugerem que a capacidade de bactérias e leveduras degradarem óleo em água do mar é

dependente da adição a cultura de fontes exógenas de nitrogênio e fósforo. E entre as fontes de nitrogênio e fósforo relatadas, uréia, sulfato de amônio e fosfato monobásico de potássio são citadas como as que mais significativamente influenciaram a produção de bioemulsificante, em meios contendo água do mar (ATLAS e BARTHA, 1972; REISFELD, ROSENBERG e GUTNICK, 1972; HOROWITZ, GUTNICK e ROSENBERG, 1975; ROSENBERG *et al.*, 1975; PALEJWALA e DESAI, 1989; JONHASON *et al.*, 1992; VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000; VANCE-HARROP *et al.*, 2000; VANCE-HARROP *et al.*, 2003).

3.2.3. Preparação do inóculo

O inóculo para produção de bioemulsificante foi preparado pela transferência asséptica de várias alçadas de uma cultura fresca de *Candida lipolytica*, crescida sobre placas de *Petri* com agar a 28° C durante 48 h, em frascos *Erlenmeyer* de 250 mL, contendo 50 mL do meio YMB (*Yeast Malt Broth*), composição idêntica a do meio YMA, sem agar. A suspensão YMB foi incubada durante 48 h a 28° C em agitador rotatório a 150 rpm. A concentração final do inóculo foi de 10⁷ células por mililitros. O inóculo foi utilizado na proporção de 1% nos meios.

3.2.4. Determinação da salinidade

As salinidades da água do mar e dos cultivos livres de células foram determinadas através de refratômetro de mão sem compensação automática de temperatura, escala entre 0 e 100‰ (ppt) de salinidade e 1000 a 1070 de gravidade específica, resolução 0,7‰. Uma amostra de água do mar estéril filtrada (tomada da superfície com auxílio de uma micropipeta) foi colocada sobre o prato do refratômetro, que foi mantido contra a luz. A leitura da salinidade em parte por mil e da gravidade específica foi realizada através do visor do refratômetro.

3.2.5. Determinação da concentração da biomassa

A determinação da concentração de biomassa (massa seca) foi realizada por gravimetria, através de filtração do meio de cultivo em membrana Milipore de 0,22 μm , secagem a 70° C até peso seco e pesagem em balança analítica.

3.2.6. Determinação da atividade de emulsificação

A atividade de emulsificação para emulsão água-em-hexadecano foi determinada de acordo com o método descrito por Cirigliano e Carman (1984). As amostras coletadas durante o crescimento dos microrganismos foram filtradas através de membrana Milipore® (diâmetro do poro de 0,22 μm). Aliquotas de 2mL do filtrado livre de células foram transferidas para tubos de ensaio (15x125 mm) com tampa e diluídas com 2 mL de solução tampão de acetato de sódio 0,1 M (pH 3,0). A seguir, adicionou-se 1mL de hexadecano, tampou-se o tubo de ensaio e a mistura foi submetida à agitação máxima, durante 2 minutos a 25° C, em agitador de tubos (vortex). A emulsão uniforme resultante foi colocada em repouso por 10 minutos. Após este período, a absorbância foi medida, sendo o meio de cultivo utilizado como branco. O método define uma unidade de emulsificação como a quantidade de emulsificante que produz uma emulsão com absorbância de 1,0 em 540nm, ou seja, a quantidade de emulsificante capaz de elevar a absorbância do sistema em uma unidade.

3.2.7. Determinação da Tensão Superficial

A tensão superficial foi determinada pelo método do anel de *Du Noy* utilizando um tensiômetro digital SIGMA 70 (KSV Instruments Ltd., Helsinki, Finland) na temperatura ambiente (25°C). A medida da tensão superficial (mN/m) foi feita através do registro da força requerida para empurrar o anel de platina no líquido metabólico livre de células, através da interface ar-líquido.

3.2.8. Otimização de meios de produção de bioemulsificantes

No presente trabalho, a otimização dos componentes do meio de produção de bioemulsificantes por *Candida lipolytica*, visando à maximização da atividade de emulsificação para emulsões água-em-hexadecano foi realizada em duas etapas. Inicialmente, um planejamento fatorial completo 2^4 foi realizado para verificar os efeitos e interações do óleo de milho, uréia, sulfato de amônio e fosfato (monobásico) de potássio sobre a atividade de emulsificação do biossurfactante produzido por *Candida lipolytica*. Em seguida, planejamento composto central para três fatores e metodologia de superfície de resposta (BOX, HUNTER e HUNTER, 1978; BOX e DRAPPER, 1987; KHURY e CORNELL, 1987; MYERS e MONTGOMERY, 1995; MONTGOMERY, 1997) foram usados para localizar os níveis ótimos de uréia, sulfato de amônio e fosfato de potássio e também para investigar os efeitos e interações destes fatores sobre a atividade de emulsificação. Convém ressaltar que, durante o processo de otimização da variável resposta atividade de emulsificação, aproveitou-se para investigar o comportamento da variável resposta concentração de biomassa, visando verificar a existência de alguma relação entre estas duas variáveis respostas e também a possibilidade de reduzir os custos de produção do bioemulsificante produzido por *Candida lipolytica*, através de aproveitamento e comercialização da biomassa. Portanto, nos dois planejamentos realizados, os efeitos dos fatores selecionados e de suas interações sobre a concentração de biomassa foram também investigados. Informações detalhadas sobre os métodos e critérios de julgamento da significância, da adequação aos dados experimentais e da capacidade preditiva dos modelos propostos no presente capítulo, encontram-se descritos detalhadamente na literatura (BOX e WETZ, 1971; BOX, HUNTER e HUNTER, 1978; BOX e DRAPPER, 1987; MONTGOMERY e RUNGER, 1994; BARROS NETO, SCARMÍNIO e BRUNS, 1996; MONTGOMERY, 1997). Análise de variância, determinações dos coeficientes de regressão e gráficos foram realizados usando o programa Statistica® versão 6.0 (Statsoft.Inc, Tulsa/OK, USA).

3.2.8.1. Planejamento Fatorial Completo 2⁴

Neste planejamento, um total de 20 experimentos foi realizado, incluindo 4 repetições no ponto central para obtenção de estimativa da variabilidade em função do erro experimental e para permitir o teste de adequação do modelo de primeira ordem. Os experimentos foram realizados em ordem sorteada para evitar que erros atípicos fossem relacionados a determinadas combinações. Os intervalos e níveis dos componentes (fatores ou variáveis independentes) estudados são apresentados na tabela 2. Óleo de milho, uréia, sulfato de amônio e fosfato monobásico de potássio, componentes de meio essenciais para o crescimento celular e para produção de bioemulsificantes foram selecionados como variáveis independentes. Os valores das concentrações de óleo de milho, uréia, sulfato de amônio e fosfato de potássio do nível inferior foram selecionados a partir de resultados obtidos por Vance-Harrop (2000) em experimentos preliminares. O valor da concentração de óleo de milho no nível superior foi selecionado visando confirmar resultado obtido por Vance-Harrop (2000) sobre efeito inibitório do aumento da concentração da fonte de carbono na produção de biopolímeros com atividade de emulsificação. Os valores selecionados para as concentrações de uréia, sulfato de amônio e fosfato de potássio nos níveis central e superior do planejamento fundamentaram-se em resultados obtidos por Vance-Harrop (2000), que sugerem efeito positivo do aumento da concentração de uréia sobre a produção de biopolímeros com atividade de emulsificação. A concentração da biomassa e a atividade de emulsificação para emulsões água-em-hexadecano foram selecionadas como variáveis respostas e estudadas, em várias combinações dos constituintes dos meios usando óleo de milho como fonte de carbono, de acordo com a matriz do planejamento fatorial 2⁴ (Tabela 4). Os componentes de cada meio foram dissolvidos em 50 ml de água do mar natural, filtrada e estéril, com salinidade de 35 ppt (gravidade específica 1025) e 50mL de água destilada. Ajuste de pH de cada um dos meios para 5,3 foi realizado. Aliquotas de 1 mL do pré-inóculo contendo 10⁷ células/mL de *Candida lipolytica*, foram usadas para inocular cada um dos 20 frascos de Erlenmeyer de 250 mL, contendo 100 mL de cada meio de produção estéril. Os frascos foram incubados em agitador rotatório a 28° C e 150 rpm, durante 127 h.

Tabela 2. Níveis e valores das variáveis independentes do planejamento fatorial completo

Variável Independente	Nível		
	-1	0	+1
Óleo de Milho (% v/v)	5,00	7,50	10,00
(NH ₂) ₂ CO (% p/v)	0,10	0,25	0,40
(NH ₄) ₂ SO ₄ (% p/v)	0,10	0,60	1,1
KH ₂ PO (% p/v)	0,68	1,36	2,04

3.2.8.2. Planejamento Composto Central para Três Fatores.

Neste segundo planejamento, um total de 17 experimentos foi realizado, incluindo 3 repetições no ponto central para obtenção de estimativa da variabilidade em função do erro experimental. Os experimentos foram realizados em ordem sorteada para evitar que erros atípicos fossem relacionados a determinadas combinações. Os intervalos e níveis dos componentes são apresentados na tabela 3. A concentração da biomassa e a atividade de emulsificação para emulsões água-em-hexadecano foram estudadas, em várias combinações dos constituintes dos meios usando óleo de milho como fonte de carbono, de acordo com a matriz do planejamento composto central 2³ (Tabela 5). Os componentes de cada meio foram dissolvidos em 50 ml de água do mar natural, filtrada e estéril, com salinidade de 35 ppt (gravidade específica 1025) e 50mL de água destilada. O pH de cada um dos meios foi ajustado para 5,3. Aliquotas de 1mL do pré-inóculo contendo 10⁷ células/mL de *Candida lipolytica*, foram usadas para inocular cada um dos 20 frascos de Erlenmeyer de 250 mL, contendo 100 mL de cada meio de produção estéril. Os frascos foram incubados em agitador rotatório a 28° C e 150 rpm, durante 127 h.

Tabela 3. Níveis e valores das variáveis independentes do planejamento fatorial central composto 2³. Óleo de milho está fixado em 5%(v/v).

Variável Independente	Nível				
	-1,68	-1	0	+1	+1,68
(NH ₂) ₂ CO (% p/v)	0,196	0,4	0,7	1,0	1,204
(NH ₄) ₂ SO ₄ (% p/v)	0,454	1,1	2,05	3,0	3,646
KH ₂ PO ₄ (% p/v)	1,3056	2,04	3,12	4,2	4,934

3.3. Resultados e Discussão

Para substituir emulsificantes sintéticos por bioemulsificantes, é necessário diminuir custos de produção e desenvolver bioemulsificantes com características diferenciadas em relação aos surfactantes químicos. A substituição de fontes de nutrientes caras por substratos alternativos e de baixo custo é uma forma eficiente de reduzir custos. Óleos vegetais, como o óleo de milho, são fontes de carbono mais econômicas que as de origem petroquímica, como o n-hexadecano e outras. A produção de bioemulsificantes, que apresentam elevadas atividades de emulsificação, em meios contendo água do mar com salinidade superior a 3%, é um diferencial importante em relação à surfactantes convencionais que normalmente são inativados em meios com concentração salina entre 2 e 3% (BOGNOLO,1999).

Apesar da quantidade limitada de informações existentes na literatura, sobre fatores influenciando a produção de bioemulsificantes em meios de cultura de baixo custo, contendo óleo vegetal e água do mar, sabe-se que fontes de nitrogênio e fósforo podem ser importantes chaves para a regulação de síntese de bioemulsificantes. Trabalhos encontrados na literatura sobre este assunto, sugerem que a capacidade de culturas mistas ou puras de bactérias ou de culturas puras de leveduras degradarem significativamente óleo em água do mar, é dependente da adição àquelas culturas de uma fonte exógena de nitrogênio e fósforo. As fontes de nitrogênio e fósforo mais citadas nestes trabalhos são uréia, sulfato de amônio e fosfato monobásico de potássio (ATLAS e BARTHA, 1972; REISFELD, ROSENBERG e GUTNICK, 1972; HOROWITZ, GUTNICK e ROSENBERG, 1975; ROSENBERG *et al.*, 1975; PALEJWALA e DESAI, 1989; JONHASON *et al.*,1992; VRDOLJAK *et al.*, 1992; BANAT, 1995; VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000; VANCE-HARROP *et al.*, 2000; VANCE-HARROP *et al.*, 2003).

Portanto, no presente estudo, a produção de bioemulsificante por *Candida lipolytica* foi otimizada usando óleo de milho como fonte de carbono e um meio basal de baixo custo constituído de água do mar diluída em 50% suplementada com uréia, sulfato de amônio e fosfato monobásico de potássio.

O uso de planejamento fatorial associado à metodologia de superfície de resposta mostrou ser bastante eficiente no desenvolvimento de um meio ótimo para produção de bioemulsificante. Apenas dois planejamentos fatoriais, envolvendo um total de 37 experimentos, foram necessários para otimização dos constituintes do meio de produção de bioemulsificante por *Candida lipolytica*. Experimentos planejados estatisticamente são bastante efetivos justamente porque os efeitos e interações das variáveis independentes (fatores) sobre a variável independente (resposta) podem ser avaliados com um número limitado de experimentos.

As matrizes codificadas e os resultados das variáveis respostas dos planejamentos experimentais supracitados são apresentados, respectivamente, nas tabelas 4 e 5. As últimas colunas das tabelas 4 e 5 mostram os valores reais das variáveis respostas, concentração de biomassa e atividade de emulsificação para emulsões água-em-hexadecano (AE).

As variáveis independentes – concentração de óleo de milho, concentração de uréia, concentração de sulfato de amônio e concentração de fosfato de potássio – estão codificadas de acordo com a equação:

$$X_i = \frac{x_i - x_0}{\Delta x} \quad (1)$$

onde X_i , x_i e x_0 são respectivamente, o valor codificado, o valor não codificado e o valor no ponto central da i -ésima variável independente. Δx é o valor do incremento de mudança.

Os resultados do planejamento fatorial 2^4 e do planejamento composto central 2^3 , apresentados nas tabelas 4 e 5 e discutidos a seguir, confirmam a capacidade de *Candida lipolytica* crescer sobre óleo de milho, em meio basal constituído por água do mar suplementada com fontes de nitrogênio e fósforo e produzir grande quantidade de biomassa e atividade de emulsificação extracelular induzida capaz de estabilizar emulsões água-em-hexadecano em diversos meios, em operação batelada.

Tabela 4. Resultados do planejamento fatorial completo 2^4 para concentração de biomassa e atividade de emulsificação para emulsões água-em-hexadecano.

Ensaio	Óleo de Milho	$(\text{NH}_2)_2\text{CO}$	$(\text{NH}_4)_2\text{SO}_4$	KH_2PO_4	Biomassa (g/L)	AE* (UAE)
1	-1	-1	-1	-1	9,550	2,161
2	+1	-1	-1	-1	13,050	2,239
3	-1	+1	-1	-1	13,400	1,897
4	+1	+1	-1	-1	10,100	1,475
5	-1	-1	+1	-1	10,850	3,691
6	+1	-1	+1	-1	7,350	2,107
7	-1	+1	+1	-1	9,450	2,167
8	+1	+1	+1	-1	12,100	2,218
9	-1	-1	-1	+1	18,900	1,739
10	+1	-1	-1	+1	14,750	2,273
11	-1	+1	-1	+1	16,200	3,182
12	+1	+1	-1	+1	21,550	3,053
13	-1	-1	+1	+1	28,750	3,614
14	+1	-1	+1	+1	17,200	3,223
15	-1	+1	+1	+1	15,900	3,727
16	+1	+1	+1	+1	21,950	3,607
17	0	0	0	0	12,550	1,337
18	0	0	0	0	14,650	1,429
19	0	0	0	0	13,550	1,341
20	0	0	0	0	11,950	1,598

* Atividade de Emulsificação (AE) medida em Unidade de Atividade de Emulsificação (UAE).

Tabela 5. Resultados do planejamento composto central 2^3 para a concentração de biomassa e atividade de emulsificação para emulsões água-em-hexadecano.

Ensaio	$(\text{NH}_2)_2\text{CO}$	$(\text{NH}_4)_2\text{SO}_4$	KH_2PO_4	Biomassa (g/L)	AE* (UAE)
1	-1	-1	-1	28,033	3,830
2	+1	-1	-1	29,144	3,806
3	-1	+1	-1	18,656	3,982
4	+1	+1	-1	20,968	3,340
5	-1	-1	+1	20,292	2,250
6	+1	-1	+1	20,693	3,770
7	-1	+1	+1	28,989	3,380
8	+1	+1	+1	31,226	3,242
9	-1,68	0	0	32,285	3,202
10	+1,68	0	0	26,908	3,640
11	0	-1,68	0	25,177	3,480
12	0	+1,68	0	28,673	3,930
13	0	0	-1,68	30,444	3,940
14	0	0	+1,68	25,105	2,920
15	0	0	0	24,821	4,436
16	0	0	0	24,743	4,408
17	0	0	0	26,863	4,194

* Atividade de Emulsificação (AE) medida em Unidade de Atividade de Emulsificação (UAE)

3.3.1. Planejamento Fatorial Completo 2⁴

3.3.1.1. Variável Resposta: Biomassa

Com relação a experimentos preliminares realizados no Núcleo de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco (VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000), os resultados apresentados na tabela 4 indicam que altas concentrações de biomassa foram obtidas sob todas as condições teste estudadas no planejamento fatorial, principalmente nos meios ricos em fósforo e em fósforo e nitrogênio. A biomassa média produzida nos 20 ensaios do planejamento 2⁴ foi igual 14,83 g/L. Os meios pobres em fósforo produziram em média 10,73 g/L, enquanto os meios ricos em fósforo produziram em média 16,83 g/L e os meios ricos em fósforo e nitrogênio produziram em média 20,26 g/L. Estes resultados e em especial o ensaio 9 da tabela 4 ratificam a já conhecida importância do fósforo para produção de biomassa. O meio que apresentou menor produção de biomassa, 7,35 g/L, foi um meio composto por 0,1 g de uréia, 1,1g de sulfato de amônio, 0,68 g de fosfato de potássio, 10 mL de óleo de milho, 50 mL de água do mar e 50mL de água destilada. O meio composto por 0,1 g de uréia, 1,1g de sulfato de amônio, 2,04 g de fosfato de potássio, 5 mL de óleo de milho, 50 mL de água do mar e 50mL de água destilada foi o mais eficiente para produção de biomassa, 28,750 g/L.

Os efeitos estimados para a variável resposta concentração de biomassa são apresentados na tabela 6. Os valores em negrito são estatisticamente significativos a um intervalo de confiança de 95%. Nas figuras 2a e 2b encontram-se ilustrados respectivamente o histograma dos efeitos das variáveis independentes do planejamento fatorial sobre a variável resposta biomassa e o diagrama de *Pareto*, com nível de significância de 95% para estimativa dos efeitos lineares, em valor absoluto. No diagrama de Pareto a magnitude de cada efeito é representada pelas colunas e a linha transversal, correspondente ao valor $p=0,05$, indica quão grande deve ser o efeito para ter significado estatístico. Pode ser observado, em ambas as figuras, que aumento da concentração de fosfato de potássio de 0,68 g/L (nível -1) para 2,04 g/L (nível +1), exerce um efeito positivo altamente significativo sobre a variável resposta concentração de biomassa. A interação entre as concentrações de óleo de milho e uréia e as interações entre as concentrações de sulfato de amônio e fosfato de potássio também exercem um efeito

positivo significativo sobre a produção de biomassa. Por outro lado, as demais variáveis independentes e interações não exercem efeito significativo sobre a concentração de biomassa.

Tabela 6. Efeitos estimados para a variável resposta concentração de biomassa.

	Efeito	Erro Padrão	t(3)	p
Média	14,688	0,265	55,464	0,000
(4)KH ₂ PO ₄	8,669	0,592	14,640	0,001
1 e 2	3,306	0,592	5,584	0,011
3 e 4	2,344	0,592	3,958	0,029
2 e 3	-1,219	0,592	-2,058	0,132
2 e 4	-1,031	0,592	-1,742	0,180
1 e 3	-0,969	0,592	-1,636	0,200
(3)(NH ₄) ₂ SO ₄	0,756	0,592	1,277	0,291
(1)Óleo Milho	-0,619	0,592	-1,045	0,373
1 e 4	-0,456	0,592	-0,771	0,497
(2)(NH ₂) ₂ CO	0,031	0,592	0,053	0,961

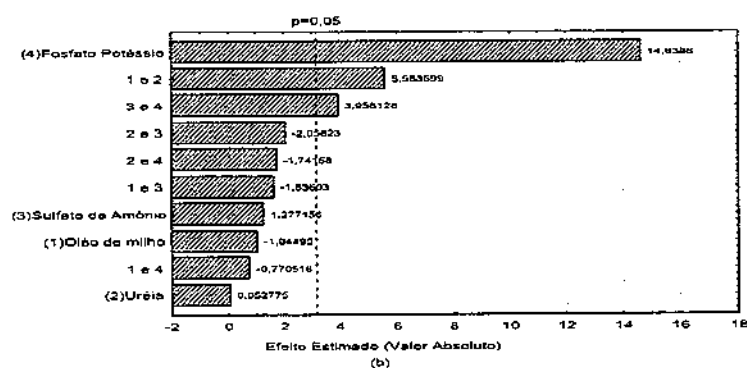
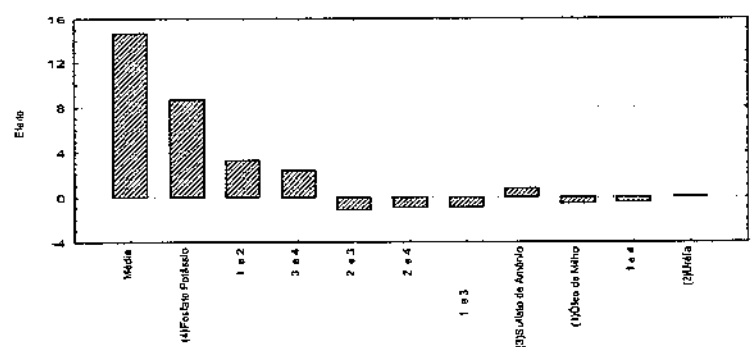


Figura 2. (a) Histograma dos efeitos das variáveis independentes sobre a variável resposta concentração de biomassa (b) Diagrama de Pareto de efeitos padronizados para o planejamento fatorial completo 2⁴. O ponto no qual os efeitos estimados foram estatisticamente significativos (em p=0,05) é indicado por uma linha tracejada vertical.

O ajuste de um modelo contendo interações até dois fatores aos dados de biomassa (Y) apresentados na tabela 4 resultou na equação:

$$Y = 14,69 - 0,30938 x_1 + 0,01562 x_2 + 0,37813 x_3 + \mathbf{4,33438} x_4 + \mathbf{1,65313} x_1 x_2 + \\ - 0,48438 x_1 x_3 - 0,22813 x_1 x_4 - 0,60938 x_2 x_3 - 0,51563 x_2 x_4 + \mathbf{1,17188} x_3 x_4 \quad (2)$$

onde os valores das variáveis independentes, concentração de óleo de milho (x_1), concentração de uréia (x_2), concentração de sulfato de amônio (x_3), e concentração de fosfato de potássio (x_4), estão na forma codificada (equação 1). Os coeficientes em negrito são estatisticamente significativos no nível de 95% de confiança.

A significância deste modelo foi verificada por análise de variância (ANOVA) e os resultados estão apresentados na tabela 7. O modelo apresenta um coeficiente de determinação satisfatório ($R^2=0,7803$) explicando 78% de variabilidade na resposta. O coeficiente de correlação de 0,8834 indica uma boa concordância entre os valores experimentais e os valores previstos pelo modelo. O teste F para o modelo mostra que o F calculado ($F_c = 3,20$) é maior que o F tabelado ($F_t(0,05; 10; 9) = 3,14$) demonstrando a significância do modelo, que, entretanto não pode ser usado para fins preditivos. O teste F para falta de ajuste mostra que o F calculado ($F_c = 12,38$) é maior que o F tabelado ($F_t(0,05; 6; 3) = 8,94$), indicando uma significativa falta de ajuste do modelo com os dados experimentais e necessidade de se realizar um planejamento fatorial estrela, para obter um modelo quadrático ou de segunda ordem que se ajuste melhor aos dados experimentais.

Tabela 7. Análise de Variância (ANOVA) para o modelo de regressão representando a biomassa.

Fonte	SQ	GL	MQ	Fc	Ft	Fc/Ft
Regressão	384,892	10	38,489	3,20	3,14	1,06
Resíduo	108,348	9	12,039			
Falta Ajuste	104,140	6	17,357	12,38	8,94	1,38
Erro puro	4,208	3	1,403			
Total	493,239	19	25,960		$R^2=0,7803$	$R=0,8834$
SQ - Soma Quadrática GL - Graus de Liberdade MQ - Média Quadrática F_c - F calculado F_t - F tabelado						

A adequação do modelo de regressão ajustado aos dados também foi verificada por meio de análise de resíduos. A análise de variância assume que as observações são independentes e normalmente distribuídas com a mesma variância em cada nível do fator. Estas suposições também foram avaliadas através de análise de resíduos. Os resíduos deixados pelo modelo (Figuras 3a e 3b), apesar de parecerem flutuar aleatoriamente em torno do valor zero, também sugerem uma ligeira curvatura, apresentando um comportamento sistemático, com valores passando de positivos para negativos e depois se tornando positivo novamente. Este padrão viola a suposição de que as observações são independentes. Isto sugere que talvez outras variáveis importantes não foram incluídas no planejamento e que o modelo talvez possa ser melhorado se um termo quadrático for incluído no mesmo. O gráfico dos valores previstos contra os valores observados (Figura 3c) confirma visualmente a moderada falta de ajuste do modelo. O gráfico de probabilidade normal (Figura 3d) não apresenta evidências que levem a rejeição da hipótese de normalidade.

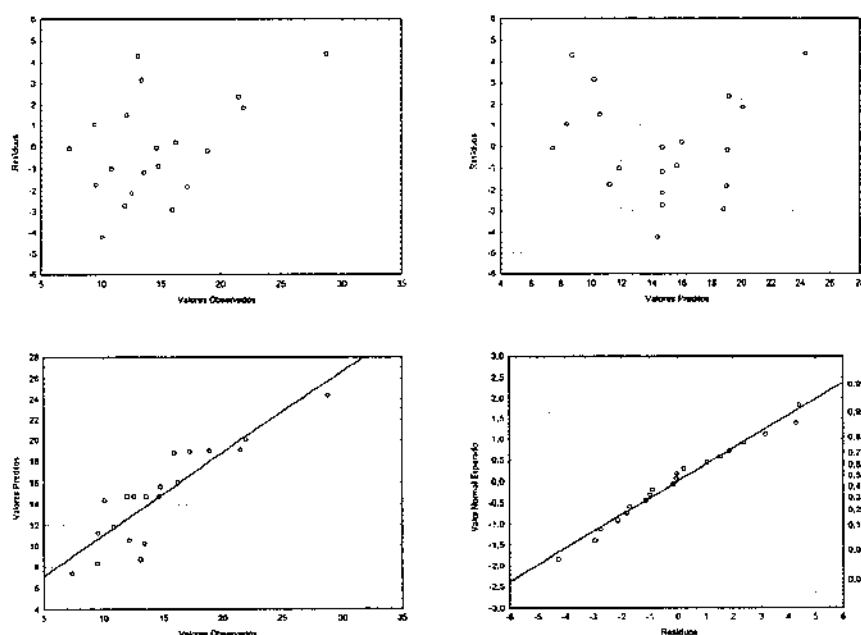


Figura 3. (a) Gráfico de resíduos contra valores observados; (b) Gráfico de resíduos contra valores previstos; (c) Gráfico de valores previstos contra valores observados; (d) Gráfico de probabilidade normal.

Superfícies de respostas (Figura 4) construídas a partir do modelo descrito pela equação 2, foram usadas para analisar as condições do meio de cultivo que levam a maior produção de biomassa.

A figura 4a mostra a superfície de resposta da concentração da biomassa em função da concentração de óleo de milho e da concentração de uréia, mantendo os valores da concentração de sulfato de amônio e de fosfato de potássio no ponto central. Verifica-se que a concentração da biomassa aumenta com a diminuição das concentrações de óleo de milho e uréia. Observa-se também que aumentando a concentração de óleo de milho e diminuindo a concentração de uréia ocorre uma diminuição na produção de biomassa.

A figura 4b mostra a superfície de resposta da concentração da biomassa em função da concentração de uréia e da concentração de fosfato de potássio, mantendo os valores da concentração de óleo de milho e da concentração de sulfato de amônio no ponto central. Verifica-se que a concentração da biomassa aumenta com a diminuição da concentração de uréia e aumento da concentração de fosfato de potássio. Observa-se também que diminuindo as concentrações de uréia e fosfato de potássio ocorre uma diminuição na produção de biomassa.

A figura 4c mostra a superfície de resposta da concentração da biomassa em função da concentração de óleo de milho e da concentração de sulfato de amônio, mantendo os valores da concentração de uréia e de fosfato de potássio no ponto central. Verifica-se que a concentração da biomassa aumenta com a diminuição da concentração de óleo de milho e com o aumento da concentração de sulfato de potássio. Observa-se também que diminuindo as concentrações de óleo de milho e sulfato de potássio ocorre uma diminuição na produção de biomassa.

A figura 4d mostra a superfície de resposta da concentração da biomassa em função da concentração de uréia e da concentração de sulfato de amônio, mantendo os valores da concentração de óleo de milho e da concentração de fosfato de potássio no ponto central. Verifica-se que a concentração da biomassa aumenta com a diminuição da concentração de uréia e aumento da concentração de sulfato de amônio. Observa-se

também que diminuindo as concentrações de uréia e sulfato de amônio ocorre uma diminuição na produção de biomassa.

A figura 4e mostra a superfície de resposta da concentração da biomassa em função da concentração de óleo de milho e da concentração de fosfato de potássio, mantendo os valores da concentração de uréia e de sulfato de amônio no ponto central. Verifica-se que a ocorrência de uma faixa ótima de concentração da biomassa para concentrações de óleo de milho variando do seu nível inferior (-1) até o ponto central (0) e para concentração de fosfato de potássio em seu nível superior (+1). Observa-se também uma faixa de concentração mínima de biomassa para a concentração de fosfato de potássio em seu nível inferior.

A figura 4f mostra a superfície de resposta da concentração da biomassa em função da concentração de sulfato de amônio e da concentração de fosfato de potássio, mantendo os valores da concentração de uréia e óleo de milho no ponto central. Verifica-se que a concentração da biomassa aumenta com o aumento das concentrações de sulfato de amônio e de fosfato de potássio, obtendo-se valores de biomassa superiores a 20 g/L. Observa-se também que aumentando a concentração de sulfato de amônio e diminuindo a concentração de fosfato de potássio ocorre uma diminuição na produção de biomassa, obtendo-se valores de biomassa inferiores a 9,5 g/L.

Os resultados obtidos para biomassa após 127 horas de cultivo, nesta primeira etapa do processo de otimização da atividade de emulsificação do biopolímero extracelular produzido por *Candida lipolytica*, também são bastante favoráveis quando comparados aos resultados (21 g/L e 10g/L) obtidos, em dois meios diferentes, por Musial, Rymowicz e Cibis (2004) no processo de otimização de produção de biomassa por *Yarrowia lipolytica*. O emprego de massa de levedura como suplemento rico em proteína para indústria de aquíicultura pode ser visto como forma de baratear custo de produção de biossurfactante por *Candida lipolytica*.

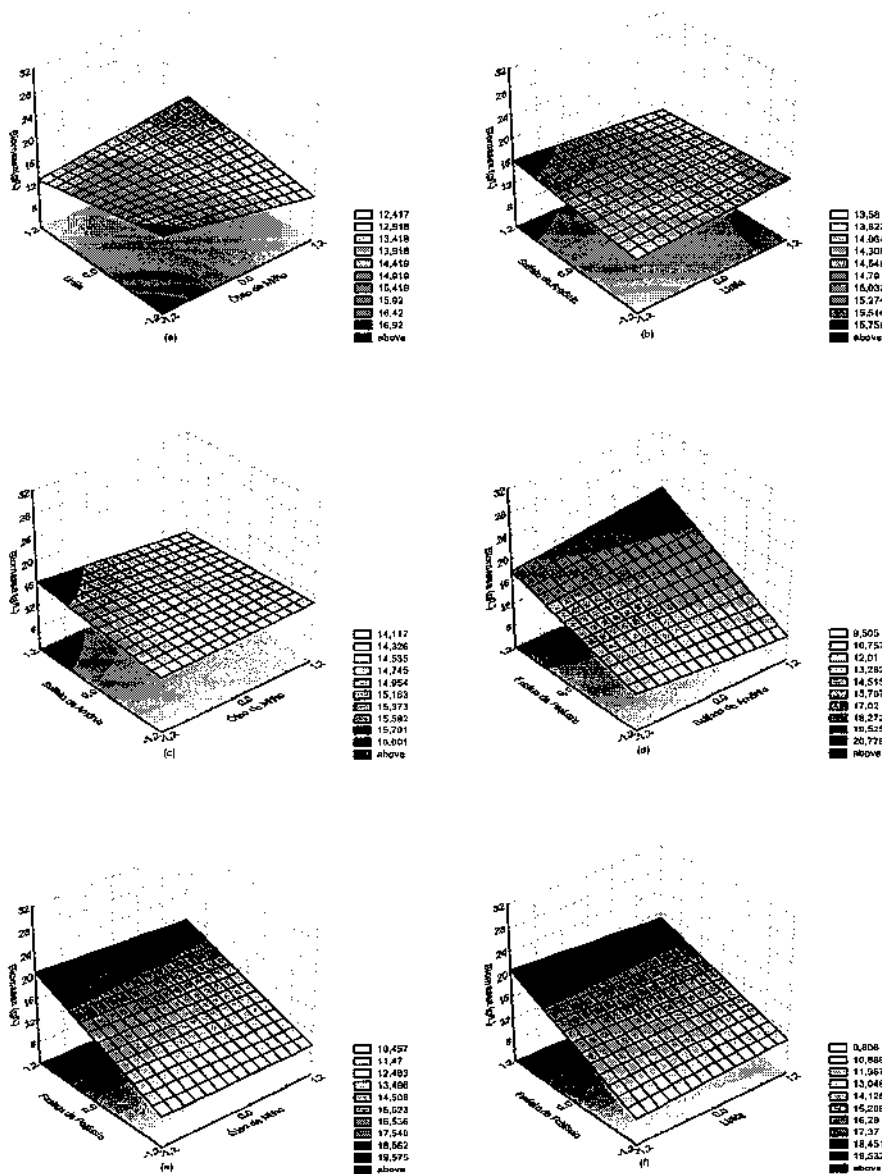


Figura 4. Superfícies de resposta e curvas de contorno obtidas para biomassa mostrando interações entre: (a) óleo de milho e uréia; (b) uréia e fosfato de potássio; (c) óleo de milho e sulfato de amônio (d) uréia e sulfato de amônio; (e) óleo de milho e fosfato de potássio e (f) sulfato de amônio e fosfato de potássio.

3.3.1.2. Variável Resposta: Atividade de Emulsificação

Os dados apresentados na tabela 4 indicam que altas atividades de emulsificação para emulsões água-em-hexadecano foram obtidas sob todas as condições teste estudadas no planejamento fatorial, principalmente nos meios ricos em fósforo e nitrogênio. Estes resultados estão de acordo com a literatura (REISFELD, ROSENBERG e GUTNICK, 1972; PALEJWALA e DESAI, 1989; VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2000; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2003, VANCE-HARROP, 2004) e demonstram a importância das fontes de nitrogênio e fósforo para a produção de bioemulsificante por *Candida lipolytica* em meio basal contendo água do mar. A atividade de emulsificação média para emulsões água-em-hexadecano produzidas a partir dos filtrados dos cultivos com 127 horas dos 20 ensaios do planejamento 2^4 foi igual a 2,40 UAE. Os meios ricos em fósforo e nitrogênio produziram em média 3,40 UAE, sendo que os meios ricos em fósforo e nitrogênio orgânico produziram em média 3,12 UAE enquanto que os meios ricos em fósforo e nitrogênio inorgânico produziram em média 3,54 UAE. As menores atividades de emulsificação foram apresentadas nos ensaios do ponto central (ensaios 17, 18, 19 e 20), compostos por 0,25 g de uréia, 0,6 g de sulfato de amônio, 1,36 g de fosfato de potássio, 7,5 mL de óleo de milho, 50 mL de água do mar e 50mL de água destilada e cuja atividade média foi igual a 1,43 UAE. A melhor atividade de emulsificação obtida foi 3,727 UAE com um meio composto com 0,4 g de uréia, 1,1 g de sulfato de amônio, 2,04 g de fosfato de potássio, 5 mL de óleo de milho, 50 mL de água destilada e 50 mL de água do mar (Tabela 3 – ensaio 15), denominado a partir deste ponto de meio YPASUC-1 (*Yeast, Potassium Phosphate, Amonium Sulfate, Urea, Corn Oil* - 1).

O ajuste de um modelo contendo interações até dois fatores aos dados de atividade de emulsificação para emulsão água-em-hexadecano (Y) apresentados na tabela 4 resultou na equação:

$$Y = 2,40 - 0,12392 x_1 + 0,01742 x_2 + 0,39592 x_3 + 0,40392 x_4 + 0,0465 x_1 x_2 + \\ - 0,13150 x_1 x_3 + 0,11083 x_1 x_4 - 0,132 x_2 x_3 - 0,32267 x_2 x_4 + 0,0945 x_3 x_4 \quad (3)$$

onde os valores das variáveis independentes, concentração de óleo de milho (x_1), concentração de uréia (x_2), concentração de sulfato de amônio (x_3), e concentração de fosfato de potássio (x_4), estão na forma codificada (equação 1). Os coeficientes em negrito são estatisticamente significativos no nível de 95% de confiança.

A tabela 8 apresenta os resultados da análise de variância do modelo obtido para a atividade de emulsificação. O modelo apresenta um coeficiente de determinação baixo ($R^2=0,58$) explicando apenas 58% de variabilidade na resposta. O coeficiente de correlação de 0,7616 indica uma concordância razoável entre os valores experimentais e os valores previstos pelo modelo. O teste F para o modelo mostra que o F calculado ($F_c = 1,24$) é menor que o F tabelado ($F_t(0,05; 10; 9) = 3,14$) demonstrando a não significância estatística do modelo. O teste F para falta de ajuste mostra que o F calculado ($F_c = 63,80$) é sete vezes maior que o F tabelado ($F_t(0,05; 6; 3) = 8,94$), indicando falta de ajuste, altamente significativa, do modelo com os dados experimentais e necessidade de se realizar um planejamento fatorial estrela, para obter um modelo quadrático ou de segunda ordem que se ajuste melhor aos dados.

Tabela 8. Análise de Variância (ANOVA) para o modelo de regressão representando a atividade de emulsificação.

Fonte	SQ	GL	MQ	Fc	Ft	Fc/Ft
Regressão	7,964	10	0,796	1,24	3,14	0,41
Resíduo	5,768	9	0,641			
Falta Ajuste	5,723	6	0,954	63,80	8,94	7,14
Erro puro	0,045	3	0,015			
Total	13,732	19	0,723		$R^2=0,58$	$R=0,7616$
SQ - Soma Quadrática GL - Graus de Liberdade MQ - Média Quadrática F _c - F calculado F _t - F tabelado						

O gráfico dos resíduos contra os valores ajustados (Figura 5a) e o gráfico dos resíduos contra os valores previstos (Figura 5b) evidenciam a falta de ajuste do modelo. Os gráficos sugerem que as suposições de independência e igualdade de variância tenham sido violadas, já que os resíduos deixados pelo modelo (Figuras 5a e 5b), encontram-se dispostos em sua maioria acima do valor zero e não aleatoriamente em torno dele. Isto sugere que a ordem dos experimentos é importante ou que outras variáveis são importantes e não foram incluídas no planejamento. Portanto, o modelo talvez possa ser melhorado se

um termo quadrático for incluído no mesmo. O gráfico dos valores previstos contra os valores observados (Figura 5c) confirma visualmente a falta de ajuste do modelo linear. Como os pontos do gráfico de probabilidade normal (Figura 5d) estão localizados aproximadamente ao longo de uma linha reta, não foram obtidas indicações de violação da hipótese de normalidade.

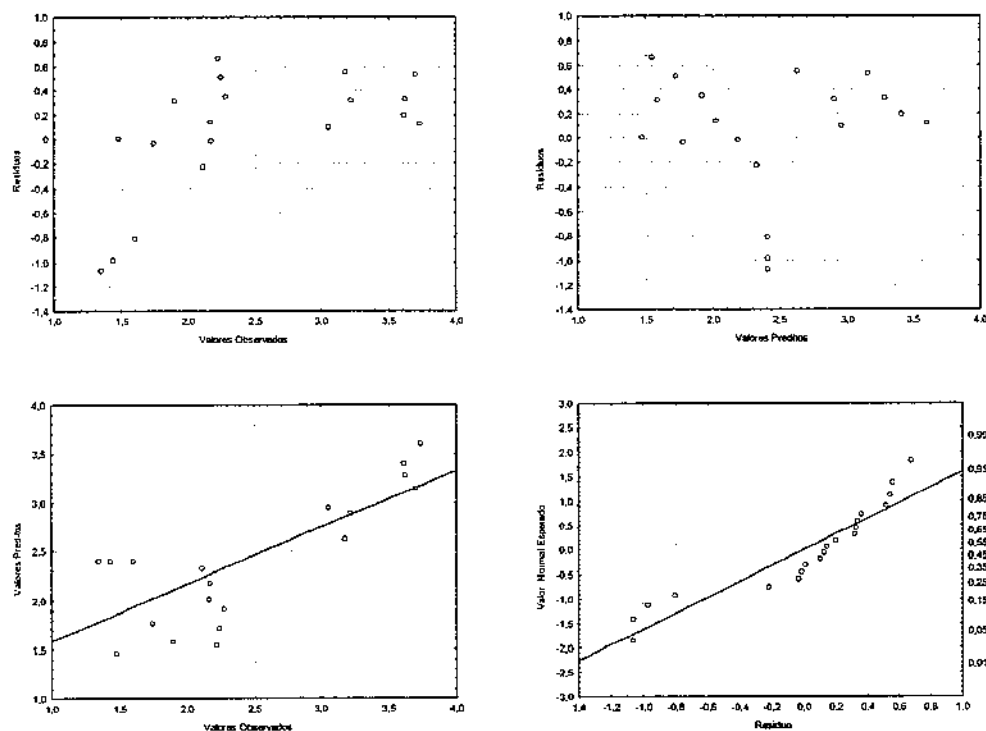


Figura 5. (a) Gráfico de resíduos contra valores observados; (b) Gráfico de resíduos contra valores previstos; (c) Gráfico de valores previstos contra valores observados; (d) Gráfico de probabilidade normal.

Como o modelo proposto anteriormente não foi satisfatório, pois apresentou falta de ajuste altamente significativa, teste de curvatura foi realizado. O ajuste de um modelo, considerando o efeito da curvatura, aos dados de atividade de emulsificação para emulsão água-em-hexadecano (Y) apresentou na tabela 4 resultou na equação:

$$Y = 2,65 - 0,12392 x_1 + 0,01742 x_2 + 0,39592 x_3 + 0,40392 x_4 + 0,0465 x_1 x_2 + \\ - 0,13150 x_1 x_3 + 0,11083 x_1 x_4 - 0,132 x_2 x_3 - 0,32267 x_2 x_4 + 0,0945 x_3 x_4 \quad (4)$$

onde os valores das variáveis independentes, concentração de óleo de milho (x_1), concentração de uréia (x_2), concentração de sulfato de amônio (x_3), e concentração de fosfato de potássio (x_4), estão na forma codificada (equação 1). Os coeficientes em **negrito** são estatisticamente significativos no nível de 95% de confiança.

A tabela 9 apresenta os resultados da análise de variância incluindo o efeito da curvatura, do modelo obtido para a atividade de emulsificação. O teste F para curvatura foi altamente significativo, já que o F calculado ($F_c = 319,28$) é cerca de trinta e duas vezes maior que o F tabelado ($F_t(0,05; 1; 3) = 10,13$). O modelo linear incluindo efeito da curvatura apresentou um elevado coeficiente de determinação ($R^2 = 0,9282$) explicando 92,82% de variabilidade na resposta. O coeficiente de correlação de 0,9634 indica uma excelente concordância entre os valores experimentais e os valores previstos pelo modelo. O teste F para o modelo mostrou que o F calculado ($F_c = 9,39$) é cerca de três vezes maior que o F tabelado ($F_t(0,05; 11; 8) = 3,32$) demonstrando significância estatística do modelo. O teste F para falta de ajuste mostrou o que o F calculado ($F_c = 12,60$) é maior que o F tabelado ($F_t(0,05; 5; 3) = 1,40$), indicando falta de ajuste significativa do modelo com relação aos dados experimentais e mantendo a necessidade de se realizar um planejamento fatorial estrela, para obter um modelo quadrático ou de segunda ordem que se ajuste melhor aos dados. Entretanto, a falta de ajuste deste modelo é aproximadamente 5 vezes menor que a do anterior.

Tabela 9. Análise de Variância (ANOVA) incluindo curvatura para o modelo de regressão linear representando a atividade de emulsificação para emulsão água-em-hexadecano.

Fonte	SQ	GL	MQ	F_c	F_t	F_c/F_t
Curvatura	4,781	1	4,781	319,78	10,13	31,57
Regressão	12,745	11	1,159	9,39	3,32	2,83
Resíduo	0,987	8	0,123			
Falta. Ajuste	0,942	5	0,188	12,60	9,01	1,40
Erro puro	0,045	3	0,015			
Total	493,239	19	25,960		$R^2=0,9282$	$R=0,9634$
SQ - Soma Quadrática GL - Graus de Liberdade MQ - Média Quadrática F_c - F calculado F_t - F tabelado						

A adequação do modelo de regressão considerando o efeito da curvatura também foi verificada por meio de análise de resíduos. Como pode ser observado - em relação aos gráficos de resíduos do modelo anterior (Figuras 5a e 5b) - os resíduos do modelo que incorporou o efeito da curvatura (Figuras 6a e 6b) apresentam uma melhor distribuição em torno do valor zero; mas ainda não apresentam um comportamento totalmente aleatório em torno do valor zero, evidenciando falta de ajuste do modelo. O gráfico dos valores previstos contra os valores observados (Figura 6c) confirma visualmente que o ajuste do modelo precisa ser melhorado. Como os pontos do gráfico de probabilidade normal (Figura 6d) estão localizados aproximadamente ao longo de uma linha reta, não foram obtidas indicações de violação da hipótese de normalidade.

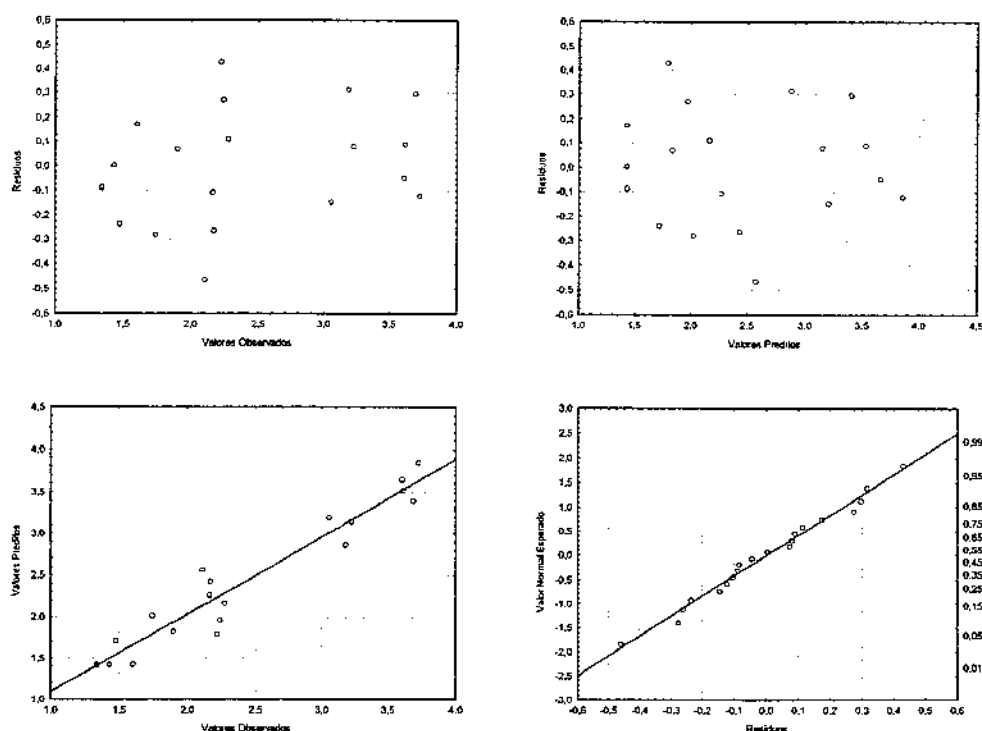


Figura 6. (a) Gráfico de resíduos contra valores observados; (b) Gráfico de resíduos contra valores previstos; (c) Gráfico de valores previstos contra valores observados; (d) Gráfico de probabilidade normal.

Os efeitos estimados para a variável resposta atividade de emulsificação para emulsão água-em-hexadecano são apresentados na tabela 10. Os valores em negrito são estatisticamente significativos a um intervalo de confiança de 95%.

Tabela 10. Efeitos estimados para a variável resposta atividade de emulsificação.

	Efeito	Erro Padrão	t(3)	p
Média	2,648	0,031	86,636	0,000
Curvatura	-2,445	0,137	-17,882	0,000
(4)KH₂PO₄	0,808	0,061	13,213	0,001
(3)(NH₄)₂SO₄	0,792	0,061	12,952	0,001
2 e 4	0,645	0,061	10,556	0,002
2 e 3	-0,264	0,061	-4,318	0,023
1 e 3	-0,263	0,061	-4,302	0,023
(1)O_MILHO	-0,248	0,061	-4,054	0,027
1 e 4	0,222	0,061	3,626	0,036
3 e 4	0,189	0,061	3,091	0,054
1 e 2	0,093	0,061	1,521	0,226
(2)(NH₂)₂CO	0,035	0,061	0,570	0,609

Nas figuras 7a e 7b encontram-se ilustrados respectivamente o histograma dos efeitos das variáveis independentes do planejamento fatorial sobre a variável resposta biomassa e o diagrama de Pareto, com nível de significância de 95% para estimativa dos efeitos lineares, em valor absoluto.

Como se pode observar a concentração de fosfato de potássio e a concentração de sulfato de amônio, nesta ordem, são as variáveis mais significativas e que mostram maior influência sobre a variável resposta atividade de emulsificação para emulsão água-em-hexadecano. As contribuições individuais da concentração de fosfato de potássio e da concentração de sulfato de amônio são positivas e significativas. O aumento da concentração de fosfato de potássio de 0,68 g/L (nível -1) para 2,04 g/L (nível +1) produz um efeito positivo significativo sobre a atividade de emulsificação para emulsão água-em-hexadecano. O aumento da concentração de sulfato de amônio de 0,1 g/L (nível -1) para 1,1 g/L (nível +1) também exerce um efeito positivo sobre a atividade de emulsificação para emulsão água-em-hexadecano.

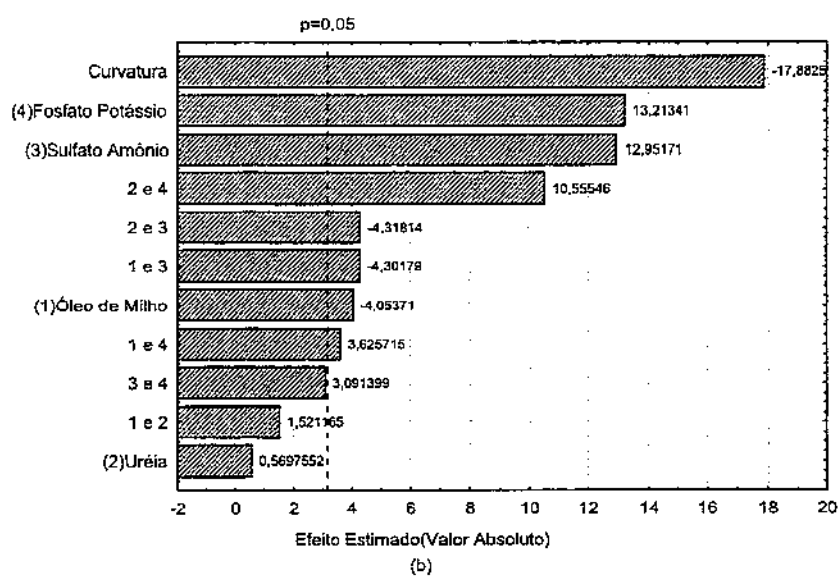
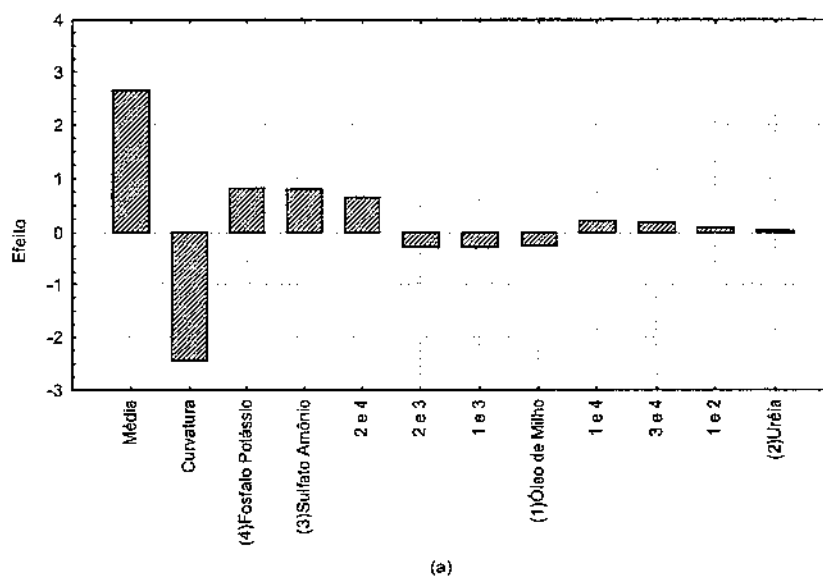


Figura 7. (a) Histograma dos efeitos das variáveis independentes sobre a variável resposta atividade de emulsificação. (b) Diagrama de Pareto de efeitos padronizados para o planejamento fatorial completo 2^4 . O ponto no qual os efeitos estimados foram estatisticamente significativos (em $p=0,05$) é indicado por uma linha tracejada vertical.

De acordo com estes resultados, na próxima etapa do processo de otimização, os níveis experimentais das variáveis concentração de fosfato de potássio e concentração de sulfato de amônio devem ser aumentados para que maiores valores de atividade de emulsificação sejam alcançados. Esses resultados sobre a importância das fontes de nitrogênio e fósforo para a produção de bioemulsificante corroboram com outros estudos, previamente discutidos, descritos na literatura. (ATLAS e BARTHA, 1972; REISFELD, ROSENBERG e GUTNICK, 1972; HOROWITZ, GUTNICK E ROSENBERG, 1975; ROSENBERG *et al.*, 1975; PALEJWALA e DESAI, 1989; JONHASON *et al.* (1992); VRDOLJAK *et al.*, 1992; BANAT, 1995; VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2000; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2003, VANCE-HARROP, 2004).

Por outro lado, o aumento da concentração de óleo de milho de 5% (nível -1) para 10% (nível +1) exerce um efeito negativo significativo sobre a atividade de emulsificação para emulsão água-em-hexadecano. Isto sugere que o óleo de milho tem um efeito repressivo sobre a atividade de emulsificação, quando usado na concentração de 10% (v/v) e um efeito indutivo quando empregado na concentração de 5%. Este resultado está de acordo com estudos relatados na literatura, sobre o uso da fonte de carbono na concentração de 5%, induzir atividade de emulsificação. Cirigliano e Carman (1984) estudando a indução de atividade de emulsificação para diferentes concentrações de n-hexadecano, durante o crescimento de *Candida lipolytica* obtiveram produção de liposom em meio YNB suplementado com 5% de hexadecano (fonte de carbono). Marçal (1991) trabalhando com diversos tipos de óleos vegetais (babaçu, coco e dendê) em diversas concentrações, também observou que a utilização da concentração de 5% da fonte de carbono induziu o aumento da produção de biopolímero com características emulsificantes. Vance-Harrop (2000) mostrou que empregando concentração de óleo de milho igual a 5%(v/v), *Candida lipolytica* foi capaz de produzir uma quantidade superior de biopolímero com atividade de emulsificação do que empregando concentração de óleo de milho igual a 10%(v/v). Baseado nesta avaliação, no próximo estágio do processo de otimização, o nível experimental do óleo de milho será mantido em seu valor inferior (5%), desde que este valor da concentração de óleo de milho causa um aumento na atividade de emulsificação.

A interação do fosfato de potássio com a uréia e a interação do fosfato de potássio com o óleo de milho produz efeitos positivos sobre a atividade de emulsificação para emulsão água-em-hexadecano. Entretanto, a interação do sulfato de amônio com a uréia e a interação do sulfato de amônio com o óleo de milho produz efeitos negativos sobre a atividade de emulsificação para emulsão água-em-hexadecano. O aumento da concentração de uréia de 0,1g/L para 0,4 g/L não exerce efeito significativo sobre a atividade de emulsificação para emulsão água-em-hexadecano.

Embora, a contribuição individual da uréia seja positiva e não significativa, a interação da uréia com o fosfato de potássio apresenta um efeito sinérgico positivo e muito significativo sobre a variável resposta atividade de emulsificação, que justifica a manutenção da uréia na segunda etapa do processo de otimização da atividade de emulsificação do bioemulsificante produzido por *Candida lipolytica*. Estudos preliminares que relatam a importância de uréia, sulfato de amônio e fosfato de potássio para a produção de bioemulsificante, também dão suporte a decisão de manter a uréia no planejamento. (ATLAS e BARTHA,1972; REISFELD, ROSENBERG e GUTNICK, 1972; HOROWITZ, GUTNICK E ROSENBERG, 1975; ROSENBERG *et al.*, 1975; PALEJWALA e DESAI, 1989; JONHASON *et al.* (1992); VRDOLJAK *et al.*,1992; BANAT, 1995; VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2000; VANCE-HARROP, BUARQUE-GUSMÃO e CAMPOS-TAKAKI, 2003, VANCE-HARROP, 2004).

A figura 7b também mostra que o efeito de curvatura cruza o nível 95% de confiança, indicando a proximidade do ponto ótimo e a necessidade de modelo quadrático e de planejamentos de segunda ordem que incorporem o efeito da curvatura.

Superfícies de respostas (Figura 8) construídas a partir do modelo descrito pela equação 4, foram usadas para analisar as condições do meio de cultivo que levam a maior atividade de emulsificação para emulsão água-em-hexadecano.

A figura 8a mostra a superfície de resposta da atividade de emulsificação em função da concentração de óleo de milho e da concentração de uréia, mantendo os valores

da concentração de sulfato de amônio e de fosfato de potássio no ponto central. Verifica-se que a atividade de emulsificação aumenta com a diminuição das concentrações de óleo de milho e uréia. Observa-se também que aumentando a concentração de óleo de milho e diminuindo a concentração de uréia ocorre uma diminuição na atividade de emulsificação.

A figura 8b mostra a superfície de resposta da atividade de emulsificação em função da concentração de uréia e da concentração de fosfato de potássio, mantendo os valores da concentração de óleo de milho e da concentração de sulfato de amônio no ponto central. Verifica-se que a atividade de emulsificação aumenta com o aumento da concentração de uréia e com o aumento da concentração de fosfato de potássio. Observa-se também que aumentando a concentração de uréia e diminuindo a concentração de fosfato de potássio ocorre uma diminuição na atividade de emulsificação.

A figura 8c mostra a superfície de resposta da atividade de emulsificação em função da concentração de óleo de milho e da concentração de sulfato de amônio, mantendo os valores da concentração de uréia e de fosfato de potássio no ponto central. Verifica-se que a atividade de emulsificação aumenta com a diminuição da concentração de óleo de milho e com o aumento da concentração de sulfato de potássio. Observa-se também ocorrência de atividade de emulsificação mínima para qualquer concentração de óleo de milho quando a concentração de sulfato de potássio encontra-se em seu nível inferior.

A figura 8d mostra a superfície de resposta da atividade de emulsificação em função da concentração de uréia e da concentração de sulfato de amônio, mantendo os valores da concentração de óleo de milho e da concentração de fosfato de potássio no ponto central. Verifica-se que a atividade de emulsificação aumenta com a diminuição da concentração de uréia e aumento da concentração de sulfato de amônio. Observa-se também que diminuindo as concentrações de uréia e sulfato de amônio ocorre uma diminuição na atividade de emulsificação para emulsões água-em-hexadecano.

A figura 8e mostra a superfície de resposta da atividade de emulsificação em função da concentração de óleo de milho e da concentração de fosfato de potássio, mantendo os valores da concentração de uréia e de sulfato de amônio no ponto central.

Verifica-se a ocorrência de uma faixa ótima de atividade de emulsificação para concentrações de óleo de milho variando do seu nível inferior (-1) até o ponto central (0) e para concentração de fosfato de potássio em seu nível superior (+1). Observa-se também uma diminuição na atividade de emulsificação quando se diminui a concentração de fosfato de potássio e aumenta a concentração de óleo de milho.

A figura 8f mostra a superfície de resposta da atividade de emulsificação em função da concentração de sulfato de amônio e da concentração de fosfato de potássio, mantendo os valores da concentração de uréia e óleo de milho no ponto central. Verifica-se que a atividade de emulsificação aumenta com o aumento das concentrações de sulfato de amônio e de fosfato de potássio, obtendo-se valores de atividade de emulsificação superiores a 3,57 UAE. Observa-se também que aumentando a concentração de sulfato de amônio e diminuindo a concentração de fosfato de potássio ocorre uma diminuição na atividade de emulsificação, obtendo-se valores de atividade de emulsificação inferiores a 1,99UAE.

A comparação das superfícies de respostas geradas pelos modelos descritos, respectivamente, pelas equações 2 e 4, para concentração da biomassa e da atividade de emulsificação, mostra que de uma maneira geral estas duas variáveis dependentes, crescem e decrescem no mesmo sentido, sendo afetadas de maneira semelhante pelas variáveis independentes. Este comportamento sugere, portanto a existência de uma relação entre crescimento de biomassa e aumento de atividade de emulsificação para emulsão água-em-hexadecano.

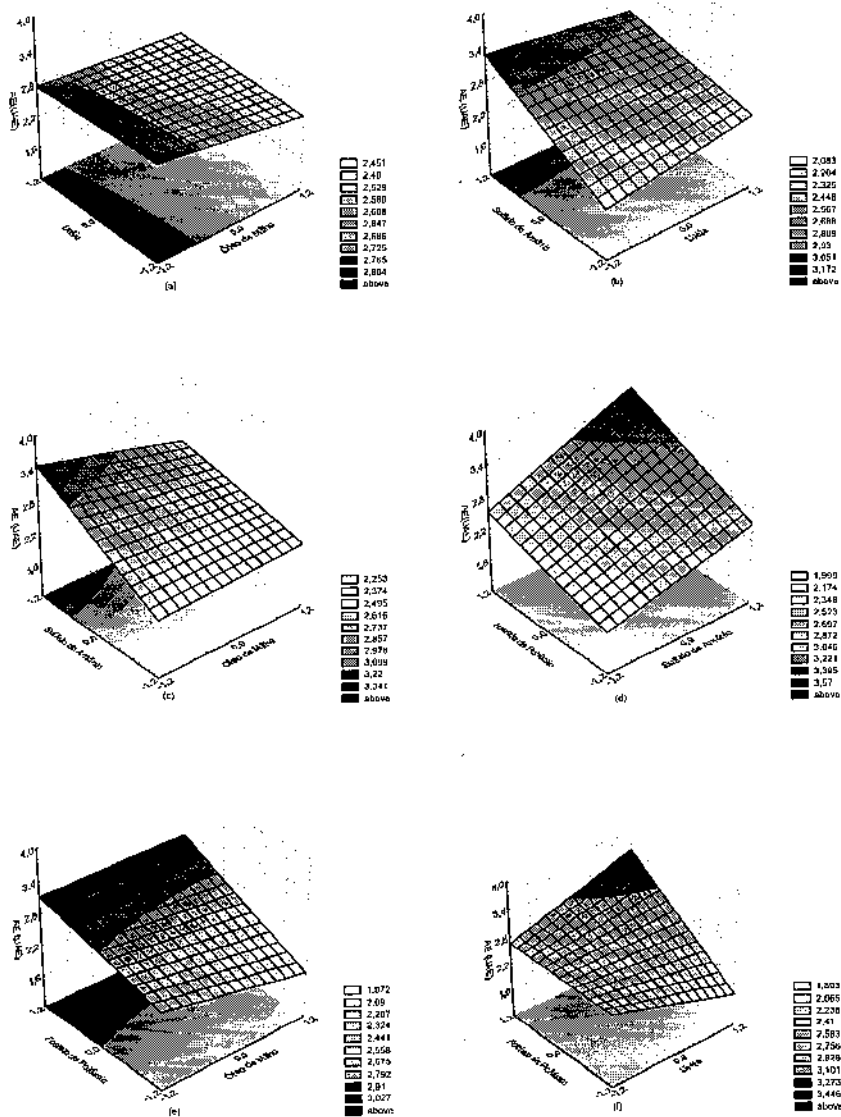


Figura 8. Superfícies de resposta obtidas para atividade de emulsificação para emulsão água-em-hexadecano mostrando interações entre: (a) uréia e fosfato de potássio (b) óleo de milho e uréia; (c) óleo de milho e fosfato de potássio; (d) sulfato de amônio e fosfato de potássio; (e) óleo de milho e sulfato de amônio e (f) uréia e sulfato de amônio.

3.3.2. Planejamento Composto Central para Três Fatores

Nessa segunda fase do processo de otimização, planejamento composto central e metodologia de superfície de resposta foram usados para ajustar um modelo de segunda ordem. A metodologia de superfície de resposta é um procedimento seqüencial. Quando um ponto na superfície de resposta está distante do ótimo, existe pouca curvatura no sistema e um modelo de primeira ordem pode ser apropriado. Entretanto, a análise dos resultados do planejamento fatorial completo 2^4 (Figura 7) indicou a existência de curvatura quadrática na resposta sobre a região explorada. Então, um planejamento composto central com mudanças nos intervalos das concentrações dos componentes do meio foi projetado, de forma a conduzir os experimentos rápida e eficientemente ao longo do caminho de melhoramentos para a vizinhança do ótimo. Desde que o principal objetivo desta etapa foi obter atividades de emulsificação para emulsões água-em-hexadecano tão altas quanto possível, os níveis dos fatores usados neste planejamento (Tabela 3) foram selecionados com base nos resultados, da variável resposta atividade de emulsificação (Tabela 4), obtidos no planejamento fatorial completo 2^4 , realizado na primeira etapa do processo de otimização. A concentração de óleo de milho foi mantida em 5% (v/v), devido ao efeito negativo significativo produzido sobre a atividade de emulsificação, quando foi aumentada para 10%, no planejamento fatorial completo 2^4 . Meios contendo 5% (v/v) de óleo de milho mostraram ser melhores meios para produção de bioemulsificantes por *Candida lipolytica*. Embora, no planejamento fatorial completo 2^4 , o efeito da concentração de uréia não tenha sido significativo, este fator foi mantido para novo estudo no planejamento composto central 2^3 devido ao efeito positivo e significativo da interação da uréia com o fosfato de potássio. Como no planejamento experimental completo 2^4 (Tabela 4) o uso do nível superior das concentrações de uréia, sulfato de amônio e fosfato de potássio resultou em efeitos positivos significativo sobre o aumento da atividade de emulsificação (Figura 7), as mesmas concentrações foram escolhidas para o nível inferior do planejamento composto central 2^3 . Empregando estes critérios, coincidentemente, as concentrações de óleo de milho, uréia, sulfato de amônio e fosfato de potássio do nível inferior do planejamento composto central 2^3 (Tabela 2) ficaram iguais a do melhor ensaio experimental do planejamento fatorial completo 2^4 (Tabela 4 – Ensaio 15).

3.3.2.1. Variável Resposta: Biomassa

Os resultados para as concentrações de biomassa de *Candida lipolytica*, obtidos após a execução do planejamento composto central 2³ estão apresentados na tabela 5. Em relação a experimentos preliminares realizados por Vance-Harrop *et al.* (1999) e Vance-Harrop, (2000) e aos ensaios realizados no planejamento fatorial completo 2⁴, altos valores de biomassa foram obtidos em todas as condições estudadas no planejamento composto central 2³. A biomassa média produzida nos 17 ensaios do planejamento composto central 2³ foi de 26,84 g/L, ou seja, 82,71 % maior que a biomassa média dos ensaios do planejamento fatorial completo 2⁴ (14,69 g/L). Os meios ricos em fósforo e nitrogênio produziram em média 28,04 g/L. A produção de biomassa no meio rico em fósforo e nitrogênio inorgânico (32,19 g/L) foi maior que a produção de biomassa no meio rico em fósforo e nitrogênio orgânico (20,69 g/L). O meio que apresentou menor produção de biomassa, 20,292 g/L, foi um meio composto por 0,4 g de uréia, 2,05g de sulfato de amônio, 4,2 g de fosfato de potássio, 5 mL de óleo de milho, 50 mL de água do mar e 50mL de água destilada. O meio composto por 0,196 g de uréia, 2,05 g de sulfato de amônio, 3,12 g de fosfato de potássio, 5 mL de óleo de milho, 50 mL de água do mar e 50 mL de água destilada foi o mais eficiente para produção de biomassa, 32,235 g/L.

Os efeitos estimados para a variável resposta concentração de biomassa são apresentados na tabela 11. Os valores em negrito são estatisticamente significativos a um intervalo de confiança de 95%. Nas figuras 9a e 9b encontram-se ilustrados respectivamente o histograma dos efeitos das variáveis independentes sobre a variável resposta biomassa e o diagrama de Pareto, com nível de significância de 95% para estimativa dos efeitos lineares, em valor absoluto. Os resultados obtidos mostram que a interação entre a concentração de sulfato de amônio e a concentração de fosfato de potássio exerceu um efeito positivo significativo sobre a produção de biomassa. O aumento da concentração de sulfato de amônio de 1,1 g/L (nível -1) para 3,0 g/L (nível +1) também favoreceu significativamente a produção de biomassa. Os demais efeitos principais e de interação da concentração de uréia, da concentração de sulfato de amônio e da concentração de fosfato de potássio sobre a biomassa não foram significativos no nível de confiança de 95%.

Tabela 11. Efeitos estimados para a variável resposta concentração de biomassa

	Efeito	Erro Padrão	t(2)	P
Média	25,600	0,693	36,963	0,001
2L e 3L	7,496	0,850	8,818	0,013
(2)(NH ₄) ₂ SO ₄ (L)	3,041	0,651	4,673	0,043
1L e 3L	-2,541	0,850	-2,989	0,096
(3)KH ₂ PO ₄ (L)	-2,370	0,651	-3,642	0,068
(NH ₂) ₂ CO (Q)	2,088	0,717	2,913	0,100
(1)(NH ₂) ₂ CO (L)	-1,667	0,651	-2,561	0,125
1L e 2L	1,504	0,850	1,769	0,219
KH ₂ PO ₄ (Q)	0,797	0,717	1,112	0,382
(NH ₄) ₂ SO ₄ (Q)	0,195	0,717	0,273	0,811

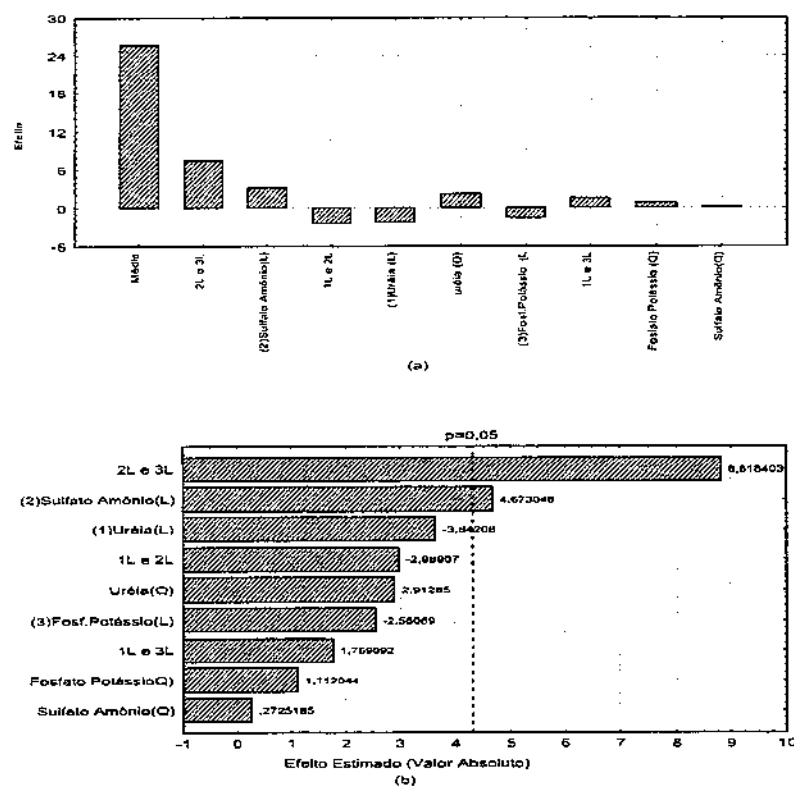


Figura 9. (a) Histograma dos efeitos das variáveis independentes sobre a variável concentração de biomassa . (b) Diagrama de Pareto de efeitos padronizados para o planejamento composto central 2³. O ponto no qual os efeitos estimados foram estatisticamente significativos (em p=0,05) é indicado por uma linha tracejada vertical.

O ajuste de um modelo de segunda ordem completo aos dados da concentração da biomassa (Y) apresentados na tabela 5 resultou na equação:

$$Y = 25,59991 - 1,18524 x_1 + 1,52075 x_2 - 0,83332 x_3 + 1,04424 x_1^2 + 0,09770 x_2^2 + 0,39869 x_3^2 - 1,27038 x_1 x_2 + 0,75188 x_1 x_3 + 3,74788 x_2 x_3 \quad (5)$$

onde os valores das variáveis independentes, concentração de uréia (x_1), concentração de sulfato de amônio (x_2), e concentração de fosfato de potássio (x_3), estão na forma codificada (equação 1). Os coeficientes em negrito são estatisticamente significativos no nível de 95% de confiança.

A tabela 12 apresenta os resultados da análise de variância do modelo quadrático obtido para a concentração da biomassa. O modelo quadrático apresenta um coeficiente de determinação relativamente alto ($R^2=0,8705$) explicando 87% de variabilidade na resposta. O coeficiente de correlação de 0,9330 indica uma ótima concordância entre os valores experimentais e os valores previstos pelo modelo. O teste F para o modelo mostra que o F calculado ($F_c = 5,23$) é maior que o F tabelado ($F_t(0,05; 9; 7) = 3,68$) demonstrando que o modelo é estatisticamente significativo, mas não é preditivo. O teste F para falta de ajuste mostrou o que o F calculado ($F_c = 3,78$) é menor que o F tabelado ($F_t(0,05; 5; 2) = 19,30$), não indicando, portanto, falta de ajuste do modelo quadrático aos dados experimentais.

Tabela 12. Análise de Variância (ANOVA) para o modelo de regressão representando a concentração da biomassa.

Fonte	SQ	GL	MQ	Fc	Ft	Fc/Ft
Regressão	203,056	9	22,562	5,23	3,68	1,42
Resíduo	30,216	7	4,317			
Falta Ajuste	27,326	5	5,465	3,78	19,30	0,20
Erro puro	2,890	2	1,445			
Total	233,272	16	14,579		$R^2=0,8705$	$R=0,9330$
SQ - Soma Quadrática GL - Graus de Liberdade MQ - Média Quadrática Fc - F calculado Ft - F tabelado						

A adequação do modelo quadrático ajustado aos dados também foi verificada por meio de análise de resíduos. O gráfico dos resíduos contra os valores observados (Figura 10a) e o gráfico dos resíduos contra os valores previstos (Figura 10b) não evidenciam a falta de ajuste do modelo nem de violação das suposições de independência e igualdade de variância, já que os resíduos deixados pelo modelo (Figuras 10a e 10b), encontram-se distribuídos aleatoriamente em torno do valor zero. O gráfico dos valores previstos contra os valores observados (Figura 10c) confirma visualmente a qualidade do ajuste. Indicações de violação da hipótese de normalidade não foram observadas através do gráfico de probabilidade normal (Figura 10d) já que os resíduos estão localizados aproximadamente ao longo de uma linha reta.

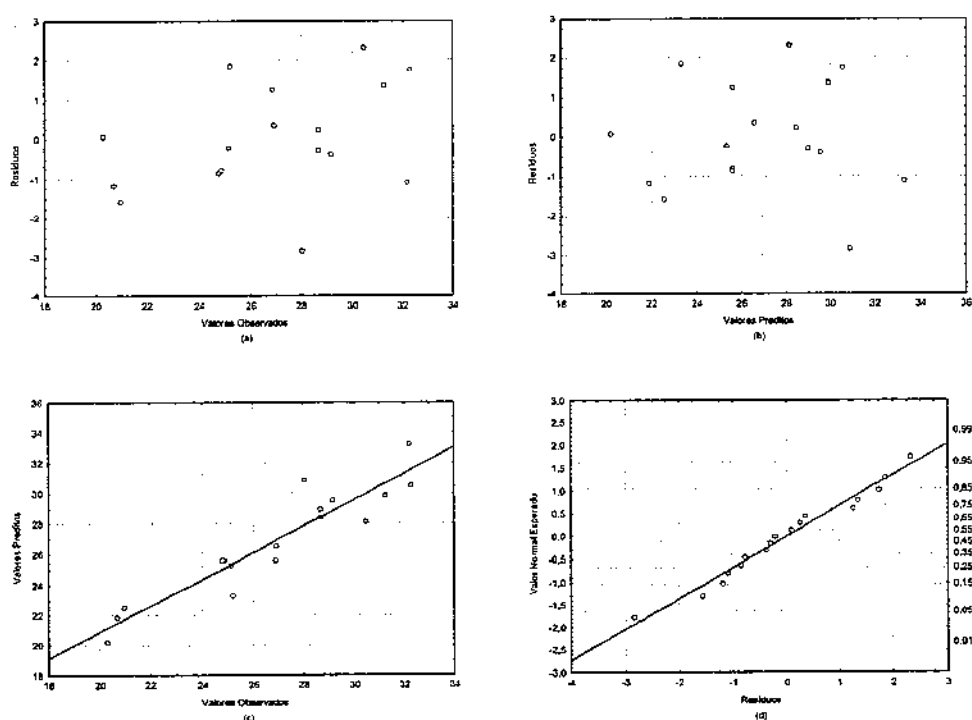


Figura 10. (a) Gráfico de resíduos contra valores observados; (b) Gráfico de resíduos contra valores previstos; (c) Gráfico de valores previstos contra valores observados; (d) Gráfico de probabilidade normal.

Superfícies de resposta (Figura 11) construídas a partir do modelo quadrático descrito pela equação 5, foram usadas para analisar as condições do meio de cultivo que levam a maior produção de biomassa.

A figura 11a mostra a superfície de resposta da concentração da biomassa em função da concentração de uréia e da concentração de sulfato de amônio, mantendo o valor da concentração de fosfato de potássio no ponto central. Verifica-se que a concentração da biomassa aumenta com a diminuição da concentração de uréia e aumento da concentração de sulfato de amônio. Observa-se também que, para concentrações de sulfato de amônio próximas do nível $-1,68$ e concentrações de uréia variando entre o nível $-1,68$ e o nível zero, a biomassa alcança concentrações em torno de 24 g/L, em uma região de mínimo.

A figura 11b mostra a superfície de resposta da concentração da biomassa em função da concentração de uréia e da concentração de fosfato de potássio, mantendo o valor da concentração de sulfato de amônio no ponto central. Verifica-se que a concentração da biomassa aumenta com a diminuição das concentrações de uréia e de fosfato de potássio para o nível $-1,68$. Observa-se também uma região de mínimo para biomassa, com concentrações em torno de 26 g/L, quando a concentração de uréia encontra-se em torno do nível $+1$ e as concentrações de fosfato de potássio aumentam do nível -1 até o nível $+1,68$.

A figura 11c mostra a superfície de resposta da concentração da biomassa em função da concentração de sulfato de amônio e da concentração de fosfato de potássio, mantendo o valor da concentração de uréia no ponto central. Verifica-se que a superfície de resposta da concentração da biomassa tem a forma de uma sela, com um vale de concentrações, em torno de 24 g/L, em seu ponto central. Quando as concentrações de sulfato de amônio e de fosfato de potássio crescem ou decrescem juntas - a partir do nível zero, em direção, respectivamente, ao nível $+1,68$ ou ao nível $-1,68$ - a concentração de biomassa aumenta atingindo valores superiores a 40 g/L. Quando a concentração de sulfato de amônio cresce e a de fosfato de potássio decresce ou vice-versa - a partir do nível zero, em direção, ao nível $+1,68$ ou ao nível $-1,68$ - a concentração de biomassa diminui atingindo valores inferiores a 11 g/L.

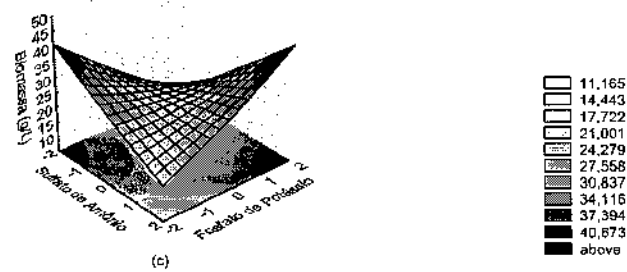
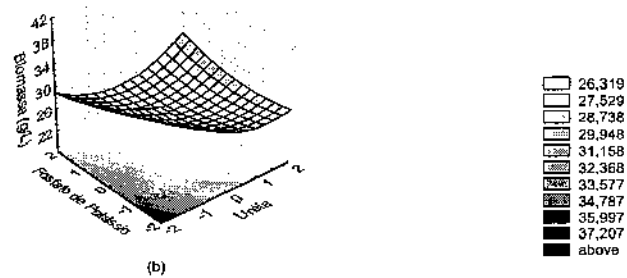
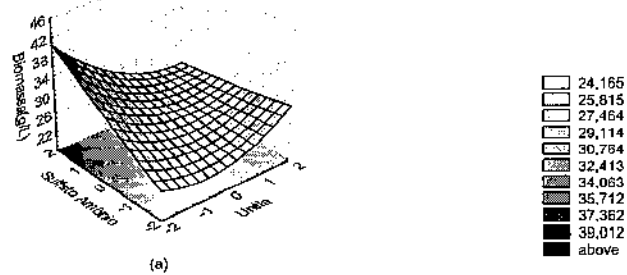


Figura 11. Superfície de resposta para biomassa mostrando as interações entre as concentrações de: a) uréia e sulfato de amônio; b) uréia e fósforo de potássio e (c) sulfato de amônio e fósforo de potássio.

Por meio da equação (5), os valores críticos (ponto de sela) codificados para as concentrações de uréia, sulfato de amônio e fosfato de potássio foram obtidos como sendo respectivamente iguais a 0,702, 0,118 e -0,174. Os valores críticos decodificados foram calculados como sendo iguais a 0,911% (p/v), 2,163% (p/v) e 2,932% (p/v), respectivamente. A máxima concentração de biomassa prevista nesta condição crítica foi 25,346 g/L. Os ensaios experimentais revelaram como sendo o melhor resultado uma concentração de biomassa de aproximadamente 32 g/L para um meio composto por 0,196% (p/v) de uréia, 2,05% (p/v) de sulfato de amônio e 3,12% (p/v) de fosfato de potássio (ensaio 9). Nos ensaios experimentais, a concentração média de biomassa foi igual a 24,89 g/L, sob as condições do ponto central: uréia 0,7% (p/v), sulfato de amônio 2,05% (p/v) e fosfato de potássio 3,12% (p/v). Resultados que confirmam a proximidade dos dados previstos pelo modelo aos dados experimentais, principalmente se levarmos em considerações as dificuldades de ordem prática existentes para separar a biomassa do óleo de milho residual.

Mesmo sem pretender otimizar a concentração de biomassa, valores elevados da mesma foram obtidos durante o processo de otimização da atividade de emulsificação do bioemulsificante produzido por *Candida lipolytica*. Os resultados médios obtidos para a concentração de biomassa (26,84 g/L), após 120 horas de cultivo, nos 17 ensaios realizados nesta segunda etapa do processo de otimização da atividade de emulsificação do biopolímero extracelular produzido por *Candida lipolytica* são bastante favoráveis quando comparados com os resultados médios obtidos para a concentração de biomassa (14,83 g/L), após 120 horas de cultivo, obtidos nos 20 ensaios do planejamento fatorial 2^4 e principalmente com os resultados de concentração de biomassa (21 g/L e 10g/L), obtidos por Musial, Rymowicz e Cibis (2004), usando dois meios de produção diferentes, em processo específico para otimização de produção de biomassa por *Yarrowia lipolytica*. Portanto, os valores de concentração de biomassa obtidos, nos experimentos de otimização da atividade de emulsificação do bioemulsificante produzido por *Candida lipolytica*, sugerem a possibilidade de redução de custos de produção, através do uso deste subproduto (massa de levedura) como mistura alimentar de proteínas de alta qualidade e baixo custo, destinada à indústria de aquicultura (BHATTACHARJEE, 1970; KIHLEBERG, 1972; RUMSEY, 1978; MARTIN *et al.*, 1993; LEE e KIM, 2001).

3.3.2.2. Variável Resposta: Atividade de Emulsificação

Os resultados para as atividades de emulsificação dos bioemulsificantes produzidos por *Candida lipolytica*, obtidos após a execução do planejamento composto central 2^3 estão apresentados na tabela 4. Os ensaios deste planejamento mostram que altos valores de atividades de emulsificação para emulsões água-em-hexadecano foram obtidos em todas as condições estudadas e que os mais altos valores de atividade de emulsificação para emulsões água-em-hexadecano foram encontrados nas condições do ponto central. A atividade de emulsificação média no nível zero foi igual a 4,346 UAE, ou seja, 262,77 % superior que a máxima atividade de emulsificação (1,198 UAE) do processo padrão – processo de produção de bioemulsificante por *Candida lipolytica*, com 120 h de cultivo, em meio YSWC-1, contendo uréia, sulfato de amônio, fosfato de potássio e 5% de óleo de milho (VANCE-HARROP, 2000). A atividade de emulsificação média dos ensaios do planejamento composto central 2^3 é igual a 3,632 UAE, 51,08 % superior a média das atividades de emulsificação (2,404 UEA) dos ensaios do planejamento fatorial completo 2^4 . Estes resultados são também bastante favoráveis quando comparados aqueles em que *Candida lipolytica* foi cultivada com diferentes fontes de carbono e componentes do meio (CIRIGLIANO e CARMAN 1984; MARÇAL, 1991; ROCHA, 1996; SARUBBO, PORTO e CAMPOS-TAKAKI, 1999; VANCE-HARROP *et al.*, 1999; VANCE-HARROP, 2000; SARUBBO *et al.*, 2001; VANCE-HARROP, GUSMÃO e CAMPOS-TAKAKI, 2003; VANCE-HARROP, 2004).

Cirigliano e Carman (1984) relataram que filtrados de cultura de *Candida lipolytica* crescida sobre meio YNB suplementado com 5% de substrato de carbono imiscível em água (glicose, hexadecano, parafina, óleo de soja, óleo de oliva, óleo de milho e óleo de algodão) por 130 horas a 27°C e 220 rpm em um agitador rotatório, produziram um agente com capacidade de emulsificar e estabilizar emulsões de hexadecano-em-água. A levedura *Candida lipolytica* quando crescida sobre meio YNB suplementado com óleo de soja produziu atividade de emulsificação máxima de 0,98 UAE. Nos meios YNB suplementados respectivamente com hexadecano e óleo de milho, as atividades de emulsificação foram iguais a 0,88 UAE. O bioemulsificante solúvel em água, purificado, chamado liposan, apresentou em sua composição 83% de carboidratos e 17% de proteínas (CIRIGLIANO e CARMAN, 1985).

Marçal (1991) relatou a produção de biopolímeros extracelulares por *Candida lipolytica* usando como fonte de carbono óleo de babaçu, óleo de coco e óleo de dendê, em diferentes concentrações. As atividades de emulsificação aumentaram progressivamente e máximos valores de atividades de emulsificação, todos inferiores a 0,5 UAE foram alcançados na fase estacionária. Quimicamente, os biopolímeros mostraram ser constituídos por proteínas, carboidratos e lipídeos.

Sarubbo, Porto e Campos-Takaki (1999) descreveram que a levedura *Candida lipolytica* crescida em meio YNB suplementado com 5% de óleo de babaçu, durante 144 horas a 28°C em agitador rotatório a 150 rpm, produziu um emulsificante extracelular durante cultivo em operação batelada e batelada alimentada. As máximas atividades de emulsificação alcançadas foram inferiores a 1,4 UAE. As moléculas dos bioemulsificantes foram constituídas de carboidratos, proteínas e lipídeos.

Sarubbo *et al.*(2001). relataram que a levedura *Candida lipolytica* produziu atividade de emulsificação extracelular induzida quando se utilizou glicose em diferentes concentrações como fonte de carbono durante cultivo em batelada a 27° C. Em todas as concentrações estudadas, atividades de emulsificação máximas, de aproximadamente 1,8 UAE, foram detectadas na fase decrescimento estacionário, após o pH atingir valores mínimos. O bioemulsificante isolado mostrou ser um biopolímero complexo constituído por proteínas , carboidratos e lipídeos.

Vance-Harrop *et al.*(1999) e Vance-Harrop(2000) usaram óleo de milho como fonte de carbono para produção de bioemulsificante por *Candida lipolytica* usando água do mar, suplementada com fonte de nitrogênio e tamponada com fosfato de potássio. Os melhores resultados foram obtidos com 144 horas de cultivo, usando o meio YSWC-1, contendo uréia, sulfato de amônio, fosfato de potássio e 5% de óleo de milho. A máxima atividade de emulsificação alcançada foi igual a 2,555 UAE. A composição bioquímica dos biopolímeros, com atividade de emulsificação, produzidos por *Candida lipolytica* indicou a presença de proteínas, carboidratos e lipídeos.

Vance-Harrop, Gusmão e Campos-Takaki (2003) relataram a produção por *Candida lipolytica* de biopolímeros extracelulares com atividade de emulsificação usando como fonte de carbono óleo de babaçu e D-glicose. Água do mar diluída em 50% suplementada com uréia, sulfato de amônio e fosfato de potássio foi usada como meio econômico. Os melhores resultados foram obtidos com 60 horas de fermentação em batelada alimentada, com o meio YSWB-2 contendo uréia, sulfato de amônio e 5% de óleo de babaçu e com o meio YSWB-3 contendo uréia, sulfato de amônio, fosfato de potássio e 5% de óleo de babaçu. Para os dois meios as máximas atividades de emulsificação alcançadas foram respectivamente iguais a 0,666 UAE e 0,158 UAE. As moléculas destes dois bioemulsificantes foram compostas por carboidratos, proteínas e lipídeos.

Embora os meios e condições de crescimento empregados por Cirigliano e Carman (1984) e outros pesquisadores discutidos anteriormente tenham sido bastante diferentes para garantir uma comparação direta, as atividades de emulsificação dos bioemulsificantes produzidos por *Candida lipolytica* no presente trabalho mostram ser significativamente superiores às atividades de emulsificação obtidas nos trabalhos supracitados. Fundamentando-se nas composições bioquímicas dos bioemulsificantes dos diversos trabalhos analisados, pode-se presumir que o bioemulsificante produzido neste estudo, e que ainda não foi caracterizado, deve ser composto por carboidratos e proteínas ou por carboidratos, lipídeos e proteínas.

Os efeitos estimados para a variável resposta atividade de emulsificação para emulsão água-em-hexadecano, nesta segunda etapa do processo de otimização da atividade de emulsificação do bioemulsificante produzido por *Candida lipolytica*, são apresentados na tabela 13. Os valores em negrito são estatisticamente significativos a um intervalo de confiança de 95%.

Nas figuras 12a e 12b encontram-se ilustrados respectivamente o histograma dos efeitos das variáveis independentes sobre a variável resposta atividade de emulsificação e o diagrama de Pareto, com nível de significância de 95% para estimativa dos efeitos, em valor absoluto.

Tabela 13. Efeitos estimados para a variável resposta atividade de emulsificação

	Efeito	Erro Padrão	t(2)	p	
	Média	4,34711	0,07627	56,99768	0,00031
	(NH2)2CO (Q)	-0,66288	0,07896	-8,39495	0,01389
	KH2PO4 (Q)	-0,65650	0,07896	-8,31418	0,01416
(3)	KH2PO4 (L)	-0,59064	0,07167	-8,24065	0,01441
	1L e 2L	-0,56900	0,09361	-6,07870	0,02601
	1L e 3L	0,51200	0,09361	5,46976	0,03184
	(NH4)2SO4 (Q)	-0,46163	0,07896	-5,84629	0,02803
	2L e 3L	0,22900	0,09361	2,44644	0,13424
(1)	(NH2)2CO (L)	0,21280	0,07167	2,96906	0,09718
(2)	(NH4)2SO4 (L)	0,15303	0,07167	2,13501	0,16631

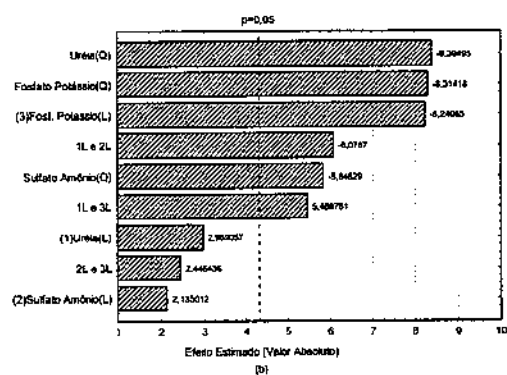
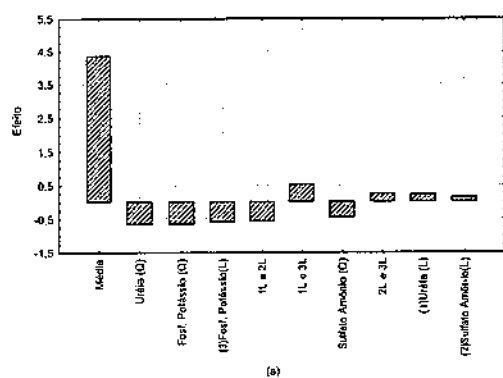


Figura 12.(a) Histograma dos efeitos das variáveis independentes sobre a variável resposta atividade de emulsificação. (b) Diagrama de Pareto de efeitos padronizados para o planejamento fatorial completo 2⁴. O ponto no qual os efeitos estimados foram estatisticamente significativos (em p=0,05) é indicado por uma linha tracejada vertical.

A análise dos resultados da tabela 13 mostra que as concentrações de uréia, fosfato de potássio e sulfato de amônio foram as variáveis que mais influenciaram a variável resposta atividade de emulsificação para emulsão água-em-hexadecano. As contribuições quadráticas da uréia, do fosfato de potássio e do sulfato de amônio são negativas e significativas. A contribuição linear do fosfato de potássio também é negativa e significativa. A contribuição linear da uréia é positiva e não significativa. Apenas a interação entre a uréia e o fosfato de potássio mostra um efeito positivo e significativo sobre a atividade de emulsificação. Entretanto este efeito é inferior aos efeitos individuais negativos e significativos das concentrações de uréia, do sulfato de amônio e do fosfato de potássio. O efeito da interação entre uréia e fosfato de potássio também é inferior ao efeito negativo e significativo da interação entre uréia e sulfato de amônio. Isto significa que aumentos na concentração de uréia, sulfato de amônio e fosfato de potássio de seus níveis inferiores para seus níveis superiores provocam uma diminuição na variável resposta, atividade de emulsificação.

As melhores condições para a atividade de emulsificação são, portanto, aquelas do nível zero do planejamento composto central para três fatores. As atividades de emulsificação aumentam quando as concentrações de uréia, sulfato de amônio e fosfato de potássio se aproximam das concentrações do nível zero e diminuem quando as concentrações de uréia, sulfato de amônio e fosfato de potássio se afastam das concentrações do nível zero. Os resultados mostram, portanto, que para atingir máxima atividade de emulsificação para emulsões água-em-hexadecano, as concentrações de uréia, sulfato de amônio e fosfato de potássio devem ser mantidas próximas as concentrações do ponto central do planejamento 2^3 .

O ajuste de um modelo de segunda ordem completo aos dados de atividade de emulsificação para emulsão água-em-hexadecano (Y) apresentados na tabela 5 resultou na equação:

$$\begin{aligned}
 Y = & 4,34719 + 0,10637 x_1 + 0,07650 x_2 - 0,29520 x_3 + \\
 & - 0,33114 x_1^2 - 0,23073 x_2^2 - 0,32796 x_3^2 + \\
 & - 0,28450 x_1 x_2 + 0,25600 x_1 x_3 - 0,11450 x_2 x_3
 \end{aligned} \tag{6}$$

onde os valores das variáveis independentes, concentração de uréia (x_1), concentração de sulfato de amônio (x_2), e concentração de fosfato de potássio (x_3), estão na forma codificada (equação 1). Os coeficientes em negrito são estatisticamente significativos no nível de 95% de confiança.

A tabela 14 apresenta os resultados da análise de variância do modelo quadrático obtido para a atividade de emulsificação. O modelo quadrático apresentou um alto coeficiente de determinação ($R^2=0,9572$) deixando de explicar apenas 4,281% de variabilidade na resposta. O coeficiente de correlação de 0,9784 indica uma excelente concordância entre os valores experimentais e os valores previstos pelo modelo. O teste F mostrou que o F calculado ($F_c = 17,40$) é aproximadamente quatro vezes maior que o F tabelado ($F_t (0,05; 9; 7) = 3,68$) demonstrando a alta significância do modelo e a possibilidade de seu uso para fazer previsões da atividade de emulsificação. O teste F para falta de ajuste mostrou o que o F calculado ($F_c = 1,99$) é muito menor que o F tabelado ($F_t (0,05; 5; 2) = 19,30$), indicando um ótimo ajuste do modelo aos dados experimentais.

Tabela 14. Análise de Variância (ANOVA) para o modelo de regressão representando a atividade de emulsificação para emulsão água-em-hexadecano.

Fonte	SQ	GL	MQ	Fc	Ft	Fc/Ft
Regressão	4,676	9	0,520	17,40	3,68	4,73
Resíduo	0,209	7	0,030			
Falta Ajuste	0,174	5	0,035	1,99	19,30	0,10
Erro puro	0,035	2	0,018			
Total	4,885	16	0,305		$R^2=0,9572$	$R=0,9784$
SQ - Soma Quadrática GL - Graus de Liberdade MQ - Média Quadrática F_c - F calculado F_t - F tabelado						

A adequação do modelo quadrático ajustado aos dados também foi verificada por meio de análise de resíduos. O gráfico dos resíduos contra os valores ajustados (Figura 13a) e o gráfico dos resíduos contra os valores previstos (Figura 13b) não evidenciam a falta de ajuste do modelo nem de violação das suposições de independência e igualdade de variância, já que os resíduos deixados pelo modelo (Figuras 13a e 13b), encontram-se distribuídos aleatoriamente em torno do valor zero. O gráfico dos valores previstos contra os valores observados (Figura 13c) confirma visualmente a qualidade do ajuste. Indicações de violação da hipótese de normalidade não foram observadas através do gráfico de

probabilidade normal (Figura 13d) já que os resíduos estão localizados aproximadamente ao longo de uma linha reta.

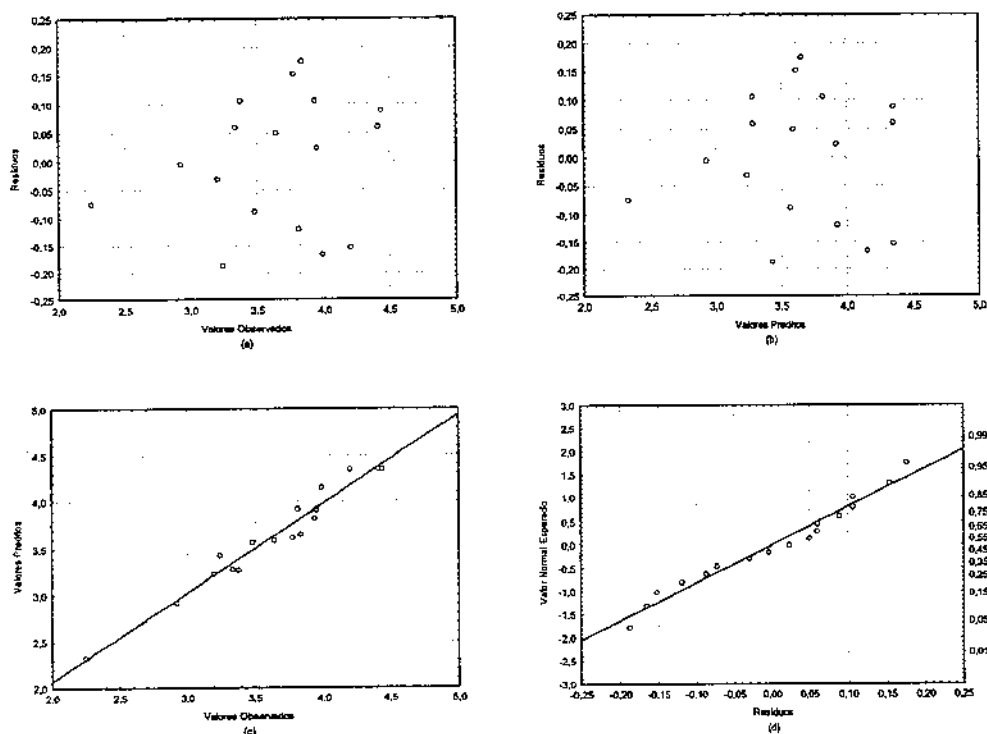


Figura 13. (a) Gráfico de resíduos contra valores observados; (b) Gráfico de resíduos contra valores previstos; (c) Gráfico de valores previstos contra valores observados; (d) Gráfico de probabilidade normal.

Superfícies de resposta (Figura 14) construídas a partir do modelo quadrático descrito pela equação 6, foram usadas para analisar as condições do meio de cultivo que levam a maior atividade de emulsificação para emulsão água-em-hexadecano.

A figura 14a mostra a superfície de resposta da atividade de emulsificação em função da concentração de uréia e da concentração de sulfato de amônio, mantendo o valor da concentração de fosfato de potássio no ponto central. Verifica-se que a atividade de emulsificação atinge uma região de valores máximos (acima de 4 UAE) quando as concentrações de uréia e sulfato de amônio movem-se na direção do ponto central (das condições do nível zero).

A figura 14b mostra a superfície de resposta da atividade de emulsificação em função da concentração de uréia e da concentração de fosfato de potássio, mantendo o valor da concentração de sulfato de amônio no ponto central. Verifica-se que a atividade de emulsificação atinge uma região de valores máximos (acima de 4 UAE) quando as concentrações de uréia e fosfato de potássio movem-se na direção do ponto central (das condições do nível zero).

A figura 14c mostra a superfície de resposta da atividade de emulsificação em função da concentração de sulfato de amônio e da concentração de fosfato de potássio, mantendo o valor da concentração de uréia no ponto central. Verifica-se que a atividade de emulsificação atinge uma região de valores máximos (acima de 4 UAE) quando as concentrações de sulfato de amônio e fosfato de potássio movem-se na direção do ponto central (das condições do nível zero).

Por meio da equação (2), os valores ótimos codificados para as concentrações de uréia, sulfato de amônio e fosfato de potássio foram obtidos como sendo respectivamente iguais a 0,052, 0,085 e -0,456. Os valores ótimos decodificados foram calculados como sendo iguais a 0,544% (p/v), 2,131% (p/v) e 2,628% (p/v), respectivamente. A máxima atividade de emulsificação prevista nesta condição ótima foi 4,415 UAE. Este meio otimizado será denominado a partir deste ponto de meio YPASUC-2 (Yeast, Potassium Phosphate, Amonium Sulfate, Urea, Corn Oil).

Os ensaios do planejamento composto central revelaram como melhor resultado uma atividade de emulsificação média de aproximadamente 4,346 UAE, sob as condições do ponto central: uréia 0,7% (p/v), sulfato de amônio 2,05% (p/v) e fosfato de potássio 3,12% (p/v). Resultados que confirmam a proximidade dos dados previstos pelo modelo aos dados experimentais. É interessante o fato do ponto crítico do modelo quadrático desenvolvido para atividade de emulsificação ser um ponto de máximo com atividade de emulsificação de 4,4415 UAE, em um meio com composição próxima a composição do ponto central do planejamento e que o ponto crítico para o modelo de biomassa seja um ponto de sela com concentração de 25,346 g/L em um meio com composição próxima a do ponto central do planejamento (Tabela 15).

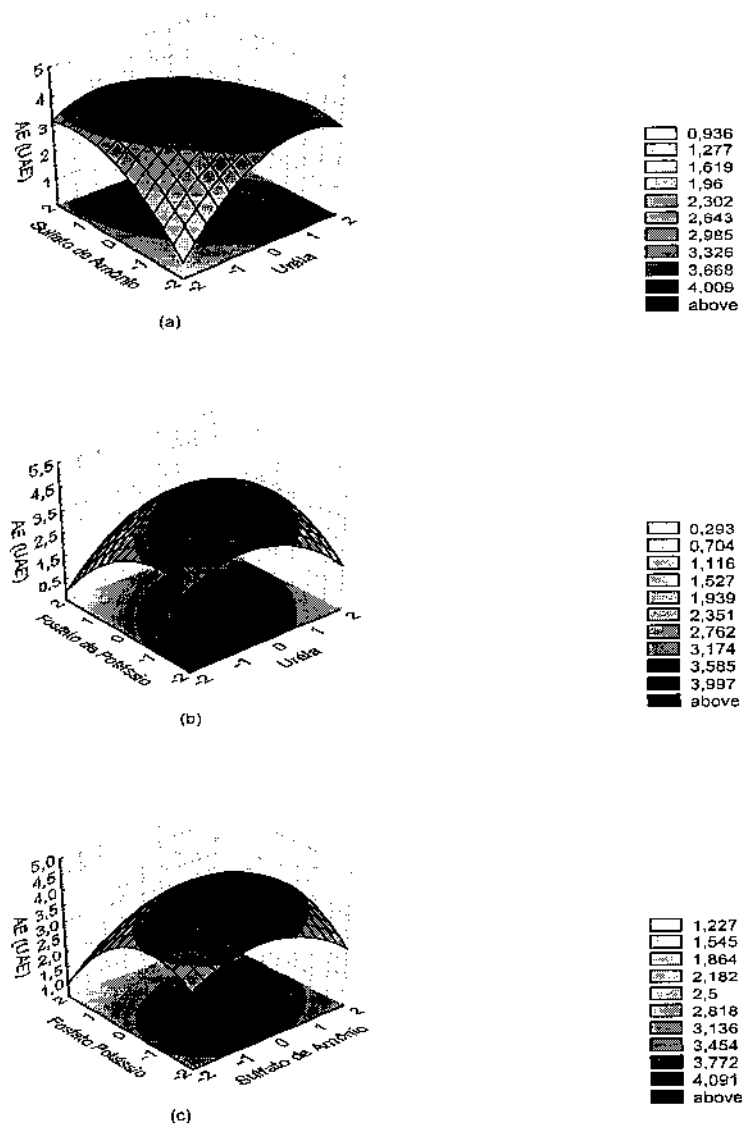


Figura 14. Superfície de Resposta para Atividade de Emulsificação mostrando as interações entre as concentrações de: (a) uréia e sulfato de amônio; (b) uréia e fosfato de potássio e (c) sulfato de amônio e fosfato de potássio.

Tabela 15. Composições do meio no ponto central, no ponto de máxima atividade de emulsificação e no ponto de sela da concentração de biomassa.

Componentes Otimizados	Composição do meio de produção		
	Ponto Máximo de Atividade de Emulsificação	Ponto central Planejamento Experimental	Ponto Sela da Concentração de Biomassa
Uréia	0,544%(p/v)	0,7%(p/v)	0,911%(p/v)
Sulfato de Amônio	2,131%(p/v)	2,05%(p/v)	2,163%(p/v)
Fosfato de Potássio	2,628%(p/v)	3,12%(p/v)	2,932%(p/v)

Validação do modelo

Para confirmar os resultados obtidos pelo modelo (Equação (6)), três experimentos foram conduzidos nas concentrações ótimas previstas para os componentes do meio. A atividade de emulsificação experimental máxima concordou perfeitamente com a atividade de emulsificação máxima prevista pelo modelo. (Tabela 16). Estes resultados mostram claramente que o modelo se ajusta muito bem aos dados experimentais e portanto descreve bem a região estudada. A diferença entre a previsão do modelo (4,415 UAE) e a média dos resultados obtidos experimentalmente (4,435 UEA) foi menor que 0,5 %. Isto valida o modelo de segunda ordem (Equação (6)) usado no presente trabalho, desde que essa diferença pode ser explicada pelos erros experimentais.

Tabela 16. Validação do modelo de superfície de resposta. Atividade de Emulsificação Média Experimental prevista versus Atividade de Emulsificação Máxima Prevista.

Repetição	Atividade de Emulsificação (UAE)
1	4,38467
2	4,498999
3	4,421625
Atividade de Emulsificação Média Experimental	4,435098
Atividade de Emulsificação Prevista p/ modelo	4,414905

A atividade de emulsificação média (4,435098 UAE) - obtida nos três experimentos realizados nas condições ótimas previstas pelo modelo - é 270,2 % superior a máxima atividade de emulsificação (1,198 UAE) obtida no processo padrão – processo de produção de bioemulsificante por *Candida lipolytica*, com inóculo de 1%, 120 h de cultivo, em meio YSWC-1, contendo uréia, sulfato de amônio, fosfato de potássio e 5% de óleo de milho (VANCE-HARROP, 2000). Resultado que confirma o potencial de uso de planejamento experimental associado à metodologia de superfície de resposta como ferramentas eficientes para otimizar meios de produção.

Três experimentos foram também realizados sob as concentrações otimizadas dos componentes do meio, para avaliar se o bioemulsificante produzido por *Candida lipolytica* apresentava propriedades surfactantes. O bioemulsificante produzido por *Candida lipolytica*, nas condições otimizadas, mostrou ser estável e reduziu em média a tensão superficial da água para 32,75 mN/m (Figuras 15, 16 e 17). Estes resultados indicam que os biopolímeros produzidos por *Candida lipolytica* são capazes tanto de estabilizar emulsões quanto de reduzir a tensão superficial. A redução da tensão superficial frequentemente tem sido usada como um critério primário para a seleção de microrganismos para produção de bioemulsificantes. Entretanto, agentes emulsificantes e dispersantes não precisam necessariamente reduzir a tensão superficial da água. O bioemulsificante extracelular *liposan* de *Candida lipolytica*, por exemplo, não é capaz de reduzir a tensão superficial da água (72,8 mN/m) embora emulsifique e estabilize efetivamente emulsão água-em-óleo (CIRIGLIANO e CARMAN, 1985). Por outro lado, apesar dos sorolipídeos de *Torulopsis bombicola* terem mostrado capacidade de reduzir tensão superficial, eles não demonstraram ser bons emulsificantes (COOPER e PADDOCK, 1984). Portanto, os resultados obtidos no presente trabalho corroboram com estudos anteriores sobre biossurfactantes que tanto baixam a tensão superficial quanto estabilizam emulsões (COOPER e GOLDENBERG, 1987; SINGER, 1985; COOPER e ZAJIC, 1980; RAPP *et al.*, 1979; ZAJIC, GUIGNARD e GERSON, 1977).

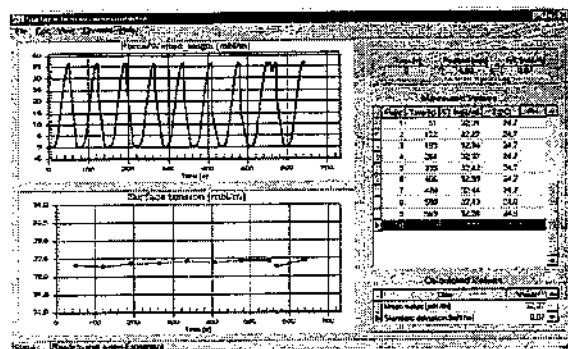


Figura 15. Tela do tensiômetro automático KSV Sigma 70, controlado por computador, registrando medida de tensão superficial do primeiro dos três experimentos realizados sob as concentrações otimizadas dos componentes do meio de produção de bioemulsificante por *Candida lipolytica*.

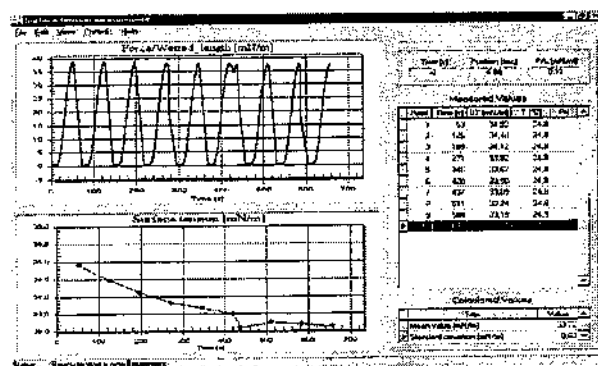


Figura 16. Tela do tensiômetro automático KSV Sigma 70, controlado por computador, registrando medida de tensão superficial do segundo dos três experimentos realizados sob as concentrações otimizadas dos componentes do meio de produção de bioemulsificante por *Candida lipolytica*.

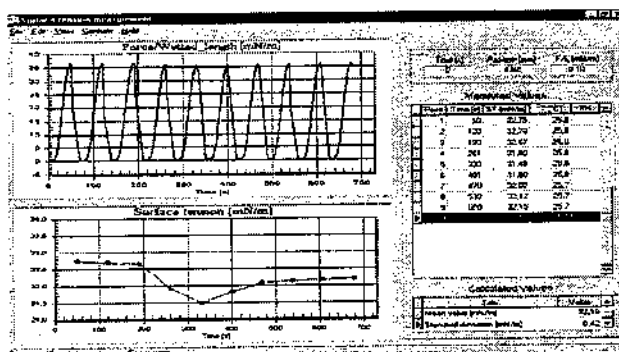


Figura 17. Tela do tensiômetro automático KSV Sigma 70, controlado por computador, registrando medida de tensão superficial do terceiro dos três experimentos realizados sob as concentrações otimizadas dos componentes do meio de produção de bioemulsificante por *Candida lipolytica*.

Vários emulsificantes comerciais foram examinados por Cirigliano e Carman (1985) e Kim et al. (2000), usando a metodologia para determinação de atividade de emulsificação descrita por Cirigliano e Carman (1984) e empregada também no presente trabalho para determinação das atividades de emulsificação dos bioemulsificantes produzidos por *Candida lipolytica*. Os resultados dos exames realizados com os emulsificantes comerciais, com o emulsificante *liposan* produzido por *Candida lipolytica* (Cirigliano e Carman, 1984) e com os biosurfactantes tipo I e II produzidos por *Norcadia sp.* L-417 (KIM et al., 2000) são apresentados na tabela 17.

Tabela 17. Comparação de propriedade de emulsificação de bioemulsificantes e emulsificantes comerciais.

Cirigliano e Carman		Kim, Lim Lee et al.	
Emulsificante	AE (UAE)	Emulsificante	AE(UAE)
Caragenina	1,75	Biosurfactant I	2,51
Tween 80	1,60	Triton N-42	1,63
Tween 20	1,35	Triton N-45	2,52
Pectina	1,21	Triton X-100	2,58
Triton X-100	1,20	Span 40	0,56
Liposan	0,75	Span 80	1,96
Cascina	0,65	Tween 40	1,84
Alginato	0,60	Tween 80	1,10
Goma arábica	0,59	SDS	0,81
Goma Ghatti	0,47	Biosurfactante II	nd
Hiroxipropilmetil celulose	0,30		
Dextran	0,22		
Gelatina	0,08		

AE – Atividade de Emulsificação UAE – Unidade de Atividade de Emulsificação

nd – atividade de emulsificação não detectada

Dos trezes agentes examinados por Cirigliano e Carman (1985), *liposan* ficou classificado em sexto lugar em sua efetividade como emulsificante. Os resultados mostram que a atividade de emulsificação do biosurfactante tipo I produzido por *Norcadia sp.* L-417 foi superior a dos oito surfactantes comerciais sintéticos examinados por Kim et al. (2000), exceto triton X-100. O bioemulsificante produzido na segunda etapa do processo de otimização do presente trabalho, apresentou atividade de emulsificação 72% maior que a apresentada pelo surfactante comercial triton X-100 examinado por Kim et al. (2000). No caso do biosurfactante tipo II produzido por *Norcadia sp.* L-417, atividade de emulsificação significativa não foi detectada. A comparação das atividades de emulsificação dos emulsificantes comerciais e de origem microbiana apresentados na tabela 17 com as atividades de emulsificação dos biosurfactantes produzidos na primeira e na segunda etapa do processo de otimização e especialmente nas condições otimizadas do

meio de produção (Tabela 16) revela a excelente qualidade dos bioemulsificantes produzidos no presente trabalho. Os resultados obtidos sugerem que os bioemulsificantes com propriedades surfactantes produzidos no presente estudo podem ser convenientes para uso na indústria de petróleo e aplicações ambientais, tais como aumento da recuperação de óleo, limpeza de tanques e decomposição de óleo espalhado em solo ou ambiente marinho.

3.4. Conclusão

Otimização estatística de componentes de meios de produção de bioemulsificantes por *Candida lipolytica* foi realizada usando planejamento fatorial e metodologia de superfície de resposta. Um modelo matemático para representar a atividade de emulsificação como função das concentrações de uréia, sulfato de potássio e fosfato de amônio foi proposto. O modelo não apresentou falta de ajuste e o coeficiente de determinação indicou que o modelo pode explicar 95,72% da variabilidade na atividade de emulsificação. De acordo com o modelo, as concentrações de uréia, sulfato de amônio e fosfato de potássio necessárias para atingir a atividade de emulsificação máxima (4,415 UAE) foram respectivamente 0,544 % (p/v), 2,131 % (p/v) e 2,628 % (p/v). A validação do modelo foi realizada sob as condições do meio otimizado. Os resultados mostraram que a atividade de emulsificação prevista pelo modelo foi similar àquela obtida experimentalmente (diferença inferior a 0,5%). A otimização aumentou a atividade de emulsificação (1,198 UAE) do processo padrão (VANCE-HARROP, 2000) de produção de bioemulsificante em 262,8% nas condições do ponto central do planejamento experimental e em 272% nas condições otimizadas do meio de produção. O bioemulsificante produzido por *Candida lipolytica*, mostrou ser estável e reduziu a tensão superficial da água para 32,75 mN/m. A metodologia de superfície de resposta demonstrou ser uma ferramenta eficiente e poderosa para otimização de componentes de meios de produção, sendo então indicada para ser usada na ampliação de escala do processo para estudar as condições operacionais do biorreator BioFlo200® (NewBrunswick). Este é o primeiro estudo sobre otimização de componentes de meio de produção de bioemulsificante por *Candida lipolytica* usando óleo de milho como fonte de carbono e água do mar suplementada como meio basal.

4. AMPLIAÇÃO DO PROCESSO DE PRODUÇÃO DE BIOEMULSIFICANTE POR *Candida lipolytica* DA ESCALA DE FRASCOS PARA ESCALA DE BIORREATOR DE BANCADA

4.1. Introdução

Após a otimização dos componentes do meio de produção de biossurfactante em escala de frascos, procedeu-se à ampliação do processo para escala de biorreator de bancada. Inicialmente, quatro ensaios preliminares foram realizados no biorreator BioFlo 2000 visando à familiarização com o equipamento e a avaliação das melhores condições para realização do planejamento fatorial. A seguir um planejamento fatorial completo 2^2 , empregando o meio de produção otimizado, foi realizado para identificar a influência das variáveis temperatura e velocidade de agitação sobre a concentração da biomassa e sobre as atividades de emulsificação do biossurfactante frente ao hexadecano, ao óleo de milho e ao óleo de canola. Curvas cinéticas de pH, de oxigênio dissolvido, de crescimento, de densidade ótica, de salinidade, de densidade, de biomassa e de atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola foram construídas para cada uma das condições do planejamento fatorial. Posteriormente, essas curvas foram suavizadas e ampliadas através de interpolação por meio de *spline* cúbico para serem usadas no treinamento e teste dos protótipos dos *softsensores* neurais.

4.2. Materiais e Métodos

4.2.1. Descrição do Equipamento

Os ensaios preliminares e os ensaios principais do planejamento fatorial 2^2 para produção de biossurfactante foram realizados em biorreator BioFlo 2000® (New Brunswick Scientific Co. Inc., New Jersey) com capacidade de 5 litros. O biorreator BioFlo 2000® permite o controle e monitoramento de temperatura, agitação, oxigênio dissolvido,

pH, nível de espuma e adição de nutrientes. Através do console principal pode-se selecionar e exibir *set-points* e valores correntes de cada parâmetro operacional (Figura 18).

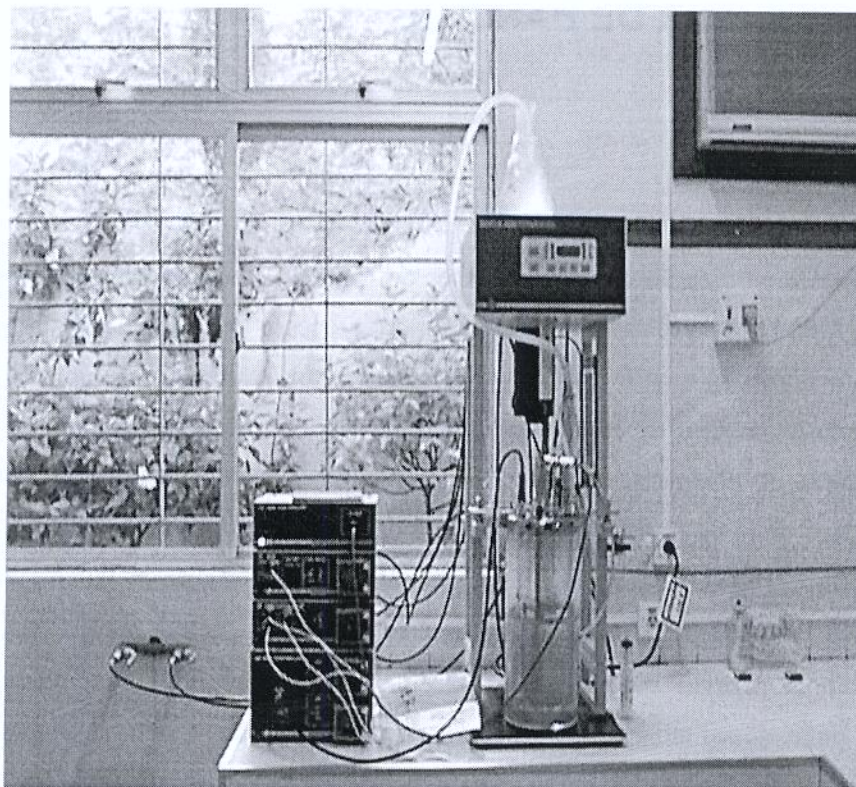


Figura 18. Biorreator BioFlo 2000 equipado com 4 módulos de controle.

4.2.2. Microrganismo e Meio de Preservação

O microrganismo usado nos experimentos realizados em escala de biorreator de bancada continuou sendo a levedura *Candida lipolytica* UCP 988 cuja procedência, meio e forma de preservação encontram-se descritos no item 3.2.1.

4.2.3. Preparação do Inóculo

4.2.3.1. Ensaios Preliminares

Para os ensaios preliminares no biorreator BioFlo 2000® o inóculo consistiu de 200mL de uma suspensão de células pré-cultivadas em incubador rotativo, em meio com composição idêntica a do ensaio que apresentou maior atividade de emulsificação no planejamento fatorial completo 2^4 (ensaio 15), exceto que no lugar de óleo de milho, empregou-se glicose a 1%. Portanto, para crescimento do inóculo foi utilizado este meio, denominado YPASUG-1 (Yeast, Potassium Phosphate, Amonium Sulfate, Urea, Glucose), composto por: glicose 1%(p/v); uréia 0,4%(p/v); sulfato de amônio 1,1%(p/v), fosfato de potássio 2,04 % (p/v), 50 mL de água do mar, 50 mL de água destilada e pH ajustado para 5,3 com solução de hidróxido de sódio a 40%. O inóculo para produção de bioemulsificante foi preparado pela transferência de alçadas de cultura fresca de *Candida lipolytica* para erlenmeyer de 500 mL contendo 200 ml de meio estéril YPASUG-1 (121°C/15min). A suspensão de células foi incubada durante 48 horas a 28° C, em agitador rotatório a 150 rpm, até a concentração final de 10^7 células por mililitro (cel/ml). Após o término do tempo de cultivo, o pré-inóculo foi transferido para o biorreator de bancada na proporção de 5% do volume de trabalho presente no biorreator.

4.2.3.2. Ensaios do Planejamento Fatorial 2^2

Para os ensaios do planejamento fatorial 2^2 no biorreator BioFlo 2000®, o inóculo consistiu de 400 mL de uma suspensão de células pré-cultivadas em incubador rotativo, em meio com composição idêntica a do meio otimizado em frascos, descrito no capítulo 3, exceto que no lugar de óleo de milho, empregou-se glicose a 1%. Portanto, para crescimento do inóculo foi utilizado este meio semelhante ao meio ótimo, denominado YPASUG-2 (Yeast, Potassium Phosphate, Amonium Sulfate, Urea, Glucose), composto por: glicose 1%(p/v); uréia 0,544%(p/v); sulfato de amônio 2,130 %(p/v), fosfato de potássio 2,628 % (p/v), 50 mL de água do mar, 50 mL de água destilada e pH ajustado para 5,3 com solução de hidróxido de sódio a 40%. O inóculo para produção de bioemulsificante foi preparado em duas etapas (Figura 19). Na primeira etapa seis alçadas de cultura de *Candida lipolytica* - crescida em placa de Petri, incubada durante 24 h a 28°

C - foram transferidas para erlenmeyer de 500 mL contendo 220 mL de meio estéril YPASUG-2 (121°C/15min). Em seguida, o frasco de erlenmeyer contendo a suspensão de células foi incubado durante 48 horas a 28° C, em agitador rotatório a 150 rpm, até a concentração final de 10^7 células por mililitro (cel/ml). Na segunda etapa, 4 frascos de erlenmeyer de 500 mL contendo 50 ml de meio estéril YPASUG-2 receberam cada um, 50 mL da suspensão de células do primeiro pré-inóculo.

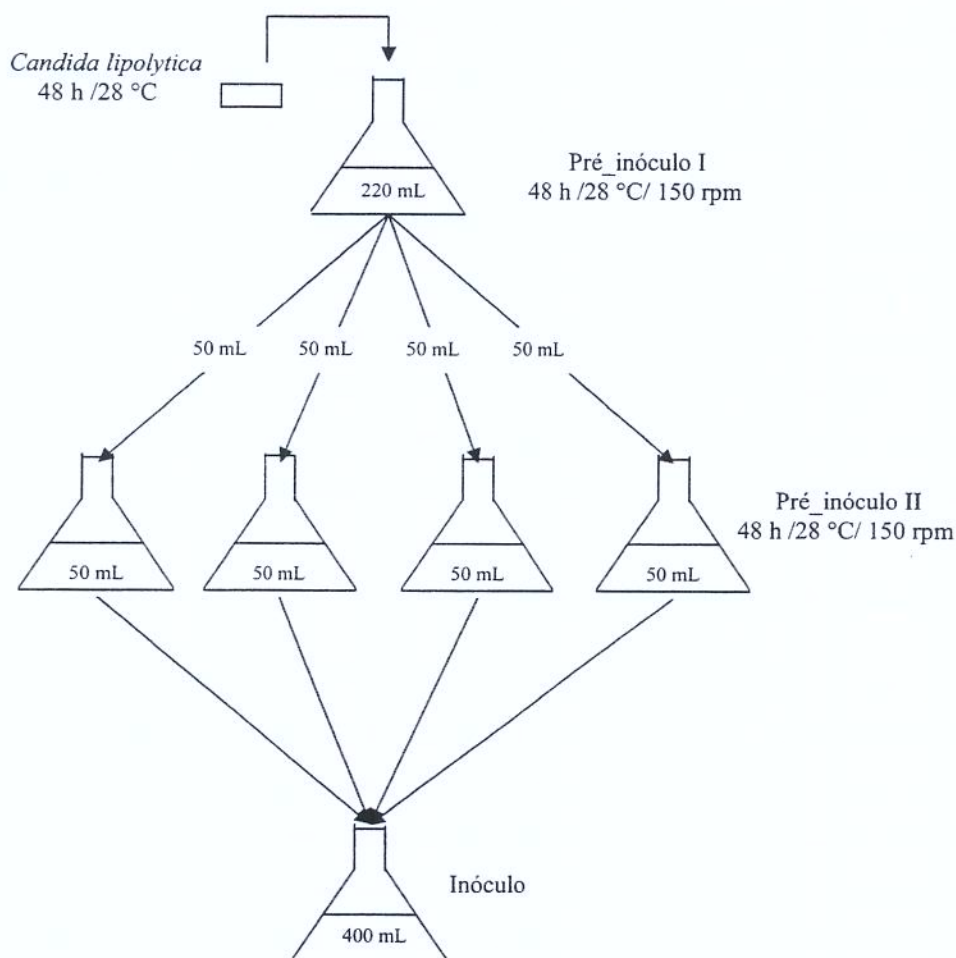


Figura 19. Esquema de repicagem e preparação do inóculo para os ensaios do planejamento fatorial realizados em biorreator de bancada

Os frascos, contendo cada um 100 mL da suspensão de células, foram incubados durante 48 horas a 28° C, em agitador rotatório a 150 rpm, até a concentração final de 10^8 células por mililitro (cel/ml). Após o término do tempo de cultivo, o segundo pré-inóculo foi transferido para o biorreator de bancada na proporção de 10 % do volume de trabalho presente no biorreator.

4.2.4. Preparação dos Meios de Crescimento e Produção

4.2.4.1. Ensaios Preliminares

Ensaios preliminares, visando à familiarização e a avaliação das condições operacionais do biorreator, foram realizados após a conclusão da primeira (planejamento fatorial completo 2^4) e paralelamente a segunda etapa (planejamento composto central com três fatores) do processo de otimização dos componentes do meio de produção de biossurfactantes em frascos. Nos ensaios preliminares utilizou-se o meio de crescimento e produção denominado YPASUC-1 (*Yeast, Potassium Phosphate, Amonium Sulfate, Urea, Corn Oil*), com composição idêntica a do ensaio que apresentou maior atividade de emulsificação no planejamento fatorial completo 2^4 (ensaio 15): óleo de milho 5% (v/v); uréia 0,4%(p/v); sulfato de amônio 1,1%(p/v), fosfato de potássio 2,04 % (p/v), água do mar e água destilada, na proporção 1:1 v/v, correspondendo a 100ml. Foram preparados 3,6 litros do meio YPASUC-1 no biorreator BioFlo 2000. O pH do meio foi ajustado para 5,3 com solução de hidróxido de sódio 40%. O biorreator contendo o meio mineral foi preparado para esterilização e autoclavado a 121°C durante 30 minutos. Um frasco de erlenmeyer de 500mL contendo 200mL de óleo de milho foi autoclavado em vapor fluente. A suplementação do meio YPASUC-1 com o óleo de milho foi realizada posteriormente em câmara de fluxo laminar.

4.2.4.2. Ensaios do Planejamento Fatorial 2^2

Nos ensaios do planejamento fatorial 2^2 utilizou-se o meio de crescimento e produção denominado YPASUC-2 (*Yeast, Potassium Phosphate, Amonium Sulfate, Urea, Corn Oil*), com composição idêntica a do meio otimizado descrito no item 3: óleo de milho 5% (v/v); uréia 0,544% (p/v); sulfato de amônio 2,130 % (p/v), fosfato de potássio 2,628 %

(p/v), água do mar e água destilada, na proporção 1:1 v/v, correspondendo a 100ml. Foram preparados 3,42 litros do meio YPASUC-2 no biorreator BioFlo 2000. O pH do meio foi ajustado para 5,3 com solução de hidróxido de sódio 40%. O biorreator contendo o meio mineral foi preparado para esterilização em autoclave a 121°C durante 30 minutos. Um frasco de erlenmeyer de 500mL contendo 180 mL de óleo de milho foi autoclavado em vapor fluente. A suplementação do meio YPASUC-2 com o óleo de milho foi realizada posteriormente em câmara de fluxo laminar.

4.2.5. Produção em Batelada de Bioemulsificante

4.2.5.1. Ensaios Preliminares

A produção em batelada de bioemulsificante teve início com a adição do óleo de milho e do inóculo ao meio basal - através de portas específicas situadas no topo do reator, em câmara de fluxo laminar. Em seguida o biorreator foi transportado para a bancada onde foi instalado e posto em operação. As condições operacionais mantidas no biorreator foram: temperatura de 28°C, aeração de 1 vvm e agitação orbital de 150 rpm. O pH e o oxigênio dissolvido não foram controlados. Para acompanhamento das curvas cinéticas - de pH, de oxigênio dissolvido, de concentração celular, de densidade ótica, de biomassa, de salinidade e de atividade de emulsificação - alíquotas de 20 ml foram coletadas com 0 h, 4 h, 18 h, 24 h, 48 h, 72 h, 96h, 120 h, 144 h e 168 h.

4.2.5.2. Ensaios do Planejamento Fatorial 2²

A produção em batelada de bioemulsificante teve início com a adição do óleo de milho e do inóculo ao meio basal - através de portas específicas situadas no topo do reator, em câmara de fluxo laminar. Em seguida o biorreator foi transportado para a bancada onde foi instalado e posto em operação. As condições operacionais de temperatura e agitação seguiram as especificações do planejamento fatorial 2² (Tabelas 18 e 19). A aeração foi mantida em 1 vvm. O pH e o oxigênio dissolvido não foram controlados. Para acompanhamento das curvas cinéticas - de pH, de oxigênio dissolvido, de concentração celular, de densidade ótica, de biomassa, de salinidade e de atividade de emulsificação -

alíquotas de 20 ml foram coletadas com 0 h, 4 h, 18 h, 24 h, 48 h, 72 h, 96h, 120 h, 144 h e 168 h.

4.2.5.3. Planejamento Fatorial 2²

Com a finalidade de avaliar os efeitos principais e interações das variáveis independentes temperatura e velocidade de agitação sobre as variáveis respostas concentração de biomassa e atividades de emulsificação do biossurfactante produzido por *Candida lipolytica* sobre o hexadecano, o óleo de milho e o óleo de canola com 120 horas, um planejamento fatorial completo 2², composto de 4 ensaios foi realizado. Para determinação do erro experimental, três ensaios no ponto central foram também conduzidos. Como nos planejamentos realizados em frascos as variáveis respostas biomassa e atividade de emulsificação para emulsão água-em-hexadecano, foram analisadas com 120 horas de cultivo, optou-se também por adotar este mesmo tempo, para analisar as variáveis respostas biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola, nos ensaios realizados em biorreator de bancada. A tabela 18 apresenta os valores reais (decodificados) das variáveis independentes, temperatura e agitação, com seus respectivos níveis codificados. Os níveis para cada variável independente (Tabela 18) foram escolhidos com base nos ensaios experimentais preliminares realizados em frascos e no biorreator. Questões operacionais, de cronograma e de custos relacionados ao resfriamento, impossibilitaram no presente trabalho, a realização de um planejamento fatorial composto central com dois fatores cobrindo intervalos maiores de temperatura e velocidade de agitação, entretanto no futuro o presente planejamento poderá ser ampliado para o referido planejamento composto central. Análise de variância, determinações dos coeficientes de regressão e gráficos foram realizados usando o programa Statistica® versão 6.0 (Statsoft.Inc, Tulsa/OK,USA).

Tabela 18 – Níveis das variáveis do planejamento fatorial completo 2².

Variáveis Independentes	Níveis		
	-1	0	1
Temperatura	28	29,5	31
Agitação	150	225	300

4.2.6. Métodos Analíticos

4.2.6.1. Determinação da concentração celular da levedura

A concentração celular foi verificada através de contagem em Câmara de Neubauer - utilizando microscopia de luz com aumento de 400x - e expressa em numero de células por mililitro (cels/ml).

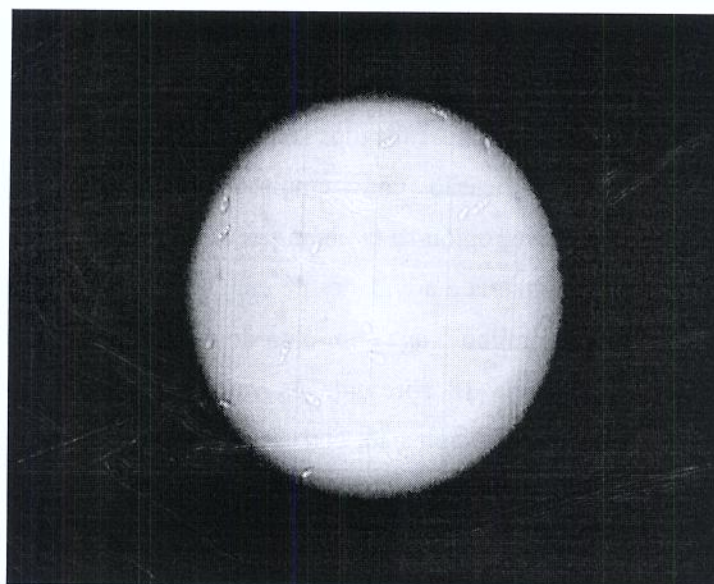


Figura 20. Leveduras *Candida lipolytica* vistas em Câmara de Neubauer

4.2.6.2. Determinação de densidade ótica

A densidade ótica das alíquotas coletadas durante o crescimento de *Candida lipolytica* foi determinada por espectrofotometria. As leituras foram realizadas em comprimento de onda de 540nm.

4.2.6.3. Determinação de salinidade

As salinidades da água do mar e dos cultivos livres de células foram determinadas conforme procedimento descrito no item 3.2.4..

4.2.6.4. Determinação da concentração da biomassa

A concentração da biomassa (massa seca) foi determinada segundo procedimento descrito na página 3.2.5..

4.2.6.5. Determinação de atividade de emulsificação

A atividade de emulsificação para emulsões água-em-hexadecano foi determinada empregando o método de Cirigliano e Carman (1984) descrito no item 3.2.6. O mesmo método foi empregado para determinação da atividade de emulsificação de emulsões água-em-óleo de milho e de emulsões água-em-óleo de canola.

4.2.6.4. Determinação da velocidade máxima de crescimento ($\mu_{\max}$) e do tempo de geração (t_g)

A velocidade específica de crescimento ($\mu_{\max}$) e o tempo de geração (t_g) foram calculados empregando as fórmulas abaixo descritas por Pirt (1975) e usadas por Marçal (1991), SARUBBO (1999) e VANCE-HARROP (2000):

$$\mu_{\max} = [\ln(X) - \ln(X_0)] / (t - t_0) \quad (7)$$

e

$$t_g = \ln(2) / \mu_{\max} \quad (8)$$

onde:

X - biomassa final;

X_0 - biomassa inicial;

t - tempo final e

t_0 - tempo inicial.

4.3. Resultados e Discussão

4.3.1. Ensaios Preliminares

Visando familiarização com o biorreator e verificação das condições de temperatura, agitação e aeração a serem usadas nos ensaios do planejamento experimental, quatro ensaios preliminares foram realizados usando o meio de crescimento e produção YPASUC-1.

O primeiro ensaio foi conduzido à temperatura de 28°C com velocidade de agitação de 150 rpm e aeração de 1vvm. O segundo ensaio foi conduzido à temperatura de 31°C com velocidade de agitação de 150 rpm e aeração de 1vvm. O terceiro ensaio foi conduzido à temperatura de 28 °C com velocidade de agitação de 150 rpm e aeração de 0,8 vvm. O quarto ensaio foi conduzido à temperatura de 28 °C com velocidade de agitação de 150 rpm e aeração foi de 2vvm. Em todos os ensaios dados de pH e oxigênio dissolvido foram registrados e determinações de atividade de emulsificação para emulsões água-em-hexadecano foram realizadas com 0h, 4h, 18h, 24h, 48h, 72h, 96h, 120h, 144h e 168 h.

Nos ensaios realizados com aeração de 0,8 vvm e 2vvm, durante as 168 horas de cultivo, nenhuma atividade de emulsificação para emulsões água-em-hexadecano foi detectada.. Nos ensaios realizados com aeração de 1vvm, foram alcançadas atividades de emulsificação similares às obtidas em frascos agitados, mas em um tempo bem maior. Por conta disto, nos ensaios do planejamento experimental, providências, como o aumento da concentração e da fração inicial do inóculo, foram tomadas no sentido de minimizar a fase *lag* e conseqüentemente provocar uma diminuição no tempo total de fermentação.

4.3.2. Ensaios do Planejamento Experimental

Os resultados do planejamento fatorial 2^2 demonstraram a capacidade de *Candida lipolytica* crescer sobre óleo de milho, em biorreator de bancada, operado em batelada e produzir grande quantidade de biomassa e atividade de emulsificação extracelular induzida capaz de estabilizar emulsões água em óleo em diversos meios. A matriz codificada do planejamento e os resultados das variáveis respostas, após 120 horas de cultivo, são apresentados na tabela 19. As variáveis independentes, temperatura e agitação, estão codificadas de acordo com a equação:

$$X_i = \frac{x_i - x_0}{\Delta x} \quad (1)$$

onde X_i , x_i e x_0 são respectivamente, o valor codificado, o valor não codificado e o valor no ponto central da i -ésima variável independente. Δx é o valor do incremento de mudança.

Tabela 19. Resultados do planejamento fatorial completo 2^2 , após 120 h de cultivo, para concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola.

Ensaio	Temperatura	Agitação	Biomassa	Atividade de Emulsificação Hexadecano	Atividade de Emulsificação Milho	Atividade de Emulsificação Canola
1	-1	-1	16,103	5,46	6,1	4,624
2	1	-1	28,88	0,00	4,63	5,04
3	-1	1	27,78	5,44	5,92	5,8
4	1	1	28,90	0,0	3,22	5,4
5	0	0	24,16	3,8	5,26	5,32
6	0	0	23,82	3,5	5,84	5,15
7	0	0	24,94	3,7	5,34	5,21

4.3.2.1. Variável Resposta: Biomassa

Os resultados para as concentrações de biomassa de *Candida lipolytica*, obtidos após a execução do planejamento composto central 2^2 estão apresentados na tabela 19. Nos ensaios realizados, os valores da concentração de biomassa variaram entre 16,103 g/L e 28,88 g/L. A biomassa média produzida com 120 horas, nos 4 ensaios do planejamento 2^2 e nas 3 repetições no ponto central, foi igual a 24,94 g/L, ou seja, apenas 7,62 % menor que a biomassa média dos ensaios do planejamento composto central 2^3 (26,84 g/L). Entretanto, a biomassa média produzida nos ensaios realizados na temperatura de 31°C (nível +1) foi de 28,89 g/L, ou seja, 7,64% maior que a biomassa média dos ensaios do planejamento composto central 2^3 (26,84 g/L). A produção média de biomassa no ponto central foi igual a 24,307 g/L. A menor produção de biomassa, 16,103 g/L, foi obtida com temperatura de 28°C e velocidade de agitação de 150 rpm (nível -1). A maior produção de biomassa, 28,9 g/L, foi obtida com temperatura de 31°C e velocidade de agitação de 300 rpm (nível +1).

Os efeitos estimados para a variável resposta concentração de biomassa são apresentados na tabela 20. Os valores em negrito são estatisticamente significativos a um intervalo de confiança de 95%. Nas figuras 21a e 21b encontram-se ilustrados respectivamente o histograma dos efeitos das variáveis independentes sobre a variável resposta biomassa e o diagrama de Pareto, com nível de significância de 95% para estimativa dos efeitos lineares, em valor absoluto. Os resultados obtidos mostram que as variáveis independentes temperatura e agitação exercem uma grande influência sobre a produção de biomassa, sendo a influência da temperatura maior que a da agitação. A interação entre as variáveis operacionais temperatura e velocidade de agitação exerce um efeito antagônico sobre a produção de biomassa.

Tabela 20. Efeitos estimados para a variável resposta concentração de biomassa

	Efeito	Erro Padrão	t(2)	p
Média	24,94043	0,21704	114,91362	0,00008
(1) Temperatura	6,94850	0,57422	12,10068	0,00676
(2) Agitação	5,84850	0,57422	10,18505	0,00950
1 e 2	-5,82850	0,57422	-10,15022	0,00957

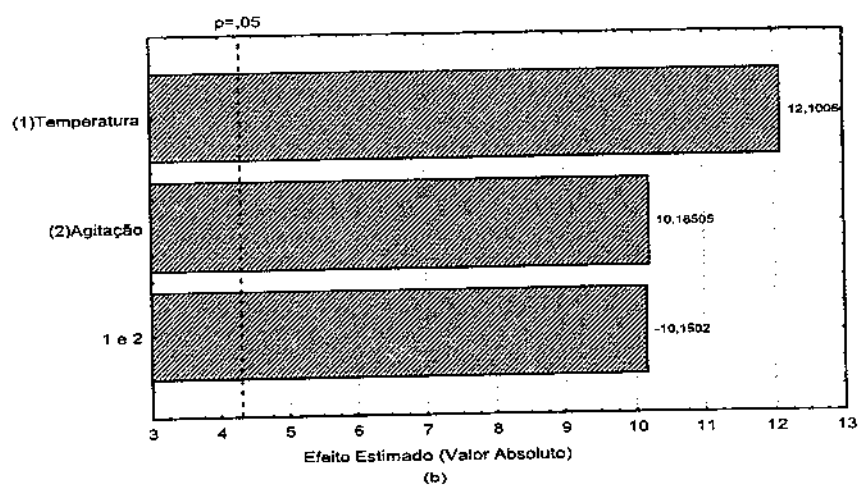
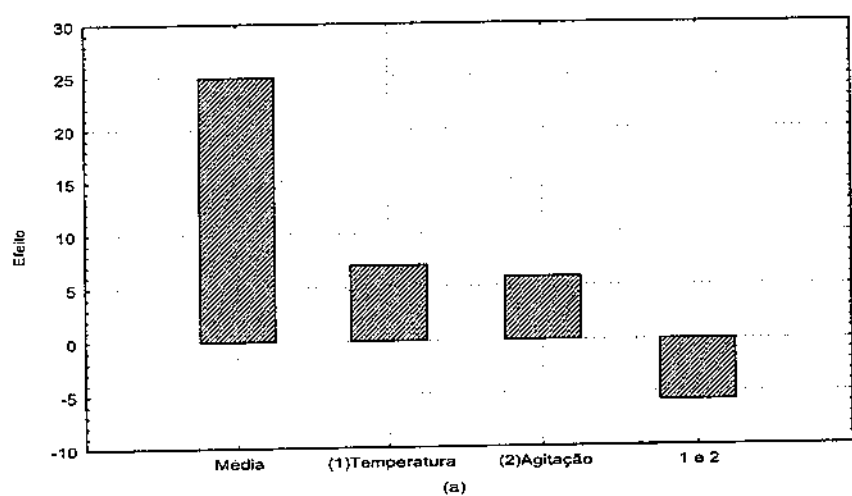


Figura 21. (a) Histograma dos efeitos das variáveis independentes sobre a variável resposta concentração de biomassa. (b) Diagrama de Pareto de efeitos padronizados para o planejamento fatorial completo 2^4 . O ponto no qual os efeitos estimados foram estatisticamente significativos (em $p=0,05$) é indicado por uma linha tracejada vertical.

O ajuste de um modelo contendo interações até dois fatores aos dados da concentração de biomassa (Y) apresentados na tabela 19 resultou na equação

$$Y = 24,94043 + 3,47425 x_1 + 2,92425 x_2 - 2,91425 x_1 x_2 \quad (9)$$

onde os valores das variáveis independentes, temperatura (x_1) e a velocidade de agitação (x_2), estão na forma codificada (equação 1). Os coeficientes em negrito são estatisticamente significativos no nível de 95% de confiança.

A tabela 21 apresenta os resultados da análise de variância do modelo obtido para a concentração da biomassa. O modelo apresentou um alto coeficiente de determinação ($R^2=0,9572$) deixando de explicar apenas 2,32% da variabilidade na resposta. O coeficiente de correlação de 0,9883 indica uma excelente concordância entre os valores experimentais e os valores previstos pelo modelo. O teste F para o modelo mostra que o F calculado ($F_c = 42,07$) é 4,5 vezes maior que o F tabelado ($F_t (0,05; 3; 3) = 9,28$), demonstrando que o modelo é altamente significativo, podendo ser usado para fazer previsões da biomassa. O teste F para falta de ajuste mostrou o que o F calculado ($F_c = 6,40$) é menor que o F tabelado ($F_t (0,05; 1; 2) = 18,51$), indicando que o modelo está muito bem ajustado aos dados experimentais.

Tabela 21. Análise de Variância (ANOVA) para o modelo de regressão representando a concentração de biomassa.

Fonte	SQ	GL	MQ	Fc	Ft	Fc/Ft
Regressão	116,458	3	38,819	42,07	9,28	4,53
Resíduo	2,768	3	0,923			
Falta Ajuste	2,109	1	2,109	6,40	18,51	0,35
Erro puro	0,659	2	0,330			
Total	119,226	6	19,871		$R^2=0,9768$	$R=0,9883$
SQ - Soma Quadrática GL - Graus de Liberdade MQ - Média Quadrática F _c - F calculado F _t - F tabelado						

A adequação do modelo quadrático ajustado aos dados também foi verificada por meio de análise de resíduos. O gráfico dos resíduos contra os valores ajustados (Figura 22a) e o gráfico dos resíduos contra os valores previstos (Figura 22b) não evidenciam a falta de ajuste do modelo nem de violação das suposições de independência e igualdade de variância, já que os resíduos deixados pelo modelo (Figuras 22a e 22b), encontram-se distribuídos aleatoriamente em torno do valor zero. O gráfico dos valores previstos contra os valores observados (Figura 22c) confirma visualmente a excelente qualidade do ajuste. Indicações de violação da hipótese de normalidade não foram observadas através do gráfico de probabilidade normal (Figura 22d) já que os resíduos estão localizados aproximadamente ao longo de uma linha reta.

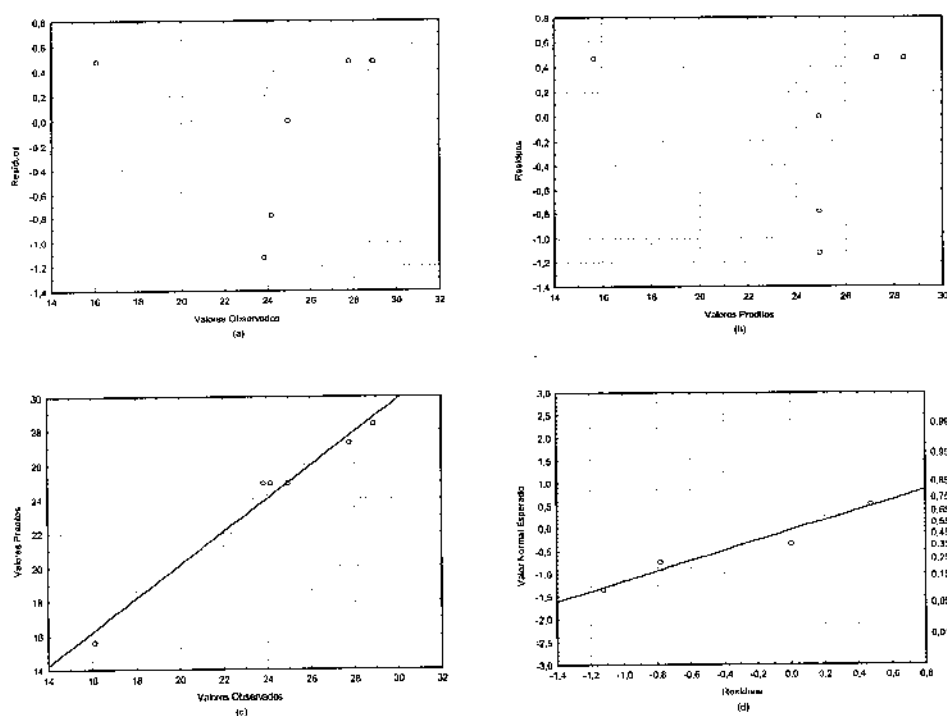


Figura 22. (a) Gráfico de resíduos contra valores observados ; (b) Gráfico de resíduos contra valores previstos; (c) Gráfico de valores previstos contra valores observados; (d) Gráfico de probabilidade normal.

Superfície de resposta (Figura 23) construída a partir do modelo descrito pela equação 7, foi usada para analisar as condições operacionais de temperatura e agitação que levam a maior produção de biomassa.

A figura 23 mostra a superfície de resposta da concentração da biomassa em função da temperatura e da velocidade de agitação. Verifica-se que a concentração da biomassa aumenta com o aumento simultâneo da temperatura e da agitação (nível +1), atingindo valores acima de 28 g/L.

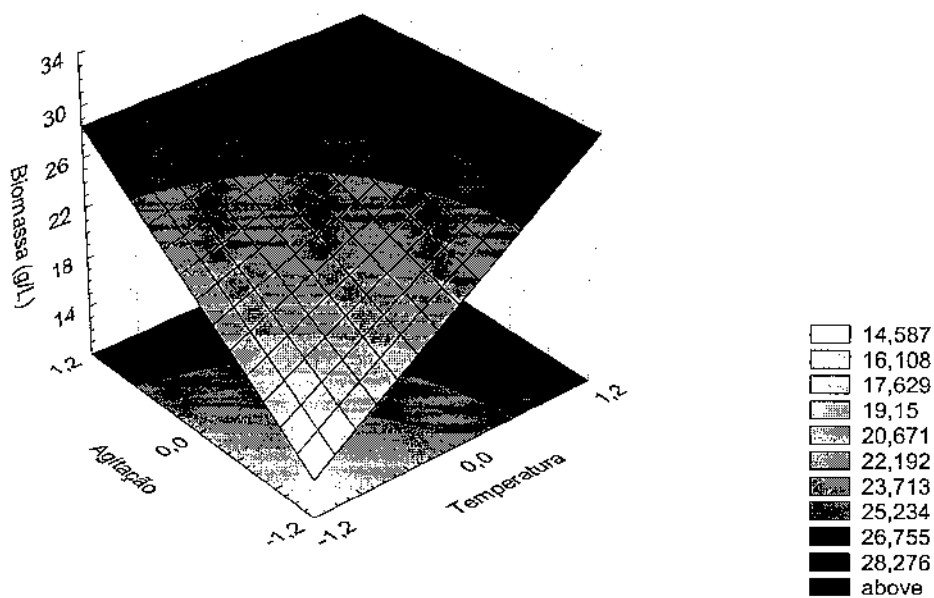


Figura 23. Superfície de resposta para concentração de biomassa mostrando a interação entre temperatura e velocidade de agitação.

4.3.2.2. Variável Resposta: Atividade de Emulsificação

Hexadecano, óleo de milho e óleo de canola foram testados para sua capacidade de serem emulsionados por bioemulsificantes produzidos por *Candida lipolytica*. Filtrados de culturas de *Candida lipolytica* crescidas sobre o meio YPASUC-2, em diferentes condições de temperatura e agitação apresentaram agentes com capacidade de emulsionar ou estabilizar emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola. Os resultados das atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola, obtidos após a execução do planejamento fatorial 2^2 , estão apresentados na tabela 19. Os ensaios deste planejamento mostram que altos valores de atividade de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola foram obtidos em todas as condições estudadas. A seguir serão analisados separadamente os resultados da atividade de emulsificação do bioemulsificante produzido por *Candida lipolytica* para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola.

4.3.2.2.1. Atividade de emulsificação para emulsões água-em-hexadecano

A atividade de emulsificação média para emulsões água-em-hexadecano - obtidas a partir dos filtrados dos cultivos com 120 horas dos 4 ensaios do planejamento 2^2 e das 3 repetições no ponto central - foi igual a 3,129 UAE. Em relação às emulsões água-em-hexadecano não foi possível determinar atividade de emulsificação na temperatura de 31°C, conforme se pode verificar na tabela 19. A atividade de emulsificação média para emulsões água-em-hexadecano obtidas a partir dos ensaios realizados a 28° C foi igual a 5,45 UAE, ou seja, 21,5% maior que a atividade de emulsificação média experimental para emulsões água-em-hexadecano (4,435 UAE) obtida a 28°C, com o meio otimizado em frascos. No ponto central do planejamento - ou seja, com temperatura de 29,5°C e velocidade de agitação de 225 rpm - foram obtidas as menores atividades de emulsificação para emulsões água-em-hexadecano, com média igual a 3,667 UAE, ou seja, 20,95% menor que a atividade de emulsificação média experimental (4,435 UAE) obtida a 28°C

com o meio otimizado em frascos. Estes resultados, por si só, já demonstram que o segundo objetivo do nosso trabalho - ampliação do processo da escala de frascos para a escala de fermentador de bancada – foi alcançado.

Os efeitos estimados para a variável resposta atividade de emulsificação para emulsões água-em-hexadecano são apresentados na tabela 22. Os valores em negrito são estatisticamente significativos a um intervalo de confiança de 95%. Nas figuras 24a e 24b encontram-se ilustrados respectivamente o histograma dos efeitos das variáveis independentes sobre a variável resposta biomassa e o diagrama de Pareto, com nível de significância de 95% para estimativa dos efeitos lineares, em valor absoluto. Os resultados obtidos mostram que a variável independente temperatura exerce uma influência negativa altamente significativa sobre a atividade de emulsificação para emulsões água-em-hexadecano. A variável independente agitação não exerce efeito significativo sobre a atividade de emulsificação para emulsões água-em-hexadecano, assim como a interação entre a agitação e a temperatura.

Tabela 22. Efeitos estimados para a variável resposta atividade de emulsificação para emulsões água-em-hexadecano.

	Efeito	Erro Padrão	t(2)	p
(1) Temperatura	-5,45000	0,15275	-35,67863	0,00078
Média	3,12857	0,05774	54,18845	0,00034
1 e 2	0,01000	0,15275	0,06547	0,95376
(2) Agitação	-0,01000	0,15275	-0,06547	0,95376

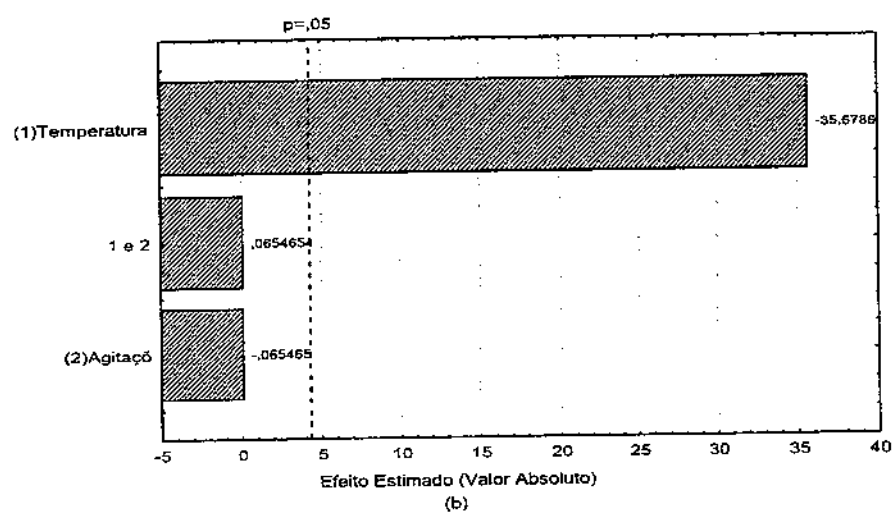
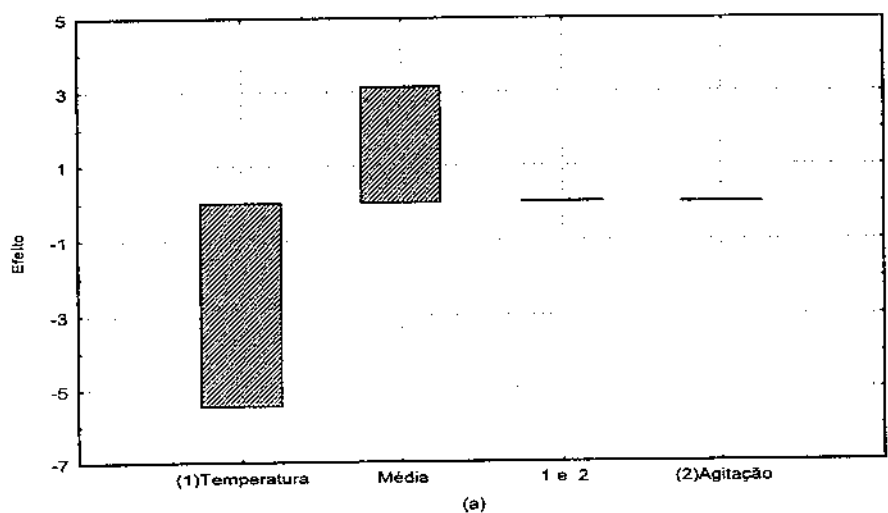


Figura 24. (a) Histograma dos efeitos das variáveis independentes sobre a variável resposta atividade de emulsificação para emulsões água-em-hexadecano . (b) Diagrama de Pareto de efeitos padronizados para o planejamento fatorial completo 2^4 . O ponto no qual os efeitos estimados foram estatisticamente significativos (em $p=0,05$) é indicado por uma linha tracejada vertical.

O ajuste de um modelo contendo interações até dois fatores aos dados da atividade de emulsificação para emulsões água-em-hexadecano (Y) apresentados na tabela 19 resultou na equação:

$$Y = 3,12857 - 2,72500 x_1 + 0,00500 x_2 - 0,00500 x_1 x_2 \quad (10)$$

onde os valores das variáveis independentes, temperatura (x_1) e a velocidade de agitação (x_2), estão na forma codificada (equação 1). Os coeficientes em negrito são estatisticamente significativos no nível de 95% de confiança.

A tabela 23 apresenta os resultados da análise de variância do modelo obtido para a atividade de emulsificação para emulsões água-em-hexadecano

O modelo apresenta um coeficiente de determinação alto ($R^2=0,9499$) explicando 94,99% de variabilidade na resposta. O coeficiente de correlação de 0,9746 indica uma excelente concordância entre os valores experimentais e os valores previstos pelo modelo. O teste F para o modelo mostra que o F calculado ($F_c = 18,46$) é cerca de 2 vezes maior que o F tabelado ($F_t (0,05; 3; 3) = 9,28$) demonstrando significância estatística do modelo. O teste F para falta de ajuste mostra que o F calculado ($F_c = 65,15$) é 3,5 vezes maior que o F tabelado ($F_t (0,05; 1; 2) = 18,51$), indicando falta de ajuste significativa do modelo com os dados experimentais e necessidade de se realizar um planejamento fatorial estrela, para obter um modelo quadrático ou de segunda ordem que se ajuste melhor aos dados.

Tabela 23. Análise de Variância (ANOVA) para o modelo de regressão representando a atividade de emulsificação para emulsões água-em-hexadecano

Fonte	SQ	GL	MQ	Fc	Ft	Fc/Ft
Regressão	29,703	3	9,901	18,96	9,28	2,04
Resíduo	1,567	3	0,522			
Falta Ajuste	1,520	1	1,520	65,15	18,51	3,52
Erro puro	0,047	2	0,023			
Total	31,269	6	5,212		$R^2=0,9499$	$R=0,9746$
SQ - Soma Quadrática GL - Graus de Liberdade MQ - Média Quadrática F _c - F calculado F _t - F tabelado						

A adequação do modelo de regressão ajustado aos dados também foi verificada por meio de análise de resíduos. Os resíduos deixados pelo modelo (Figuras 25a e 25b), não se distribuem aleatoriamente em torno do valor zero, apresentando um comportamento sistemático, com valores passando de positivos para negativos e depois se tornando positivo novamente. Este padrão viola a suposição de que as observações são independentes. Isto sugere que talvez outras variáveis importantes não foram incluídas no planejamento e que o modelo talvez possa ser melhorado se um termo quadrático for incluído no mesmo. O gráfico dos valores previstos contra os valores observados (Figura 25c) confirma visualmente a falta de ajuste do modelo. O gráfico de probabilidade normal (Figura 25d) não apresenta evidências que levem a rejeição da hipótese de normalidade.

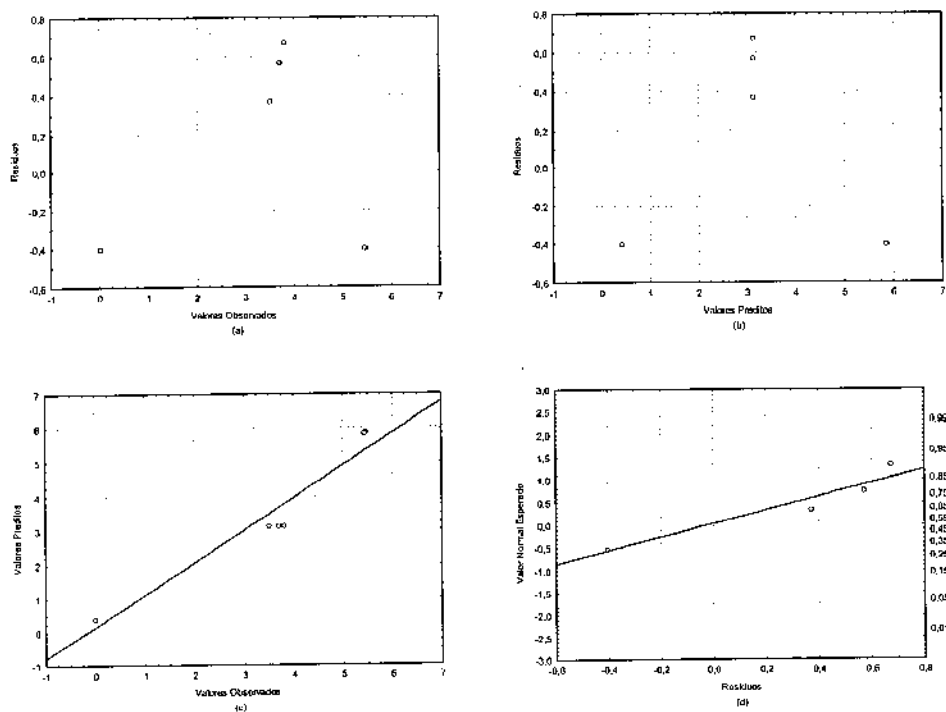


Figura 25. (a) Gráfico de resíduos contra valores observados; (b) Gráfico de resíduos contra valores previstos; (c) Gráfico de valores previstos contra valores observados; (d) Gráfico de probabilidade normal.

Superfície de resposta (Figura 26) construída a partir do modelo descrito pela equação 8, foi usada para analisar as condições operacionais de temperatura e agitação que levam a maior atividade de emulsificação para emulsões água-em-hexadecano.

A figura 26 mostra a superfície de resposta da atividade de emulsificação para emulsões água-em-hexadecano em função da temperatura e da velocidade de agitação. Verifica-se que a atividade de emulsificação para emulsões água-em-hexadecano aumenta com a diminuição da temperatura (nível -1) não importando o nível da agitação.

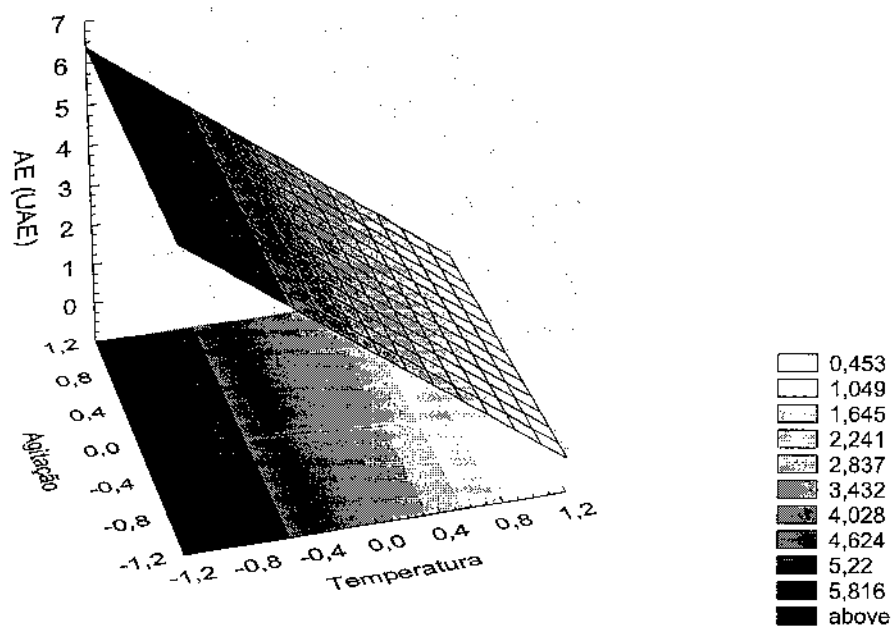


Figura 26. Superfície de resposta para atividade de emulsificação para emulsões água-em-hexadecano mostrando a interação entre temperatura e velocidade de agitação.

4.3.2.2.2. Atividade de emulsificação para emulsões água-em-óleo de milho

A atividade de emulsificação média para emulsões água-em-óleo de milho - obtidas a partir dos filtrados dos cultivos com 120 horas dos 4 ensaios do planejamento 2² e das 3 repetições no ponto central - foi igual a 5,187 UAE, ou seja, 65,80% maior que a atividade de emulsificação média para emulsões água-em-hexadecano obtidas nas mesmas condições (3,129 UAE). A atividade de emulsificação média para emulsões água-em-óleo de milho obtidas a partir dos filtrados dos cultivos com 120 horas dos 4 ensaios do planejamento 2² foi igual a 4,968 UAE, ou seja, 82,29% maior que a atividade de emulsificação média para emulsões água-em-hexadecano obtidas nas mesmas condições (2,725 UAE) e 8,85% menor que a atividade de emulsificação média para emulsões água-em-hexadecano obtidas a partir dos filtrados dos cultivos com 120 horas dos ensaios realizados na temperatura de 28°C (5,45 UAE). No ponto central do planejamento a atividade de emulsificação média para emulsões água-em-óleo de milho foi igual a 5,48 UAE, ou seja, 49,45% maior que a atividade de emulsificação média para emulsões água-em-hexadecano obtidas nas mesmas condições (3,667 UAE) e 0,55 % maior que a atividade de emulsificação média para emulsões água-em-hexadecano obtidas a partir dos filtrados dos cultivos com 120 horas dos ensaios realizados na temperatura de 28°C (5,45 UAE).

Os efeitos estimados para a variável resposta atividade de emulsificação para emulsões água-em-óleo de milho são apresentados na tabela 24. Os valores em **negrito** são estatisticamente significativos a um intervalo de confiança de 95%. Nas figuras 27a e 27b encontram-se ilustrados respectivamente o histograma dos efeitos das variáveis independentes sobre a variável resposta biomassa e o diagrama de Pareto, com nível de significância de 95% para estimativa dos efeitos lineares, em valor absoluto. Os resultados obtidos mostram que a variável independente temperatura exerce uma influência negativa altamente significativa sobre a atividade de emulsificação para emulsões água-em-óleo de milho. A variável independente agitação não exerce efeito significativo sobre a atividade de emulsificação para emulsões água-em-óleo de milho, assim como a interação entre a agitação e a temperatura.

Tabela 24. Efeitos estimados para a variável resposta atividade de emulsificação para emulsões água-em-óleo de milho.

	Efeito	Erro Padrão	t(2)	p
Média	5,18743	0,11880	43,66392	0,00052
(1)Temperatura	-2,08400	0,31432	-6,63009	0,02200
(2)Agitação	-0,79400	0,31432	-2,52605	0,12744
1 e 2	-0,61400	0,31432	-1,95339	0,19000

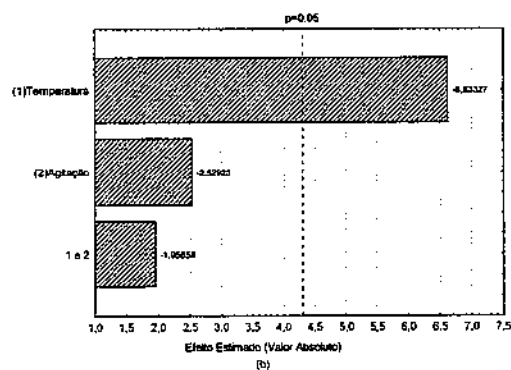
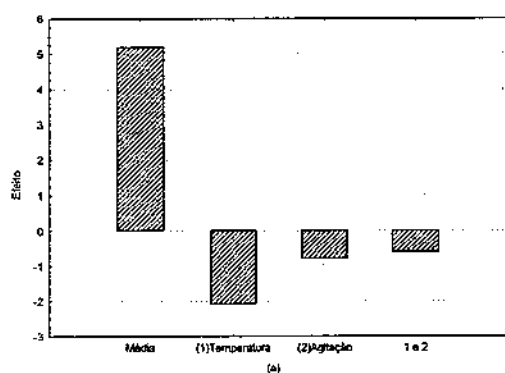


Figura 27. (a) Histograma dos efeitos das variáveis independentes sobre a variável resposta atividade de emulsificação para emulsões água-em-óleo de milho. (b) Diagrama de Pareto de efeitos padronizados para o planejamento fatorial completo 2^4 . O ponto no qual os efeitos estimados foram estatisticamente significativos (em $p=0,05$) é indicado por uma linha tracejada vertical.

O ajuste de um modelo contendo interações até dois fatores aos dados de atividade de emulsificação para emulsões água-em-óleo de milho (Y), apresentados na tabela 19 resultou na equação:

$$Y = 5,18743 - 1,04200 x_1 - 0,39700 x_2 - 0,30700 x_1 x_2 \quad (11)$$

onde os valores das variáveis independentes, temperatura (x_1) e a velocidade de agitação (x_2), estão na forma codificada (equação 1). Os coeficientes em negrito são estatisticamente significativos no nível de 95% de confiança

A tabela 25 apresenta os resultados da análise de variância do modelo obtido para a atividade de emulsificação para emulsões água-em-hexadecano. O modelo apresenta um coeficiente de determinação relativamente alto ($R^2=0,8921$) explicando 89,21% de variabilidade na resposta. O coeficiente de correlação de 0,9445 indica uma ótima concordância entre os valores experimentais e os valores previstos pelo modelo. Entretanto, o teste F para o modelo mostra que o F calculado ($F_c = 8,27$) é menor que o F tabelado ($F_t(0,05; 3; 3) = 9,28$) demonstrando a não significância estatística do modelo. O teste F para falta de ajuste mostra que o F calculado ($F_c = 4,55$) é menor que o F tabelado ($F_t(0,05; 1; 2) = 18,51$), não indicando falta de ajuste significativa do modelo com os dados experimentais.

Tabela 25. Análise de Variância (ANOVA) para o modelo de regressão representando a atividade de emulsificação para emulsões água-em-óleo de milho.

Fonte	SQ	GL	MQ	Fc	Ft	Fc/Ft
Regressão	5,350	3	1,783	8,27	9,28	0,89
Resíduo	0,647	3	0,216			
Falta Ajuste	0,449	1	0,449	4,55	18,51	0,25
Erro puro	0,198	2	0,099			
Total	5,997	6	1,000		$R^2=0,8921$	$R=0,9445$
SQ - Soma Quadrática GL - Graus de Liberdade MQ - Média Quadrática				F_c - F calculado F_t - F tabelado		

A adequação do modelo de regressão ajustado aos dados também foi verificada por meio de análise de resíduos. Apesar do resultado do teste F não indicar falta de ajuste do modelo, os resíduos deixados pelo modelo (Figuras 28a e 28b), não apresentam uma distribuição aleatória em torno do valor zero, apresentando um comportamento sistemático, com valores passando de positivos para negativos e depois se tornando positivo novamente. Este padrão viola a suposição de que as observações são independentes. Isto sugere que talvez outras variáveis importantes não foram incluídas no planejamento e que o modelo talvez possa ser melhorado se um termo quadrático for incluído no mesmo. O gráfico dos valores previstos contra os valores observados (Figura 28c) confirma visualmente uma certa falta de ajuste do modelo. O gráfico de probabilidade normal (Figura 28d) não apresenta evidências que levem a rejeição da hipótese de normalidade.

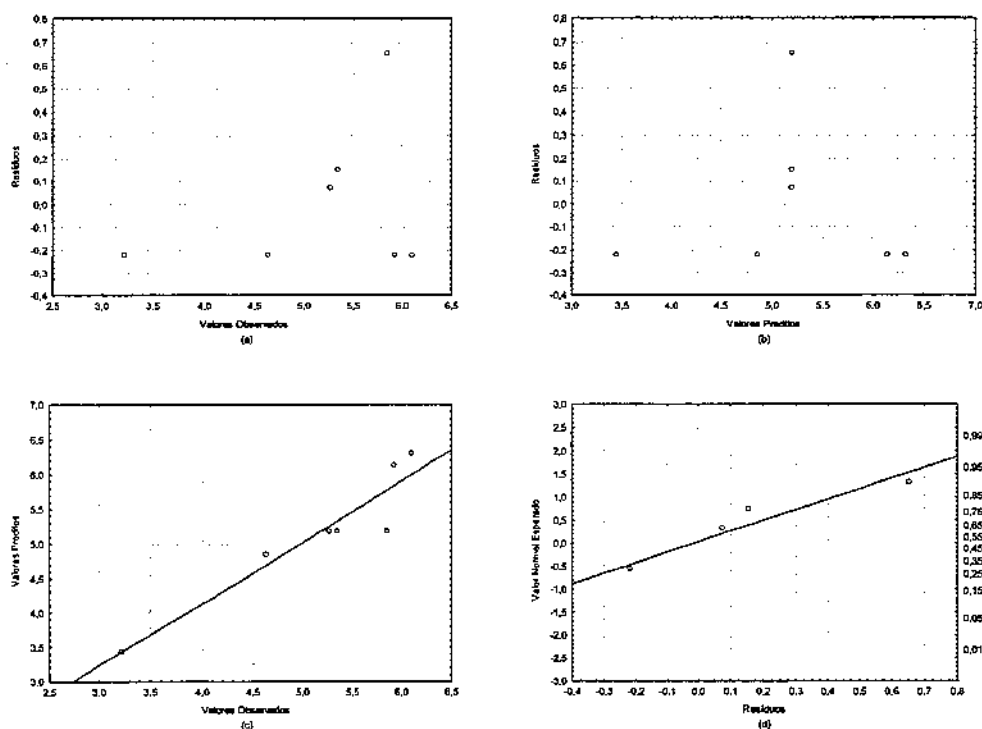


Figura 28. (a) Gráfico de resíduos contra valores observados; (b) Gráfico de resíduos contra valores previstos; (c) Gráfico de valores previstos contra valores observados; (d) Gráfico de probabilidade normal.

Superfície de resposta (Figura 29) construída a partir do modelo descrito pela equação 9, foi usada para analisar as condições operacionais de temperatura e agitação que levam a maior atividade de emulsificação para emulsões água-em-óleo de milho.

A figura 29 mostra a superfície de resposta da atividade de emulsificação para emulsões água-em-óleo de milho em função da temperatura e da velocidade de agitação. Verifica-se que a atividade de emulsificação para emulsões água-em-óleo de milho aumenta com a diminuição da temperatura (nível -1) não importando o nível da agitação.

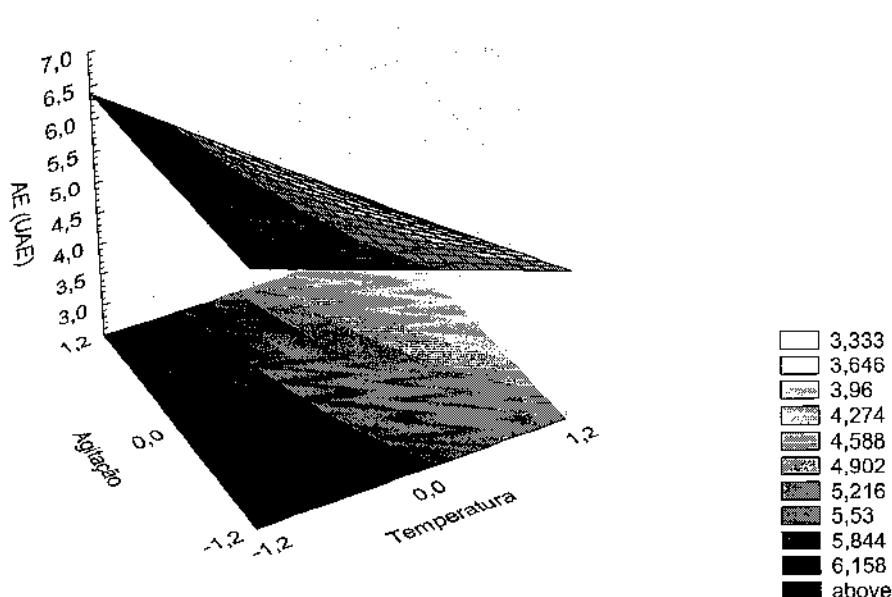


Figura 29. Superfície de resposta para atividade de emulsificação para emulsões água-em-óleo de milho mostrando a interação entre temperatura e velocidade de agitação.

4.3.2.2.3. Atividade de emulsificação para emulsões água-em-óleo de canola

A atividade de emulsificação média para emulsões água-em-óleo de canola - obtidas a partir dos filtrados dos cultivos com 120 horas dos 4 ensaios do planejamento 2² e das 3 repetições no ponto central - foi igual a 5,221 UAE, ou seja, respectivamente, 66,87% e 0,64% maior que as atividades de emulsificação médias para emulsões água-em-hexadecano (3,129 UAE) e para emulsões água-em-óleo de milho (5,187 UAE), obtidas nas mesmas condições. A atividade de emulsificação média para emulsões água-em-óleo de canola obtidas a partir dos filtrados dos cultivos com 120 horas dos 4 ensaios do planejamento 2² foi igual a 5,216 UAE, ou seja, respectivamente, 91,41% e 5% maior que as atividades de emulsificação médias para emulsões água-em-hexadecano (2,725 UAE) e para emulsões água-em-óleo de milho (4,968 UAE), obtidas nas mesmas condições e 4,29% menor que a atividade de emulsificação média para emulsões água-em-hexadecano obtidas a partir dos filtrados dos cultivos com 120 horas dos ensaios realizados na temperatura de 28°C (5,45 UAE). No ponto central do planejamento a atividade de emulsificação média para emulsões água-em-óleo de canola foi igual a 5,227 UAE, ou seja, 42,55% maior que a atividade de emulsificação média para emulsões água-em-hexadecano obtidas nas mesmas condições (3,667 UAE) e 4,10% menor que a atividade de emulsificação média para emulsões água-em-hexadecano obtidas a partir dos filtrados dos cultivos com 120 horas dos ensaios realizados na temperatura de 28°C (5,45 UAE).

Os efeitos estimados para a variável resposta atividade de emulsificação para emulsões água-em-canola são apresentados na tabela 26. Os valores em negrito são estatisticamente significativos a um intervalo de confiança de 95%. Nas figuras 30a e 30b encontram-se ilustrados respectivamente o histograma dos efeitos das variáveis independentes sobre a variável resposta biomassa e o diagrama de Pareto, com nível de significância de 95% para estimativa dos efeitos lineares, em valor absoluto. Os resultados obtidos mostram que a variável independente agitação exerce uma influência positiva significativa sobre a atividade de emulsificação para emulsões água-em-canola. A variável independente temperatura não exerce efeito significativo sobre a atividade de emulsificação para emulsões água-em-canola. A interação entre a agitação e a temperatura exerce um efeito negativo significativo sobre a atividade de emulsificação para emulsões água-em-canola.

Tabela 26. Efeitos estimados para a variável resposta atividade de emulsificação para emulsões água-em-óleo de canola.

	Efeito	Erro Padrão	t(2)	p
Média	5,22057	0,03259	160,20470	0,00004
(2)Agitação	0,76800	0,08622	8,90778	0,01237
1 e 2	-0,40800	0,08622	-4,73226	0,04187
(1)Temperatura	0,00800	0,08622	0,09279	0,93453

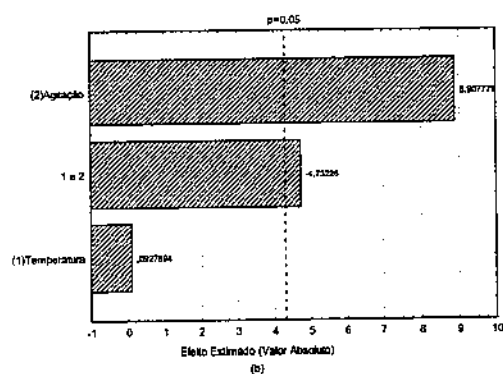
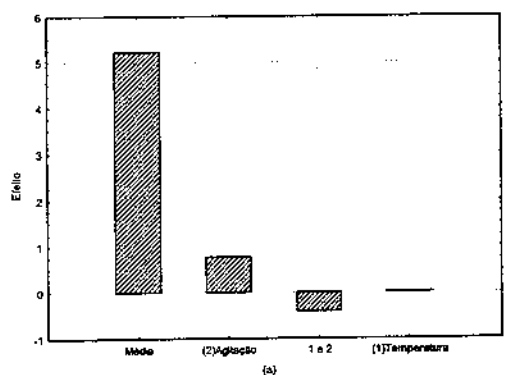


Figura 30. (a) Histograma dos efeitos das variáveis independentes sobre a variável resposta atividade de emulsificação para emulsões água-em-canola. (b) Diagrama de Pareto de efeitos padronizados para o planejamento fatorial completo 2^4 . O ponto no qual os efeitos estimados foram estatisticamente significativos (em $p=0,05$) é indicado por uma linha tracejada vertical.

O ajuste de um modelo contendo interações até dois fatores aos dados da atividade de emulsificação para emulsões água-em-canola (Y) apresentados na tabela 19 resultou na equação:

$$Y = 5,22057 + 0,00400 x_1 + 0,38400 x_2 - 0,20400 x_1 x_2 \quad (12)$$

onde os valores das variáveis independentes, temperatura (x_1) e a velocidade de agitação (x_2), estão na forma codificada (equação 1). Os coeficientes em **negrito** são estatisticamente significativos no nível de 95% de confiança.

A tabela 27 apresenta os resultados da análise de variância do modelo obtido para a atividade de emulsificação para emulsões água-em-canola. O modelo apresenta um coeficiente de determinação muito alto ($R^2=0,9805$) deixando de explicar apenas 1,95% de variabilidade na resposta. O coeficiente de correlação de 0,9902 indica uma excelente concordância entre os valores experimentais e os valores previstos pelo modelo. O teste F para o modelo mostra que o F calculado ($F_c = 18,46$) é 5,41 vezes maior que o F tabelado ($F_t(0,05; 3; 3) = 9,28$) demonstrando a alta significância estatística do modelo. O teste F para falta de ajuste mostra que a relação entre o F calculado ($F_c=0,03$) e o F tabelado ($F_t(0,05; 1; 2) = 18,51$), é praticamente igual a zero, indicando um excelente ajuste do modelo com os dados experimentais.

Tabela 27. Análise de Variância (ANOVA) para o modelo de regressão representando a atividade de emulsificação para emulsões água-em-canola

Fonte	SQ	GL	MQ	Fc	Ft	Fc/Ft
Regressão	0,756	3	0,252	50,22	9,28	5,41
Resíduo	0,015	3	0,005			
Falta Ajuste	0,000	1	0,000	0,03	18,51	0,00
Erro puro	0,015	2	0,007			
Total	0,771	6	0,129		$R^2=0,9805$	$R=0,9902$
SQ - Soma Quadrática GL - Graus de Liberdade MQ - Média Quadrática F_c - F calculado F_t - F tabelado						

A adequação do modelo de regressão ajustado aos dados também foi verificada por meio de análise de resíduos. O gráfico dos resíduos contra os valores ajustados (Figura 31a) e o gráfico dos resíduos contra os valores previstos (Figura 31b) não evidenciam falta de ajuste do modelo nem violação das suposições de independência e igualdade de variância, já que os resíduos deixados pelo modelo (Figuras 31a e 31b), encontram-se distribuídos aleatoriamente em torno do valor zero. O gráfico dos valores previstos contra os valores observados (Figura 31c) confirma visualmente a excelente qualidade do ajuste. Indicações de violação da hipótese de normalidade não foram observadas através do gráfico de probabilidade normal (Figura 31d) já que os resíduos estão localizados aproximadamente ao longo de uma linha reta.

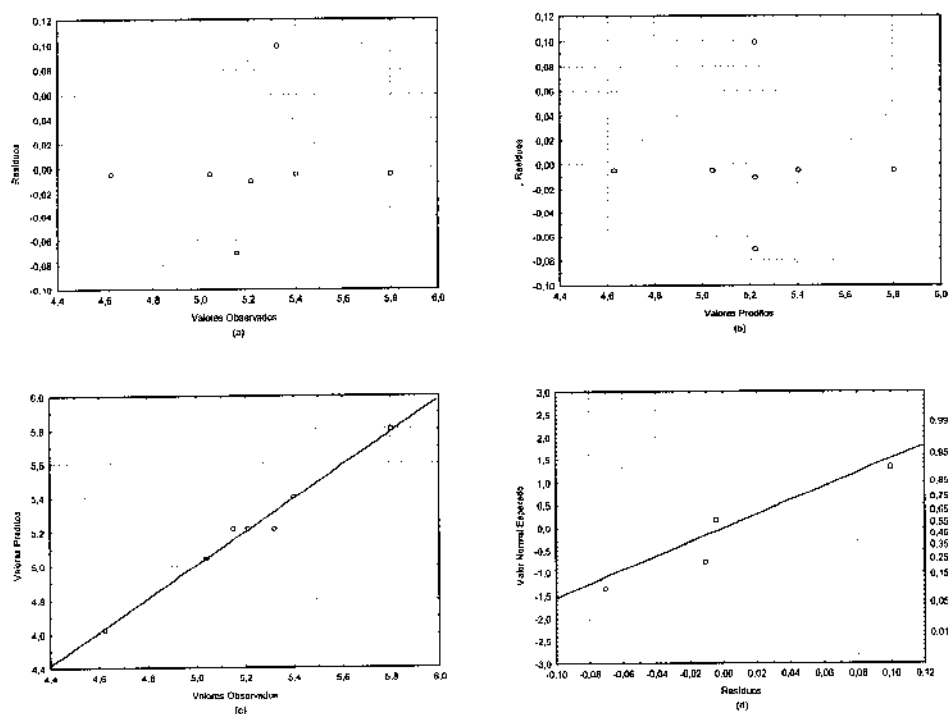


Figura 31. (a) Gráfico de resíduos contra valores observados; (b) Gráfico de resíduos contra valores previstos; (c) Gráfico de valores previstos contra valores observados; (d) Gráfico de probabilidade normal.

Superfície de resposta (Figura 32) construída a partir do modelo descrito pela equação 10, foi usada para analisar as condições operacionais de temperatura e agitação que levam a maior atividade de emulsificação para emulsões água-em-canola.

A figura 32 mostra a superfície de resposta da atividade de emulsificação para emulsões água-em-canola em função da temperatura e da velocidade de agitação. Verifica-se que a atividade de emulsificação para emulsões água-em-canola aumenta com o aumento da agitação (nível +1) e diminuição da temperatura (nível -1).

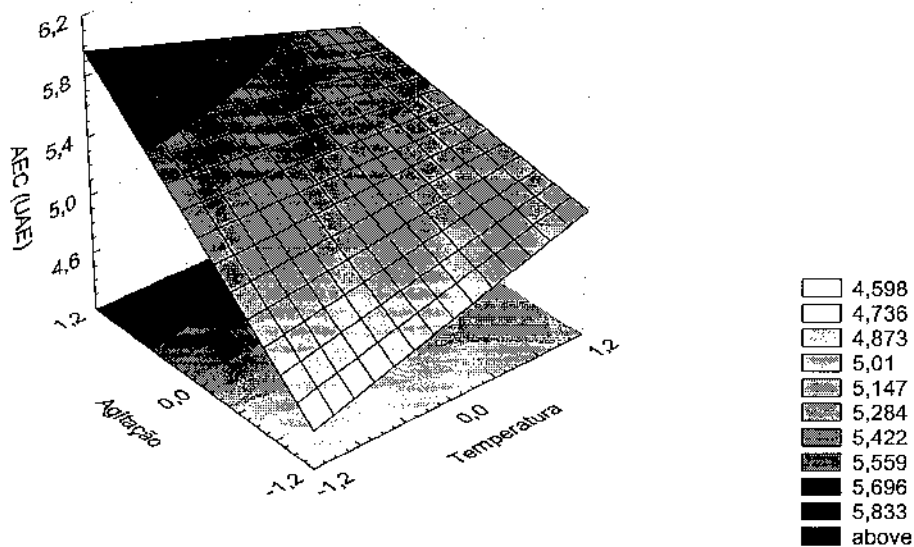


Figura 32. Superfície de resposta para atividade de emulsificação para emulsões água-em-canola mostrando a interação entre temperatura e velocidade de agitação.

4.3.2.3. Monitoramento do processo de produção de biossurfactante

Durante os ensaios do planejamento, medidas *on-line* de pH e oxigênio dissolvido e análises *off-line* de densidade ótica, de salinidade e densidade do cultivo livre de células, de biomassa e de atividade de emulsificação foram realizadas com 0, 4, 18, 24, 48, 72, 96, 120, 144 e 168 horas. Curvas cinéticas destas variáveis foram construídas para serem usadas como padrões de entrada e saída do *softsensor* neural que irá fazer a inferência da concentração de biomassa e da atividade de emulsificação do biossurfactante produzido por *Candida lipolytica*.

4.3.2.3.1. Padrões de entrada

Os perfis conjuntos de pH, oxigênio dissolvido, densidade ótica e salinidade do líquido metabólico livre de células, para todos os sete ensaios do planejamento fatorial 2^2 são apresentados na Figura 33 e constituem os **padrões de entrada** do *softsensor* neural.

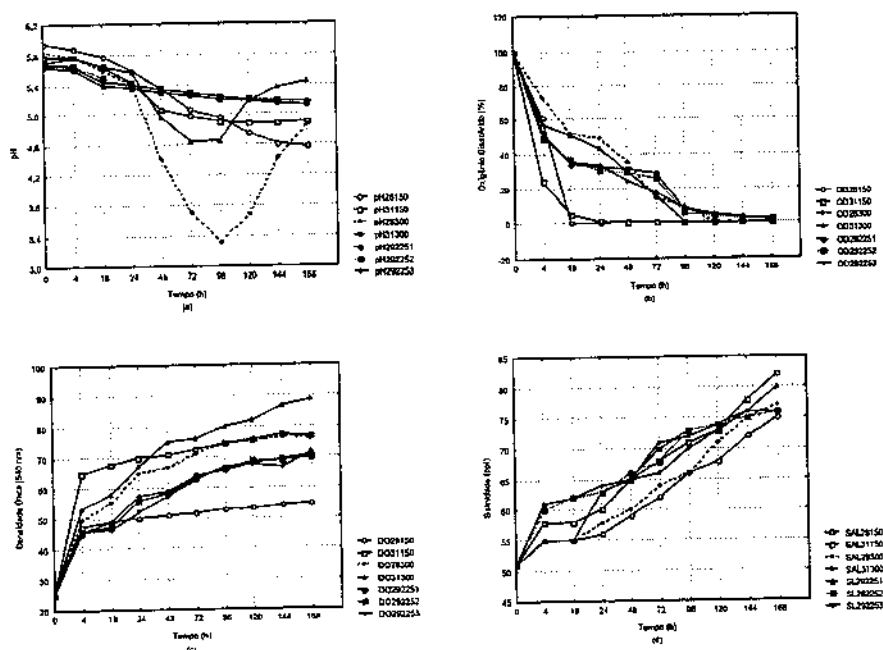


Figura 33. Comparação dos perfis de (a) pH, (b) oxigênio dissolvido, (c) densidade ótica e (d) salinidade do líquido metabólico livre de células nas diversas condições do planejamento fatorial 2^2 .

Perfis de pH

A comparação dos perfis de pH é apresentada na figura 33a. Geralmente quando leveduras crescem, o pH do meio extracelular diminui com o tempo. A atividade enzimática é crucial para a sobrevivência das células de levedura, por regular tudo desde crescimento e respiração até divisão celular, portanto o pH do meio deve ser mantido de forma que as enzimas permaneçam intactas e trabalhando em uma taxa relativamente alta. Se o pH do meio se tornar muito baixo ou muito alto, as células de levedura não serão capazes de manter seus níveis ótimos de pH, e isto pode prejudicar a atividade enzimática e até provocar a morte das células. No ensaio realizado a 28°C e 150 rpm, o pH sofreu uma queda gradativa de 5,95 para 4,57 após 168h. No ensaio realizado a 31°C e 150 rpm, o pH apresentou um comportamento similar, passando de 5,78 para 4,9 após 168h. Entretanto, com a mudança da velocidade de agitação para 300 rpm ocorreu uma nítida mudança no perfil do pH, tornando-se parabólico. No ensaio realizado a 28°C, o pH caiu de um valor inicial de 5,85 para um valor mínimo de 3,67 com 120h, tempo a partir do qual passou a aumentar progressivamente, alcançando o valor de 4,86 com 168h. O mesmo comportamento, um pouco mais suave, pôde ser observado no ensaio realizado a 31°C, onde o pH caiu de 5,7 para um mínimo de 4,65 entre 72h e 96 h, quando então passou a aumentar progressivamente, atingindo o valor 5,44 com 168h. O decréscimo do pH deve ser resultado da produção por *Candida lipolytica* de ácidos orgânicos, tais como o ácido cítrico (ARMILIATO, 2000; CORZO e REVAH, 1999, RANE e SIMS, 1996; REVAH e LEBEAUT, 1989), ou do uso de compostos básicos tais como amônia ou a absorção pelo meio de gás carbônico produzido pelas leveduras (THANH e SIMARD, 1973). O aumento subsequente do pH deve ser causado pela utilização de ácidos e de proteólise (CORZO e REVAH, 1999, RANE e SIMS, 1996; REVAH e LEBEAUT, 1989) e /ou pelo aumento da velocidade de agitação. Os ensaios realizados - em triplicata, a 29,5°C e 225rpm - apresentaram uma variação de pH relativamente pequena, quando comparada às variações de pH dos outros ensaios, passando do valor inicial médio de 5,68 para o valor médio 5,15 em 168h.

Perfis de oxigênio dissolvido

A aeração é um dos parâmetros que mais influenciam o desenvolvimento dos processos metabólicos dos microrganismos. Algumas espécies de *Candida* podem se multiplicar em meios convencionais tanto aerobicamente, como também sob condições estritamente anaeróbicas (ROSE e HARISSON, 1970). A aeração fornece o oxigênio necessário para os microrganismos, e a agitação mantém uniformes as condições dentro do biorreator. Juntas, a aeração e a agitação são importantes na promoção efetiva de transferência de massa para o meio líquido no biorreator.

Na figura 33b é apresentada a comparação dos perfis de oxigênio dissolvido. Apesar da taxa de aeração ser mantida em 1 vvm durante todo o tempo de cultivo, baixos níveis de oxigênio dissolvidos foram atingidos. No ensaio realizado a 28°C e 150 rpm, a concentração de oxigênio dissolvido caiu de 100% para 0,1% após 18h. No ensaio realizado a 31°C e 150 rpm, a concentração de oxigênio dissolvido, apresentou um comportamento similar, passando de 100% para 5,2 após 18h. Estes decréscimos acentuados das concentrações de oxigênio dissolvido podem ser explicados pelo fornecimento insuficiente de oxigênio para as relativamente altas taxas de respiração microbial no início do cultivo, associado a baixa taxa de agitação. Com a mudança da velocidade de agitação para 300 rpm, a queda na concentração de oxigênio dissolvido tornou-se menos brusca, haja vista que mais agitação produz mais dispersão de gás, e mais dispersão de gás produz mais transferência de massa de oxigênio. No ensaio realizado a 28°C e 300 rpm, a concentração de oxigênio dissolvido caiu de 100% para 9,1% com 96h. Comportamento similar foi também observado no ensaio realizado a 31°C e 300 rpm, com a concentração de oxigênio dissolvido caindo de 100% para 0,5% em 96h. Observando-se os perfis dos três ensaios realizados, nas condições do ponto central do planejamento, 29,5°C e 225rpm - pode-se verificar que o nível de oxigênio dissolvido, caiu progressivamente de 100% até atingir cerca de 30% em média após 18h, entrando em um patamar que se estendeu até 72h, quando passou novamente a diminuir até atingir concentração média de aproximadamente 2 % com 168h.

Os resultados obtidos nos ensaios realizados mostram que os aumentos das taxas de agitação de 150 rpm para 225 rpm e de 150 rpm para 300 rpm, aumentam a taxa de transferência de massa do oxigênio para o meio líquido do biorreator, estando de acordo com resultados da literatura (HENSIRISAK, 1997; AHMAD, HOLLAND e McKAY, 1994; HUMPHREY, 1967; DARLINGTON, 1964; STROHM, DALE e PEPPLER, 1959; KAROW, BARTHOLOMEW e SFAT, 1953) que demonstram que os níveis de oxigênio dissolvido no meio de cultura são afetados pela velocidade de agitação, taxa de fluxo de gás e pressão parcial de oxigênio no gás de aeração.

Perfis de densidade ótica

Perfis de densidade ótica dos ensaios realizados são apresentados na figura 33c. Analisando a figura 33c percebe-se que o perfil de crescimento celular (densidade ótica) do ensaio realizado a 28°C e 150 rpm é inferior aos perfis de crescimento dos ensaios realizados a 29,5° C e 225 rpm, e que estes por sua vez são inferiores aos perfis dos ensaios realizados respectivamente, a 28°C e 300 rpm e 31°C e 300 rpm. Estes resultados demonstram que o aumento da temperatura e da velocidade agitação, de uma forma geral, favorecem o crescimento celular. O ensaio realizado a 31° C e 150 rpm, foi a exceção, com seu perfil situando-se entre o perfil do ensaio realizado a 28°C e 150 rpm e os perfis dos ensaios situados a 29,5° C e 225 rpm. Tal comportamento sugere que o aumento da temperatura tem uma influência maior sobre o crescimento que o aumento da agitação.

O padrão de crescimento do ensaio realizado a 28° C e 150 rpm, é bem menor que o dos outros ensaios. Exames dos perfis de oxigênio dissolvido também revelam a inadequada quantidade de oxigênio dissolvido neste ensaio. A causa do decréscimo na produção de biomassa deste ensaio em relação aos outros, deve-se a queda de oxigênio dissolvido provocado pela baixa agitação. Haja vista que, que baixa solubilidade de oxigênio associado com baixas taxas de transferência de oxigênio resulta frequentemente em crescimento e produtividade reduzidos. A condição de oxigênio limitado afeta seriamente o metabolismo de crescimento aeróbico das células de leveduras. Foi relatado por Gelinas e Goulet (1991) que um decréscimo na disponibilidade de oxigênio no meio de

crescimento causou danos celulares, especialmente a membrana citoplasmática e mitocôndria, em leveduras de padaria. Os danos causaram mau funcionamento na parte do metabolismo aeróbico. Com a mudança de ambiente aeróbico para o anaeróbico, as células de leveduras não puderam continuar usando o substrato para crescimento celular e passaram, então, a converter fermentativamente a glicose para gás carbônico.

Perfis de salinidade

A figura 33d mostra a comparação entre os perfis de salinidade do líquido metabólico livre de células obtidos durante os ensaios do planejamento. Conforme se pode observar os perfis de salinidade dos ensaios do planejamento apresentam um comportamento aproximadamente linear. As curvas partem de um valor inicial de 57 ppt e crescem linearmente até alcançarem, com 168h, valores variando entre 75 ppt e 82 ppt (média de 77,43 ppt). As curvas de salinidade do líquido metabólico livre de células dos ensaios realizados a 28°C, apresentaram as menores taxas de variação de salinidade ($0,143 \text{ h}^{-1}$ para agitação de 150 rpm e $0,155 \text{ h}^{-1}$ para agitação de 300rpm), enquanto as dos ensaios realizados a 31°C, apresentaram as maiores taxas de variação de salinidade ($0,185 \text{ h}^{-1}$ para agitação de 150 rpm e $0,173 \text{ h}^{-1}$ para agitação de 300rpm). As curvas de salinidade do líquido metabólico livre de células dos ensaios realizados a 29,5°C e 225 rpm apresentaram em média taxa de variação de salinidade igual a $0,149 \text{ h}^{-1}$. Os resultados mostram que quanto maior a temperatura e menor a agitação maior a salinidade dos líquidos metabólicos livres de células dos ensaios realizados. A observação dos perfis de densidade ótica e dos perfis de salinidade dos líquidos metabólicos livres de células dos ensaios realizados sugere a existência de uma associação positiva entre o crescimento celular e a salinidade do filtrado do cultivo, explicada provavelmente pelo fato das leveduras terem adquirido tolerância ao sal.

4.3.2.3.2. Padrões de saída

Condições operacionais tais como pH, aeração, agitação e temperatura exercem influência sobre a produção de bioemulsificante através de seus efeitos sobre o crescimento e atividade das células.

Os perfis de biomassa e os perfis de atividade de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola obtidos durante os ensaios do planejamento experimental 2² e que constituem os **padrões de saída** do *softsensor* neural são apresentados na figura 34.

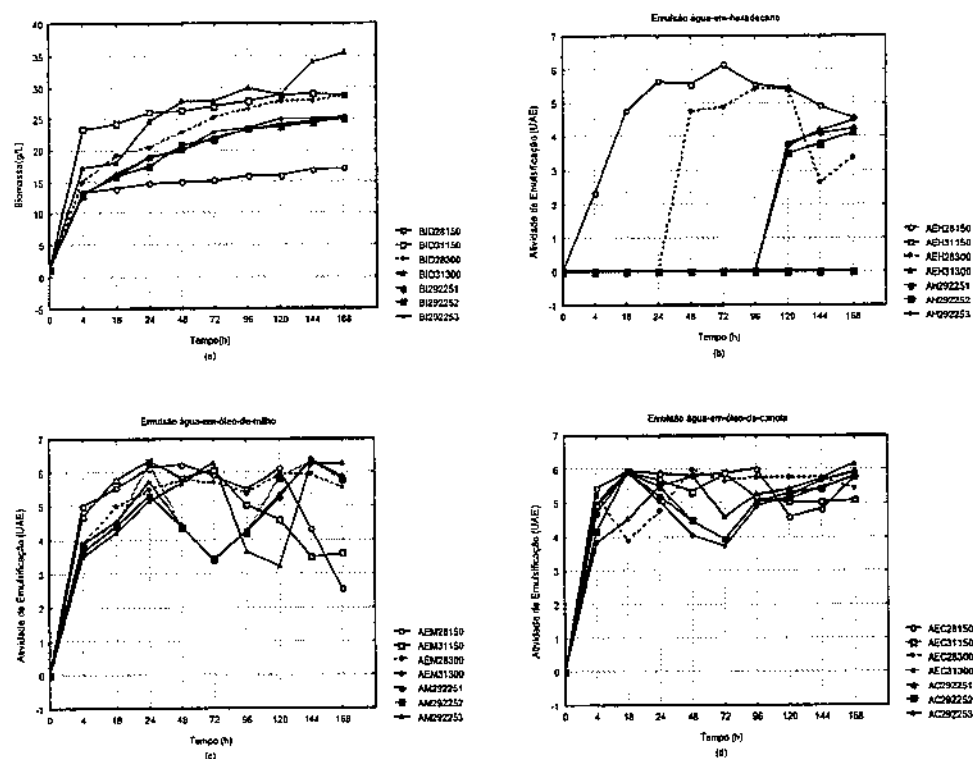


Figura 34. Comparação dos perfis de (a) biomassa e dos perfis de atividades de emulsificações para (b) emulsões água-em-hexadecano, (c) água-em-óleo de milho e (d) água-em-óleo de canola nas diversas condições do planejamento fatorial 2².

Perfis de Biomassa

A Figura 34a apresenta os perfis de biomassa dos ensaios realizados no planejamento experimental completo 2². Aplica-se aqui, a mesma análise de crescimento celular realizada empregando perfis de densidade ótica (Figura 33c).

Em todos os ensaios o crescimento iniciou sem *lag* fase e atingiu a fase exponencial quase que imediatamente, como pretendido pelo procedimento de inoculação em duas etapas. As velocidades máximas de crescimento (μ_{max}) para os ensaios realizados a 28° C e 150 rpm, 31°C e 150 rpm, 28°C e 300 rpm, 31°C e 300rpm e 29,5 ° C e 225 rpm foram respectivamente iguais a 0,59 h⁻¹; 0,74 h⁻¹; 0,63 h⁻¹; 0,66 h⁻¹ e 0,60 h⁻¹ com tempos de geração (t_g) respectivamente iguais a 1,17h; 0,94h; 1,11h; 1,05h e 1,16 h.

Valores de biomassa superiores a 12 g/L após 4 horas de cultivo foram obtidos em todos os ensaios do planejamento. O ensaio realizado a 28°C e 150 rpm foi o que produziu menor quantidade de biomassa, 17,11 g/L, após 120 horas de cultivo. O ensaio realizado a 28° C com velocidade de 300rpm produziu 28,83 g/L de biomassa. Os ensaios realizados a 31°C, foram os que apresentaram maior produção de biomassa. O ensaio realizado, na temperatura de 31° C com velocidade de agitação de 150 rpm, apresentou biomassa igual a 28,77 g/L e o ensaio realizado na mesma temperatura com velocidade de 300rpm produziu 35,46 g/L de biomassa. Os ensaios realizados a 29,5°C e 225 rpm produziram em média 25,14 g/L de biomassa. Estes resultados mostram que para as faixas de temperaturas e velocidades de agitação estudadas, os aumentos da temperatura e da agitação favoreceram a produção de biomassa.

As concentrações de biomassa reportadas neste estudo foram mais altas que alguns valores publicados na literatura. Musiał, Rymowick e Cibis (2004) relataram que durante processo de otimização de produção de biomassa por *Yarrowia lipolytica*, usando o método pulso, concentrações de biomassa iguais a 32,5 g/L e 17,5 g/L foram obtidas respectivamente com dois diferentes meios de cultura, I e II após 240 h. Fazendo uso dos resultados obtidos com os meios de cultura I e II, a seguinte composição ótima do meio, para alcançar 10g/L de biomassa, foi determinada: 2,1 g/L de sulfato de amônio, 0,15 g/L

de fosfato dihidrogênio de potássio, 0,06 g/L de sulfato de magnésio heptahidratado e 1,2g/L de extrato de levedura. Usando um planejamento composto central com 4 fatores e metodologia de superfície de resposta para a produção de 10g/L de biomassa após 72 h, Musial, Rymowick e Cibis (2004) chegaram a seguinte composição para o meio ótimo: 25,7 g/l de óleo rapeseed, 4,2 g/L de sulfato de amônio, 0,28 g/L de fosfato dihidrogênio de potássio, 0,122 g/L de sulfato de magnésio heptahidratado e 2,1g/L de extrato de levedura.

Durante a produção de lipase por *Yarrowia lipolytica*, Corzo e Revah (1999) obtiveram concentração de biomassa igual a 12,2 g/L após 60 h, com um meio incubado a 30° C, composto por 20g/L de glicose, 2g/l de uréia, 10g/L de óleo de oliva ; 1,0 g/l de fosfato monobásico de potássio, 0,5 g de sulfato de magnésio heptahidratado; 0,1 g/L de cloreto de cálcio; 0,1 g/L de cloreto de sódio; 0,5 mg/L de borato de hidrogênio; 0,0040 mg/L de sulfato de cobre pentahidratado; 0,1 mg/L de iodeto de potássio; 0,2 mg/L d cloreto de ferro tetra hidratado; 0,4 mg/Lde sulfato de zinco hepta hidratado; 0,4 mg/L de sulfsto de magnésio hidratado; 4 µg/L de mio-inositol; 8 µg/L de biotina e 200 µg/L de tiamina.

Em estudo sobre os efeitos da agitação em fermentador operado em modo de batelada alimentada na produção de ácido cítrico por *Candida lipolytica*, Crolla e Kennedy (2004) relataram que em 96 horas a produção de biomassa atingiu o valor máximo de 11,5 g/L aproximadamente para todas as velocidades de agitação testadas (400, 800,1000 e1200 rpm). A composição do meio usado para crescimento da levedura e produção de ácido cítrico foi a seguinte: 2,0 g/L de uréia, 0,4 g/L de sulfato de magnésio heptahidratado; 6,0 g/L de carbonato de cálcio; 0,75 g/L de fosfato monobásico de potássio; 0,25 mg/L de tiamina; 0,65 mg/L de nitrato férrico nonahidratado; 1,2 mg/l de sulfato de zinco heptahidratado; 0,31 mg/L de sulfato de cobre pentahidratado; 0,27 mg/L de sulfato de manganês monohidratado e 157 mL/L de Norpar-15. A composição do meio usado para produção de ácido cítrico em batelada alimentada foi a seguinte: 0,4 g/L de sulfato de magnésio heptahidratado; 0,75 g/L de fosfato monobásico de potássio; 0,25 mg/L de tiamina; 0,65 mg/L de nitrato férrico nonahidratado; 1,2 mg/l de sulfato de zinco

heptahidratado; 0,31 mg/L de sulfato de cobre pentahidratado; 0,27 mg/L de sulfato de manganês monohidratado e 166 mL/L de Norpar-15.

As altas concentrações de biomassa obtidas neste trabalho sugerem, portanto, que este subproduto possa vir a ser aproveitado e comercializado como suplemento alimentar em aquicultura, visando redução do custo de produção do bioemulsificante com propriedades surfactantes produzido por *Candida lipolytica*

Perfis de Atividade de Emulsificação

A análise conjunta dos perfis de atividade de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola, apresentados nas figuras 34b, 34c e 34d, e dos perfis de biomassa, apresentado na figura 34a, sugere a existência de uma associação parcial entre crescimento e produção de bioemulsificantes, haja vista que apesar do aumento inicial da atividade de emulsificação paralelo ao crescimento, a produção de emulsificante continuou ao longo da fase estacionária. Pode-se também verificar que a produção de biomassa e de bioemulsificantes, após 4 horas de cultivo, fim da fase exponencial de crescimento, foram altas e influenciadas pela temperatura.

Sob todas as condições testadas, valores de atividades de emulsificação, para emulsões água-em-óleo de milho (Figura 34c) e água-em-óleo de canola (Figura 34d), superiores a 3 UAE foram verificados após 4 horas de cultivo. Após 24 horas valores superiores a 5 UAE foram alcançados, permanecendo praticamente nesta faixa até o fim do cultivo. As atividades de emulsificação para emulsões água-em-óleo de milho (Figura 34c), após 4 horas de cultivo, variaram dentro de uma faixa de 3 a 6 UAE. Por outro lado, atividades de emulsificação para emulsões água-em-óleo de milho (Figura 34c), após 4 horas de cultivo, variaram dentro de uma faixa de 4 a 6 UAE.

Todas as curvas de emulsificação apresentaram um comportamento bifásico ou diáuxico, corroborando com estudos de ROSENBERG *et al.* (1979), CIRIGLIANO e

CARMAN (1884), MARÇAL (1991), SARUBBO(1997) e BRANDÃO(2001). A primeira fase ocorreu ao final do crescimento exponencial e a segunda na fase estacionária tardia.

Para as curvas de atividade de emulsificação para emulsão água-em-óleo de milho - obtidas em 28° C e 150 rpm, 31° C e 150 rpm, 28° C e 300rpm, 31° C e 300rpm e 29,5 ° C e 225 rpm, replicatas 1, 2 e 3 - os primeiros picos ocorreram em 48h, 24 h, 18h, 72 h e 24 h, 24 h e 24 h e apresentaram valores de 6,24 UAE; 6,34 UAE; 5,78 UAE; 6,3 UAE; 5,32 UAE; 6,287 UAE e 5,724 UAE, respectivamente. Para estas mesmas curvas, nas mesmas condições, os segundos picos ocorreram em 120h, 72 h, 144h, 168 h, 144 h, 144 h e 144 h e apresentaram valores de 6,1 UAE; 6,08 UAE; 5,94 UAE; 6,26 UAE; 6,38 UAE; 6,315 UAE e 6,34 UAE, respectivamente.

Para as curvas de atividade de emulsificação para emulsão água-em-óleo de canola - obtidas em 28° C e 150 rpm, 31° C e 150 rpm, 28° C e 300rpm, 31° C e 300rpm e 29,5 ° C e 225 rpm, replicatas 1, 2 e 3 - os primeiros picos de emulsificação ocorreram em 18 h, 18 h, 4 h, 48 h, 18 h , 18 h e 18 h e apresentaram valores de 5,96 UAE; 5,92 UAE; 5,3 UAE; 5,81 UAE; 5,934 UAE; 5,925 UAE e 5,94 UAE, respectivamente. Para estas mesmas curvas, nas mesmas condições, os segundos picos ocorreram em 96 h, 72 h, 96-144 h, 168 h, 168 h, 168 h e 168 h e apresentaram valores de 6,020 UAE; 5,88 UAE; 5,8 UAE; 6,18 UAE; 5,809 UAE; 5,812 UAE e 5,920 UAE, respectivamente.

Com exceção dos ensaios realizados na temperatura de 31° C que não apresentaram nenhuma atividade de emulsificação para emulsões água-em-hexadecano (Figura 34b), nos demais ensaios do planejamento, após 4 horas de cultivo, as atividades de emulsificação para emulsões água-em-hexadecano variaram dentro de um faixa de 2 a 6 UAE.

Na temperatura de 28° C, com velocidade de agitação de 150 rpm, atividades de emulsificação para emulsões água-em-hexadecano (Figura 34b) superiores a 2 UAE foram detectadas com 4 horas de cultivo. Valores superiores a 5 UAE foram obtidos a partir de 24 horas de cultivo. Nestas condições, a curva de atividade de emulsificação também

apresentou comportamento diáuxico, apresentando um pico de 5,64 UAE com 24 horas e um pico de 6,14 UAE com 72 horas.

A curva de atividade de emulsificação para emulsões água-em-hexadecano obtida em 28° C e 300 rpm (Figura 34b), também apresentou um comportamento bifásico. Nenhuma atividade foi detectada até 24 horas. Entretanto, com 48 horas detectou-se uma atividade de emulsificação de 4,782 UAE que aumentou até 5,44 UAE em 120 horas. A seguir, a atividade decaiu até 2,66 UAE em 144 horas, passando então a aumentar até 3,38 UAE em 168 horas.

Por outro lado, a curva de atividade de emulsificação para emulsões água-em-hexadecano (Figura 34b) obtida nas condições do ponto central do planejamento 2² (29,5° C e 225 rpm), não apresentou comportamento bifásico. Nos 3 ensaios realizados nesta condição, até 96 horas nenhuma atividade foi detectada. Atividades de emulsificação em torno de 4UAE foram detectadas entre 120 e 168 horas de cultivo.

Comparando os perfis de biomassa (figura 34a) e os perfis de atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola (Figuras 34b, 34c e 34d), observa-se que apesar de máximo crescimento ter ocorrido na temperatura de 31°C, maior produção de biopolímeros com atividade de emulsificação para os três tipos de emulsões-água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola - ocorreu na temperatura de 28° C. Estes resultados são semelhantes aos obtidos por Deshpande e Daniels (1995), que estudando a produção de biossurfactante por *Candida bombicola* usando gordura animal, obtiveram máximo crescimento em 30° C e máxima produção de sorolipídeos em 27° C.

4.4. Conclusão

O emprego do meio YPASUC-2 otimizado em frascos - usando planejamento fatorial e metodologia de superfície de resposta - na ampliação do processo de produção de bioemulsificante para escala de biorreator de bancada produziu bons resultados. A atividade de emulsificação média para emulsões água-em-hexadecano - obtidas a partir de filtrados de cultivos realizados em biorreator a 28° C com velocidades de agitação de 150 rpm e 300 rpm - foi igual a 5,45 UAE, ou seja, 21,5% maior que a atividade de emulsificação média experimental para emulsões água-em-hexadecano (4,435 UAE) obtidas a partir de filtrados de cultivos realizados em frascos, a 28°C e 150 rpm, com o meio otimizado. Naquelas mesmas condições (temperatura de 28°C com velocidades de agitação de 150 rpm e 300 rpm), foram obtidas atividades de emulsificação médias para emulsões água-em-óleo de milho e água-em-óleo de canola de respectivamente 6,010 UAE e 5,212 UAE. Através do planejamento fatorial 2² realizado e de metodologia de superfície de resposta verificou-se que de uma maneira geral, enquanto o aumento de temperatura exerceu um efeito positivo significativo, o aumento da velocidade de agitação produziu um efeito negativo sobre as variáveis respostas biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola. Portanto, chegou-se a conclusão que nas faixas de temperatura e velocidade de agitação estudadas, para obter-se altas atividades de emulsificações para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola e também alta concentração de biomassa a melhor condição operacional de temperatura e agitação foi 28°C e 300 rpm. Entretanto, as mais altas atividades de emulsificação para emulsões água-em-hexadecano e água-em-óleo de milho foram obtidas na temperatura de 28°C e agitação de 150 rpm. Otimizações posteriores não apenas das condições operacionais de temperatura e velocidade de agitação, mas também do meio de produção e de outras condições operacionais como aeração e concentração de oxigênio dissolvido certamente conduzirão a melhores condições de produção de bioemulsificantes e biomassa.

5. DESENVOLVIMENTO DE SOFTSENSOR BASEADO EM REDES NEURAIS PARA ESTIMAÇÃO E PREVISÃO DE BIOMASSA E DE ATIVIDADE DE EMULSIFICAÇÃO DE BIOEMULSIFICANTE PRODUZIDO POR *Candida lipolytica*

5.1. Introdução

Atualmente, o Núcleo de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco não dispõe de instrumentação que forneça dados *on-line* da concentração de biomassa e da atividade de emulsificação. É por esta razão que um esquema de estimação destas variáveis foi desenvolvido. Com conhecimento mais freqüente do processo disponível, uma estrutura de controle apropriada poderá ser utilizada para melhorar diretamente a produção de biossurfactante. O regime atual de operação, através de análises *off-line* para determinar resultados da condição da fermentação é uma estratégia conservativa. A detecção da proximidade da produção ótima de biossurfactante é dificultada pela coleta infreqüente de amostras. Assim, para aumentar a freqüência de informação rotineiramente disponível da fermentação será necessária uma mudança de análises *off-line* de biomassa e da atividade de emulsificação para medidas ou estimação *on-line*.

Pesquisas em controle e otimização de bioprocessos têm revelado a importância de estudos em escala de laboratório bem sucedidos na redução do tempo e do número de modificações necessárias à operação do processo em escala industrial. Entretanto, a variabilidade inerente dos sistemas biológicos faz com que o processo necessite de ajustes de operação predeterminados em resposta a variações de desempenho. A não modificação das condições operacionais resultará inevitavelmente em perda de produtividade, a não ser que reajustes *on-line* de operação sejam realizados. Infelizmente semelhantes ajustes são dificultados pela falta de sensores *on-line* para indicação do desempenho do biorreator (Di MASSIMO *et al.*, 1992). Uma alternativa para a instalação de instrumentação adicional é uma utilização mais efetiva das informações medidas já disponíveis. Um benefício adicional de se utilizar apenas medidas diretas disponíveis é a inexistência de problemas

associados com a instalação e uso de sensores locais adicionais. *Softwares* sensores fornecem um modo elegante e efetivo para melhorar a utilização de informação disponível correntemente. Eles são primeiramente um meio pelo qual inferência pode ser feita sobre o estado do processo através da fusão de informação disponível *on-line* e/ou *off-line* e informação histórica do processo.

5.2. Materiais e Métodos

O conhecimento de relações e interdependências do sistema é inerente à tecnologia do *softsensor*. A extração deste conhecimento e a forma como é utilizado é a chave da eficiência da tecnologia de *softwares* sensores (MONTAGUE, MORRIS e THAM, 1992). Portanto, a seguir encontra-se descrita a metodologia adotada no desenvolvimento de *softsensor*, baseado em rede neural, para estimar a concentração de biomassa e a atividade de emulsificação do bioemulsificante produzido por *Candida lipolytica*, durante o curso da fermentação, a partir de variáveis primárias do processo.

5.2.1. Metodologia de desenvolvimento dos *softsensores* neurais

No desenvolvimento do *softsensores* neurais adotamos a metodologia mais empregada na revisão de literatura realizada (LINKO e ZHU, 1991; LINKO e ZHU, 1992; KARIM e RIVERA 1992a,b; LINKO e ZHU, 1995; WANG, LUOPA e LINKO, 1995; LINKO, RAJALAHTI e ZHU, 1995; ZHU, RAJALAHTI e LINKO, 1996; LINKO, LUOPA e ZHU, 1997; LINKO e ZHU, 1999; LINKO, LUOPA e ZHU, 1997):

1. Definição do número R de entradas, através da determinação das medidas disponíveis (*on-line* ou *off-line* infrequente) para serem usadas pelo *softsensor*.
2. Definição do número S_2 , de saídas, através da determinação de quais variáveis do processo que devem ser estimadas e/ou previstas.
3. Definição do conjunto de treinamento consistindo de Q pares de vetores de entrada/saída.

4. Seleção da configuração da rede para R entradas e S_2 saídas. Especificação do número S_1 de nós da(s) camada(s) escondida(s).
5. Treinamento *off-line* da rede usando o conjunto de treinamento
6. Teste da rede pela apresentação de um conjunto de vetores de entrada ainda não apresentado à rede e observação dos vetores de saída previstos.
7. Cálculo da diferença do erro entre as saídas previstas e as saídas reais. Obtenção do erro quadrado médio (SSE – *Sum of Squared Errors*) como um índice de desempenho para a rede.
8. Teste de aceitabilidade do erro quadrado médio. Se o erro for aceitável, deve-se continuar. Caso contrário deve-se usar uma configuração diferente e repetir o algoritmo a partir do item 4.
9. Armazenamento das matrizes de pesos $IW1$ e bias $b1$ da camada escondida e das matrizes de pesos $LW2$ e bias $b2$ da camada de saída da configuração selecionada.
10. Uso *on-line* da rede já treinada e previsões das variáveis do processo desejadas.

5.2.2. Pré-Processamento dos Dados

Por estar diretamente ligado a quantidade e qualidade dos dados que compõem os padrões de aprendizagem da rede, o pré-processamento dos dados constituiu-se em uma das chaves para o bom desempenho do *softsensor* neural. Os ensaios do planejamento fatorial 2^2 descrito no capítulo anterior forneceram vários conjuntos de dados sobre o comportamento do processo, em diferentes condições operacionais, que foram usados no treinamento dos protótipos do *softsensor* neural. Os conjuntos de dados de treinamento foram cuidadosamente obtidos de forma que fornecessem informação suficiente e representativa sobre o comportamento do sistema.

5.2.2.1. Construção da base de dados

Durante a ampliação do processo de produção de bioemulsificante da escala de frascos para escala de biorreator de bancada, discutida no capítulo anterior, dados foram gerados em diferentes condições de temperatura e agitação para fornecer informações do processo. Quatro ensaios foram realizados respectivamente em 28°C e 150 rpm, 31 ° C e 150 rpm, 28°C e 300 rpm e 31 ° C e 300 rpm e três ensaios (triplicata) foram realizados em 29° C e 225 rpm. Em cada um dos sete ensaios, os valores das variáveis de entrada - pH, oxigênio dissolvido (OD), densidade ótica (DO) e salinidade do líquido metabólico livre de células (SAL) - e das variáveis de saída – concentração de biomassa (BIO) e atividades de emulsificação para emulsões água-em-hexadecano (AEH), água-em-óleo-de-milho (AEM) e água-em- óleo-de-canola (AEC) foram medidos a cada 0h, 4h, 18h, 24h, 48h, 72h, 96h, 120h, 144h e 168h. Portanto, em cada ensaio foram geradas 8 curvas cinéticas (uma para cada variável) constituídas por 10 pontos cada.

Análises gráfica e estatística de cada curva experimental foram realizadas visando especialmente à identificação e eliminação de dados contendo valores extremos ou *outliers* (MONTEGOMERY e RUNGER, 1994). As curvas cinéticas experimentais obtidas após as eliminações e substituição dos valores extremos através de interpolação por meio de *spline* cúbica estão ilustradas nas Figuras 33 e 34.

Devido à pequena quantidade de dados experimentais, para cada ensaio, a partir das 8 curvas cinéticas experimentais foram geradas mais duas repetições de cada curva (figura 35), atribuindo-se desvios aleatórios de até 5% a cada uma das curvas originais. A seguir, devido aos diferentes e relativamente grandes intervalos de amostragem, cada curva cinética foi suavizada e ampliada para **Q** pontos igualmente espaçados - através de interpolação usando *spline* cúbica. As três repetições das curvas de cada ensaio foram gravadas em arquivos específicos para treinamento, validação e teste das redes neurais. A suavização dos dados experimentais relativamente esparsos foi necessária não apenas para eliminar os ruídos relacionados com os erros de medida, mas também para expandir, por interpolação os dados a serem usados nos conjuntos de treinamento, teste e validação.

O conjunto formado pelos R valores das variáveis de entrada no mesmo instante de tempo foi considerado como um padrão de entrada. O conjunto formado pelos $S2$ valores das variáveis de saída, no mesmo instante de tempo foi considerado como um padrão de saída da base de dados. Como foram geradas 3 repetições, cada repetição contendo $7*Q$ padrões (Q padrões de entrada ou saída de cada ensaio), obteve-se uma base de dados históricos do processo com um total de $(7*Q)*3$ padrões de entrada e um total de $(7*Q)*3$ padrões de saída na base de dados.

Resultados prévios analisados no capítulo anterior (Figura 34) revelaram que apesar de máximo crescimento ter ocorrido na temperatura de 31°C , maior produção de biopolímeros com atividade de emulsificação para os três tipos de emulsões-água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola - ocorreu na temperatura de 28°C . Nas faixas de temperatura e velocidade de agitação estudadas, para obter-se altas atividades de emulsificações para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola e também alta concentração de biomassa, a melhor condição operacional de temperatura e agitação foi 28°C e 300 rpm.

Entretanto, considerando também o ponto de vista econômico, a melhor condição operacional foi 28°C e 150 rpm. Nesta condição além da redução de custo advinda da operação do processo em temperatura e velocidade de agitação mais baixas, as mais altas atividades de emulsificação para emulsões água-em-hexadecano e água-em-óleo de milho foram obtidas e altas atividades de emulsificação para emulsão água-em-óleo de canola e produção de biomassa foram também alcançadas. Portanto, apesar de ser ter uma base de dados constituída por um total de $(7*Q)*3$ padrões de entrada e um total de $(7*Q)*3$ padrões de saída, os seis protótipos dos *softsensores* propostos a seguir, foram treinados e testados para prever a biomassa e as atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola, usando apenas $(1*Q)*3$ padrões de entrada e $(1*Q)*3$ padrões de saída da base histórica do processo na condição operacional 28°C e 150 rpm. Logo, conjuntos com Q igual a 100 e 1000 padrões foram gerados para treinamento, teste e validação do protótipo II e conjuntos com Q igual a 100 padrões foram gerados para treinamento, teste e validação dos demais protótipos.

Base de Dados (5*N*3 padrões)

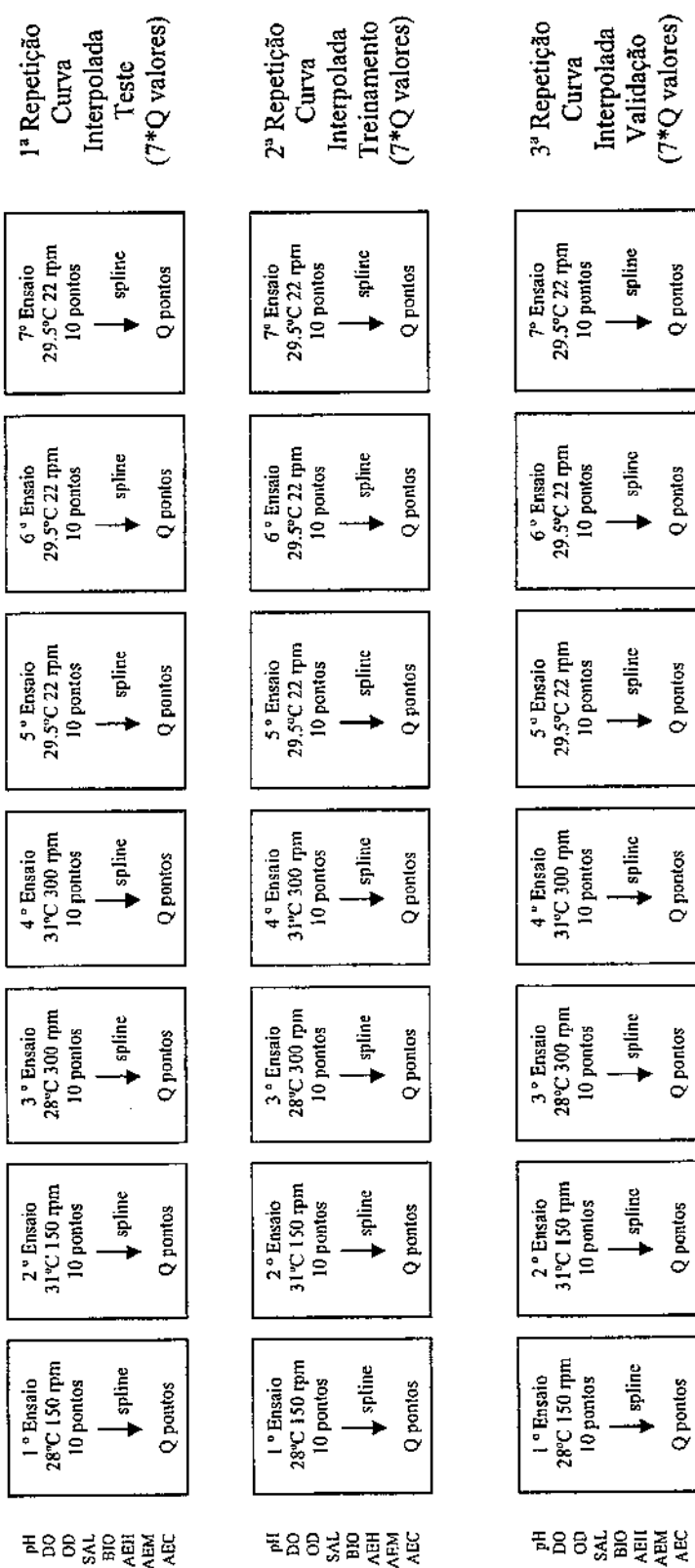


Figura 35. Diagrama do procedimento de criação, ampliação e suavização da base de dados e de sua divisão em conjuntos de treinamento, teste e validação.

5.2.2.2. Divisão do Conjunto de Padrões em Treinamento, Validação e Teste

O total de padrões foi dividido em três conjuntos conforme ilustrado na figura 35: o de treinamento, formado por uma das repetições geradas a partir das curvas cinéticas originais; o de validação, constituído por outra repetição gerada a partir das curvas cinéticas originais e o de teste, contendo os dados das curvas cinéticas originais. Desta forma cada rede neural, recebeu as curvas completas das variáveis de entrada e saída. Visando evitar problemas já que redes neurais não possuem capacidade de extrapolação, os dados no conjunto de treinamento abrangeram o intervalo completo dos valores experimentais ($0h \leq t \leq 168h$), enquanto os dados dos conjuntos de teste e validação abrangeram subintervalos do conjunto de treinamento ($1h \leq t \leq 167h$ e $2h \leq t \leq 166h$ respectivamente).

Como na literatura de redes neurais não existe ainda uma padronização sobre o uso dos termos *conjunto de treinamento*, *conjunto de validação* e *conjunto de teste*, os mesmos foram usados no presente trabalho, no sentido empregado por Bishop (1995) e Ripley (1996), duas referências indiscutíveis em redes neurais:

Conjunto de treinamento: Conjunto de exemplos usados para aprendizagem, isto é para ajustar os parâmetros (pesos) do *softsensor* neural. Várias redes são treinadas pela minimização de uma função de erro apropriada definida com respeito a um *conjunto de treinamento*.

Conjunto de validação: Conjunto de exemplos usados para sintonizar os parâmetros do *softsensor* neural, isto é, para escolher o número de unidades escondidas. Qualquer conjunto de dados que é usado para escolher a melhor entre duas ou mais redes é, por definição, um *conjunto de validação*, e o erro obtido sobre o conjunto de validação é otimisticamente enviesado. Para selecionar a rede com melhor desempenho entre diferentes redes neurais, o método mais simples é avaliar a função erro usando dados que são independentes daqueles usados no treinamento. O desempenho das redes é então comparado pela avaliação da função de erro usando um *conjunto de validação* independente, e a rede tendo o menor erro com relação ao *conjunto de validação* é selecionada. Desde que este procedimento ele mesmo pode conduzir a algum sobreajuste

para o *conjunto de validação*, o desempenho da rede selecionada deve ser confirmado pela avaliação de seu desempenho sobre um terceiro conjunto independente de dados denominado *conjunto de teste*.

Conjunto de teste: Um conjunto de exemplos usados apenas para avaliar (estimar) o desempenho (generalização) de uma rede neural inteiramente especificada. O ponto chave é que um *conjunto de teste*, por definição, nunca é usado para escolher entre duas ou mais redes, de forma que o erro sobre o conjunto de teste possa fornecer uma estimativa não enviesada do erro de generalização (assumindo que o conjunto de teste é representativo da população, etc.).

5.2.4.3. Normalização dos conjuntos de treinamento, teste e validação.

Para evitar problemas com a função de ativação sigmoideal usada no treinamento, os conjuntos de padrões de entrada e saída empregados foram normalizados para o intervalo $[0,1; 0,9]$. Teoricamente, os intervalos das variáveis normalizadas deveriam ser próximos ao intervalo $[0,0; 1,0]$. Na prática, o intervalo $[0,1; 0,9]$ é frequentemente usado para evitar problemas com ambientes de computação digital envolvendo funções de transferência não linear tais como a função sigmóide usada no presente trabalho. Embora a função sigmóide aceite qualquer valor real como entrada, valores absolutos muito altos produzem como saída valores muito próximos a 1, não havendo, portanto grandes alterações na saída ao variar-se a entrada. Quando os valores de entrada são altos, mesmo pequenos erros podem ser grandes em valores absolutos, forçando os neurônios a trabalharem em regiões onde a derivada da função sigmóide é praticamente nula. Como o erro que é retropropagado durante o treinamento é diretamente proporcional a esta derivada, o processo de treinamento poderia ficar paralisado e a rede não convergiria.

5.2.4. Arquitetura e topologia dos *softsensores* neurais

A questão chave na utilização de redes neurais artificiais é a sua capacidade de generalização, ou seja, a confiabilidade das estimativas das redes quando lhe são apresentados dados não utilizados em seu treinamento. Esta capacidade é principalmente afetada pela arquitetura da rede, pela quantidade de dados disponíveis e pela complexidade do problema em estudo. Obviamente, não se teve como controlar a complexidade do problema. Portanto, teve-se que determinar a melhor arquitetura de rede e ao mesmo tempo obter um número adequado de dados de treinamento.

A arquitetura da rede é muito importante para as capacidades de convergência e generalização das RNAs. A escolha de uma arquitetura adequada, nem tão grande e nem tão pequena, é a principal responsável pelo sucesso de uma aplicação. Para se obter uma boa capacidade de generalização, deve-se projetar a rede colocando dentro dela o máximo conhecimento possível sobre o problema (e.g. a topologia do espaço das entradas) e limitar o número de conexões adequadamente. Existem várias técnicas que tentam encontrar a arquitetura ideal através da diminuição de neurônios redundantes e do compartilhamento de pesos. Existe também a possibilidade de fazer uma busca no espaço de todas as arquiteturas possíveis. Para isso, deve-se treinar cada uma das arquiteturas separadamente e então avaliá-las com uma função erro apropriada que incorpore o desempenho e o número de unidades. A busca também pode ser feita através de um algoritmo genético, tal que as boas arquiteturas encontradas em uma triagem sejam mantidas para serem combinadas com boas arquiteturas de outras triagens. Contudo, estes tipos de busca se mostram pouco práticos para aplicações que utilizam grandes RNAs, onde cada treinamento exige um alto esforço computacional. A verdade é que a arquitetura ideal das RNAs ainda continua sendo determinada principalmente pela experimentação manual e pela experiência e sensibilidade do projetista (ABELÉM, A.J.G. 1994).

Durante a ampliação do processo de produção de biossurfactante da escala de frascos para escala de fermentador de bancada diferentes variáveis do processo foram investigadas. Apenas as variáveis consideradas mais informativas como pH, concentração de oxigênio dissolvido, densidade ótica, salinidade do filtrado do cultivo, biomassa e

atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola foram incluídas nos padrões de entrada e saída dos *softsensors* propostos a seguir, baseados em redes neurais *feedforward* com uma camada escondida e uma camada de saída.

O número de entradas dos *softsensors* propostos foi definido pelo número de variáveis possíveis de serem medidas on-line no presente (pH e oxigênio dissolvido) e em futuro próximo (densidade ótica e salinidade do líquido metabólico livre de células) no Núcleo de Pesquisa de Ciências Ambientais da Universidade Católica de Pernambuco. O número de neurônios nas camadas de saídas dos *softsensors* propostos foi definido pelo número de variáveis do processo a serem estimadas isolada ou simultaneamente (biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola). Foi assumido que as variáveis selecionadas para cada protótipo do bioprocessamento forneceram informação suficiente à rede neural sobre o estado corrente do processo e também que todos os conjuntos de dados tiveram as mesmas concentrações iniciais dos inóculos e a mesma viabilidade celular inicial.

Apenas uma camada escondida foi usada no estudo realizado, já que foi demonstrado por Cybenko (1989) que uma camada escondida é suficiente para aproximar qualquer função não linear, desde que exista um número suficiente de nós escondidos. O número de neurônios na camada escondida está relacionado com o desempenho da convergência da função de erro de saída durante o processo de treinamento. Uma camada escondida constituída por poucos neurônios poderá limitar a capacidade da rede neural para modelar o processo e uma camada escondida constituída por neurônios a mais poderá permitir liberdade excessiva para o ajuste dos pesos, resultando em aprendizagem do ruído presente na base de dados usada para o treinamento (LINKO e ZHU, 1992; ZHU, RAJALAHTI e LINKO, 1996; LINKO, LUOPA e ZHU, 1997). A especificação do número de neurônios ótimos na(s) camada(s) escondida(s) é difícil de ser feita e depende do tipo e complexidade do problema. Geralmente, o número ótimo de neurônios é determinado através do método de tentativa e erro, no qual várias redes neurais com diferentes topologias podem ser treinadas e as precisões do treinamento comparadas. Para selecionar o número de nós na camada escondida, normalmente diferentes números de nós são propostos e avaliados de acordo com o critério do erro quadrado médio, sendo selecionadas

as configurações que fornecem menor erro. A função de transferência usada nos neurônios da camada escondida foi a sigmóide por ser a mais empregada e a que mais se aproxima das características dos neurônios biológicos reais (Cowan, 1967). A função de transferência usada na camada de saída foi a linear, por permitir a obtenção de resultados no intervalo de menos infinito a mais infinito ($-\infty, +\infty$).

Seis protótipos de *softsensors* neurais *feedforward*, ilustrados nas figuras 36 a 40 foram propostos, treinados e testados visando encontrar a(s) rede(s) que fornecessem(m) erro(s) mínimo(s) em estimações e previsões de biomassa e de atividade de emulsificação para emulsões água-em-hexadecano, água-em-óleo de canola e água-em-óleo de milho. Os protótipos do *softsensor* neural foram treinados com dados históricos do processo de produção de biossurfactante por *Candida lipolytica*, apresentados em forma de perfis cinéticos nas figuras 33 e 34 do capítulo anterior e obtidos no ensaio do planejamento fatorial para ampliação de escala do processo, realizado a 28° C e 150 rpm. O primeiro protótipo tem como entradas pH, concentração de oxigênio dissolvido, densidade ótica e salinidade do cultivo livre de células, sendo as duas primeiras variáveis medidas *on-line* através de sensores e as outras duas variáveis medidas através de análises *off-line*. Os cinco últimos protótipos têm como entradas, apenas as variáveis pH e concentração de oxigênio dissolvido, medidas *on-line*.

A presença de *bias* nas redes propostas visou prevenir um possível encahlamento (*trapping*) em um mínimo local no gradiente descendente produzido e permitir a aproximação tanto de funções pares como ímpares, além de ajudar na convergência rápida para um melhor ponto sobre a superfície de erro, com a finalidade de alcançar um valor de erro mínimo (McCLELLAND e RUMELHART, 1988).

Protótipo I - *Softsensor* para inferência de concentração de biomassa e de atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola.

Redes neurais de topologia 4- S_1 -4, com S_1 representando o número de neurônios na camada escondida, foram construídas, tendo como entradas pH, concentração de oxigênio dissolvido, densidade ótica e salinidade do cultivo livre de células e como saídas concentração de biomassa e atividades de emulsificação água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola.

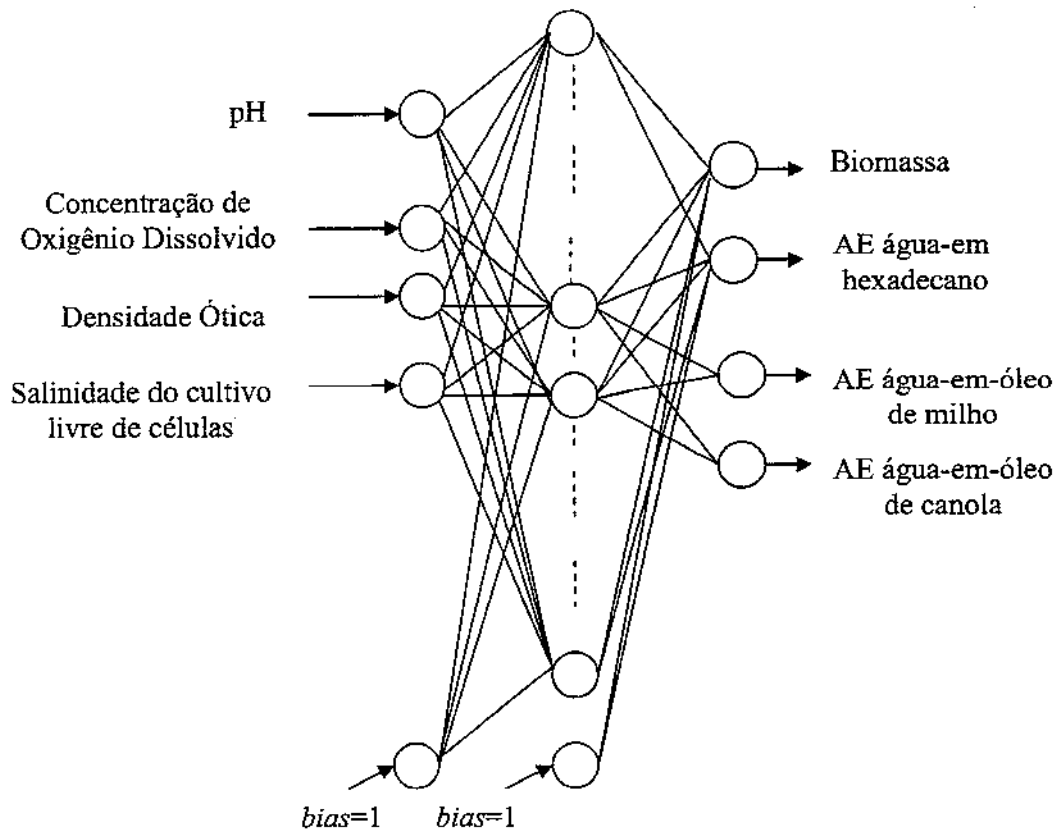


Figura 36.. Protótipo de *softsensor* neural I.

Protótipo II - *Softsensor* para inferência de concentração de biomassa e de atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola.

Redes neurais de topologia 2- S_1 -4, com S_1 representando o número de neurônios na camada escondida, foram construídas, tendo como entradas, pH e concentração de oxigênio dissolvido e como saídas concentração de biomassa e atividades de emulsificação água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola.

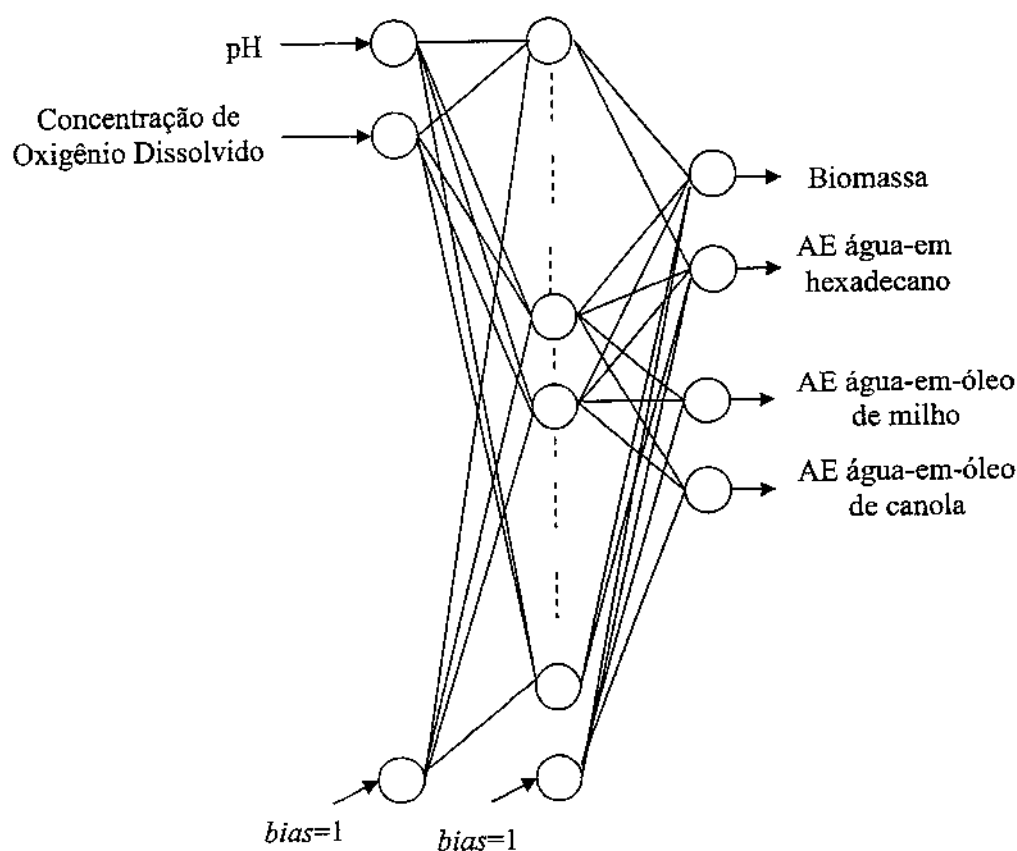


Figura 37. Protótipo de *softsensor* neural II.

Protótipo III - *Softsensor* para inferência de biomassa

Redes neurais de topologia 2- S_1 -1, com S_1 representando o número de neurônios na camada escondida, foram construídas, tendo como entradas, pH e concentração de oxigênio dissolvido e como saída concentração de biomassa.

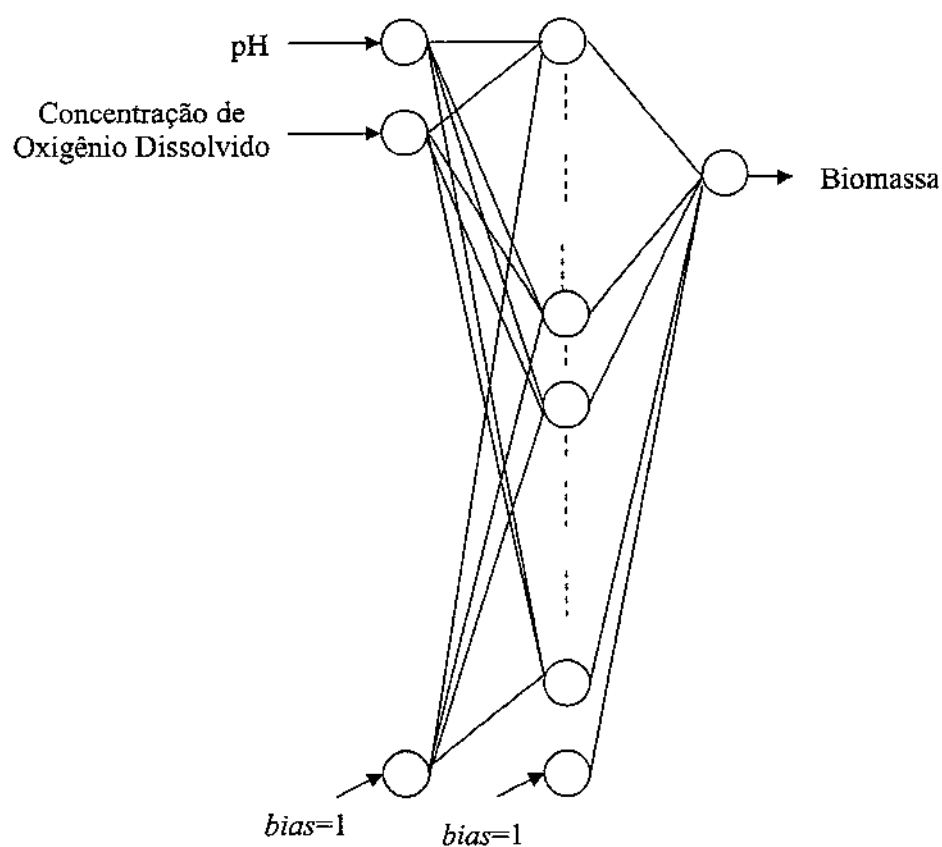


Figura 38. Protótipo de *softsensor* neural III.

Protótipo IV - *Softsensor* para inferência de atividade de emulsificação para emulsões água-em-hexadecano

Redes neurais de topologia 2- S_1 -1, com S_1 representando o número de neurônios na camada escondida, foram construídas, tendo como entradas, pH e concentração de oxigênio dissolvido e como saída atividade de emulsificação para emulsão água-em-hexadecano.

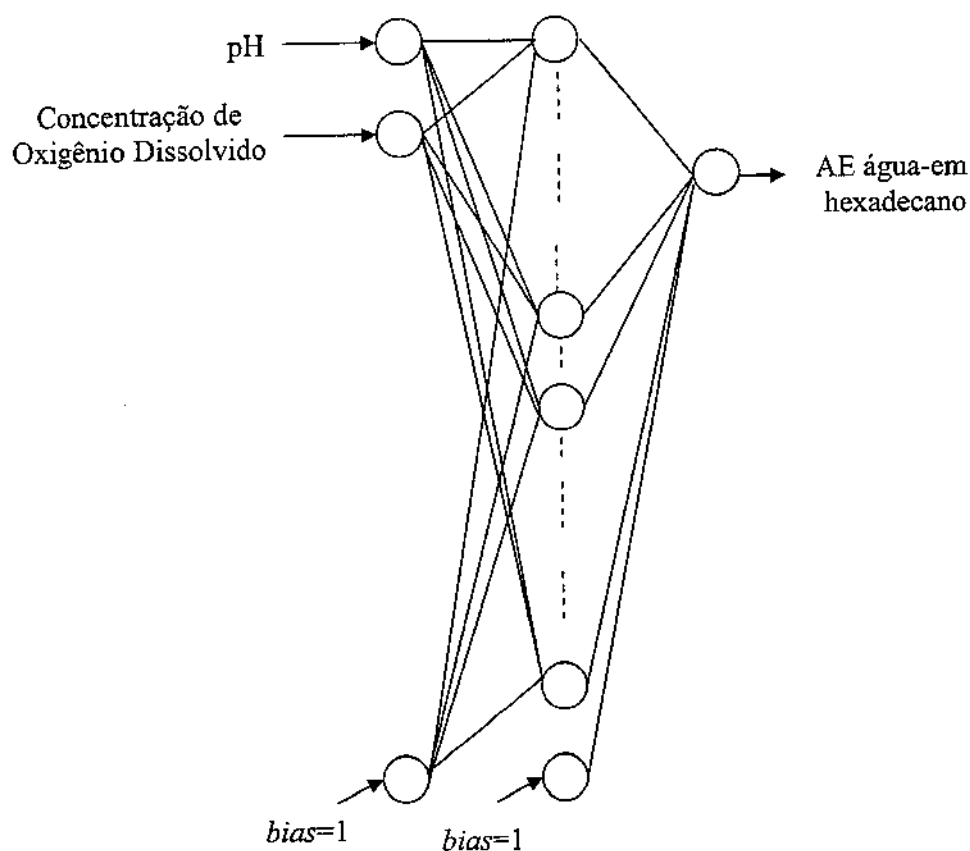


Figura 39. Protótipo de *softsensor* neural IV.

Protótipo V - *Softsensor* para inferência de atividade de emulsificação para emulsões água-em-óleo de milho

Redes neurais de topologia 2-S₁-1, com S₁ representando o número de neurônios na camada escondida, foram construídas, tendo como entradas, pH e concentração de oxigênio dissolvido e como saída atividade de emulsificação para emulsão água-em-óleo de milho.

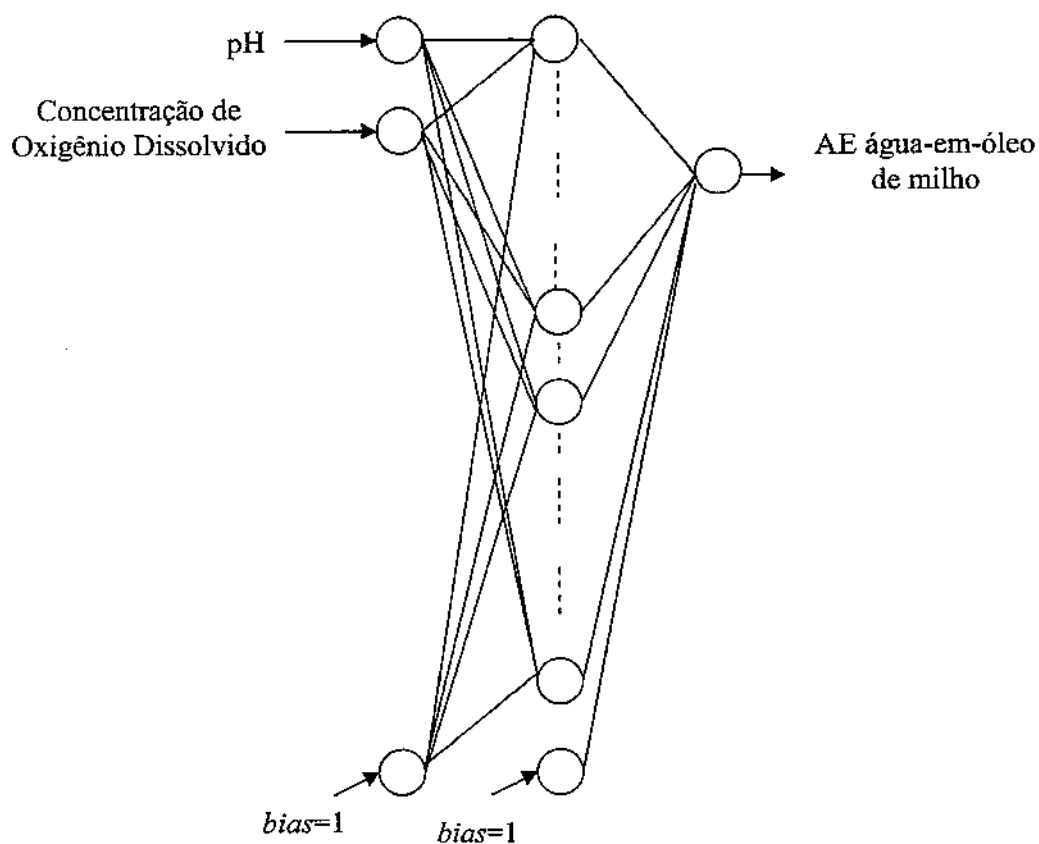


Figura 40. Protótipo de *softsensor* neural V.

Protótipo VI - *Softsensor* para inferência de atividade de emulsificação para emulsões água-em-óleo de canola

Redes neurais de topologia 2- S_1 -1, com S_1 representando o número de neurônios na camada escondida, foram construídas, tendo como entradas, pH e concentração de oxigênio dissolvido e como saída atividade de emulsificação para emulsão água-em-óleo de canola.

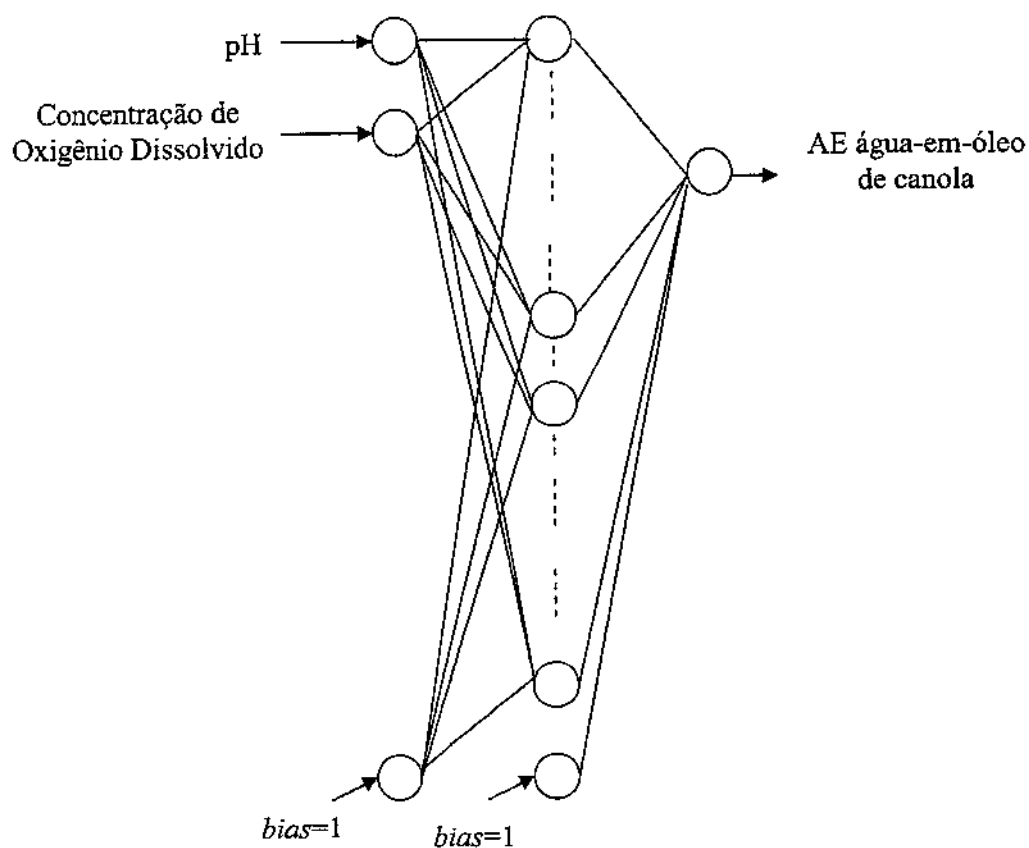


Figura 41. Protótipo de *softsensor* neural VI.

5.2.3. Ambiente de programação e treinamento dos protótipos

Os protótipos foram desenvolvidos e treinados no ambiente de computação técnico-científica MATLAB da The Mathworks Inc., com apoio do sistema denominado NeuronArTech, desenvolvido neste trabalho, usando comandos e funções predefinidas do Neural Network *Toolbox* da linguagem MATLAB. Redes neurais multicamadas *feedforward* (Figura 42) com Q padrões de entrada formados por R variáveis e Q padrões de saída formados por $S2$ variáveis foram implementadas em MATLAB como aproximadores universais de funções, tendo função de transferência sigmóide (*logsig*) em sua camada escondida e função de transferência linear (*purelin*) em sua camada de saída. Esta combinação, ilustrada na figura 36 é especial já que redes neurais com estas funções de transferência podem aproximar qualquer função (com um número finito de descontinuidades) com precisão arbitrária através do ajuste das matrizes de pesos (IW^1 e LW^2) e *bias* (b^1 e b^2) das camadas escondidas e de saída. O único requisito é que a camada escondida tenha bastantes neurônios $S1$. Mais neurônios se fazem necessários a medida que a complexidade da função sendo ajustada aumenta. Na figura 42 a matriz p representa a matriz de entrada de dados, as matrizes n^1 e n^2 representam respectivamente as entradas para as funções de transferência sigmóide e linear e a^1 e a^2 representam respectivamente as saídas das funções de transferência sigmóide e linear.

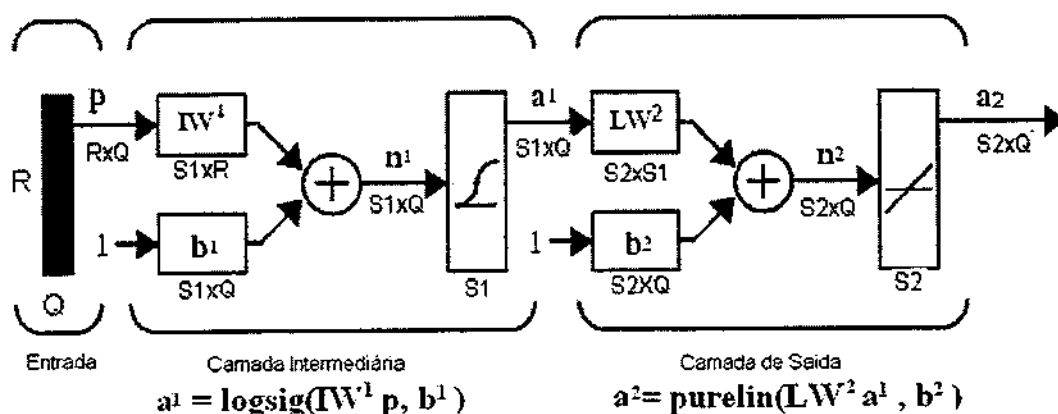


Figura 42. Rede neural multicamada *feedforward* (alimentação progressiva).

5.2.4. Algoritmo de treinamento dos protótipos

Nos últimos quinze anos, o algoritmo mais popular e amplamente usado para treinamento de redes neurais, em aplicações em engenharia, vem sendo o algoritmo de retropropagação (*backpropagation algorithm*). O algoritmo de retropropagação usa o método de gradiente descendente para minimizar uma função objetivo, igual ao quadrado médio da diferença entre o valor desejado e o valor real da saída da rede. Várias aplicações de redes neurais artificiais usando algoritmo de retropropagação têm sido relatadas na literatura (BHAT e McAVOY, 1990; VAN BREUSEGEM, THIBAUT e CHÉRU, 1991; BHAT e McAVOY, 1992; NAHAS, HENSON e SEBORG, 1992; KARIM E RIVERA, 1992ab; ALBIOL *et al.*, 1995; YET-POLE, WU e LIU, 1996; LINKO, LUOPA e ZHU, 1997).

Devido às pequenas taxas de aprendizagem usadas para obter uma convergência estável, o algoritmo de retropropagação pode tornar-se extremamente lento. Conseqüentemente, outros algoritmos foram e continuam sendo propostos para melhorar a taxa e a estabilidade do algoritmo de retropropagação. Como exemplos podem ser citados o algoritmo de *chemotaxis* (BREMERTMAN e McCLELLAND, 1989 *apud* Di MASSIMO *et al.*, 1992, Di MASSIMO *et al.*, 1992, GLASSEY *et al.*, 1994), o algoritmo de gradiente conjugado (KARIM e RIVIERA, 1992a,b) e vários métodos quase Newton (VAN BREUSEGEM, THIBAUT e CHÉRU, 1991; LATRILLE, CORRIEU e THIBAUT, 1994), merecendo especial destaque pelo seu uso crescente (MARQUARDT, 1963; THOLUDUR e RAMIREZ, 1996; FRATTINI, PEDROSA e PEREIRA, 1999; SHENE, DIEZ e BRAVO, 1999; THOLUDUR, RAMIREZ e McMILLAN, 2000), o algoritmo de otimização de Levenberg-Marquardt para pesquisa do erro mínimo, que inicia usando o método do gradiente descendente e depois muda para o método de Gauss-Newton, tornando-se assim mais rápido e preciso na vizinhança da solução ótima.

O *MATLAB Neural Network Toolbox* (DEMUTH E BEALE, 2001) fornece uma suíte de algoritmos de treinamento que foi empregada no desenvolvimento do *softsensor* neural. A função '*trainbr*' que implementa um algoritmo de retropropagação baseado em Levenberg-Marquardt (HAGAN e MENHAJ, 1992) em conjunção com regularização bayescana (MacKAY, 1994) foi empregada no treinamento dos protótipos, visando diminuir

o número de iterações necessárias para convergência em relação ao algoritmo de retropropagação e melhorar a generalização do *softsensor* neural.

A generalização diz respeito à habilidade das RNAs de reconhecer padrões fora do conjunto de treinamento e é a principal propriedade que deve ser exigida de um sistema neural. Ela determina a quantidade de dados necessária para treinar o sistema, a fim de que este responda corretamente para padrões fora do conjunto de treinamento. Apesar das várias aplicações bem sucedidas de RNAs descritas na literatura (CHAKRABORTY, 1992; AZZOF, 1993; PODDIG, 1993; TANG, 1994), as condições para que boas generalizações sejam encontradas ainda não são totalmente conhecidas. Estudos teóricos citados em Refenes (1992) têm mostrado que a probabilidade de generalizações corretas depende do tamanho do espaço das hipóteses (i.e. número total de redes consideradas), do tamanho do espaço das soluções (i.e. conjunto de RNA que fornece boas soluções) e do número de exemplos do treinamento. Se o espaço de hipóteses é muito grande e/ou o número de exemplos de treinamento é muito pequeno, existirá um vasto número de RNAs, as quais são consistentes para o conjunto de treinamento. Porém somente uma pequena porção deste conjunto se encontra no espaço de soluções corretas, fazendo com que seja provável uma generalização pobre.

De uma forma geral, a diminuição dos parâmetros livres (número de conexões) da rede aumenta a probabilidade de uma boa generalização. Contudo, isto deve ser feito sem reduzir demais o tamanho da rede, já que esta redução pode implicar na incapacidade da rede em computar a função desejada. Outro ponto importante é a escolha de um conjunto de treinamento bastante representativo, apresentando as relações mais importantes que devem ser aprendidas pela RNAs. A grande maioria dos pesquisadores concorda que uma boa representação dos dados é um ponto de fundamental importância para o sucesso de uma aplicação. Sabe-se que para uma RNA ter um bom desempenho, ela deve ser grande o suficiente para aprender o problema (convergir), mas pequena o bastante para generalizar bem (WASSERMAN, 1993; ABELÉM, A.J.G., 1994).

5.2.5. Procedimentos de treinamento, validação e teste

Para cada protótipo proposto, o número de neurônios escondidos foi variado de 1 até 20 e para cada uma destas vinte arquiteturas usando o conjunto de treinamento, o procedimento de aprendizagem (treinamento) foi realizado dez vezes, com diferentes pesos randômicos iniciais.

O objetivo do treinamento de cada um dos protótipos de rede neural foi ajustar os pesos das interconexões, de tal forma que a aplicação do conjunto de entradas produzisse o conjunto de saídas desejadas. As diferenças entre as saídas desejadas e os valores previstos pela rede forneceram os erros de previsão. Iterações foram realizadas para minimizar o erro total de previsão pela modificação dos valores dos pesos das conexões até a convergência ser alcançada. Os critérios de parada definidos para treinamento dos *softsensores* neurais foram soma dos erros quadráticos de 10^{-3} ou 1000 ciclos de treinamento, o que fosse atingido mais rápido.

Após cada treinamento, a capacidade dos protótipos do *softsensor* neural simularem as dinâmicas do processo de produção de biossurfactante por *Candida lipolytica*, foi averiguada. Para tanto se usou os dados do conjunto de validação para obter indicativos sobre a adequação do modelo ajustado, uma vez que modelos precisos - usando o conjunto de dados de treinamento - não são uma evidência confiável sobre a validade do modelo, desde que eles não fornecem indicações sobre a generalidade da rede. Os protótipos com maiores capacidades de generalização foram selecionados.

Conjuntos de teste foram usados após as etapas de treinamento e validação para ratificar os desempenhos (capacidades de generalização) dos protótipos dos *softsensores* selecionados, e assegurar seus usos como modelos preditivos.

5.2.6. Medidas de avaliação do desempenho dos protótipos

Os critérios estatísticos apresentados a seguir foram usados para avaliar o desempenho dos diferentes protótipos de *softsensores* propostos:

- Soma dos Erros Quadráticos (*Sum of Square Error - sse*)

Mede o erro entre o valor real registrado (t_i) e a estimação da rede (a_i) para um conjunto de Q padrões constituído por $S2$ variáveis de saída.

$$sse = \sum_{i=1}^Q (t_i - a_i)^2 \quad (13)$$

O *sse* além de ser utilizado para avaliar os resultados dos protótipos de *softsensores* neurais, também foi usado durante o treinamento, no controle do grau de convergência das redes, isto é, mostrando o quanto as redes neurais aprenderam o conjunto de treinamento apresentado a elas.

- Erro Quadrático Médio (*Mean Square Error - mse*)

Mede a exatidão da previsão em termos de valores absolutos. É uma medida de erro que penaliza mais os erros de maior magnitude.

$$mse = \frac{1}{Q \times S2} \sum_{i=1}^Q (t_i - a_i)^2 \quad (14)$$

- Erro Quadrático Médio Modificado (*Modified Mean Square Error – msereg*)

A função de desempenho típica usada para treinar redes neurais *feedforward* é a soma dos erros quadráticos (*mse*). Entretanto, é possível melhorar a generalização

modificando-se a função de desempenho sse pela adição de um termo que consiste na soma dos quadrados dos pesos e bias (ssw)

$$msereg = \alpha sse + \beta ssw \quad (15)$$

$$\text{sendo que : } ssw = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (16)$$

$$\alpha = \frac{\gamma}{2ssw} \quad (17)$$

$$\beta = \frac{np - \gamma}{2sse} \quad (18)$$

e np e γ (ou $gamk$) são respectivamente o número total de parâmetros da rede e o número de parâmetros efetivamente utilizados pela rede (FORESEE e HAGAN, 1997; DEMUTH e BEALE, 2001). O parâmetro γ é a medida de quantos parâmetros da rede são efetivamente usados na redução da função erro e pode variar de 0 até np .

- Erro Quadrático Médio Normalizado (*Normalized Mean Square Error - nmse*)

Mede a qualidade da previsão. Um valor de $nmse$ igual a zero indica perfeita previsão, enquanto um valor de $nmse$ igual a um corresponde simplesmente a previsão da média. Basicamente, o erro quadrático médio normalizado compara o erro quadrático médio do preditor com o erro quadrático médio de outro preditor mais simples (a média dos valores alvos). O preditor avaliado é considerado aceitável se possuir o $nmse$ menor que a unidade, caso contrário ele terá desempenho inferior ao preditor mais simples citado.

$$nmse = \frac{\sum_{i=1}^Q (t_i - a_i)^2}{\sum_{i=1}^Q (t_i - \bar{a})^2} \quad (19)$$

- Raiz do Erro Quadrático Médio Normalizado (*Root Mean Square Error - rmse*)

O *rmse* é uma medida absoluta do erro, com características semelhantes a do *mse*, muito usada por Engenheiros Elétricos e de Comunicações. E como o *mse* é uma medida de erro que penaliza mais os erros de maior magnitude. Se o modelo é perfeito o *rmse* será igual a zero. Usualmente o modelo não é perfeito. Quanto menor o *rmse* melhor o desempenho do modelo. O intervalo do *rmse* varia de zero a infinito, com zero correspondendo ao ideal.

$$rmse = \sqrt{\frac{1}{QxS2} \sum_{i=1}^Q (t_i - a_i)^2} \quad (20)$$

- Raiz do Erro Quadrático Médio Normalizado (*Normalized Root Mean Square Error - nrmse*)

Por ser uma medida de erro relativa, obtida dividindo o *rmse* entre os valores desejados e os valores preditos pelo *rmse* entre os valores desejados e a média destes valores desejados, o *nrmse* permite que se determine o quanto o erro é realmente grande ou pequeno.

$$nrmse = \frac{\sqrt{\frac{1}{QxS2} \sum_{i=1}^Q (t_i - a_i)^2}}{\sqrt{\frac{1}{QxS2} \sum_{i=1}^Q (t_i - \bar{a})^2}} \quad (21)$$

- Erro Absoluto Médio (*Mean Absolute Error - mae*)

Calcula o erro absoluto em função do número de observações. A diferença básica entre *mae* e *mse* é que o *mse* eleva o erro ao quadrado, penalizando erros maiores com muito mais rigor.

$$mae = \frac{1}{Q \times S2} \sum_{i=1}^Q |t_i - a_i| \quad (22)$$

- Erro Relativo Médio (*Mean Relative Error - mre*)

É uma medida do erro relativo médio percentual e, portanto permite melhores comparações de resultados entre diferentes redes neurais treinadas.

$$mre = \frac{1}{Q \times S2} \sum_{i=1}^Q \left| \frac{(t_i - a_i)}{t_i} \right| * 100 \quad (23)$$

- Coeficiente de Correlação (*R*)

Também chamado de coeficiente de correlação de Pearson mede o grau da correlação (e a direcção dessa correlação - se positiva ou negativa) entre duas variáveis. Este coeficiente, normalmente assume apenas valores entre -1 e 1.

- $R = 1$ Significa uma correlação perfeita positiva entre as duas variáveis.
- $R = -1$ Significa uma correlação negativa perfeita entre as duas variáveis.
- $R = 0$ Significa que as duas variáveis não dependem linearmente uma da outra. No entanto, pode existir uma outra dependência que seja "não linear". Assim, o resultado $R = 0$ deve ser investigado por outros meios.

- Coeficiente de Determinação (R^2)

O quadrado do coeficiente de correlação de Pearson é chamado de **coeficiente de determinação** ou simplesmente R^2 . É uma medida da proporção da variabilidade em uma variável que é explicada pela variabilidade da outra. É muito utilizado para medir o ajuste entre os dados reais e os definidos pelo modelo. Mede quanto à variação nos dados reais pode ser explicada pela modelo neural.

$$R^2 = 1 - \left(\frac{\sum (t_i - a_i)^2}{\sum (t_i - \bar{a})^2} \right) \quad (24)$$

- Média dos coeficientes de correlação médios das variáveis de saída ou coeficiente de ajuste/generalização e robustez ($medmed_R$).

Definido neste trabalho. Calcula a média dos coeficientes de correlação médios das $S2$ variáveis de saída simuladas com o conjunto de treinamento ou com o conjunto de teste após cada um dos nt treinamentos realizados para uma determinada configuração de rede neural.

$$medmed_R = \frac{1}{nt} \sum_{i=1}^{nt} \frac{1}{S2} \sum_{j=1}^{S2} r_{ij} \quad (25)$$

Mede o grau de ajuste simultâneo das variáveis de saída, a capacidade de generalização e a robustez de uma determinada configuração de rede neural. Quanto mais próximo de 1 melhor o ajuste (caso o gráfico seja feito com as amostras do conjunto de treinamento) ou a capacidade de generalização (caso o gráfico seja feito com as amostras do conjunto de validação). Quanto mais próximo a 1 maior a robustez da rede, ou seja, maior o número de simulações com o conjunto de treinamento ou simulações com o conjunto de validação com alto grau de ajuste simultâneo das variáveis de saída.

- Coeficiente de determinação global (R_g^2)

Definido neste trabalho como o quadrado da média dos coeficientes de correlação médios das variáveis de saídas da rede.

$$R_g^2 = (\text{medmed_r})^2 \quad (26)$$

Também mede o ajuste, a capacidade de generalização e a robustez de uma determinada topologia de rede. Também varia entre 0 e 1.

- Visualização

O uso de exame visual de gráficos das saídas do processo e do modelo sobre o conjunto de validação como índice de desempenho alternativo de modelos neurais é recomendado por Doerthy et al. (1997). Pesquisa recente de cientistas da NASA (TAYLOR, DARRAH e MOATS, 2003), também ressalta a importância da visualização como técnica útil para alcançar o objetivo de entender inteiramente o desempenho de redes neurais.

5.2.7. Critério de seleção dos melhores *softsensores* neurais

Medidas de desempenho da rede quando testadas com dados do conjunto de validação usualmente avaliam a qualidade do modelo neural. A melhor configuração de rede neural para cada protótipo de *softsensor* proposto foi escolhida como aquela que na simulação com o conjunto de validação para o menor número de neurônios escondidos possível apresentasse o maior coeficiente de ajuste/generalização e robustez (medmed_R) ou equivalentemente o maior coeficiente correlação global (R_g^2). O objetivo da seleção dos modelos foi achar a arquitetura de rede capaz de fornecer o melhor equilíbrio entre os dados ajustados e a complexidade do modelo a fim de evitar sobreajuste e otimizar o desempenho da generalização.

5.3. Resultados e Discussão

Produção de produtos com alta qualidade e baixo impacto ambiental encontram-se entre as metas prioritárias das indústrias modernas. Após a criação de leis governamentais e políticas internacionais que tendem a estreitar os limites do teor de poluentes de produtos e emissões, visando diminuir os riscos ambientais, em termos de poluição de água, solo e ar, o interesse mundial das indústrias com relação ao monitoramento, controle e otimização dos padrões de qualidade de processos e produtos tem aumentado continuamente. Entretanto, a implementação de estratégias de monitoramento e controle extensivas, muitas vezes confronta-se com a inexistência e/ou altos custos de dispositivos para monitoramento e controle *on-line* do processo. Como consequência, um número crescente de pesquisas vem sendo realizado visando o desenvolvimento de técnicas capazes de reduzir os custos dos sensores, através do desenvolvimento de algoritmos capazes de emular o comportamento desses dispositivos, produzindo estimativas em tempo real de dados não mensuráveis com base em dados operacionais disponíveis, ou seja, através do desenvolvimento de *softsensores*.

No caso específico do processo de produção de bioemulsificante por *Candida lipolytica* implementado no Núcleo de Pesquisa em Ciências Ambientais, o emprego de *softsensores* baseados em redes neurais - para estimação e predição de variáveis chaves do processo, como biomassa e atividade de emulsificação medidas através de análises *off-line*, demoradas e de baixa frequência - contribuirá para redução de custos do processo. A implementação dos *softsensores* desenvolvidos tornará o custo global do bioemulsificante produzido mais competitivo em relação aos emulsificantes sintéticos que chegam a ser cerca de 3 a 10 vezes mais baratos.

Dados experimentais de pH, oxigênio dissolvido, salinidade, densidade ótica e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola obtidos com 0h, 4h, 18h, 24h, 48h, 72h, 96h, 120h, 144h e 168h, em ensaio realizado a 28°C e temperatura de 150 rpm, foram usados no desenvolvimento dos *softsensores*. O desenvolvimento dos *softsensores* envolveu basicamente três fases: treinamento, validação e teste. Os seis protótipos de *softsensores* propostos (Figuras 36 a

41) foram treinados, validados e testados com conjuntos de 100 padrões de entrada-saída suavizados e ampliados a partir dos dados experimentais, conforme descrição feita no item 5.2.2.. Para estudar o efeito do número de padrões de entrada-saída sobre o desempenho dos *softsensores*, o segundo protótipo também foi treinado, validado e testado com conjuntos constituídos por 1000 padrões de entrada-saída.

Para cada um dos seis protótipos de *softsensores* propostos, o número de neurônios escondidos foi variado de 1 até 20. E para cada topologia 10 treinamentos foram executados. Cada treinamento foi interrompido a cada 1000 épocas e as saídas calculadas pelas redes comparadas com as saídas desejadas gerando um sinal de erro. Caso a soma dos erros quadráticos atingisse o valor de 10^{-3} antes de 1000 épocas o treinamento era também interrompido. Nas duas situações, o sinal de erro era transmitido de volta através da rede (retropropagação) e os pesos das conexões atualizados com o objetivo de se diminuir o erro entre a entrada e a saída e armazenados em arquivos. Através das análises destes resultados usando os critérios especificados anteriormente para seleção dos melhores *softsensores* neurais determinou-se para cada protótipo de *softsensor* proposto a melhor entre 20 arquiteturas avaliadas. Após os treinamentos, as arquiteturas que apresentaram maiores capacidades de generalização usando o conjunto de validação foram selecionadas para fazer parte do espaço das soluções (conjunto de RNA que fornecem boas soluções) do protótipo analisado. Esta análise foi muito importante desde que a capacidade de predição da rede neural é limitada pelo espaço das hipóteses (i.e. número total de redes consideradas) e pelo espaço geométrico coberto pelo conjunto de treinamento. Após a identificação dos pesos e bias das melhores configurações dos *softsensores* propostos, suas utilizações como modelos preditivos foram testadas usando conjuntos de testes.

Os resultados obtidos durante os treinamentos, validações e teste dos seis protótipos de *softsensores* baseados em redes neurais propostos neste trabalho encontram-se apresentados a seguir.

5.3.1. Treinamentos e validações dos protótipos de *softsensors* neurais

Todas as avaliações dos desempenhos dos protótipos de *softsensors* propostos no presente trabalho foram realizadas após os treinamentos, empregando conjuntos de validação. O objetivo das validações foi determinar o número de neurônios escondidos necessários para que cada protótipo apresentasse a maior capacidade de generalização possível, quando testado com dados não previamente utilizados.

5.3.1.1. Protótipo I – *Softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola

Resultados prévios obtidos durante a ampliação do processo de produção de bioemulsificante da escala de frascos para escala de biorreator de bancada, discutidos no Capítulo 4, revelaram a importância do pH, do oxigênio dissolvido, da densidade ótica e da salinidade do líquido metabólico livre de células para o processo de produção de bioemulsificante por *Candida lipolytica*. Portanto, medidas *on-line* de pH e de oxigênio dissolvido e medidas *off-line* de densidade ótica e de salinidade do líquido metabólico livre de células obtidas a 28° C e 150 rpm foram usadas como entradas para o protótipo I (Figura 36). A biomassa e as atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola foram selecionadas como variáveis de saída.

A inclusão da densidade ótica e da salinidade do líquido metabólico livre de células no padrão de entrada do *softsensor* é pouco realística, desde que se assume que estas duas variáveis estão disponíveis *on-line*, o que não é o caso no momento, mas que futuramente poderia vir a ser com a aquisição de um turbidímetro e de um salinômetro. Entretanto, ambas foram incluídas como entradas para o *softsensor* neural proposto, com a finalidade de avaliar os efeitos das mesmas sobre as variáveis do padrão de saída.

Para verificar a capacidade de predição do protótipo I, topologias com número de neurônios escondidos variando de 1 a 20 foram treinadas e validadas 10 vezes cada uma.

Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de treinamento estão apresentados na tabela 28.

Tabela 28. Protótipo I - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios(*SI*) na camada escondida, obtidos usando o conjunto de treinamento.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	1,86E+00	4,64E-03	6,81E-02	9,68E-04	5,03E-02	8,61E+00	7,806425E-01	6,094027E-01
2	7,87E-01	1,97E-03	4,44E-02	6,30E-04	2,83E-02	4,94E+00	9,114028E-01	8,306551E-01
3	2,46E-01	6,14E-04	2,48E-02	3,52E-04	1,67E-02	2,97E+00	9,778112E-01	9,561147E-01
4	4,96E-02	1,24E-04	1,11E-02	1,58E-04	9,00E-03	1,21E+00	9,983268E-01	9,966565E-01
5	6,58E-03	1,65E-05	4,06E-03	5,76E-05	3,02E-03	3,81E-01	9,994447E-01	9,988898E-01
6	1,24E-03	3,09E-06	1,76E-03	2,50E-05	1,36E-03	1,71E-01	9,998678E-01	9,997356E-01
7	9,97E-04	2,49E-06	1,58E-03	2,24E-05	1,22E-03	1,60E-01	9,999000E-01	9,998001E-01
8	9,98E-04	2,49E-06	1,58E-03	2,24E-05	1,31E-03	1,65E-01	9,999012E-01	9,998025E-01
9	9,89E-04	2,47E-06	1,57E-03	2,23E-05	1,14E-03	1,46E-01	9,999039E-01	9,998077E-01
10	9,95E-04	2,49E-06	1,58E-03	2,24E-05	1,15E-03	1,45E-01	9,999041E-01	9,998083E-01
11	9,94E-04	2,48E-06	1,58E-03	2,24E-05	1,15E-03	1,48E-01	9,999051E-01	9,998101E-01
12	9,98E-04	2,49E-06	1,58E-03	2,24E-05	1,15E-03	1,46E-01	9,999057E-01	9,998113E-01
13	9,98E-04	2,50E-06	1,58E-03	2,24E-05	1,21E-03	1,55E-01	9,999042E-01	9,998085E-01
14	9,92E-04	2,48E-06	1,57E-03	2,24E-05	1,20E-03	1,54E-01	9,999057E-01	9,998114E-01
15	9,85E-04	2,46E-06	1,57E-03	2,23E-05	1,11E-03	1,43E-01	9,999029E-01	9,998059E-01
16	9,91E-04	2,48E-06	1,57E-03	2,24E-05	1,16E-03	1,47E-01	9,999051E-01	9,998103E-01
17	9,95E-04	2,49E-06	1,58E-03	2,24E-05	1,13E-03	1,41E-01	9,999053E-01	9,998107E-01
18	9,75E-04	2,44E-06	1,56E-03	2,22E-05	1,20E-03	1,51E-01	9,999022E-01	9,998043E-01
19	9,85E-04	2,46E-06	1,57E-03	2,23E-05	1,13E-03	1,43E-01	9,999055E-01	9,998110E-01
20	9,93E-04	2,48E-06	1,58E-03	2,24E-05	1,09E-03	1,37E-01	9,999034E-01	9,998067E-01

Entre as 20 topologias testadas para o protótipo I, a topologia com 14 neurônios na camada escondida foi a que apresentou maior número de simulações, com o conjunto de treinamento, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9999057 e R_g^2 = 0,9998114).

Conforme se pode observar através das figuras 43 e 44, as trajetórias das funções de erro são típicas do processo de treinamento de qualquer rede neural, onde existe um rápido declínio do erro com o aumento do número de neurônios escondidos.

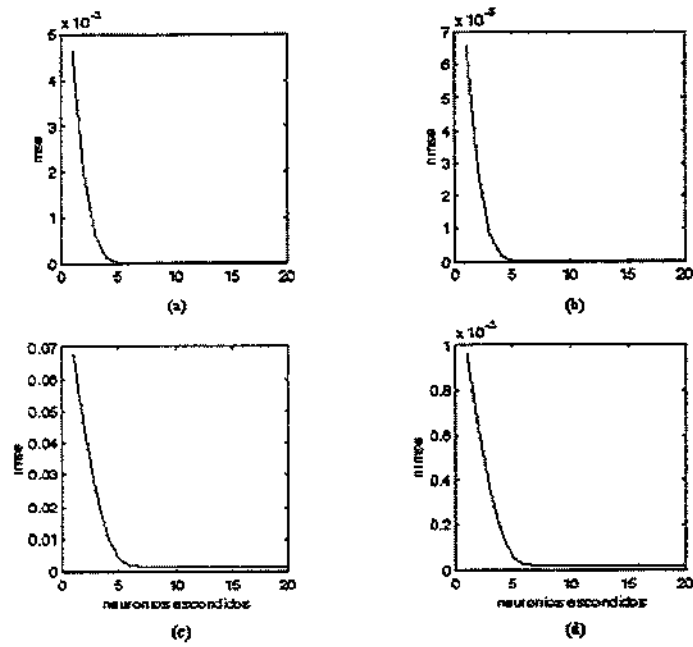


Figura 43. Protótipo I - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nrmse$ usando conjunto de treinamento em função do número de neurônios escondidos.

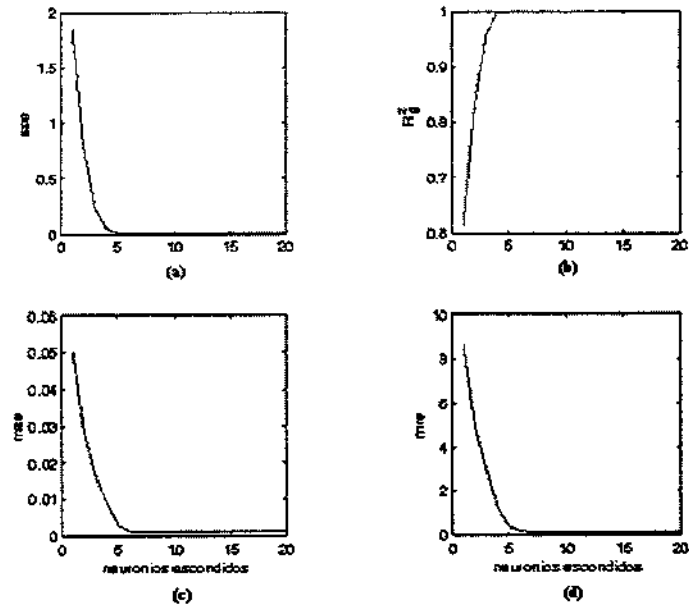


Figura 44. Protótipo I - Índices de avaliação do desempenho (a) sse (b) R_g^2 (c) mae e (d) mre usando conjunto de treinamento em função do número de neurônios escondidos.

Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação estão apresentados na tabela 29.

Tabela 29. Protótipo I - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	3,25E+00	8,13E-03	9,02E-02	1,27E-03	6,25E-02	1,15E+01	7,136419E-01	5,092847E-01
2	1,62E+00	4,05E-03	6,36E-02	8,97E-04	4,41E-02	7,86E+00	8,794099E-01	7,733618E-01
3	9,33E-01	2,33E-03	4,83E-02	6,81E-04	3,74E-02	6,33E+00	9,618113E-01	9,250810E-01
4	1,17E+00	2,91E-03	5,40E-02	7,61E-04	3,62E-02	5,38E+00	9,798977E-01	9,601996E-01
5	6,24E-01	1,56E-03	3,95E-02	5,56E-04	3,02E-02	4,57E+00	9,292467E-01	8,634994E-01
6	6,93E-01	1,73E-03	4,16E-02	5,87E-04	2,86E-02	4,28E+00	9,357589E-01	8,756447E-01
7	8,50E-01	2,13E-03	4,61E-02	6,49E-04	3,41E-02	5,05E+00	9,256034E-01	8,567417E-01
8	1,27E+00	3,19E-03	5,64E-02	7,94E-04	4,32E-02	6,11E+00	9,331133E-01	8,707004E-01
9	8,24E-01	2,06E-03	4,54E-02	6,40E-04	3,37E-02	5,00E+00	9,440900E-01	8,913059E-01
10	1,87E+00	4,67E-03	6,83E-02	9,60E-04	5,14E-02	7,08E+00	9,193939E-01	8,452851E-01
11	7,94E-01	1,98E-03	4,45E-02	6,27E-04	3,37E-02	4,96E+00	9,435313E-01	8,902514E-01
12	9,89E-01	2,47E-03	4,97E-02	7,00E-04	3,71E-02	5,29E+00	9,517411E-01	9,058112E-01
13	8,75E-01	2,19E-03	4,68E-02	6,58E-04	3,16E-02	4,78E+00	9,518448E-01	9,060085E-01
14	9,86E-01	2,47E-03	4,97E-02	6,99E-04	3,91E-02	5,66E+00	9,626505E-01	9,266960E-01
15	7,42E-01	1,85E-03	4,31E-02	6,07E-04	3,12E-02	4,65E+00	9,508583E-01	9,041316E-01
16	6,99E-01	1,75E-03	4,18E-02	5,89E-04	3,03E-02	4,54E+00	9,584842E-01	9,186920E-01
17	1,46E+00	3,66E-03	6,05E-02	8,51E-04	4,84E-02	6,77E+00	9,376156E-01	8,791231E-01
18	9,06E-01	2,26E-03	4,76E-02	6,70E-04	3,75E-02	5,38E+00	9,501304E-01	9,027477E-01
19	1,38E+00	3,44E-03	5,87E-02	8,25E-04	4,71E-02	6,59E+00	9,495927E-01	9,017263E-01
20	8,03E-01	2,01E-03	4,48E-02	6,32E-04	3,30E-02	4,87E+00	9,473605E-01	8,974919E-01

Entre as 20 topologias testadas para o protótipo I, a topologia com 4 neurônios na camada escondida (4-4-4) foi a que apresentou maior número de simulações, com o conjunto de validação, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9798977 e R_g^2 = 0,9601996), ou seja, a que apresentou maior robustez e maior capacidade de generalização global. A topologia com 3 neurônios na camada de saída (4-3-4), apesar de apresentar erros menores, apresentou capacidade de generalização e robustez (um menor número de simulações com o conjunto de validação com altos coeficientes de generalização) inferiores (*medmed_R*=0,9618113 e R_g^2 = 0,9250810).

Conforme se pode observar através das figuras 45 e 46, as trajetórias das funções de erro usando o conjunto de validação não são tão suaves quanto às obtidas usando o conjunto de treinamento. Percebe-se que apesar do rápido declínio do erro até o quinto neurônio escondido, a partir daí o erro oscila muito assim como o coeficiente de determinação global que apesar das oscilações mantém-se superior a 0,8.

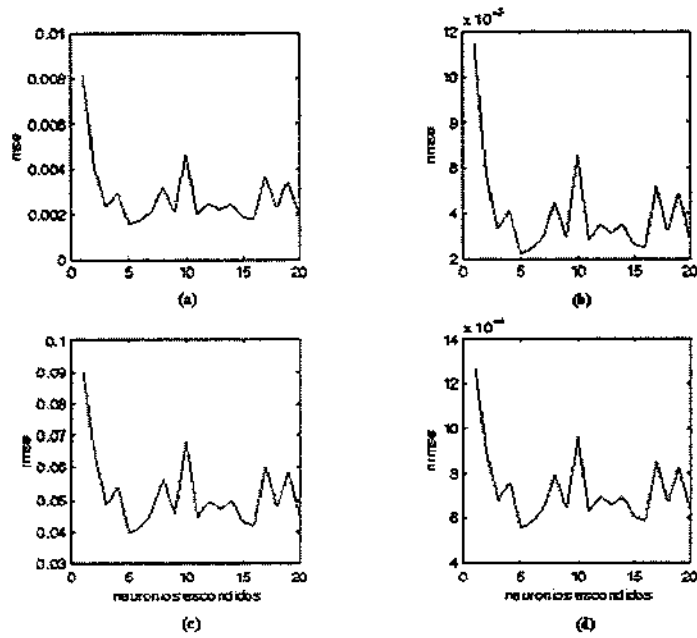


Figura 45. Protótipo I - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nrmse$ usando conjunto de validação em função do número de neurônios escondidos.

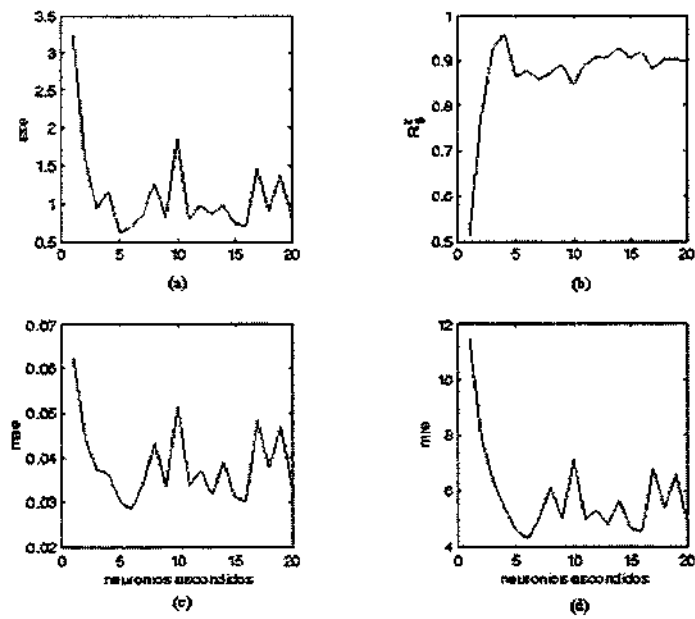


Figura 46. Protótipo I - Índices de avaliação do desempenho (a) sse (b) Rg^2 (c) mae e (d) mre usando conjunto de validação em função do número de neurônios escondidos.

Este comportamento oscilatório sugeriu que talvez houvesse problemas com a qualidade dos dados de entrada ou que talvez as entradas do protótipo não tivessem sido adequadamente escolhidas e não fossem tão importantes para a estimação das variáveis do padrão de saída.

O desempenho do protótipo I para a topologia 4-4-4 usando o conjunto de dados de treinamento pode ser avaliado através das figuras 47 e 48 e comparado através das figuras 49 e 50 com o desempenho do protótipo I para a topologia 4x4x4 usando o conjunto de dados de validação.

Conforme se pode observar através das figuras 47 e 48, as curvas previstas pelo protótipo I com topologia 4-4-4 para as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola ajustaram-se quase perfeitamente aos dados do conjunto de treinamento (coeficientes de correlação R praticamente iguais a 1 e R_g^2 igual a 0,9966565).

Conforme se pode observar através das figuras 49 e 50, as curvas previstas pelo protótipo I com topologia 4-4-4 para as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola também apresentaram excelente ajuste aos dados do conjunto de validação (coeficientes de correlação R superiores a 0,96 e R_g^2 igual a 0,9601996).

Este resultado mostra que o protótipo I com topologia 4-4-4 apresenta excelente capacidade de generalização e também indica que provavelmente o comportamento oscilatório do erro para os protótipos com números de neurônios escondidos superiores a cinco, deve estar associado à qualidade dos dados de entrada e não a adequação das variáveis de entrada escolhidas.

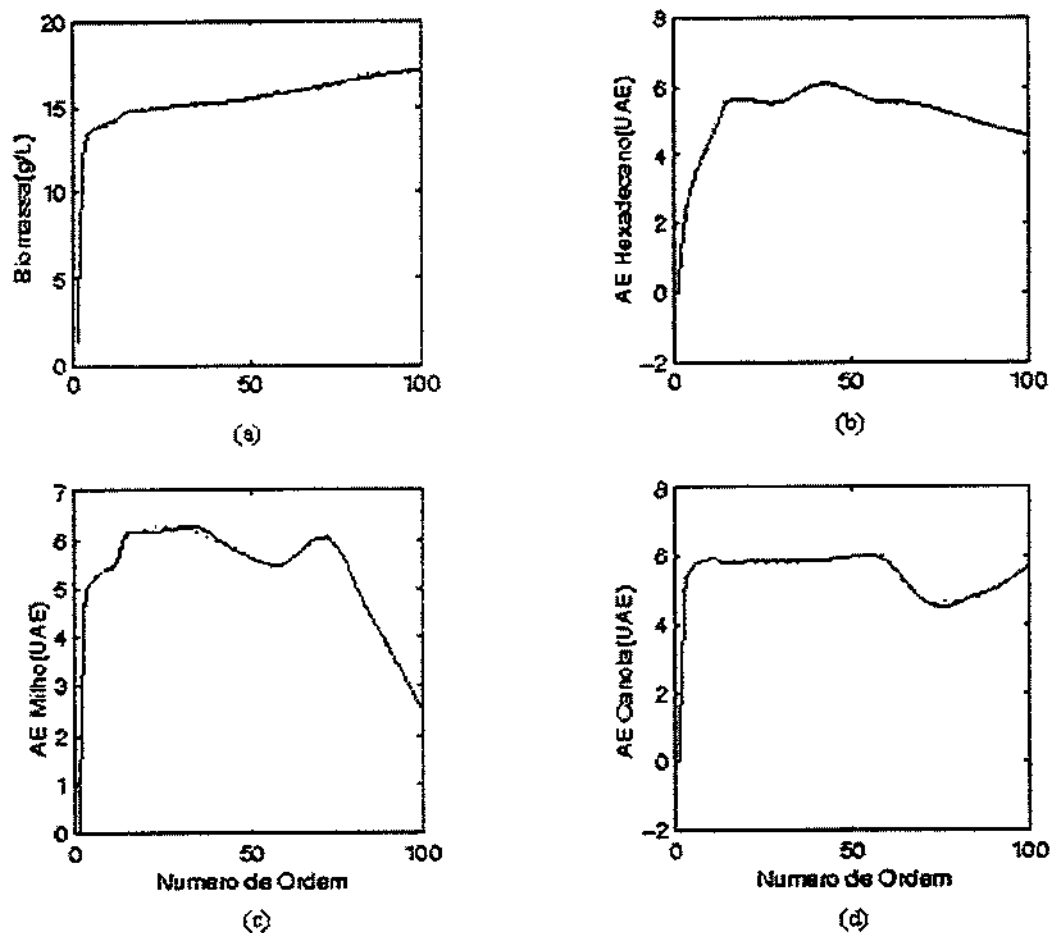


Figura 47. Protótipo I - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de treinamento.

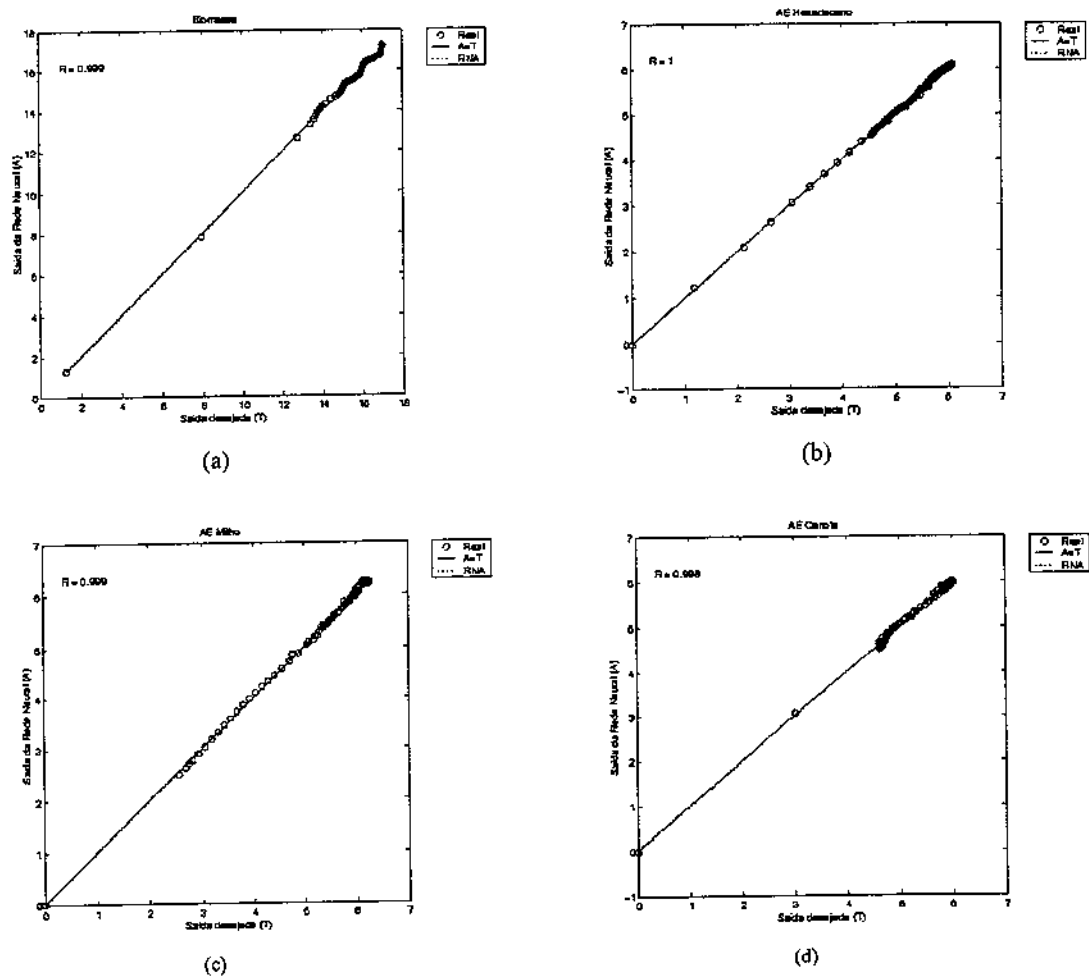


Figura 48. Protótipo 1 - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-óleo de canola usando dados do conjunto de treinamento.

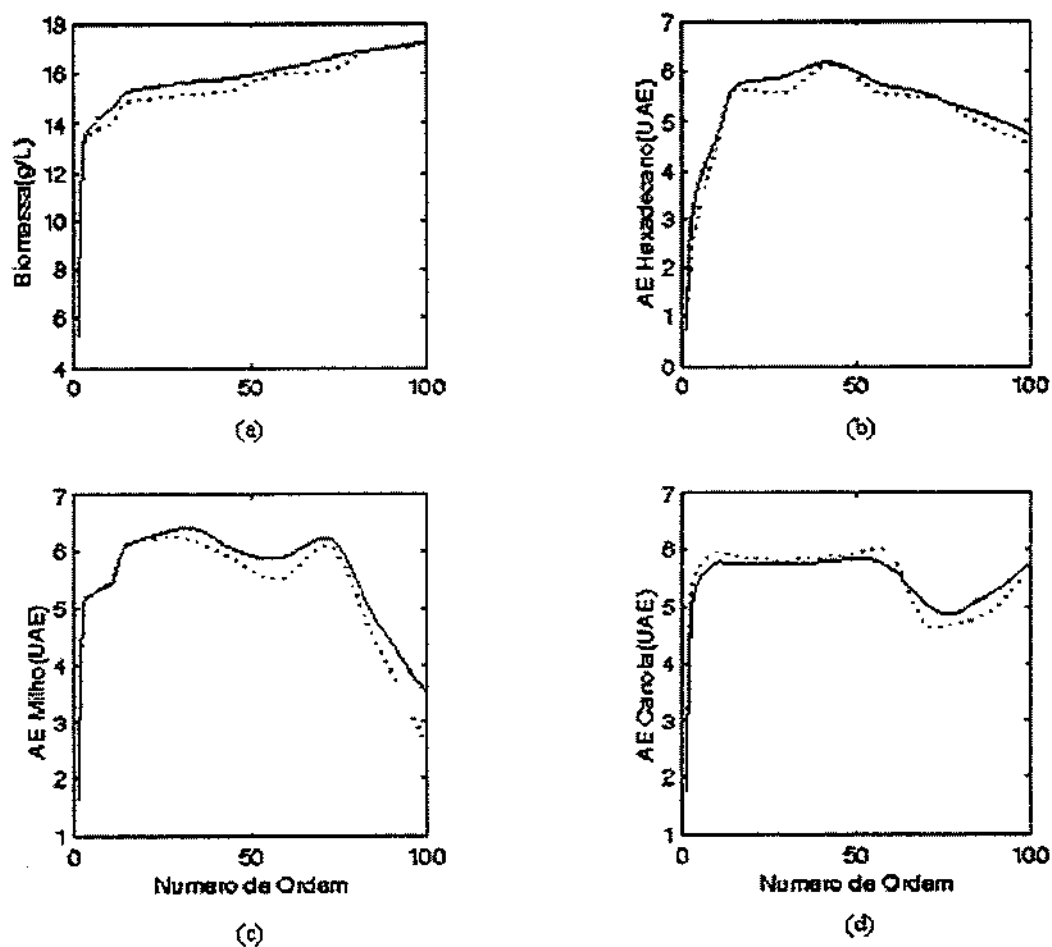
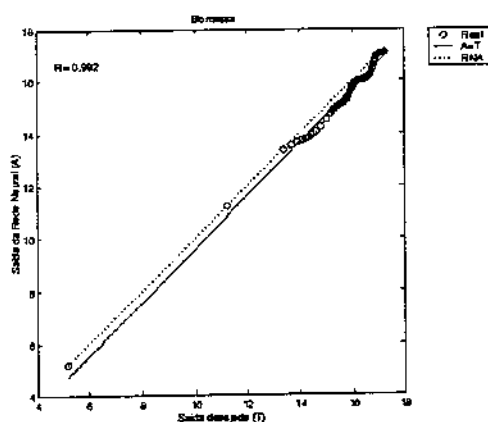
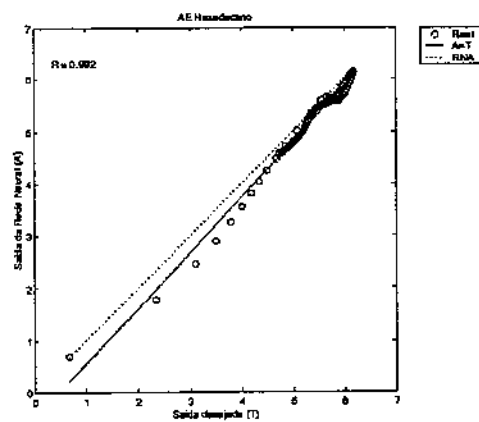


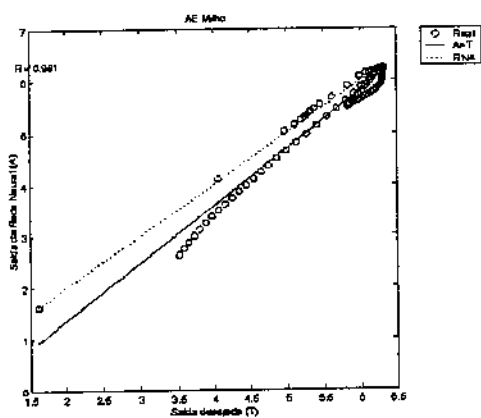
Figura 49. Protótipo I - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de validação.



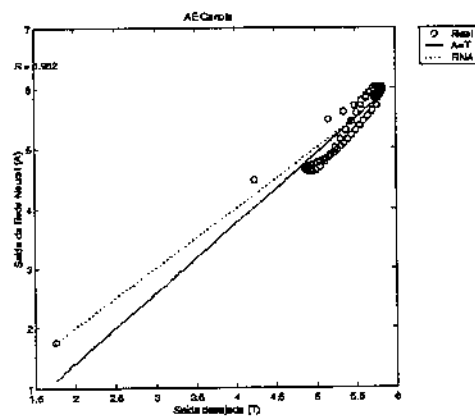
(a)



(b)



(c)



(d)

Figura 50. Protótipo I - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de validação.

A variação com o número de neurônios escondidos do número de épocas e dos parâmetros de treinamento usados no algoritmo de Levenberg-Marquardt associado com regularização bayesiana (*sse*, *mse*, *ssx*, *np* e *gamk*) pode ser analisada através da tabela 30 e da figura 51.

Segundo Foresee e Hagan (1997), a comparação após o treinamento do número total de parâmetros da rede (*np*) com o número dos parâmetros efetivamente utilizados (*gamk*) permite avaliar se o número de neurônios escondidos é suficiente para representar realmente a função. Se o número efetivo de parâmetros (*gamk*) for muito próximo do número total de parâmetros da rede (*np*), provavelmente a rede não é grande o suficiente para representar a função, sendo necessário aumentar o número de neurônios escondidos e treinar novamente a rede. Se a rede maior tiver o mesmo número de parâmetros efetivos (*gamk*), então o número de neurônios escondidos da rede anterior era suficiente, senão mais neurônios devem ser adicionados à camada escondida. Outra análise sugerida por Foresee e Hagan (1997) para verificar a consistência dos resultados é observar o número de neurônios escondidos a partir dos quais os parâmetros de treinamento *sse*, *ssw* e *gamk* começam a se estabilizar. O problema com estes critérios é que nem sempre a melhor rede selecionada usando parâmetros de treinamento é a que apresenta maior capacidade de generalização quando dados não usados no treinamento são apresentados a mesma.

Os resultados apresentados na tabela 30 e na figura 51, assim como as análises das tabelas e principalmente dos gráficos dos índices de avaliação de desempenho do protótipo I usando os conjuntos de treinamento e de validação em função do número de neurônios escondidos (Tabelas 28 e 29 e Figuras 43 a 46) foram usados para confirmar o número ótimo de neurônios na camada de saída. Na validação do protótipo, a topologia com menor número de neurônios escondidos, a partir da qual o erro começou a estabilizar e que apresentou maior coeficiente de determinação global (R_g^2) foi selecionada para ser usada na implementação *on-line* do *softsensor*.

Na tabela 30 e no gráfico 51 pode-se observar que a partir de 7 neurônios escondidos, os parâmetros de treinamento *sse*, *ssx* e *gamk* começam a estabilizar. Entretanto, este número de neurônios escondidos não correspondeu ao número de neurônios

que na validação, apresentou maior coeficiente de generalização e maior coeficiente de robustez.

Portanto, a seleção do número ótimo de neurônios de um determinado protótipo foi sempre feita usando o conjunto de validação. Os maiores coeficientes de generalização e robustez muitas vezes estão associados a altos números de neurônios escondidos, o que pode indicar sobreajuste da rede. Portanto, a topologia determinada como a melhor, foi aquela que apresentou na validação os maiores coeficientes de generalização e robustez e menor erro para o menor número de neurônios escondidos possível.

A figura 52 apresenta a evolução dos parâmetros de treinamento *sse*, *ssx* e *gamk* em função do número de épocas, para a topologia 4-4-4, selecionada como a melhor do protótipo I.

Tabela 30. Protótipo I - Parâmetros de treinamento da rede (*época*, *sse*, *mse*, *ssx*, *np* e *gamk*) em função do número de neurônios (*SI*) na camada escondida.

<i>SI</i>	<i>época</i>	<i>sse</i>	<i>mse</i>	<i>ssx</i>	<i>np</i>	<i>gamk</i>
1	72	1,86E+00	4,64E-03	4,06E+01	13	1,16E+01
2	98	7,87E-01	1,97E-03	9,15E+01	22	1,82E+01
3	116	2,46E-01	6,14E-04	3,26E+02	31	2,83E+01
4	803	4,96E-02	1,24E-04	3,95E+03	40	3,64E+01
5	548	6,58E-03	1,65E-05	1,40E+03	49	4,56E+01
6	1000	1,24E-03	3,09E-06	1,38E+03	58	5,45E+01
7	579	9,97E-04	2,49E-06	1,53E+03	67	6,28E+01
8	440	9,98E-04	2,49E-06	7,95E+02	76	6,76E+01
9	201	9,89E-04	2,47E-06	5,51E+02	85	6,75E+01
10	277	9,95E-04	2,49E-06	7,27E+02	94	7,24E+01
11	267	9,94E-04	2,48E-06	5,96E+02	103	7,11E+01
12	189	9,98E-04	2,49E-06	5,21E+02	112	7,25E+01
13	316	9,98E-04	2,50E-06	6,61E+02	121	7,26E+01
14	148	9,92E-04	2,48E-06	5,34E+02	130	7,37E+01
15	121	9,85E-04	2,46E-06	6,08E+02	139	7,63E+01
16	214	9,91E-04	2,48E-06	5,84E+02	148	7,14E+01
17	165	9,95E-04	2,49E-06	6,30E+02	157	7,82E+01
18	115	9,75E-04	2,44E-06	6,56E+02	166	7,86E+01
19	112	9,85E-04	2,46E-06	6,78E+02	175	7,93E+01
20	112	9,93E-04	2,48E-06	6,23E+02	184	7,61E+01

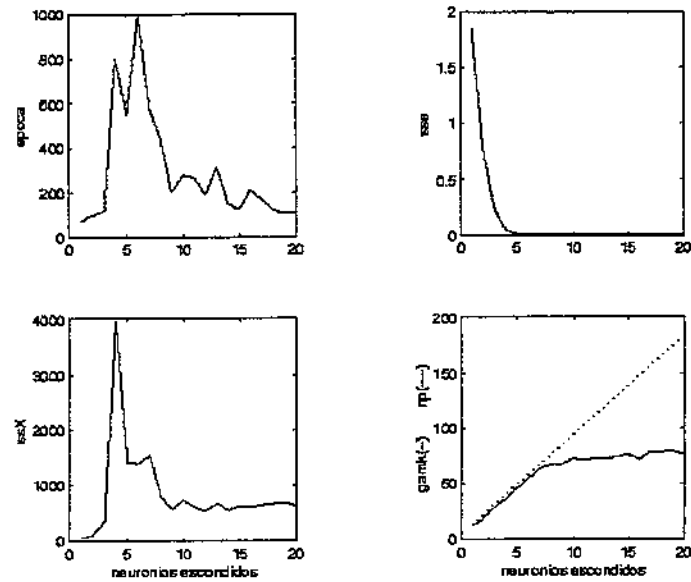


Figura 51. Protótipo I - Época e parâmetros de treinamento (sse , ssx e $gamk$) em função do número de neurônios escondidos.

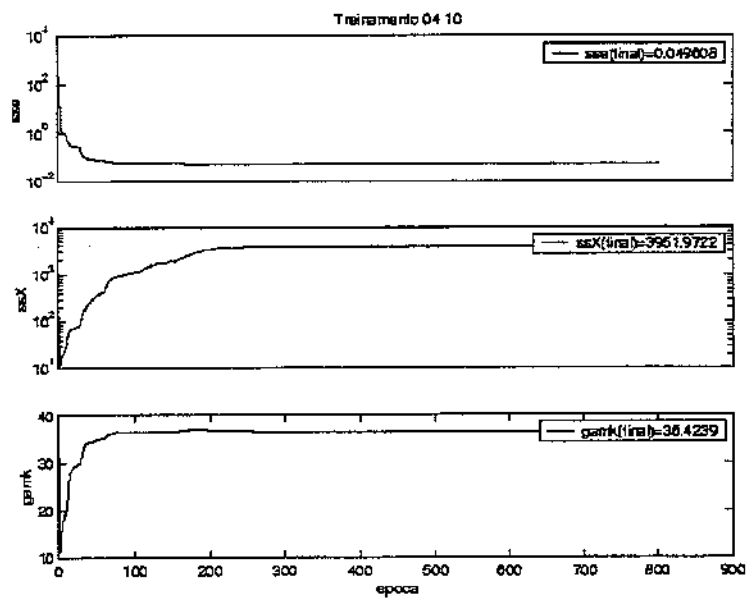


Figura 52. Protótipo I - Parâmetros de treinamento (sse , ssx e $gamk$) em função do número de épocas.

A topologia 4-4-4 do protótipo I foi capaz de fornecer boas estimativas da concentração de biomassa e das atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola, usando o conjunto de validação. Logo, um *softsensor* neural completamente especificado para estimação das variáveis de estado pode ser implementado *on-line* usando as matrizes de pesos (IW^1 e LW^2) e bias ($b1$ e $b2$) da camada escondida e da camada de saída apresentadas nas tabelas 31 e 32. O *softsensor* neural I pode ser representado pela seguinte equação :

$$a = \text{purelin}(LW^2 \logsig(IW^1 p + b1) + b2) \quad (27)$$

onde: a - é a matriz dos padrões de saída do *softsensor* I

purelin - é a função de transferência linear usada na camada de saída

$\logsig$ - é a função de transferência sigmoidal usada na camada escondida

p - é a matriz dos padrões de entrada do *softsensor* I

Tabela 31. Matrizes de pesos e bias da camada escondida do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo I com 4 neurônios na camada escondida).

Camada escondida				
Matriz de pesos (IW^1)				bias (b^1)
-7,35E+00	-3,37E-01	2,26E+00	-7,34E+00	8,00E+00
-5,54E+00	-1,77E+01	-6,75E-01	4,21E+00	3,16E-01
6,22E+00	4,73E-01	-4,04E+00	4,81E+00	-5,16E+00
-6,17E+00	4,08E+00	-8,97E-01	4,64E+00	-1,75E+00

Tabela 32. Matrizes de pesos e bias da camada de saída do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo I com 4 neurônios na camada escondida).

Camada de saída				
Matriz de pesos (LW^2)				bias (b^2)
-1,22E+00	-1,17E+00	-1,67E+00	1,17E+00	8,98E-01
2,80E+00	2,92E+00	3,44E-01	-2,88E+00	2,33E+00
4,82E-01	-1,51E+01	-1,82E+00	1,36E+01	-1,03E+00
-2,92E+00	8,46E+00	-2,01E+00	-8,11E+00	-9,38E-01

5.3.1.2. Protótipo II – *Softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola

Como foi visto, anteriormente, durante a ampliação do processo de produção de bioemulsificante por *Candida lipolytica*, todas as variáveis que demonstraram influenciar a concentração de biomassa e as atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola foram incluídas no padrão de entrada do protótipo I. Portanto, a configuração do protótipo I consistiu de 4 entradas (pH, oxigênio dissolvido do protótipo I, densidade ótica e salinidade do líquido metabólico livre de células), *SI* neurônios escondidos e 4 saídas (biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola). No protótipo II uma redução do número de variáveis de entrada foi realizada como objetivo de desenvolver um protótipo cujas entradas fossem formadas apenas por variáveis medidas *on-line*. Portanto, a configuração do protótipo II consistiu de 2 entradas (pH e oxigênio dissolvido), *SI* neurônios escondidos e 4 saídas (biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola). Essa redução de variáveis de entrada além de reduzir o número de parâmetros do *softsensor*, permitiu analisar o grau de influência da densidade ótica e da salinidade do líquido metabólico livre de células sobre o conjunto das variáveis do padrão de saída.

Para verificar a capacidade de predição do protótipo II, topologias com número de neurônios escondidos (*SI*) variando de 1 a 20, foram treinadas e validadas 10 vezes cada uma. Para estudar o efeito do número de padrões de entrada-saída sobre seu desempenho, o protótipo II foi treinado, validado e testado com conjuntos constituídos por 100 e 1000 padrões de entrada-saída. Os resultados obtidos são analisados e discutidos a seguir para os casos de apresentação de 100 ou 1000 padrões de entrada ao protótipo II.

5.3.1.2.1. Protótipo II _ 100 padrões

Vinte diferentes topologias do protótipo II foram avaliadas usando conjunto de treinamento composto por 100 padrões de entrada-saída. Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando conjunto de treinamento estão apresentados na tabela 33.

Tabela 33. Protótipo II (100 padrões) - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de treinamento.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	2,21E+00	1,11E-02	1,05E-01	7,47E-04	1,05E-01	1,72E+01	7,580181E-01	5,745915E-01
2	8,64E-01	4,32E-03	6,57E-02	4,67E-04	6,11E-02	1,05E+01	9,082734E-01	8,249606E-01
3	3,60E-01	1,80E-03	4,24E-02	3,02E-04	4,62E-02	7,34E+00	9,642023E-01	9,296862E-01
4	1,50E-01	7,52E-04	2,74E-02	1,95E-04	3,02E-02	3,85E+00	9,922619E-01	9,845836E-01
5	3,49E-02	1,75E-04	1,32E-02	9,39E-05	1,38E-02	1,75E+00	9,965959E-01	9,932034E-01
6	1,54E-02	7,71E-05	8,78E-03	6,24E-05	9,42E-03	1,19E+00	9,982344E-01	9,964719E-01
7	1,33E-02	6,63E-05	8,14E-03	5,78E-05	8,49E-03	1,15E+00	9,988087E-01	9,976188E-01
8	1,41E-03	7,03E-06	2,65E-03	1,88E-05	2,87E-03	3,76E-01	9,992718E-01	9,985441E-01
9	1,00E-03	5,00E-06	2,24E-03	1,59E-05	2,36E-03	3,18E-01	9,996733E-01	9,993467E-01
10	1,19E-03	5,95E-06	2,44E-03	1,73E-05	2,47E-03	3,26E-01	9,997892E-01	9,995785E-01
11	2,24E-03	1,12E-05	3,35E-03	2,38E-05	3,32E-03	4,24E-01	9,998463E-01	9,996926E-01
12	9,98E-04	4,99E-06	2,23E-03	1,59E-05	2,24E-03	2,87E-01	9,999149E-01	9,998299E-01
13	1,00E-03	5,00E-06	2,24E-03	1,59E-05	2,31E-03	2,96E-01	9,998992E-01	9,997985E-01
14	9,95E-04	4,98E-06	2,23E-03	1,58E-05	2,29E-03	2,97E-01	9,999044E-01	9,998087E-01
15	9,98E-04	4,99E-06	2,23E-03	1,59E-05	2,20E-03	2,84E-01	9,999164E-01	9,998328E-01
16	9,93E-04	4,97E-06	2,23E-03	1,58E-05	2,26E-03	2,94E-01	9,999160E-01	9,998321E-01
17	9,98E-04	4,99E-06	2,23E-03	1,59E-05	2,20E-03	2,90E-01	9,999170E-01	9,998341E-01
18	9,98E-04	4,99E-06	2,23E-03	1,59E-05	2,12E-03	2,72E-01	9,999182E-01	9,998363E-01
19	9,97E-04	4,98E-06	2,23E-03	1,59E-05	2,22E-03	2,86E-01	9,999187E-01	9,998373E-01
20	9,90E-04	4,95E-06	2,23E-03	1,58E-05	2,11E-03	2,72E-01	9,999184E-01	9,998368E-01

Entre as 20 topologias testadas para o protótipo II, a topologia com 19 neurônios na camada escondida foi a que apresentou maior número de simulações, com o conjunto de treinamento, com maiores coeficientes de correlação para as 4 variáveis de saída simultaneamente (*medmed_R*=0,9999187 e R_g^2 = 0,9998373).

Conforme se pode observar através das figuras 53 e 54, as trajetórias das funções de erro são suaves e apresentam um rápido declínio do erro com o aumento dos neurônios escondidos, começando a se estabilizar a partir de 5 neurônios. A curva do coeficiente de determinação global apresenta um comportamento oposto (Figura 54b).

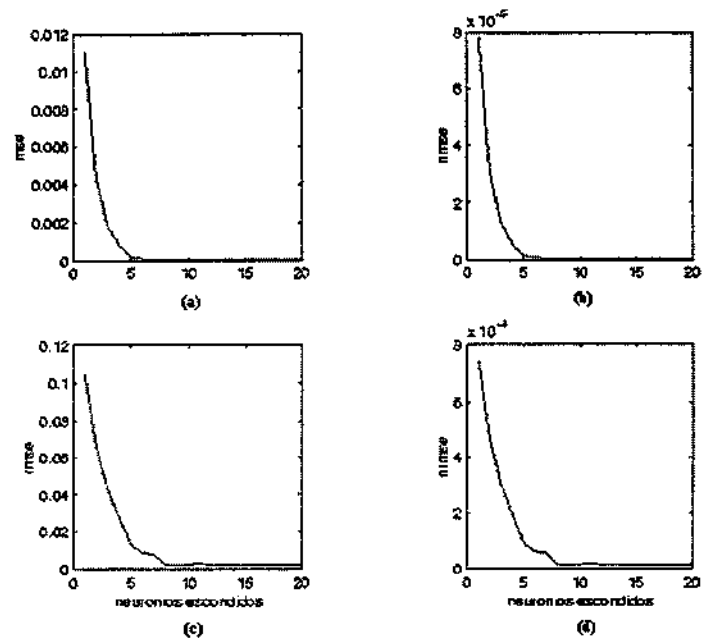


Figura 53. Protótipo II (100 padrões) - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nrmse$ usando conjunto de treinamento em função do número de neurônios escondidos.

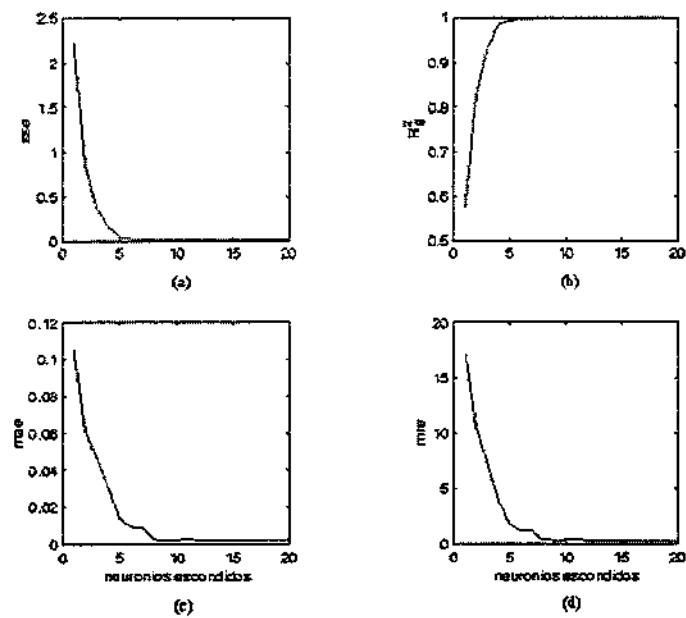


Figura 54. Protótipo II (100 padrões) - Índices de avaliação do desempenho (a) sse (b) Rg^2 (c) mae e (d) mre usando conjunto de treinamento em função do número de neurônios escondidos.

Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação estão apresentados na tabela 34.

Tabela 34. Protótipo II (100 padrões) - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida (*SI*), obtidos usando o conjunto de validação.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	3,71E+00	1,86E-02	1,36E-01	9,60E-04	1,31E-01	2,32E+01	6,705957E-01	4,496986E-01
2	1,61E+00	8,04E-03	8,97E-02	6,32E-04	8,78E-02	1,57E+01	8,812312E-01	7,765684E-01
3	9,10E-01	4,55E-03	6,75E-02	4,76E-04	7,40E-02	1,23E+01	9,445304E-01	8,921377E-01
4	6,82E-01	3,41E-03	5,84E-02	4,12E-04	6,37E-02	9,47E+00	9,738704E-01	9,484235E-01
5	4,83E-01	2,41E-03	4,91E-02	3,46E-04	4,94E-02	7,53E+00	9,792347E-01	9,589005E-01
6	4,56E-01	2,28E-03	4,78E-02	3,37E-04	4,63E-02	7,14E+00	9,815547E-01	9,634497E-01
7	4,56E-01	2,28E-03	4,77E-02	3,37E-04	4,70E-02	7,32E+00	9,831673E-01	9,666179E-01
8	4,59E-01	2,29E-03	4,79E-02	3,38E-04	4,95E-02	7,58E+00	9,821329E-01	9,645851E-01
9	4,01E-01	2,01E-03	4,48E-02	3,16E-04	4,55E-02	7,01E+00	9,837770E-01	9,678172E-01
10	5,52E-01	2,76E-03	5,25E-02	3,71E-04	5,59E-02	8,28E+00	9,821694E-01	9,646566E-01
11	5,29E-01	2,64E-03	5,14E-02	3,63E-04	5,31E-02	8,06E+00	9,828450E-01	9,659843E-01
12	4,63E-01	2,32E-03	4,81E-02	3,39E-04	4,76E-02	7,40E+00	9,832326E-01	9,667464E-01
13	5,14E-01	2,57E-03	5,07E-02	3,57E-04	5,10E-02	7,80E+00	9,810844E-01	9,625267E-01
14	4,96E-01	2,48E-03	4,98E-02	3,51E-04	5,13E-02	7,86E+00	9,831169E-01	9,665189E-01
15	4,87E-01	2,44E-03	4,94E-02	3,48E-04	4,97E-02	7,66E+00	9,832209E-01	9,667233E-01
16	4,29E-01	2,14E-03	4,63E-02	3,26E-04	4,44E-02	6,91E+00	9,823045E-01	9,649222E-01
17	5,75E-01	2,88E-03	5,36E-02	3,78E-04	5,51E-02	8,32E+00	9,822645E-01	9,648436E-01
18	4,76E-01	2,38E-03	4,88E-02	3,44E-04	4,89E-02	7,53E+00	9,824401E-01	9,651885E-01
19	4,63E-01	2,32E-03	4,81E-02	3,39E-04	4,94E-02	7,60E+00	9,828030E-01	9,659017E-01
20	4,54E-01	2,27E-03	4,76E-02	3,36E-04	4,76E-02	7,34E+00	9,831366E-01	9,665576E-01

Entre as 20 topologias testadas para o protótipo II, a topologia com 9 neurônios na camada escondida (topologia 2-9-4) foi a que apresentou maior número de simulações, com o conjunto de validação, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9837770 e R_g^2 =0,9678172). Entretanto, a topologia contendo 7 neurônios na camada escondida (topologia 2-7-4) – também poderia ser a de 6 - foi selecionada como a melhor por ser mais simples (menor número de neurônios) e apresentar capacidade de generalização e robustez (*medmed_R*=0,9831673 e R_g^2 =0,9666179) praticamente iguais as da rede com 9 neurônios na camada escondida.

Através das figuras 55 e 56, percebe-se que as curvas das funções de erro do protótipo II usando o conjunto de validação apresentam um rápido declínio do erro com o aumento dos neurônios escondidos e começam a se estabilizar a partir de 5 neurônios. Neste caso, as trajetórias das funções de erro do protótipo II usando o conjunto de validação não são tão suaves quanto às obtidas usando o conjunto de treinamento. Esta é uma situação onde aumento do número de neurônios escondido não melhora a capacidade preditiva da rede,

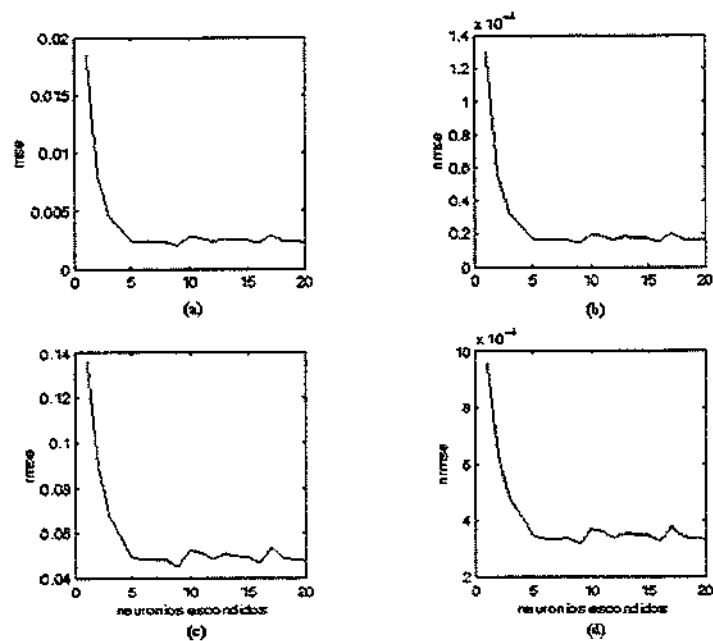


Figura 55. Protótipo II (100 padrões) - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nrmse$ usando conjunto de validação em função do número de neurônios escondidos

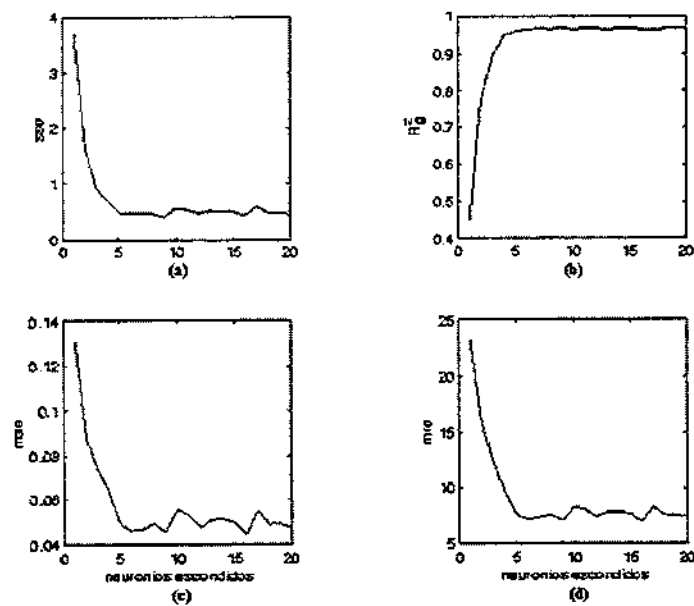


Figura 56. Protótipo II (100 padrões) - Índices de avaliação do desempenho (a) sse (b) Rg^2 (c) mae e (d) mre usando conjunto de validação em função do número de neurônios escondidos.

possivelmente devido ao mapeamento de não linearidades de alta ordem, que podem ser consideradas como ruídos.

As figuras 57 e 58 permitem observar que as curvas previstas pelo protótipo II com topologia 2-7-4 para as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola ajustaram-se quase perfeitamente aos dados do conjunto de treinamento (coeficientes de correlação R aproximadamente iguais a 1 e $R_g^2=0,9976181$). Conforme se pode observar através das figuras 59 e 60, as curvas previstas pelo protótipo II com topologia 2-7-4 para as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola também apresentaram bom ajuste aos dados do conjunto de validação (coeficientes de correlação R superiores a 0,96 e $R_g^2=0,9666179$).

Como foi visto anteriormente, a eliminação das variáveis - densidade ótica e salinidade do líquido metabólico livre de células - do protótipo I resultou no protótipo II. A comparação dos desempenhos dos dois protótipos (comparação da tabela 29 com a tabela 34, dos gráficos 45 e 46 com os gráficos 55 e 56 e dos gráficos 49 e 50 com os gráficos 59 e 60) permite observar que a remoção das variáveis - densidade ótica e salinidade do líquido metabólico livre de células - do padrão de entrada do protótipo I, resultou em uma pequena diminuição dos erros (*sse*, *mse*, *nmse*, *rmse*, *nrmse*, *mae* e *mre*) e em pequeno aumento dos coeficientes de generalização e robustez.

Este resultado ratifica, portanto, a conclusão tirada anteriormente sobre a adequação das entradas do protótipo I e sobre a existência de ruídos nos dados de densidade ótica e salinidade do líquido metabólico livre de células. Este resultado também ratifica a robustez do protótipo I com topologia 4-4-4, pois mesmo trabalhando com duas variáveis com ruído provenientes de análises *off-line* manteve sua capacidade de generalização e apresentou excelente ajuste aos dados do conjunto de validação. Também ficou demonstrado que o aumento do número de variáveis no padrão de entrada não necessariamente melhora a precisão da rede neural.

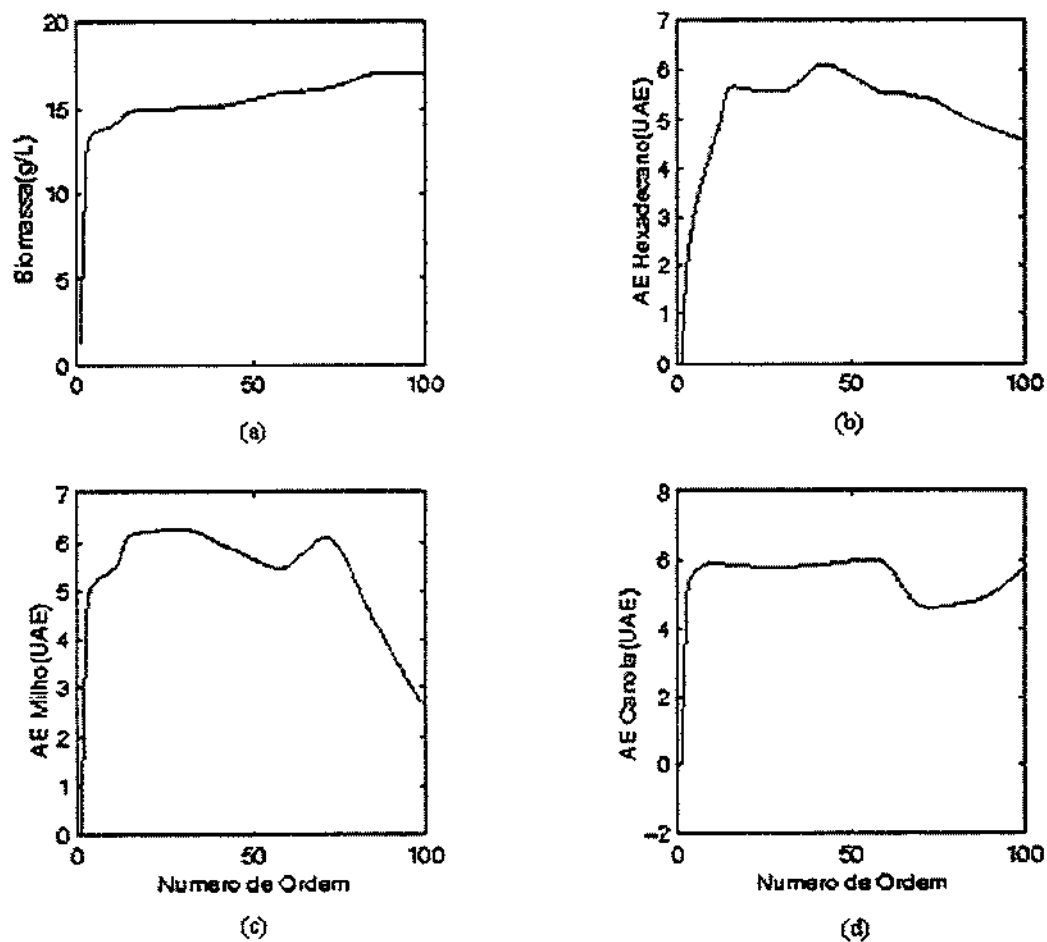


Figura 57. Protótipo II (100 padrões) - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de treinamento.

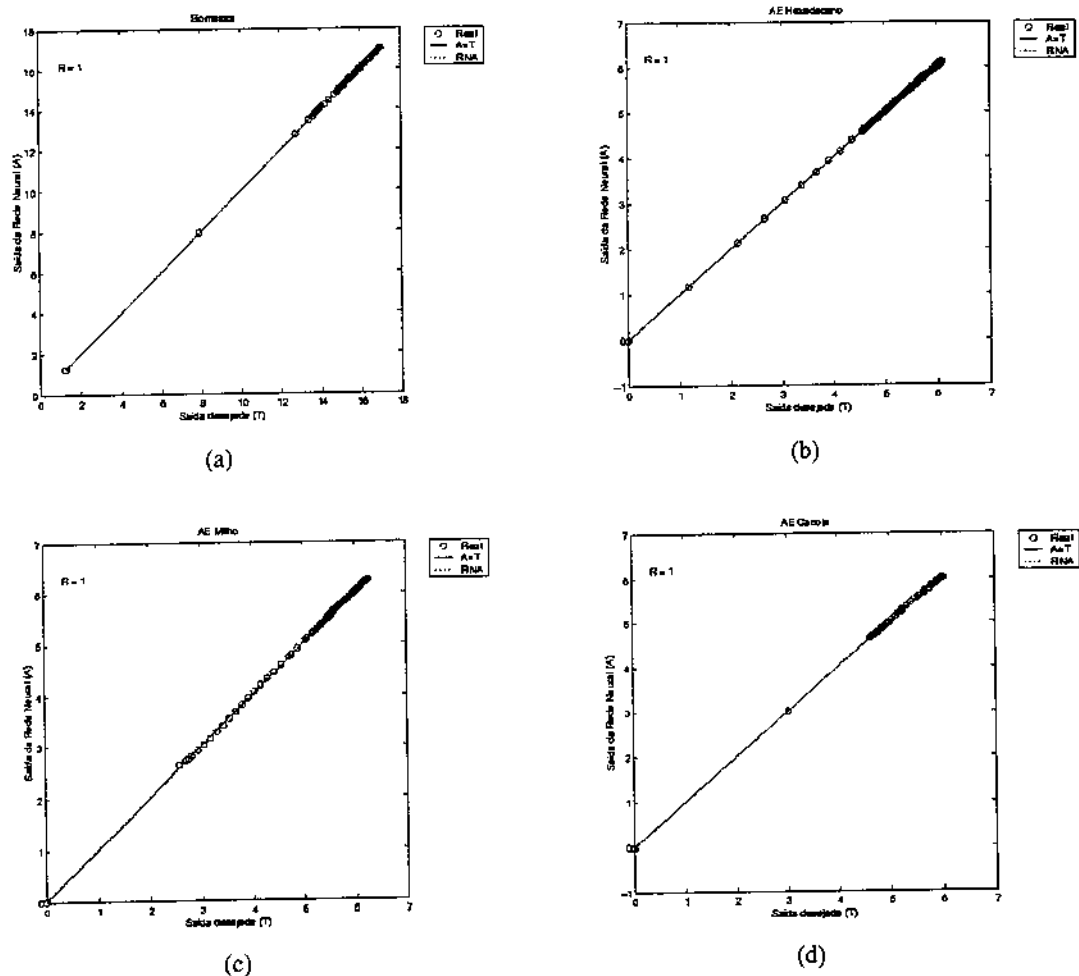


Figura 58. Protótipo II (100 padrões) - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de treinamento.

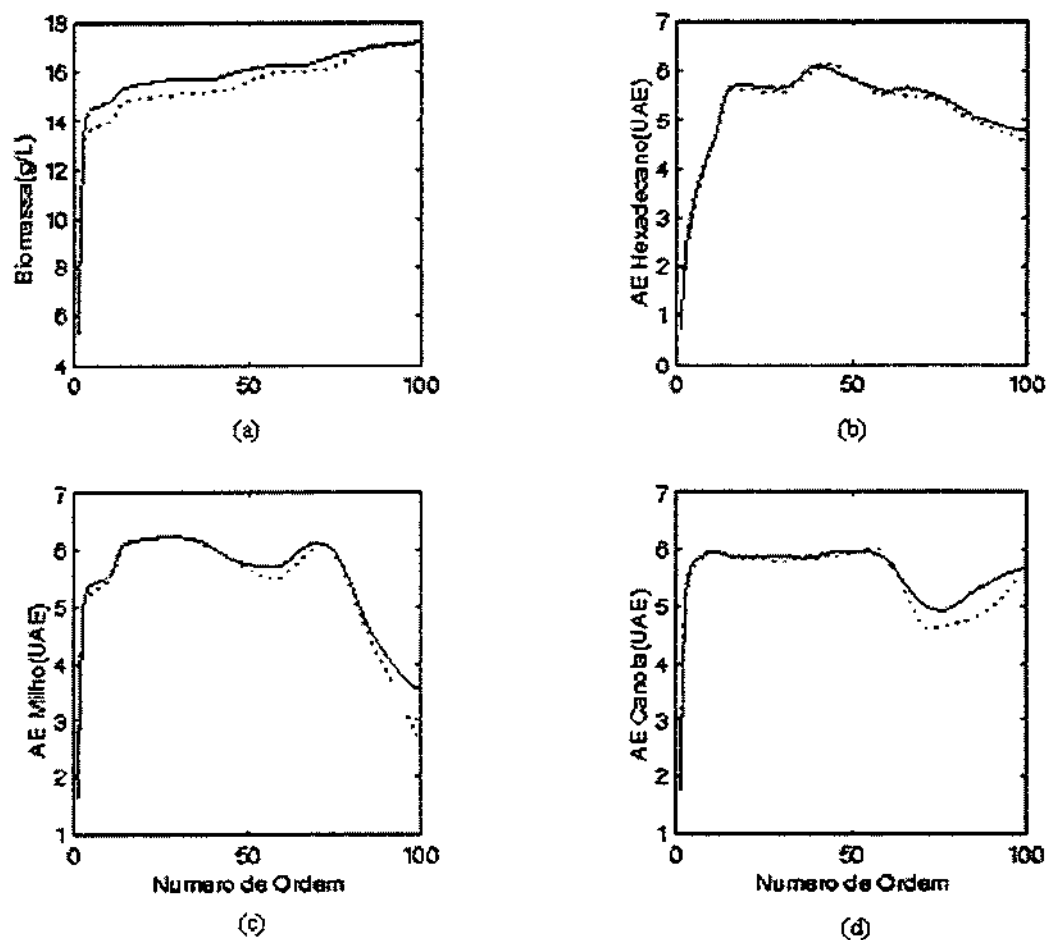
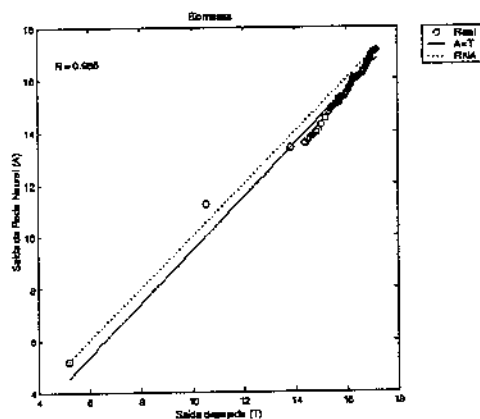
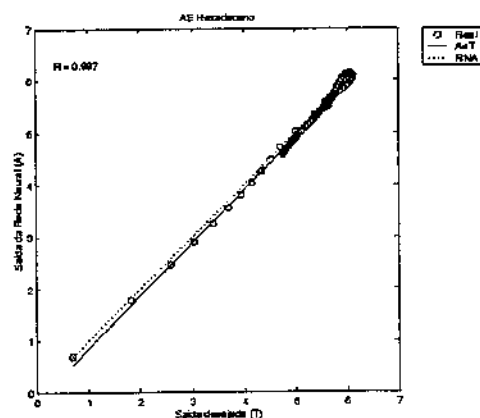


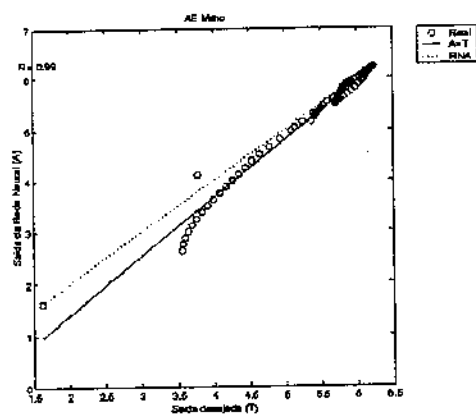
Figura 59. Protótipo II (100 padrões) - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de validação.



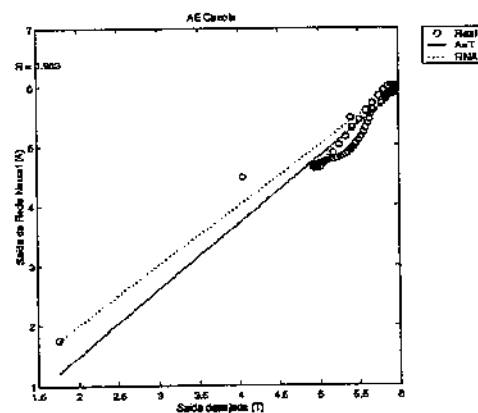
(a)



(b)



(c)



(d)

Figura 60. Protótipo II (100 padrões) - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de validação.

A variação do número de épocas e dos parâmetros de treinamento (*sse*, *mse*, *ssx*, *np* e *gamk*) com o número de neurônios escondidos (*SI*) pode ser analisada através da tabela 35 e da figura 61.

O caso em estudo é um exemplo de uma situação que demonstra a necessidade de outros critérios para seleção da melhor topologia para uma determinada configuração de rede, diferentes daqueles sugeridos por Foresee e Hagan (1997). A determinação da melhor topologia do protótipo II - através da observação do número de neurônios escondidos a partir do qual o número dos parâmetros efetivamente utilizados (*gamk*) começa a se estabilizar - sugerida por Foresee e Hagan (1997) levaria a seleção de uma topologia com 15 ou 16 neurônios escondidos.

A outra análise sugerida por Foresee e Hagan (1997) de observar o número de neurônios escondidos a partir dos quais os parâmetros de treinamento *sse*, *ssw* e *gamk* começam a se estabilizar, levaria a seleção de uma topologia com 5 neurônios caso se considerasse apenas a estabilização do parâmetro *sse*. Entretanto, a análise do gráfico do parâmetro *ssx* versus o número de neurônios escondidos mostra que o parâmetro *ssx* não estabiliza no intervalo de 1 a 20 neurônios.

Outro problema - com a utilização das sugestões de Foresee e Hagan (1997) para determinar melhor topologia para uma determinada configuração de rede - é que nem sempre a melhor rede selecionada usando parâmetros de treinamento é a que apresenta maior capacidade de generalização quando testada usando dados de um conjunto de validação.

A figura 62 apresenta a evolução dos parâmetros de treinamento *sse*, *ssx* e *gamk* em função do número de épocas, para a topologia 2-7-4. A topologia 2-7-4 foi identificada como a melhor do protótipo II, usando o critério de selecionar como melhor topologia aquela com menor número de neurônios escondidos, a partir da qual o erro começou a estabilizar e que apresentou maior coeficiente de determinação global (R_g^2) usando o conjunto de validação.

Tabela 35. Protótipo II (100 padrões) - Parâmetros de treinamento da rede (*época*, *sse*, *mse*, *ssx*, *np* e *gamk*) em função do número de neurônios (*SI*) na camada escondida.

<i>SI</i>	<i>época</i>	<i>sse</i>	<i>mse</i>	<i>ssx</i>	<i>np</i>	<i>gamk</i>
1	66	2,21E+00	1,11E-02	1,82E+02	11	10,76
2	130	8,64E-01	4,32E-03	5,10E+02	18	15,82
3	149	3,60E-01	1,80E-03	1,97E+03	25	22,58
4	122	1,50E-01	7,52E-04	2,46E+03	32	28,87
5	391	3,49E-02	1,75E-04	1,85E+03	39	36,19
6	791	1,54E-02	7,71E-05	2,01E+03	46	41,82
7	458	1,33E-02	6,63E-05	1,09E+04	53	47,38
8	1000	1,41E-03	7,03E-06	1,08E+04	60	55,59
9	984	1,00E-03	5,00E-06	1,18E+04	67	59,79
10	1000	1,19E-03	5,95E-06	1,09E+04	74	66,80
11	1000	2,24E-03	1,12E-05	6,17E+03	81	70,02
12	546	9,98E-04	4,99E-06	1,08E+04	88	72,38
13	647	1,00E-03	5,00E-06	9,03E+03	95	75,33
14	814	9,95E-04	4,98E-06	7,77E+03	102	81,20
15	973	9,98E-04	4,99E-06	7,12E+03	109	90,44
16	358	9,93E-04	4,97E-06	8,47E+03	116	84,63
17	641	9,98E-04	4,99E-06	9,81E+03	123	89,42
18	243	9,98E-04	4,99E-06	1,04E+04	130	89,67
19	511	9,97E-04	4,98E-06	1,06E+04	137	90,46
20	523	9,90E-04	4,95E-06	8,54E+03	144	96,90

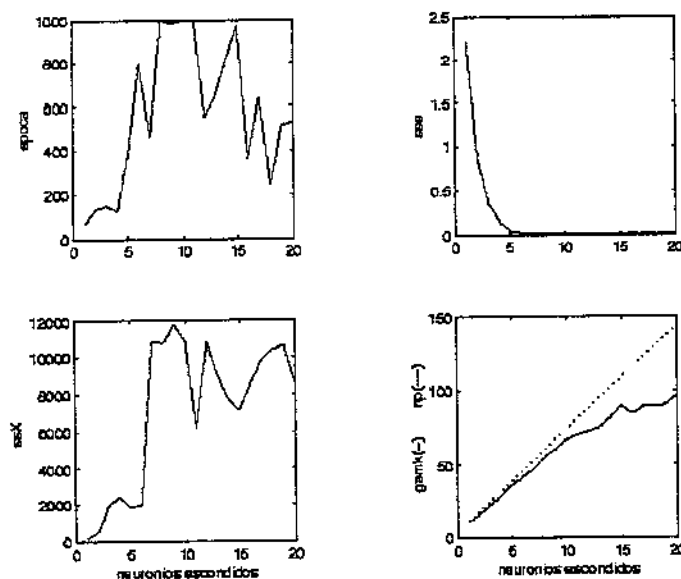


Figura 61. Protótipo II (100 padrões) - Época e parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de neurônios escondidos.

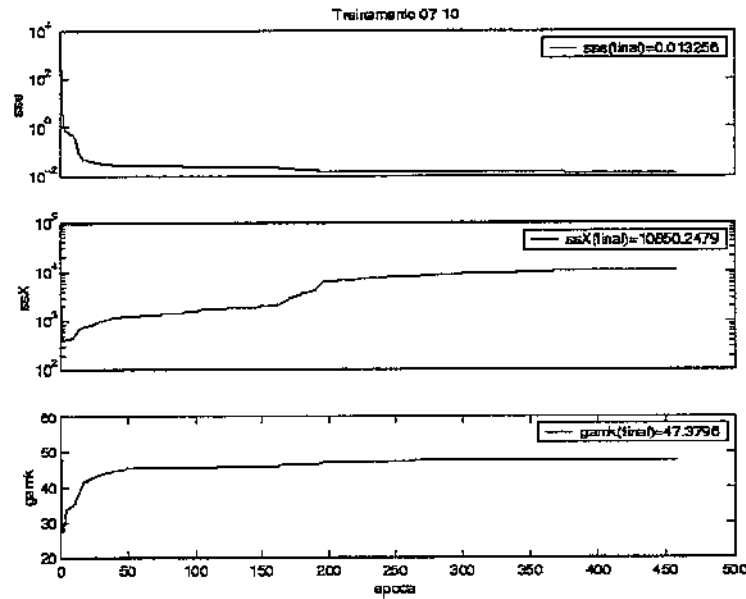


Figura 62. Protótipo II (100 padrões) - Parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de épocas.

A topologia 2-7-4 do protótipo II foi capaz de fornecer excelentes estimativas da concentração de biomassa e das atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola, usando o conjunto de validação. Logo, um *softsensor* neural completamente especificado para estimação das variáveis de estado pode ser implementado *on-line* usando as matrizes de pesos (IW^1 e LW^2) e bias ($b1$ e $b2$) da camada escondida e da camada de saída apresentadas nas tabelas 36 e 37. O *softsensor* neural II pode ser representado pela seguinte equação :

$$a = \text{purelin}(LW^2 \logsig(IW^1 p + b1) + b2) \quad (28)$$

onde: a - é a matriz dos padrões de saída do *softsensor* II
 purelin - é a função de transferência linear usada na camada de saída
 $\logsig$ - é a função de transferência sigmoideal usada na camada escondida
 p - é a matriz dos padrões de entrada do *softsensor* II

Tabela 36. Matrizes de pesos e bias da camada escondida do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo II com 7 neurônios na camada escondida).

Camada escondida		
Matriz de pesos (IW^d)		bias (b^d)
5,28E+00	-7,13E-01	-2,06E+00
8,83E+00	-3,40E+00	-7,85E-01
-6,19E+00	-9,17E-02	6,83E+00
-5,90E+01	1,36E+00	2,16E+01
-7,96E+00	-1,33E+01	1,15E+00
5,84E+00	-2,04E+01	1,05E+01
-1,01E+01	7,24E-01	4,35E+00

Tabela 37. Matrizes de pesos e bias da camada de saída do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo II com 7 neurônios na camada escondida).

Camada de saída							
Matriz de pesos (LW^s)							bias (b^s)
-3,01E-03	1,45E-01	5,62E-01	4,22E-02	5,87E-01	6,40E-01	-4,62E-02	-5,13E-01
4,33E+00	-2,90E+00	5,16E+00	-1,41E-01	-3,63E+00	1,12E-01	2,04E+00	-4,86E+00
3,98E-01	-7,24E+00	1,87E+00	-1,15E-01	-1,49E+01	6,82E-01	-2,42E-01	5,40E+00
6,23E-01	4,99E+00	1,43E-01	8,18E-02	9,43E+00	7,41E-01	5,71E-01	-5,54E+00

5.3.1.2.2. Protótipo II - 1000 padrões

A qualidade e a quantidade dos dados do conjunto de treinamento são capazes de afetar o desempenho de *softsensores* baseados em redes neurais artificiais. Resultados referentes ao protótipo I e ao protótipo II (treinado com 100 padrões de entrada-saída), discutidos anteriormente, comprovaram a influência da qualidade dos dados do conjunto de treinamento sobre o desempenho de *softsensores* neurais. Para investigar o efeito do tamanho do conjunto de treinamento sobre o desempenho de *softsensores* neurais, vinte diferentes topologias do protótipo II foram avaliadas usando conjunto de treinamento composto por 1000 padrões de entrada-saída. Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando conjunto de treinamento estão apresentados na tabela 38.

Tabela 38. Protótipo II (1000 padrões) - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de treinamento.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	2,04E+01	1,02E-02	1,01E-01	7,15E-04	9,84E-02	1,52E+01	6,699446E-01	4,488257E-01
2	7,73E+00	3,87E-03	6,22E-02	4,40E-04	6,00E-02	9,05E+00	8,867078E-01	7,862507E-01
3	3,39E+00	1,70E-03	4,12E-02	2,92E-04	3,30E-02	5,56E+00	9,569759E-01	9,158029E-01
4	5,99E-01	3,00E-04	1,73E-02	1,23E-04	1,84E-02	2,40E+00	9,917160E-01	9,835006E-01
5	2,52E-01	1,26E-04	1,12E-02	7,96E-05	1,24E-02	1,60E+00	9,955312E-01	9,910824E-01
6	1,57E-01	7,83E-05	8,85E-03	6,27E-05	9,51E-03	1,24E+00	9,976385E-01	9,952826E-01
7	7,57E-02	3,78E-05	6,15E-03	4,36E-05	6,03E-03	8,36E-01	9,983366E-01	9,966760E-01
8	1,07E-01	5,35E-05	7,31E-03	5,18E-05	7,52E-03	1,02E+00	9,989742E-01	9,979495E-01
9	3,85E-02	1,92E-05	4,39E-03	3,11E-05	4,60E-03	6,01E-01	9,993313E-01	9,986631E-01
10	1,65E-02	8,24E-06	2,87E-03	2,03E-05	3,13E-03	4,07E-01	9,996614E-01	9,993228E-01
11	2,47E-02	1,23E-05	3,51E-03	2,49E-05	3,25E-03	4,23E-01	9,998393E-01	9,996787E-01
12	1,50E-02	7,48E-06	2,73E-03	1,94E-05	3,02E-03	3,95E-01	9,999056E-01	9,998112E-01
13	2,20E-02	1,10E-05	3,32E-03	2,35E-05	3,44E-03	4,54E-01	9,999048E-01	9,998097E-01
14	3,83E-03	1,92E-06	1,38E-03	9,80E-06	1,37E-03	1,89E-01	9,999552E-01	9,999105E-01
15	2,24E-03	1,12E-06	1,06E-03	7,49E-06	1,07E-03	1,39E-01	9,999583E-01	9,999165E-01
16	1,95E-03	9,74E-07	9,87E-04	6,99E-06	9,17E-04	1,15E-01	9,999578E-01	9,999156E-01
17	3,71E-03	1,85E-06	1,36E-03	9,64E-06	1,46E-03	1,85E-01	9,999703E-01	9,999406E-01
18	2,01E-03	1,00E-06	1,00E-03	7,09E-06	9,99E-04	1,35E-01	9,999737E-01	9,999474E-01
19	3,90E-03	1,95E-06	1,40E-03	9,88E-06	1,44E-03	1,92E-01	9,999696E-01	9,999391E-01
20	1,82E-03	9,08E-07	9,53E-04	6,75E-06	9,25E-04	1,21E-01	9,999797E-01	9,999594E-01

Entre as 20 topologias testadas para o protótipo II, a topologia com 20 neurônios na camada escondida foi a que apresentou maior número de simulações, com o conjunto de treinamento, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9999187 e R_g^2 = 0,9998373).

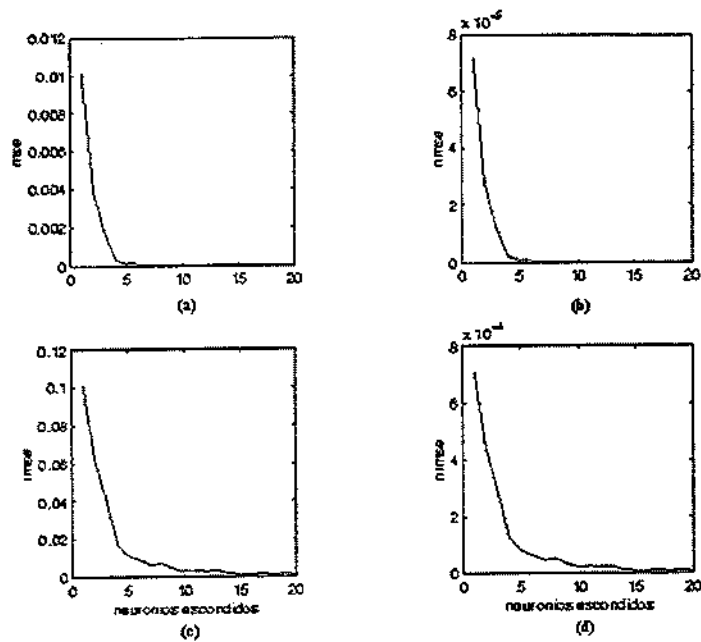


Figura 63. Protótipo II (1000 padrões) - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nrmse$ usando conjunto de treinamento em função do número de neurônios escondidos.

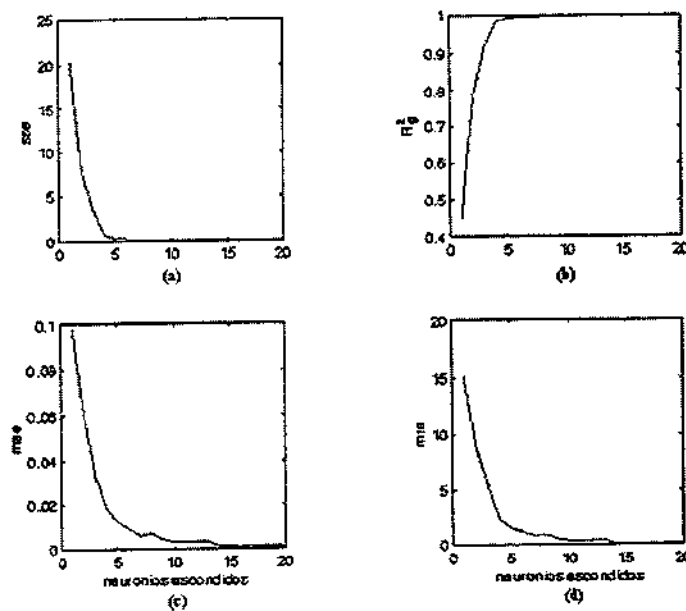


Figura 64. Protótipo II (1000 padrões) - Índices de avaliação do desempenho (a) sse (b) Rg^2 (c) mae e (d) mre usando conjunto de treinamento em função do número de neurônios escondidos.

Conforme se pode observar através das figuras 63 e 64, as trajetórias das funções de erro de treinamento *sse*, *mse* e *nmse* são suaves e apresentam um rápido declínio do erro com o aumento dos neurônios escondidos, começando a se estabilizar a partir de 5 neurônios. As trajetórias das funções de erro de treinamento *rmse*, *nrmse*, *mae* e *mre* são menos suaves, apresentam um rápido declínio do erro com o aumento dos neurônios escondidos, mas começam a se estabilizar a partir de 14 neurônios. A curva do coeficiente de determinação global cresce rapidamente com o aumento do número de neurônios da camada escondida e estabiliza em torno de 1 a partir de 5 neurônios escondidos (Figura 64b).

Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação estão apresentados na tabela 39.

Tabela 39. Protótipo II (1000 padrões) - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	3,26E+01	1,63E-02	1,28E-01	8,97E-04	1,21E-01	2,01E+01	5,720499E-01	3,272410E-01
2	1,45E+01	7,23E-03	8,50E-02	5,97E-04	8,26E-02	1,33E+01	8,620508E-01	7,431315E-01
3	7,69E+00	3,84E-03	6,20E-02	4,36E-04	6,40E-02	1,04E+01	9,401653E-01	8,839108E-01
4	5,23E+00	2,61E-03	5,11E-02	3,59E-04	5,30E-02	8,06E+00	9,712100E-01	9,432489E-01
5	4,87E+00	2,44E-03	4,94E-02	3,47E-04	5,14E-02	7,89E+00	9,741779E-01	9,490226E-01
6	4,64E+00	2,32E-03	4,82E-02	3,38E-04	4,85E-02	7,48E+00	9,767553E-01	9,540509E-01
7	4,30E+00	2,15E-03	4,64E-02	3,26E-04	4,63E-02	7,15E+00	9,773317E-01	9,551771E-01
8	4,55E+00	2,27E-03	4,77E-02	3,35E-04	4,91E-02	7,49E+00	9,673602E-01	9,357857E-01
9	4,17E+00	2,08E-03	4,57E-02	3,21E-04	4,48E-02	6,94E+00	9,769776E-01	9,544853E-01
10	4,32E+00	2,16E-03	4,65E-02	3,26E-04	4,79E-02	7,33E+00	9,759559E-01	9,524900E-01
11	5,18E+00	2,59E-03	5,09E-02	3,58E-04	4,88E-02	7,63E+00	9,766715E-01	9,538873E-01
12	4,85E+00	2,43E-03	4,93E-02	3,46E-04	5,36E-02	8,07E+00	9,780569E-01	9,565954E-01
13	1,10E+01	5,49E-03	7,41E-02	5,21E-04	6,82E-02	9,62E+00	9,663601E-01	9,338518E-01
14	4,69E+00	2,35E-03	4,84E-02	3,40E-04	4,86E-02	7,49E+00	9,767847E-01	9,541083E-01
15	4,90E+00	2,45E-03	4,95E-02	3,48E-04	4,77E-02	7,40E+00	9,761472E-01	9,528634E-01
16	4,66E+00	2,33E-03	4,83E-02	3,39E-04	4,71E-02	7,24E+00	9,788455E-01	9,581384E-01
17	3,85E+00	1,92E-03	4,39E-02	3,08E-04	4,26E-02	6,63E+00	9,772901E-01	9,550960E-01
18	4,86E+00	2,43E-03	4,93E-02	3,46E-04	5,09E-02	7,77E+00	9,759615E-01	9,525009E-01
19	4,68E+00	2,34E-03	4,84E-02	3,40E-04	4,97E-02	7,62E+00	9,775538E-01	9,556113E-01
20	5,45E+00	2,73E-03	5,22E-02	3,67E-04	5,46E-02	8,18E+00	9,758153E-01	9,522155E-01

Entre as 20 topologias testadas para o protótipo II, a topologia com 16 neurônios na camada escondida (topologia 2-16-4) foi a que apresentou maior número de simulações, com o conjunto de validação, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9788455 e R_g^2 =0,9581384). Entretanto, a topologia contendo 7 neurônios na camada escondida (topologia 2-7-4) foi selecionada

como a melhor por ser mais simples (menor número de neurônios) e apresentar capacidade de generalização e robustez ($medmed_R=0,9773317$ e $R_g^2=0,9551771$) praticamente iguais a da rede com 16 neurônios na camada escondida. A rede com 6 neurônios na camada escondida também seria uma boa escolha.

As comparações das figuras 65 e 66 com as figuras 63 e 64 permitem verificar que as trajetórias das funções de erro de validação do protótipo II, usando 1000 padrões de entrada-saída, não são tão suaves quanto às trajetórias das funções de erro de treinamento. Através das figuras 65 e 66, percebe-se que as curvas das funções de erro de validação do protótipo II com 1000 padrões de entrada-saída, semelhantemente as de 100 padrões (figuras 55 e 56), também apresentam um rápido declínio do erro com o aumento dos neurônios escondidos e começam a se estabilizar a partir de 5 neurônios. Entretanto, para 15 neurônios escondidos, as curvas das funções de erro de validação do protótipo II usando 1000 padrões, apresentam um brusco aumento do erro.

As curvas dos coeficientes de determinações globais do protótipo II usando 100 e 1000 padrões de entrada-saída (Figura 56d e 66d), apresentam comportamento similares, aumentam continuamente com o aumento do número de neurônios escondidos e estabilizam em torno de 0,9 a partir de 5 neurônios.

Devido ao aumento do número de padrões de entrada-saída, podia-se esperar que o protótipo treinado com 1000 padrões apresentasse menores erros e maior capacidade de generalização que o protótipo treinado com 100 padrões. Entretanto, quando o protótipo II foi treinado com 100 padrões, apresentou melhor desempenho do que quando foi treinado usando 1000 padrões. Isto sugere que *softsensores* neurais relativamente simples podem ser implementados no processo de produção de biossurfactante por *Candida lipolytica*.

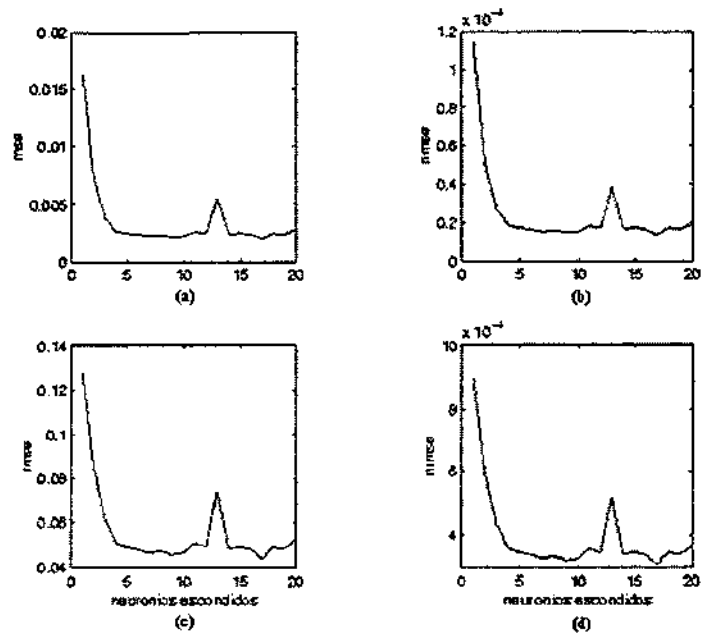


Figura 65. Protótipo II (1000 padrões) - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nrmse$ usando conjunto de validação em função do número de neurônios escondidos.

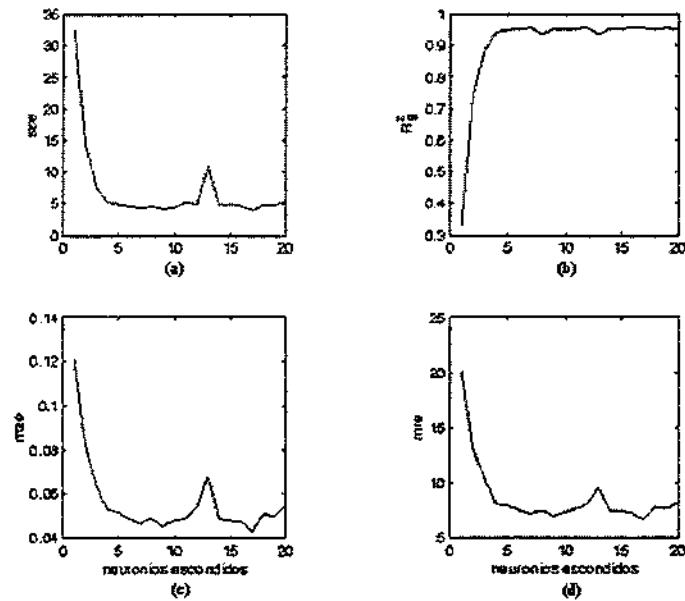


Figura 66. Protótipo II (1000 padrões) - Índices de avaliação do desempenho (a) sse (b) R_g^2 (c) mae e (d) mre usando conjunto de validação em função do número de neurônios escondidos.

Através das figuras 67 e 68 pode-se observar que as curvas previstas pelo protótipo II, com topologia 2-7-4, para as variáveis de saída - concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola - ajustaram-se quase perfeitamente aos dados do conjunto de treinamento composto por 1000 padrões (coeficientes de correlação R aproximadamente iguais a 1 e $R_g^2 = 0,9966770$).

Conforme se pode observar pelas figuras 69 e 70, as curvas previstas pelo protótipo II com topologia 2-7-4 para as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola também apresentaram bom ajuste aos dados do conjunto de validação composto por 1000 padrões (coeficientes de correlação R superiores a 0,94 e $R_g^2 = 0,9551771$). Entre as curvas previstas pelo protótipo II com topologia 2-7-4 as que apresentaram pior ajuste aos dados experimentais foram as curvas de atividade de emulsificação para emulsões água-em óleo de canola e água-em óleo de milho, nesta ordem.

Um bom critério para avaliar a influência do tamanho do conjunto de dados de treinamento é o coeficiente de determinação global (R_g^2). Os coeficientes de determinação global do protótipo II com topologia 2-7-4, treinado com 100 padrões e 1000 padrões foram respectivamente iguais 0,9666179 e 0,9551771. A superioridade do desempenho do protótipo II com topologia 2-7-4, treinado com 100 padrões sobre o protótipo II com topologia 2-7-4, treinado com 1000 padrões também pode ser verificada de outras formas. As curvas de saída previstas pelo protótipo II, com topologia 2-7-4, treinado com 100 padrões, ajustam-se melhor aos dados experimentais do que as curvas de saída previstas pelo protótipo II, com topologia 2-7-4, treinado com 1000 padrões, conforme pode ser verificado visualmente através das figuras 60 e 70. A tabela 40 e a figura 71 mostram que o treinamento do protótipo II com topologia 2-7-4 treinado com 1000 padrões atingiu 1000 épocas. Por outro lado, a tabela 35 e a figura 61 mostram que o treinamento do protótipo II com topologia 2-7-4, treinado com 100 padrões foi encerrado com apenas 458 épocas. A figura 72 mostra a evolução dos parâmetros de treinamento (sse , ssx e $gamk$) do protótipo II com topologia 2-7-4 em função do número de épocas.

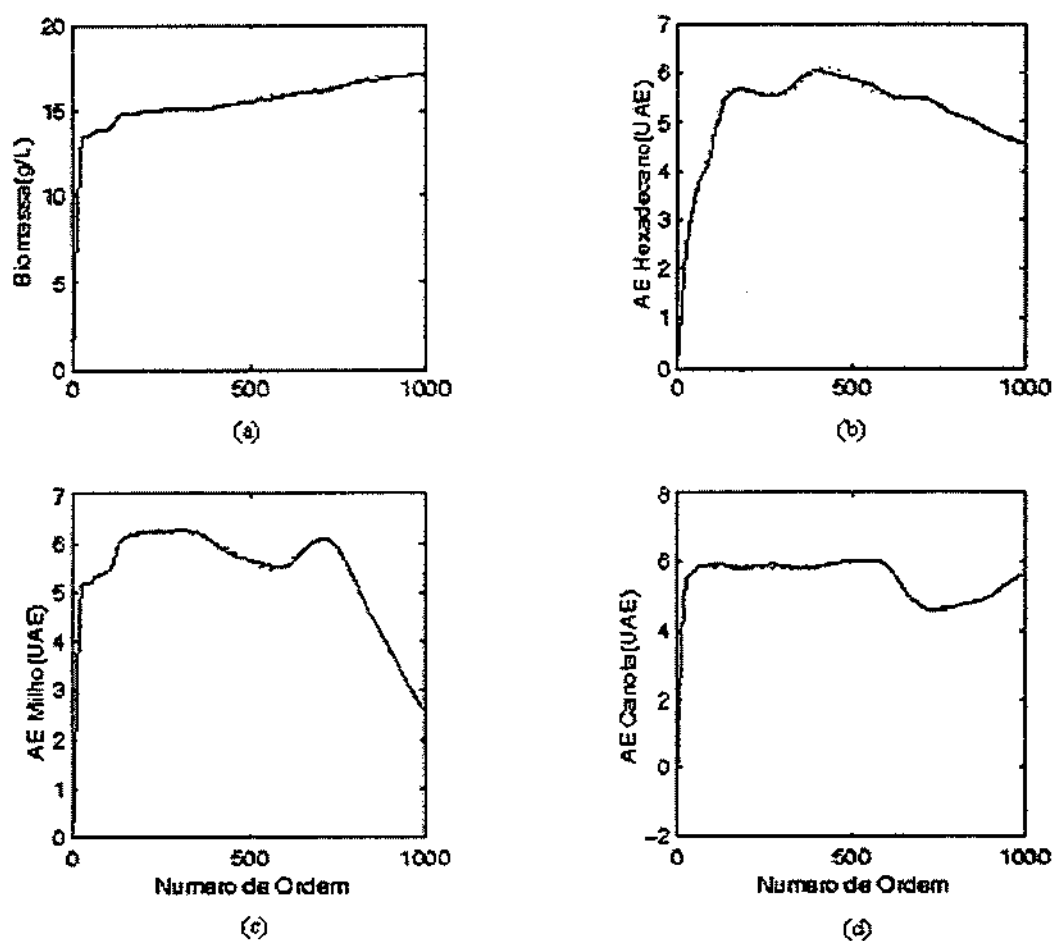
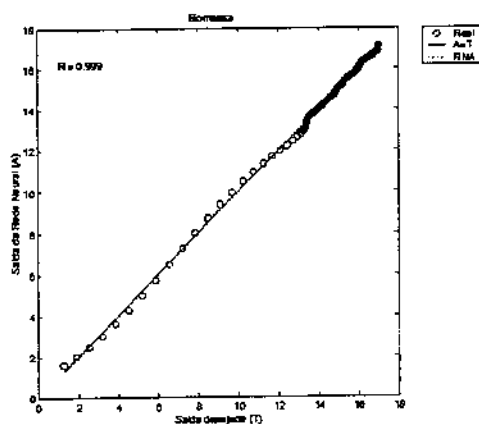
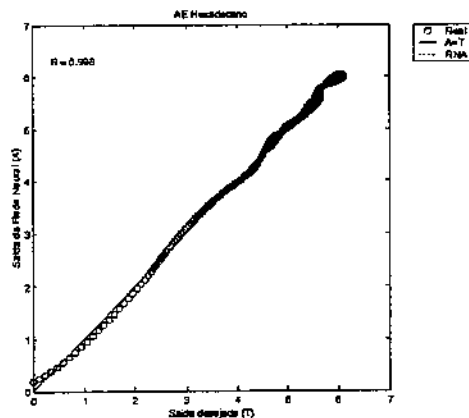


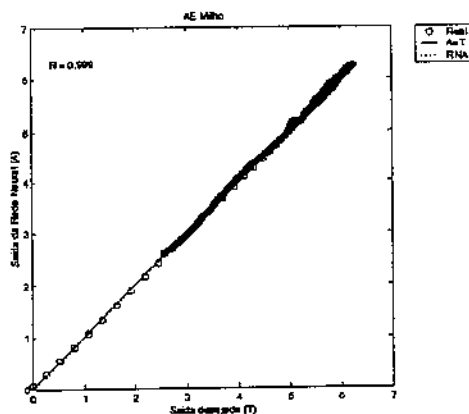
Figura 67. Protótipo II (1000 padrões) - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de treinamento.



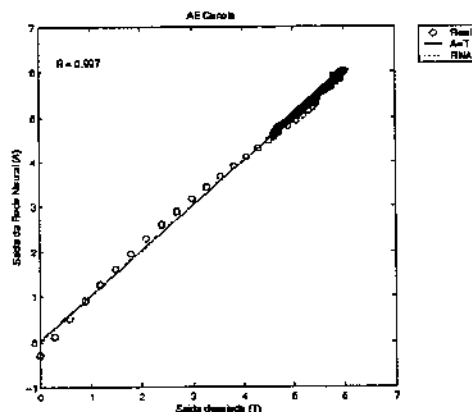
(a)



(b)



(c)



(d)

Figura 68. Protótipo II (1000 padrões) - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de treinamento.

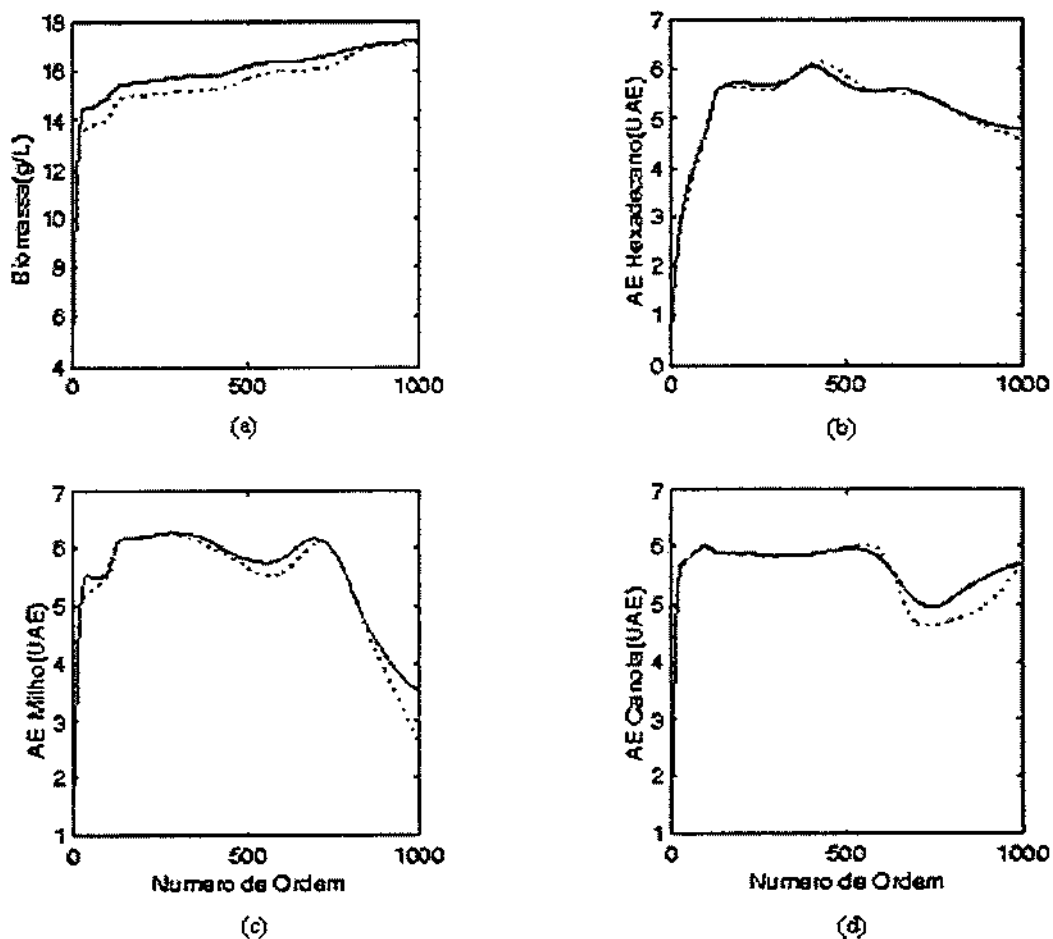


Figura 69. Protótipo II (1000 padrões) - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de validação.

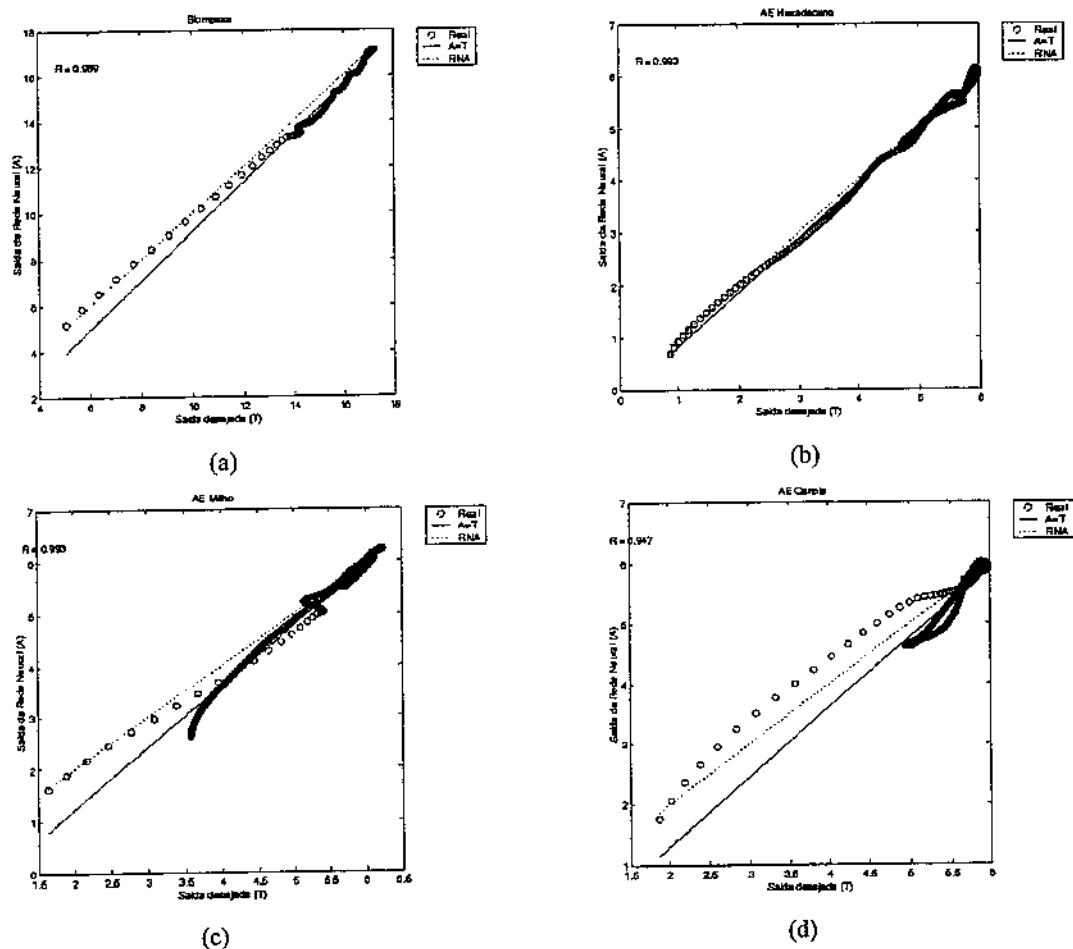


Figura 70. Protótipo II (1000 padrões) - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de validação.

Tabela 40. Protótipo II (1000 padrões) - Parâmetros de treinamento da rede (*época*, *sse*, *mse*, *ssx*, *np* e *gamk*) em função do número de neurônios (*SI*) na camada escondida.

<i>SI</i>	<i>época</i>	<i>sse</i>	<i>mse</i>	<i>ssx</i>	<i>np</i>	<i>gamk</i>
1	56	2,04E+01	1,02E-02	4,99E+02	11	1,10E+01
2	102	7,73E+00	3,87E-03	3,20E+03	18	1,51E+01
3	138	3,39E+00	1,70E-03	2,41E+03	25	2,20E+01
4	478	5,99E-01	3,00E-04	1,36E+03	32	2,83E+01
5	1000	2,52E-01	1,26E-04	5,33E+03	39	3,77E+01
6	831	1,57E-01	7,83E-05	4,30E+03	46	4,50E+01
7	1000	7,57E-02	3,78E-05	2,31E+04	53	4,94E+01
8	1000	1,07E-01	5,35E-05	1,00E+04	60	5,58E+01
9	1000	3,85E-02	1,92E-05	1,14E+04	67	6,03E+01
10	1000	1,65E-02	8,24E-06	1,45E+04	74	7,00E+01
11	1000	2,47E-02	1,23E-05	1,85E+04	81	7,27E+01
12	1000	1,50E-02	7,48E-06	1,31E+04	88	8,10E+01
13	1000	2,20E-02	1,10E-05	3,71E+04	95	8,20E+01
14	1000	3,83E-03	1,92E-06	1,54E+04	102	9,33E+01
15	1000	2,24E-03	1,12E-06	1,81E+04	109	9,98E+01
16	1000	1,95E-03	9,74E-07	2,15E+04	116	1,05E+02
17	1000	3,71E-03	1,85E-06	1,86E+04	123	1,06E+02
18	1000	2,01E-03	1,00E-06	1,90E+04	130	1,12E+02
19	1000	3,90E-03	1,95E-06	1,87E+04	137	1,15E+02
20	1000	1,82E-03	9,08E-07	1,75E+04	144	1,29E+02

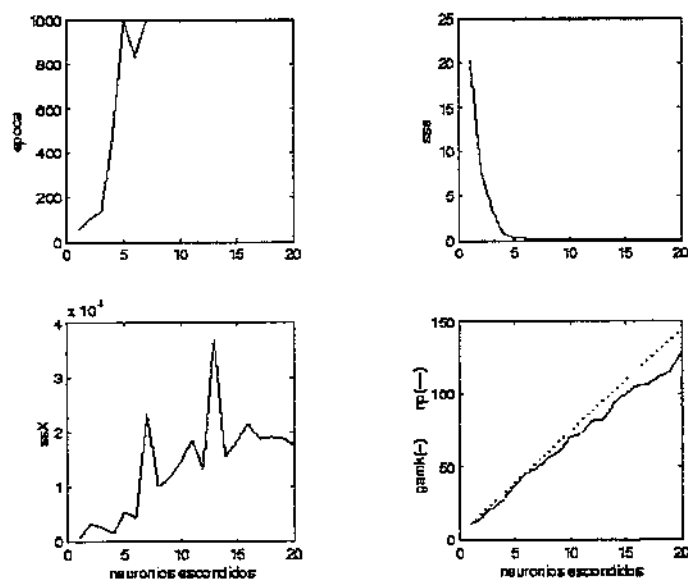


Figura 71. Protótipo II (1000 padrões) - Época e parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de neurônios escondidos

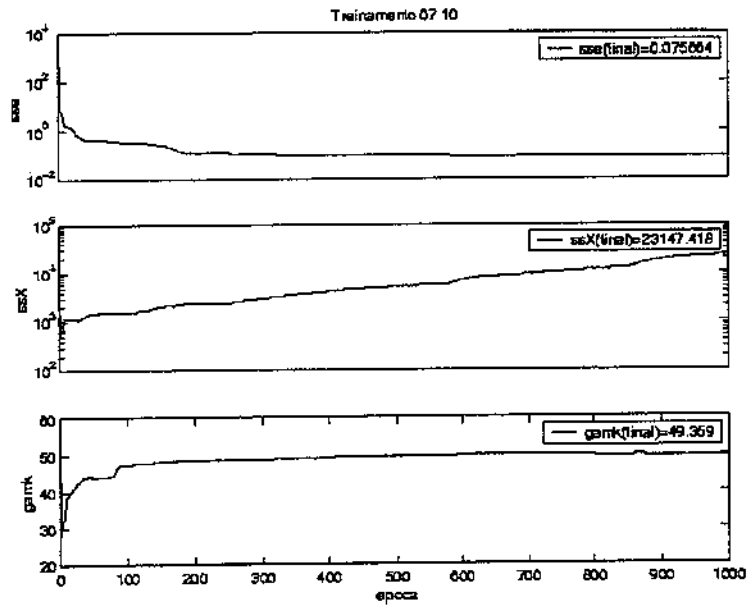


Figura 72. Protótipo II (1000 padrões) - Parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de épocas.

A topologia 2-7-4 do protótipo II treinado com 1000 padrões foi capaz de fornecer boas estimativas da concentração de biomassa e das atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola, usando o conjunto de validação. Logo, um *softsensor* neural completamente especificado para estimação das variáveis de estado pode ser implementado *on-line* usando as matrizes de pesos (IW^1 e LW^2) e bias ($b1$ e $b2$) da camada escondida e da camada de saída apresentadas nas tabelas 41 e 42. O *softsensor* neural II também pode ser representado pela seguinte equação

$$a = \text{purelin}(LW^2 \logsig(IW^1 p + b1) + b2) \quad (29)$$

onde: a - é a matriz dos padrões de saída do *softsensor* II
 purelin - é a função de transferência linear usada na camada de saída
 $\logsig$ - é a função de transferência sigmoidal usada na camada escondida
 p - é a matriz dos padrões de entrada do *softsensor* II

Tabela 41. Matrizes de pesos e bias da camada escondida do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 1000 padrões entrada-saída (Protótipo II com 7 neurônios na camada escondida).

Camada escondida			
Matriz de pesos (IW^1)		bias (b^1)	
1,70E+01	1,35E+01	-2,56E+01	
-7,13E+00	1,34E+00	3,21E+00	
4,31E+00	-3,45E-01	-5,69E+00	
-8,85E+01	1,70E+00	3,25E+01	
1,14E+01	-4,88E+00	-1,58E+00	
-8,58E+00	9,16E-01	3,89E+00	
8,05E+00	1,86E+01	-2,15E+00	

Tabela 42. Matrizes de pesos e bias da camada de saída do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 1000 padrões entrada-saída (Protótipo II com 7 neurônios na camada escondida).

Camada de saída							
Matriz de pesos (LW^2)						bias (b^2)	
-6,91E-01	-3,72E-01	-1,53E+00	3,48E-02	6,24E-02	2,13E-01	-4,12E-01	1,27E+00
-1,30E-01	-9,70E+00	-1,15E+01	-9,39E-02	-1,39E+00	7,72E+00	1,96E+00	1,66E+00
-7,29E-01	-1,58E+00	-4,55E+00	-7,15E-02	-4,20E+00	1,12E+00	9,22E+00	-3,69E+00
-7,99E-01	-9,00E-01	-9,48E-03	5,51E-02	2,81E+00	7,57E-01	-5,41E+00	3,51E+00

5.3.1.3. Protótipo III – *Softsensor* para inferência da concentração de biomassa

Os protótipos III, IV, V e VI apresentados e discutidos a seguir, têm as mesmas entradas do protótipo II, pH e de oxigênio dissolvido, mas têm como saída apenas uma das variáveis de saída do protótipo II. Eles foram construídos com o objetivo de investigar se melhores resultados poderiam ser obtidos usando diferentes protótipos para cada uma das variáveis de saída.

Medidas *on-line* de pH e de oxigênio dissolvido obtidas a 28° C e 150 rpm foram usadas como entradas para o protótipo III (Figura 38). A biomassa foi selecionada como variável de saída. Vinte diferentes topologias do protótipo III foram avaliadas usando conjunto de treinamento composto por 100 padrões de entrada-saída. Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando conjunto de treinamento estão apresentados na tabela 43.

Tabela 43. Protótipo III - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de treinamento.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	1,15E-01	5,75E-04	2,40E-02	2,21E-04	1,25E-02	2,08E+00	9,352500E-01	8,746926E-01
2	4,32E-03	2,16E-05	4,65E-03	4,28E-05	2,82E-03	3,53E-01	9,969760E-01	9,939611E-01
3	3,36E-03	1,68E-05	4,10E-03	3,78E-05	2,30E-03	2,75E-01	9,982917E-01	9,965864E-01
4	9,69E-04	4,85E-06	2,20E-03	2,03E-05	1,05E-03	1,30E-01	9,991430E-01	9,982866E-01
5	9,85E-04	4,92E-06	2,22E-03	2,05E-05	1,24E-03	1,52E-01	9,993221E-01	9,986447E-01
6	9,42E-04	4,71E-06	2,17E-03	2,00E-05	1,25E-03	1,60E-01	9,994760E-01	9,989522E-01
7	9,73E-04	4,86E-06	2,21E-03	2,03E-05	1,19E-03	1,45E-01	9,994660E-01	9,989322E-01
8	9,80E-04	4,90E-06	2,21E-03	2,04E-05	1,29E-03	1,58E-01	9,994736E-01	9,989474E-01
9	9,95E-04	4,97E-06	2,23E-03	2,06E-05	1,25E-03	1,54E-01	9,994759E-01	9,989521E-01
10	9,66E-04	4,83E-06	2,20E-03	2,03E-05	1,15E-03	1,39E-01	9,994617E-01	9,989237E-01
11	9,61E-04	4,80E-06	2,19E-03	2,02E-05	1,15E-03	1,41E-01	9,994822E-01	9,989646E-01
12	9,29E-04	4,65E-06	2,16E-03	1,99E-05	1,12E-03	1,35E-01	9,995002E-01	9,990006E-01
13	9,98E-04	4,99E-06	2,23E-03	2,06E-05	1,23E-03	1,48E-01	9,994689E-01	9,989381E-01
14	9,72E-04	4,86E-06	2,20E-03	2,03E-05	1,17E-03	1,39E-01	9,995006E-01	9,990014E-01
15	9,59E-04	4,80E-06	2,19E-03	2,02E-05	1,18E-03	1,43E-01	9,994900E-01	9,989802E-01
16	8,21E-04	4,11E-06	2,03E-03	1,87E-05	1,17E-03	1,58E-01	9,995777E-01	9,991556E-01
17	9,94E-04	4,97E-06	2,23E-03	2,06E-05	1,22E-03	1,47E-01	9,995649E-01	9,991299E-01
18	9,40E-04	4,70E-06	2,17E-03	2,00E-05	1,15E-03	1,38E-01	9,996391E-01	9,992783E-01
19	9,84E-04	4,92E-06	2,22E-03	2,05E-05	1,29E-03	1,66E-01	9,995743E-01	9,991487E-01
20	3,28E-04	1,64E-06	1,28E-03	1,18E-05	6,59E-04	7,93E-02	9,996499E-01	9,993000E-01

Entre as 20 topologias do protótipo III avaliadas, a topologia com 20 neurônios na camada escondida foi a que apresentou maior número de simulações com o conjunto de treinamento com maiores coeficientes de correlação para as 4 variáveis de saída simultaneamente ($medmed_R=0,9996499$ e $R_g^2=0,9993000$).

Conforme se pode observar através das figuras 73 e 74, as trajetórias das funções de erro de treinamento são suaves e apresentam um rápido declínio do erro com o aumento dos neurônios escondidos, começando a se estabilizar a partir do quarto neurônio escondido. A curva do coeficiente de determinação global é uma curva suave (Figura 74b), que cresce rapidamente até o quarto neurônio e então estabiliza em torno de 1, demonstrando a excelente capacidade de ajuste da maior parte das topologias do protótipo III avaliadas usando dados do conjunto de treinamento.

Entretanto, ajuste preciso sobre dados de treinamento não é uma evidência confiável da validade do modelo, desde que não fornece indicações sobre a capacidade de generalização do modelo. A decisão sobre o número ótimo de neurônios escondidos só pode ser tomada após o treinamento usando um conjunto de validação. Redes com muitos neurônios na camada escondida podem memorizar os padrões de treinamento perdendo capacidade de generalização, mas por outro lado se forem muito pequenas, perdem flexibilidade, necessitando de muito mais épocas para atingir a precisão desejada. Portanto, a melhor rede não é necessariamente aquela que apresenta melhor ajuste das saídas calculadas aos dados do conjunto de treinamento, mas sim a menor rede com maior capacidade de generalização quando testada com dados não apresentados durante o treinamento.

Os resultados dos índices de avaliação de desempenho (sse , mse , $rmse$, $nrmse$, mae , mre , $medmed_R$, R_g^2) em função do número de neurônios (SI) na camada escondida, obtidos usando o conjunto de validação estão apresentados na tabela 44.

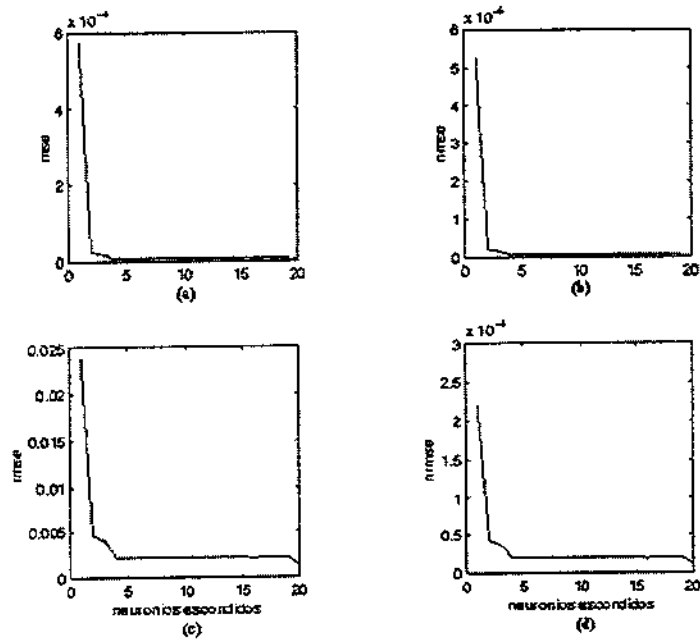


Figura 73. Protótipo III - Índices de avaliação do desempenho (a) *mse* (b) *nmse* (c) *rmse* e (d) *nrmse* usando conjunto de treinamento em função do número de neurônios escondidos.

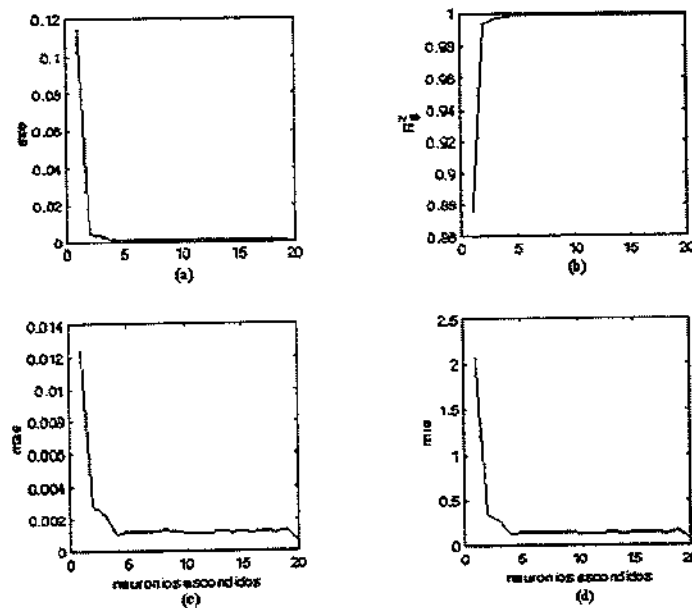


Figura 74. Protótipo III - Índices de avaliação do desempenho (a) *sse* (b) Rg^2 (c) *mae* e (d) *mre* usando conjunto de treinamento em função do número de neurônios escondidos.

Tabela 44. Protótipo III - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	2,58E-01	1,29E-03	3,59E-02	3,29E-04	2,25E-02	3,45E+00	8,985324E-01	8,073605E-01
2	1,04E-01	5,22E-04	2,28E-02	2,09E-04	1,39E-02	1,83E+00	9,810090E-01	9,623786E-01
3	1,04E-01	5,20E-04	2,28E-02	2,09E-04	1,39E-02	1,83E+00	9,818684E-01	9,640656E-01
4	8,90E-02	4,45E-04	2,11E-02	1,93E-04	1,20E-02	1,61E+00	9,844187E-01	9,690801E-01
5	8,72E-02	4,36E-04	2,09E-02	1,91E-04	1,18E-02	1,59E+00	9,834987E-01	9,672697E-01
6	9,91E-02	4,95E-04	2,23E-02	2,04E-04	1,36E-02	1,80E+00	9,830866E-01	9,664592E-01
7	9,29E-02	4,65E-04	2,16E-02	1,97E-04	1,22E-02	1,63E+00	9,828102E-01	9,659159E-01
8	9,95E-02	4,98E-04	2,23E-02	2,04E-04	1,37E-02	1,81E+00	9,837411E-01	9,677465E-01
9	1,08E-01	5,39E-04	2,32E-02	2,13E-04	1,45E-02	1,90E+00	9,834101E-01	9,670953E-01
10	8,71E-02	4,36E-04	2,09E-02	1,91E-04	1,21E-02	1,61E+00	9,842049E-01	9,686593E-01
11	9,58E-02	4,79E-04	2,19E-02	2,00E-04	1,26E-02	1,69E+00	9,845070E-01	9,692540E-01
12	9,89E-02	4,94E-04	2,22E-02	2,04E-04	1,34E-02	1,78E+00	9,852506E-01	9,707188E-01
13	1,06E-01	5,28E-04	2,30E-02	2,11E-04	1,43E-02	1,88E+00	9,850383E-01	9,703005E-01
14	9,19E-02	4,59E-04	2,14E-02	1,96E-04	1,34E-02	1,76E+00	9,858648E-01	9,719294E-01
15	1,01E-01	5,07E-04	2,25E-02	2,06E-04	1,39E-02	1,83E+00	9,844301E-01	9,691026E-01
16	1,03E-01	5,16E-04	2,27E-02	2,08E-04	1,43E-02	1,89E+00	9,854585E-01	9,711284E-01
17	9,87E-02	4,94E-04	2,22E-02	2,03E-04	1,37E-02	1,80E+00	9,851439E-01	9,705085E-01
18	1,03E-01	5,13E-04	2,27E-02	2,08E-04	1,43E-02	1,87E+00	9,848147E-01	9,698600E-01
19	9,62E-02	4,81E-04	2,19E-02	2,01E-04	1,37E-02	1,80E+00	9,850275E-01	9,702791E-01
20	9,47E-02	4,74E-04	2,18E-02	1,99E-04	1,34E-02	1,77E+00	9,859054E-01	9,720094E-01

Entre as 20 topologias testadas para o protótipo III, a topologia com 20 neurônios na camada escondida (topologia 2-20-1) foi a que apresentou maior número de simulações, com o conjunto de validação, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9859054 e R_g^2 =0,9720094). Entretanto, a topologia contendo 4 neurônios na camada escondida (topologia 2-4-1), foi selecionada como a melhor por ser mais simples (menor número de neurônios) e apresentar capacidade de generalização e robustez (*medmed_R* =0,9844187 e R_g^2 = 0,9690801) praticamente iguais a da rede com 20 neurônios na camada escondida. No presente trabalho, a decisão sobre a melhor topologia de um determinado protótipo foi sempre tomada em favor da topologia com maior capacidade preditiva e com menor número de neurônios na camada escondida.

As curvas de evolução dos erros de validação do protótipo III (figuras 75 e 76) apresentam um rápido declínio do erro com o aumento dos neurônios escondidos, começam a se estabilizar a partir de 4 neurônios e são mais ruidosas que as trajetórias das funções de erro de treinamento do protótipo III. A curva do coeficiente de determinação global cresce rapidamente até o quarto neurônio quando estabiliza em torno de 0,97.

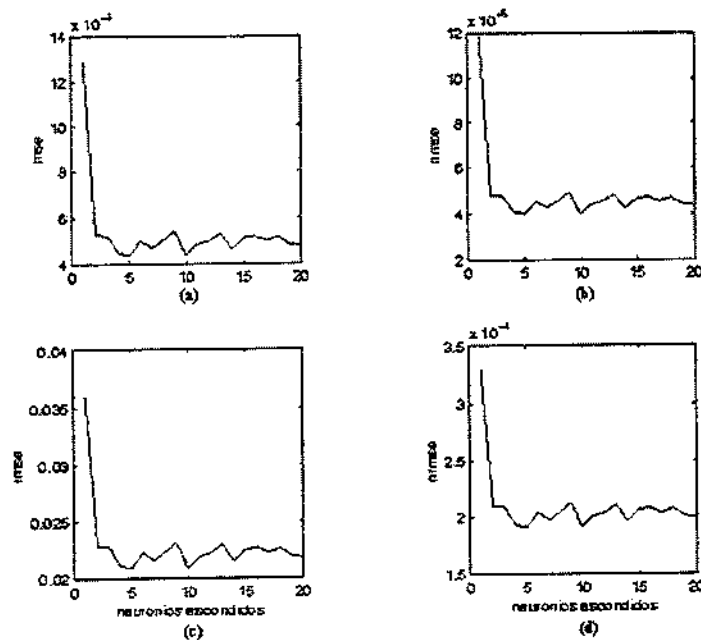


Figura 75. Protótipo III - Índices de avaliação do desempenho (a) *mse* (b) *nmse* (c) *rmse* e (d) *nrmse* usando conjunto de validação em função do número de neurônios escondidos.

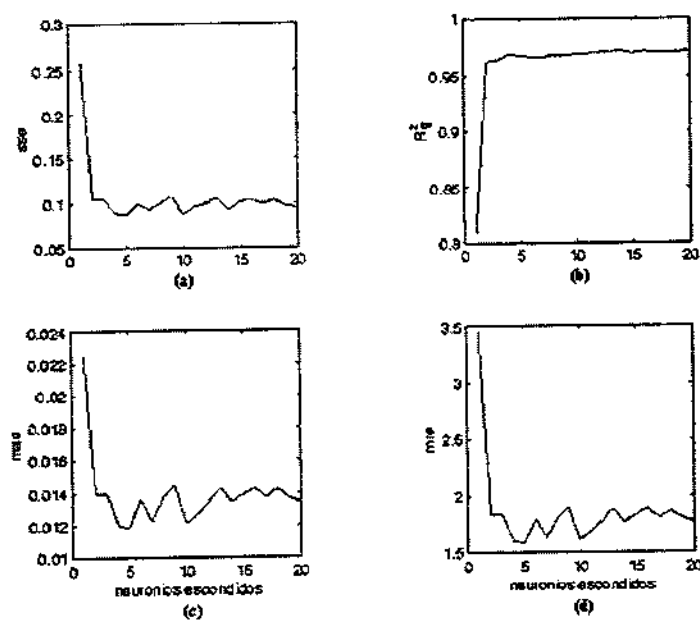


Figura 76. Protótipo III - Índices de avaliação do desempenho (a) *sse* (b) Rg^2 (c) *mae* e (d) *mre* usando conjunto de validação em função do número de neurônios escondidos.

Através da figura 77 pode-se verificar visualmente o excelente ajuste dos valores de saídas experimentais com os valores preditos, pelo protótipo III com topologia 2-4-1, para a concentração de biomassa, usando o conjunto de treinamento (Figura 77a) e o conjunto de teste (Figura 77b). Os gráficos de paridade (ajuste linear) entre os valores de saídas experimentais e os valores preditos pelo protótipo III com topologia 2-4-1 para a concentração de biomassa, usando dados do conjunto de treinamento e do conjunto de teste, podem ser comparados através das figuras 77c e 77d, permitindo observar que as curvas prevista pelo protótipo III com topologia 2-4-1 para a concentração de biomassa ajustaram-se quase perfeitamente aos dados do conjunto de treinamento ($R=0,999$ e $R_g^2=0,982866$) e aos dados do conjunto de validação ($R=0,986$ e $R_g^2=0,9690801$).

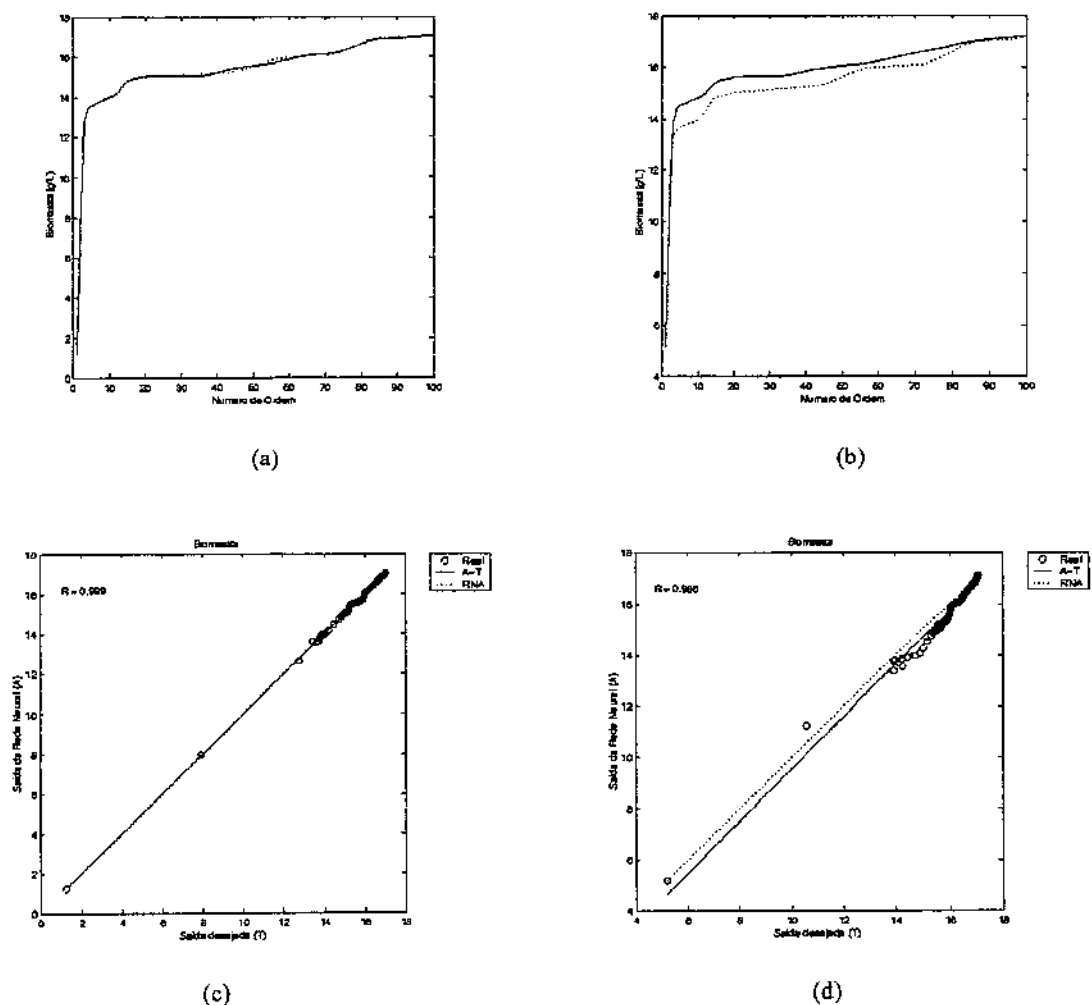


Figura 77. Protótipo III - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para concentração de biomassa (a) usando dados do conjunto de treinamento (b) usando dados do conjunto de validação. Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para concentração de biomassa (c) usando dados do conjunto de treinamento (d) usando dados do conjunto de validação.

As variações no número de épocas e nos parâmetros de treinamento (*sse*, *mse*, *ssx*, *np* e *gamk*) do protótipo III, em função do número de neurônios escondidos (*SI*) podem ser analisadas através da tabela 45 e da figura 78.

Tabela 45. Protótipo III - Parâmetros de treinamento da rede (*época*, *sse*, *mse*, *ssx*, *np* e *gamk*) em função do número de neurônios (*SI*) na camada escondida,.

<i>SI</i>	<i>época</i>	<i>sse</i>	<i>mse</i>	<i>ssx</i>	<i>np</i>	<i>gamk</i>
1	240	1,15E-01	5,75E-04	9,30E+01	5,00E+00	3,85E+00
2	93	4,32E-03	2,16E-05	5,51E+02	9,00E+00	8,27E+00
3	513	3,36E-03	1,68E-05	7,51E+02	1,30E+01	1,02E+01
4	234	9,69E-04	4,85E-06	1,61E+03	1,70E+01	1,46E+01
5	104	9,85E-04	4,92E-06	7,30E+02	2,10E+01	1,58E+01
6	463	9,42E-04	4,71E-06	9,35E+02	2,50E+01	1,66E+01
7	144	9,73E-04	4,86E-06	1,00E+03	2,90E+01	1,73E+01
8	79	9,80E-04	4,90E-06	1,38E+03	3,30E+01	1,73E+01
9	68	9,95E-04	4,97E-06	1,47E+03	3,70E+01	1,93E+01
10	124	9,66E-04	4,83E-06	6,62E+02	4,10E+01	1,70E+01
11	66	9,61E-04	4,80E-06	3,58E+03	4,50E+01	2,03E+01
12	63	9,29E-04	4,65E-06	4,19E+03	4,90E+01	2,07E+01
13	20	9,98E-04	4,99E-06	2,61E+03	5,30E+01	2,19E+01
14	59	9,72E-04	4,86E-06	5,48E+03	5,70E+01	2,36E+01
15	28	9,59E-04	4,80E-06	6,63E+03	6,10E+01	2,31E+01
16	7	8,21E-04	4,11E-06	8,98E+03	6,50E+01	2,32E+01
17	6	9,94E-04	4,97E-06	1,02E+04	6,90E+01	2,64E+01
18	5	9,40E-04	4,70E-06	8,15E+03	7,30E+01	2,50E+01
19	4	9,84E-04	4,92E-06	1,26E+04	7,70E+01	2,42E+01
20	8	3,28E-04	1,64E-06	1,08E+04	8,10E+01	2,77E+01

No caso em estudo, a determinação da melhor topologia do protótipo III, através da observação do número de neurônios escondidos a partir do qual o número dos parâmetros efetivamente utilizados (*gamk*) começa a se estabilizar - sugerida por Foresee e Hagan (1997) - coincide com o número de neurônios escondidos determinado com o critério de seleção adotado no presente trabalho. O parâmetro *sse* também começa a se estabilizar a partir do quarto neurônio escondido. O parâmetro *ssx* cresce continuamente até o quarto neurônio, onde apresenta certa estabilização até o décimo neurônio, sofre um rápido decréscimo e passa em seguida a crescer continuamente até o vigéssimo neurônio escondido, não estabilizando ao mesmo tempo em que os parâmetros *sse* e *gamk* como sugerido por Foresee e Hagan (1997). A evolução dos parâmetros de treinamento *sse*, *ssx* e *gamk* em função do número de neurônios escondidos encontra-se ilustrada na figura 78.

A evolução dos parâmetros de treinamento *sse*, *ssx* e *gamk* em função do número de épocas, para a topologia 2-4-1 - selecionada como a melhor topologia do protótipo III - é apresentada na figura 79

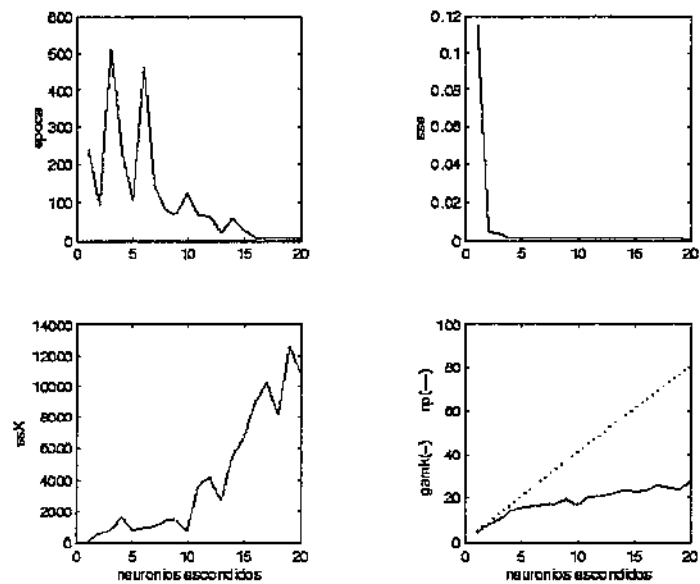


Figura 78. Protótipo III - Época e parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de neurônios escondidos.

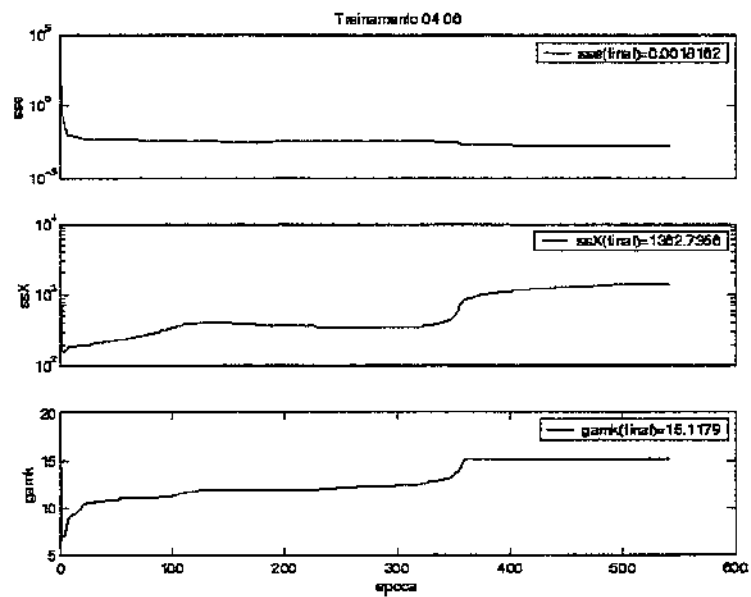


Figura 79. Protótipo III - Parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de épocas.

A topologia 2-4-1 do protótipo III foi capaz de fornecer excelentes estimativas da concentração de biomassa usando o conjunto de validação. Logo, um *softsensor* neural para estimação de biomassa representado pela equação (30) pode ser implementado *on-line* usando as matrizes de pesos (IW^1 e LW^2) e bias ($b1$ e $b2$) da camada escondida e da camada de saída apresentadas nas tabelas 46 e 47.

$$a = \text{purelin}(LW^2 \text{logsig}(IW^1 p + b1) + b2) \quad (30)$$

onde: a - é a matriz dos padrões de saída do *softsensor* III
purelin - é a função de transferência linear usada na camada de saída
logsig - é a função de transferência sigmoideal usada na camada escondida
 p - é a matriz dos padrões de entrada do *softsensor* III

Tabela 46. Matrizes de pesos e bias da camada escondida do *softsensor* para inferência de biomassa, treinado com 100 padrões entrada-saída (Protótipo III com 4 neurônios na camada escondida).

Camada escondida		
Matriz de pesos (IW^1)		bias (b^1)
-4,25E+00	-2,40E-01	3,16E+00
-2,78E+00	1,95E+01	-8,26E+00
-4,43E+00	-8,13E+00	4,06E+00
1,37E-01	-9,37E+00	4,53E-01

Tabela 47. Matrizes de pesos e bias da camada de saída do *softsensor* para inferência de biomassa, treinado com 100 padrões entrada-saída (Protótipo III com 4 neurônios na camada escondida).

Camada de saída				
Matriz de pesos (LW^2)				bias (b^2)
1,64E+01	4,08E-01	-1,52E+01	1,27E+01	-5,11E+00

5.3.1.4. Protótipo IV – *Softsensor* para inferência de atividade de emulsificação para emulsões água-em-hexadecano

Medidas *on-line* de pH e de oxigênio dissolvido obtidas a 28° C e 150 rpm foram usadas como entradas para o protótipo IV (Figura 39). A atividade de emulsificação para emulsões água-em-hexadecano foi selecionada como variável de saída. Vinte diferentes topologias do protótipo IV foram avaliadas usando conjunto de treinamento composto por 100 padrões de entrada-saída. Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando conjunto de treinamento estão apresentados na tabela 48.

Tabela 48. Protótipo IV - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de treinamento.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	2,59E-01	1,30E-03	3,60E-02	3,54E-03	1,94E-02	2,49E+00	9,078956E-01	8,242744E-01
2	2,61E-02	1,30E-04	1,14E-02	1,12E-03	6,33E-03	7,99E-01	9,907941E-01	9,816729E-01
3	2,05E-02	1,03E-04	1,01E-02	9,96E-04	5,59E-03	7,02E-01	9,935550E-01	9,871515E-01
4	3,98E-03	1,99E-05	4,46E-03	4,38E-04	2,37E-03	3,27E-01	9,987994E-01	9,976003E-01
5	3,05E-03	1,52E-05	3,90E-03	3,84E-04	1,98E-03	2,41E-01	9,995428E-01	9,990859E-01
6	9,92E-04	4,96E-06	2,23E-03	2,19E-04	1,25E-03	1,60E-01	9,996907E-01	9,993816E-01
7	9,99E-04	5,00E-06	2,24E-03	2,20E-04	1,20E-03	1,54E-01	9,996922E-01	9,993845E-01
8	9,95E-04	4,98E-06	2,23E-03	2,19E-04	1,15E-03	1,44E-01	9,996919E-01	9,993839E-01
9	9,99E-04	4,99E-06	2,23E-03	2,20E-04	1,23E-03	1,49E-01	9,996894E-01	9,993790E-01
10	9,99E-04	4,99E-06	2,23E-03	2,20E-04	1,15E-03	1,38E-01	9,996896E-01	9,993794E-01
11	9,98E-04	4,99E-06	2,23E-03	2,20E-04	1,12E-03	1,35E-01	9,996904E-01	9,993809E-01
12	9,96E-04	4,98E-06	2,23E-03	2,19E-04	1,18E-03	1,45E-01	9,996914E-01	9,993829E-01
13	1,00E-03	5,00E-06	2,24E-03	2,20E-04	1,19E-03	1,46E-01	9,996903E-01	9,993808E-01
14	1,00E-03	5,00E-06	2,24E-03	2,20E-04	1,17E-03	1,43E-01	9,996900E-01	9,993801E-01
15	9,92E-04	4,96E-06	2,23E-03	2,19E-04	1,11E-03	1,35E-01	9,996901E-01	9,993802E-01
16	9,97E-04	4,98E-06	2,23E-03	2,19E-04	1,10E-03	1,33E-01	9,996924E-01	9,993849E-01
17	9,74E-04	4,87E-06	2,21E-03	2,17E-04	1,12E-03	1,42E-01	9,996923E-01	9,993846E-01
18	9,98E-04	4,99E-06	2,23E-03	2,20E-04	1,07E-03	1,27E-01	9,996921E-01	9,993843E-01
19	9,90E-04	4,95E-06	2,22E-03	2,19E-04	1,20E-03	1,46E-01	9,996914E-01	9,993829E-01
20	9,19E-04	4,60E-06	2,14E-03	2,11E-04	1,04E-03	1,28E-01	9,996947E-01	9,993896E-01

Entre as 20 topologias do protótipo IV avaliadas, a topologia com 20 neurônios na camada escondida foi a que apresentou maior número de simulações, com o conjunto de treinamento, com maiores coeficientes de correlação para as 4 variáveis de saída simultaneamente (*medmed_R*=0,9996947 e R_g^2 =0,9993896). Nas figuras 80 e 81 encontram-se ilustrados os comportamentos das funções de erro de treinamento e do coeficiente de determinação global em função do número de neurônios escondidos.

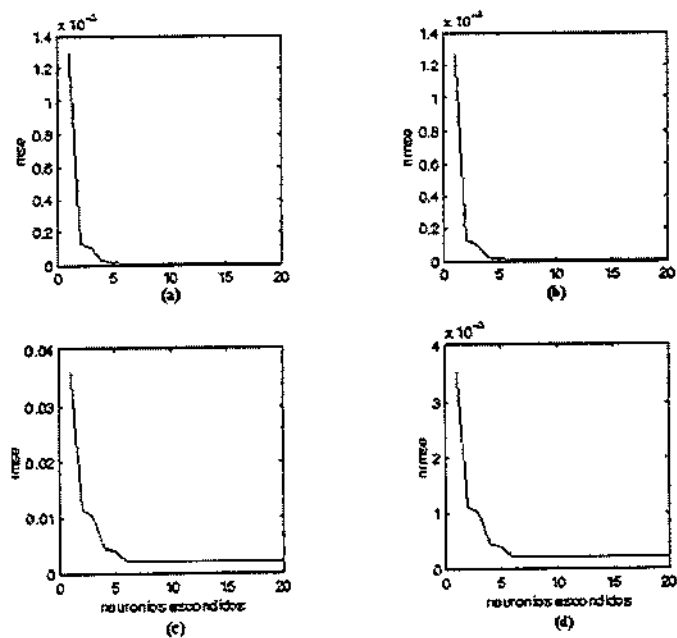


Figura 80. Protótipo IV - Índices de avaliação do desempenho (a) *mse* (b) *nmse* (c) *rmse* e (d) *nrmse* usando conjunto de treinamento em função do número de neurônios escondidos.

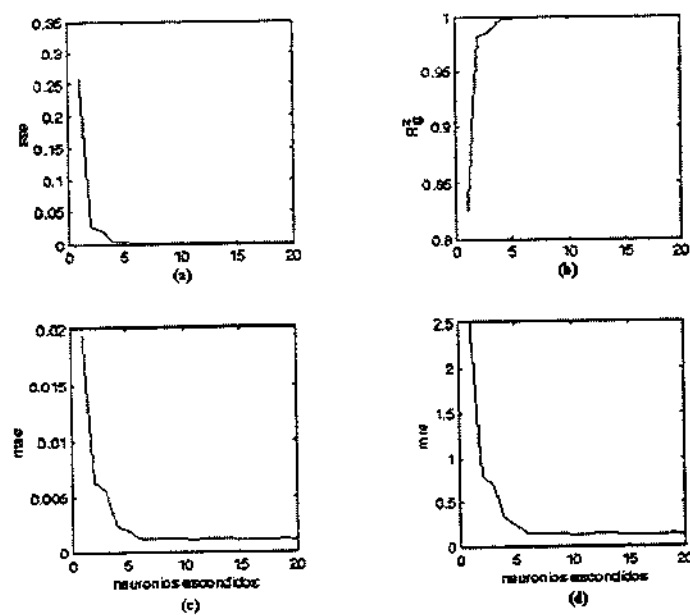


Figura 81. Protótipo IV - Índices de avaliação do desempenho (a) *sse* (b) R_g^2 (c) *mae* e (d) *mre* usando conjunto de treinamento em função do número de neurônios escondidos.

Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação estão apresentados na tabela 49.

Tabela 49. Protótipo IV - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	3,50E-01	1,75E-03	4,18E-02	4,08E-03	2,12E-02	2,85E+00	8,837952E-01	7,810939E-01
2	5,48E-02	2,74E-04	1,65E-02	1,61E-03	1,00E-02	1,33E+00	9,859688E-01	9,721344E-01
3	4,79E-02	2,39E-04	1,55E-02	1,51E-03	9,27E-03	1,26E+00	9,891409E-01	9,783998E-01
4	2,30E-02	1,15E-04	1,07E-02	1,04E-03	6,61E-03	9,18E-01	9,957302E-01	9,914786E-01
5	2,29E-02	1,14E-04	1,07E-02	1,04E-03	6,76E-03	9,25E-01	9,970555E-01	9,941197E-01
6	1,67E-02	8,34E-05	9,13E-03	8,89E-04	5,89E-03	8,15E-01	9,974312E-01	9,948690E-01
7	1,74E-02	8,70E-05	9,33E-03	9,07E-04	5,74E-03	8,01E-01	9,975107E-01	9,950276E-01
8	1,94E-02	9,71E-05	9,85E-03	9,60E-04	6,32E-03	8,80E-01	9,972701E-01	9,945476E-01
9	1,61E-02	8,04E-05	8,97E-03	8,73E-04	5,76E-03	7,88E-01	9,966925E-01	9,933960E-01
10	1,80E-02	8,98E-05	9,48E-03	9,22E-04	5,86E-03	8,13E-01	9,966398E-01	9,932908E-01
11	1,83E-02	9,15E-05	9,57E-03	9,31E-04	6,04E-03	8,34E-01	9,964737E-01	9,929599E-01
12	1,75E-02	8,74E-05	9,35E-03	9,09E-04	5,78E-03	7,77E-01	9,972004E-01	9,944087E-01
13	2,04E-02	1,02E-04	1,01E-02	9,83E-04	5,82E-03	8,25E-01	9,964555E-01	9,929235E-01
14	1,88E-02	9,41E-05	9,70E-03	9,45E-04	6,21E-03	8,60E-01	9,964279E-01	9,928686E-01
15	2,61E-02	1,30E-04	1,14E-02	1,11E-03	7,00E-03	9,97E-01	9,968295E-01	9,936691E-01
16	1,83E-02	9,17E-05	9,58E-03	9,32E-04	5,83E-03	8,14E-01	9,967087E-01	9,934283E-01
17	1,85E-02	9,25E-05	9,62E-03	9,36E-04	5,89E-03	8,33E-01	9,968947E-01	9,937991E-01
18	3,20E-02	1,60E-04	1,27E-02	1,24E-03	7,95E-03	1,08E+00	9,970606E-01	9,941298E-01
19	2,88E-02	1,44E-04	1,20E-02	1,17E-03	7,21E-03	1,01E+00	9,967993E-01	9,936089E-01
20	2,06E-02	1,03E-04	1,02E-02	9,88E-04	5,94E-03	8,40E-01	9,968389E-01	9,936877E-01

Entre as 20 topologias testadas para o protótipo IV, a topologia com 7 neurônios na camada escondida (topologia 2-7-1) foi a que apresentou maior número de simulações com o conjunto de validação com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9975107 e R_g^2 = 0,9950276). Entretanto, a topologia contendo 6 neurônios na camada escondida (topologia 2-6-1) foi selecionada como a melhor por ser mais simples (menor número de neurônios) e apresentar capacidade de generalização e robustez (*medmed_R*=0,9974312 e R_g^2 = 0,9948690) praticamente iguais a da rede com 7 neurônios na camada escondida.

As curvas dos erros de validação do protótipo IV (figuras 82 e 83) apresentam um rápido declínio com o aumento do número de neurônios escondidos e então começam a se estabilizar a partir de 4 neurônios. A curva do coeficiente de determinação em função do número de neurônios escondido tem um comportamento oposto, cresce rapidamente até o quarto neurônio quando estabiliza em torno de 1.

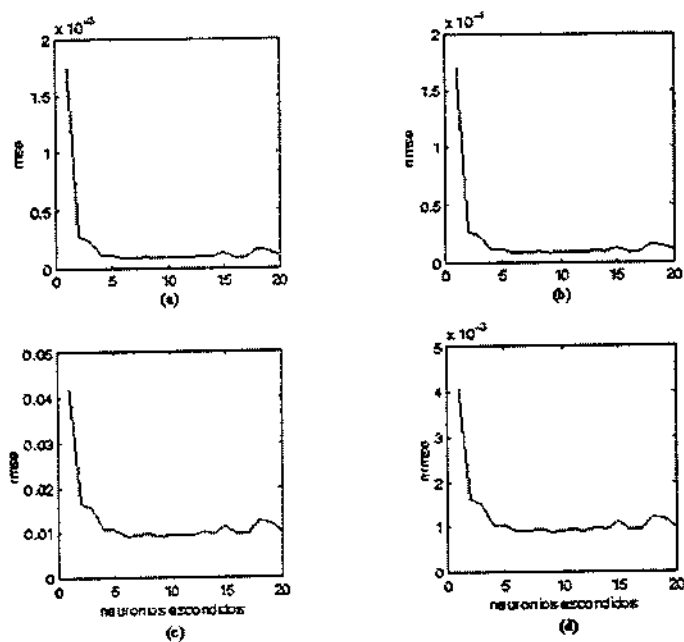


Figura 82. Protótipo IV - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nrmse$ usando conjunto de validação em função do número de neurônios escondidos.

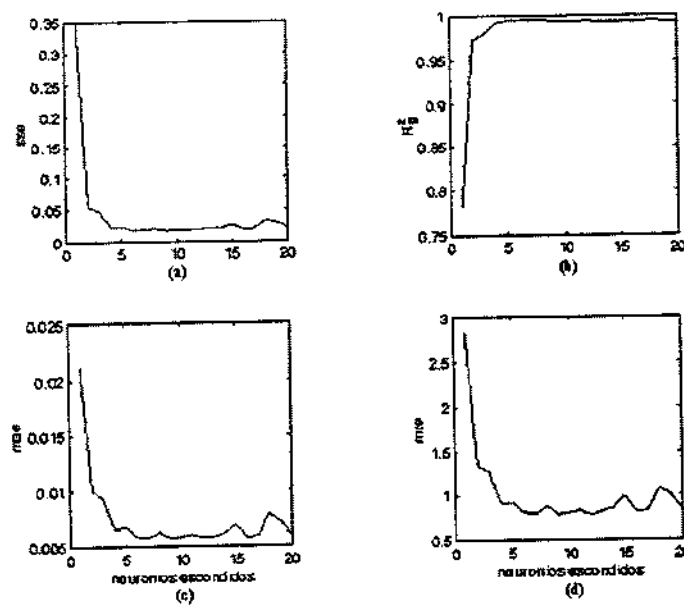


Figura 83. Protótipo IV - Índices de avaliação do desempenho (a) sse (b) Rg^2 (c) mae e (d) mre usando conjunto de validação em função do número de neurônios escondidos.

Os valores de atividade de emulsificação para emulsões água-em-hexadecano, estimados pelo protótipo IV com topologia 2-6-1, apresentaram um ajuste perfeito aos dados do conjunto de treinamento conforme pode ser verificado através das figuras 84a e 84c ($R=1$ e $R_g^2=0,9993816$). Os valores de atividade de emulsificação para emulsões água-em-hexadecano, estimados pelo protótipo IV com topologia 2-6-1, apresentaram um ajuste quase perfeito aos dados do conjunto de validação conforme pode ser verificado através das figuras 84b e 84d ($R=0,998$ e $R_g^2=0,9948690$).

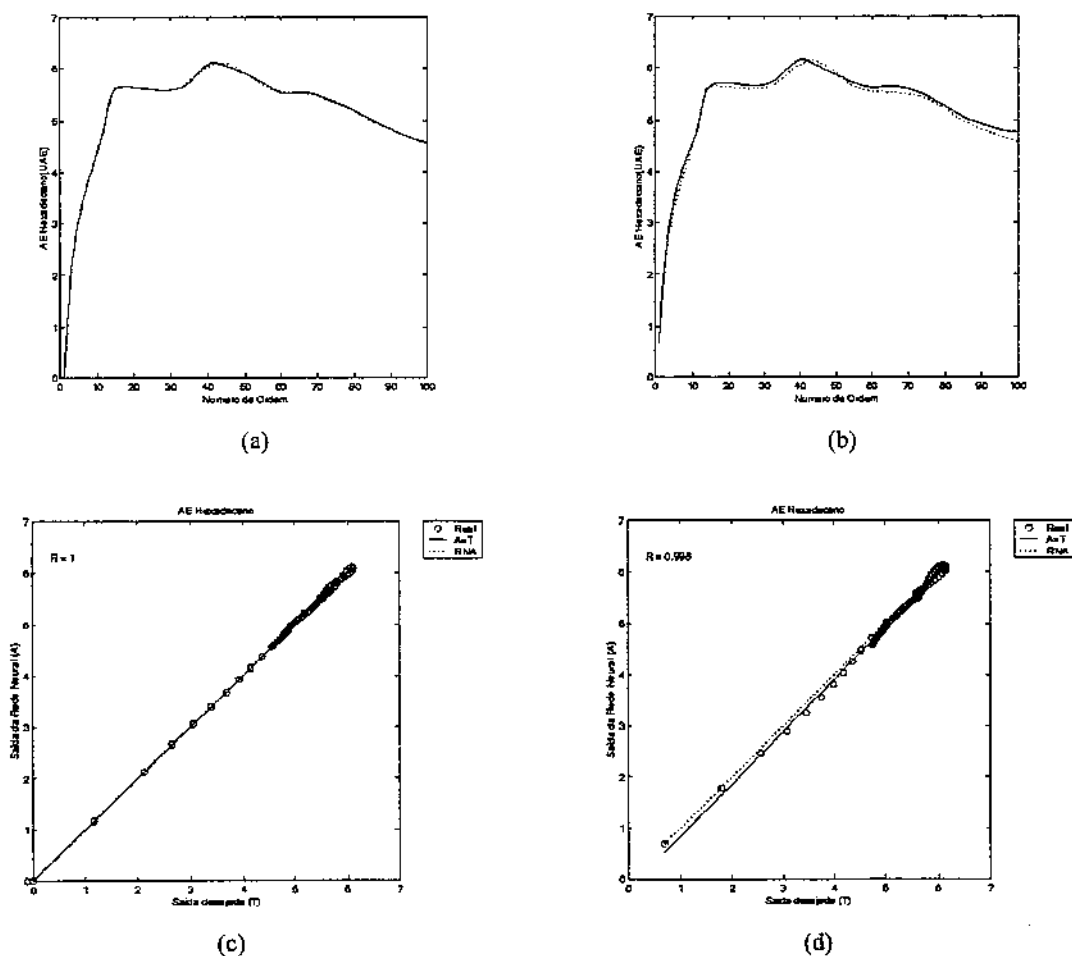


Figura 84. Protótipo IV - Comparação (perfis cinéticos) entre valores de saída experimentais (—) e gerados pela rede (---) para atividade de emulsificação para emulsão água-em-hexadecano (a) usando dados do conjunto de treinamento (b) usando dados do conjunto de validação. Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para concentração de biomassa (c) usando dados do conjunto de treinamento (d) usando dados do conjunto de validação.

A variação do número de épocas e dos parâmetros de treinamento (*sse*, *mse*, *ssx*, *np* e *gamk*) do protótipo IV em função do número de neurônios escondidos pode ser analisada através da tabela 50 e da figura 85.

Tabela 50. Protótipo IV - Parâmetros de treinamento da rede (*época*, *sse*, *mse*, *ssx*, *np* e *gamk*) em função do número de neurônios (*SI*) na camada escondida.

<i>SI</i>	<i>época</i>	<i>sse</i>	<i>mse</i>	<i>ssx</i>	<i>np</i>	<i>gamk</i>
1	19	2,59E-01	1,30E-03	6,41E+02	5,00E+00	4,47E+00
2	98	2,61E-02	1,30E-04	2,37E+02	9,00E+00	6,94E+00
3	282	2,05E-02	1,03E-04	1,96E+02	1,30E+01	9,72E+00
4	1000	3,98E-03	1,99E-05	9,24E+02	1,70E+01	1,30E+01
5	1000	3,05E-03	1,52E-05	9,73E+02	2,10E+01	1,55E+01
6	489	9,92E-04	4,96E-06	2,73E+03	2,50E+01	1,88E+01
7	528	9,99E-04	5,00E-06	2,25E+03	2,90E+01	1,84E+01
8	264	9,95E-04	4,98E-06	2,58E+03	3,30E+01	2,02E+01
9	491	9,99E-04	4,99E-06	3,08E+03	3,70E+01	2,08E+01
10	244	9,99E-04	4,99E-06	2,72E+03	4,10E+01	2,13E+01
11	179	9,98E-04	4,99E-06	4,61E+03	4,50E+01	2,28E+01
12	190	9,96E-04	4,98E-06	4,40E+03	4,90E+01	2,52E+01
13	148	1,00E-03	5,00E-06	5,72E+03	5,30E+01	2,50E+01
14	86	1,00E-03	5,00E-06	5,83E+03	5,70E+01	2,57E+01
15	67	9,92E-04	4,96E-06	6,59E+03	6,10E+01	2,77E+01
16	94	9,97E-04	4,98E-06	9,22E+03	6,50E+01	2,74E+01
17	101	9,74E-04	4,87E-06	9,99E+03	6,90E+01	2,63E+01
18	154	9,98E-04	4,99E-06	1,17E+04	7,30E+01	2,72E+01
19	205	9,90E-04	4,95E-06	1,36E+04	7,70E+01	2,83E+01
20	70	9,19E-04	4,60E-06	1,44E+04	8,10E+01	2,91E+01

Os parâmetros *sse* e *gamk* começam a se estabilizar a partir do sexto neurônio escondido. Neste caso, a melhor topologia selecionada usando o critério sugerido por Foresee e Hagan (1997) – coincidiu com a determinada usando o critério de seleção adotado no presente trabalho. O parâmetro *ssx*, no entanto cresce continuamente não estabilizando junto com o *sse* e o *gamk* como sugerem Foresee e Hagan (1997).

A figura 86 apresenta a evolução dos parâmetros de treinamento *sse*, *ssx* e *gamk* em função do número de épocas, para a topologia 2-6-1, selecionada como a melhor do protótipo IV.

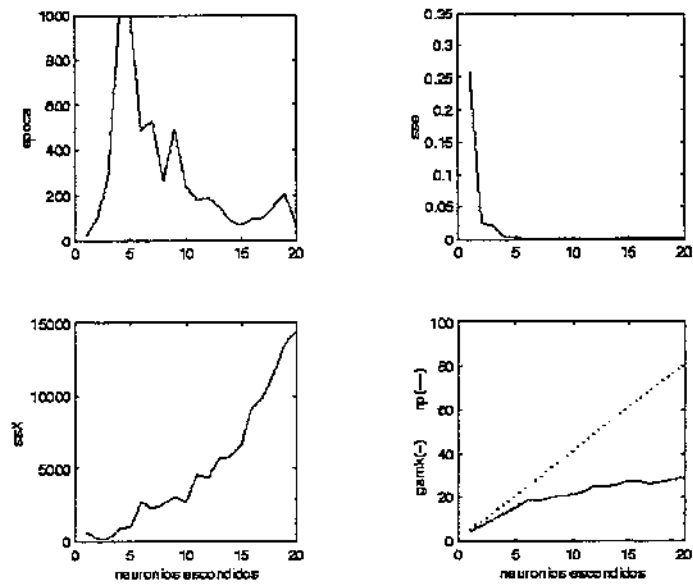


Figura 85. Protótipo IV - Época e parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de neurônios escondidos.

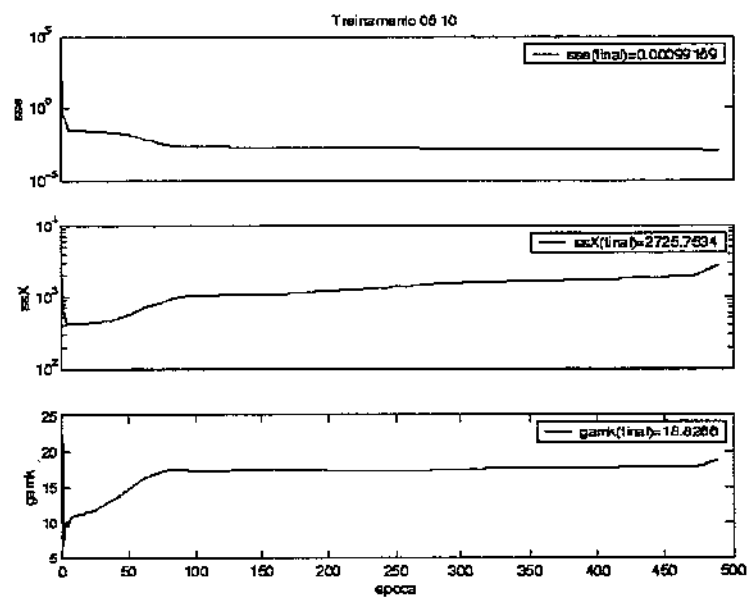


Figura 86. Protótipo IV - Parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de épocas.

A topologia 2-6-1 do protótipo IV foi capaz de fornecer excelentes estimativas da atividade de emulsificação para emulsões água-em-hexadecano usando o conjunto de validação. Logo, um *softsensor* neural para estimação de atividade de emulsificação para emulsões água-em-hexadecano, representado pela equação (31), pode ser implementado *on-line* usando as matrizes de pesos (IW^1 e LW^2) e bias ($b1$ e $b2$) da camada escondida e da camada de saída apresentadas nas tabelas 46 e 47.

$$a = \text{purelin}(LW^2 \logsig(IW^1 p + b1) + b2) \quad (31)$$

onde: a - é a matriz dos padrões de saída do *softsensor* IV

purelin - é a função de transferência linear usada na camada de saída

logsig - é a função de transferência sigmoideal usada na camada escondida

p - é a matriz dos padrões de entrada do *softsensor* IV

Tabela 51. Matrizes de pesos e bias da camada escondida do *softsensor* para inferência da atividade de emulsificação para emulsões água-em-hexadecano treinado com 100 padrões entrada-saída (Protótipo IV com 6 neurônios na camada escondida).

Camada escondida		
Matriz de pesos (IW^1)		bias (b^1)
1,78E+00	-5,28E+00	1,47E+00
-3,05E+01	-1,14E-01	1,00E+01
-1,16E+01	-1,22E+00	1,09E+01
-3,21E+01	7,53E-01	1,04E+01
3,69E+00	-5,33E+00	-5,23E-02
-1,10E+01	2,76E+00	3,53E+00

Tabela 52. Matrizes de pesos e bias da camada de saída do *softsensor* para inferência da atividade de emulsificação para emulsões água-em-hexadecano treinado com 100 padrões entrada-saída (Protótipo IV com 6 neurônios na camada escondida).

Camada de saída					
Matriz de pesos (LW^2)					bias (b^2)
-4,77E+00	-4,41E+00	1,48E+00	3,83E+00	4,84E+00	1,91E+00
					-5,59E-01

5.3.1.5. Protótipo V – *Softsensor* para inferência de atividade de emulsificação para emulsões água-em-óleo de milho

Medidas *on-line* de pH e de oxigênio dissolvido obtidas a 28° C e 150 rpm foram usadas como entradas para o protótipo V (Figura 40). A atividade de emulsificação para emulsões água-em-óleo de milho foi selecionada como variável de saída. Vinte diferentes topologias do protótipo V foram avaliadas usando conjunto de treinamento composto por 100 padrões de entrada-saída. Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando conjunto de treinamento encontram-se na tabela 53.

Tabela 53. Protótipo V - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de treinamento.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	1,52E-01	7,62E-04	2,76E-02	2,52E-03	1,56E-02	1,92E+00	9,617842E-01	9,250288E-01
2	1,15E-01	5,73E-04	2,39E-02	2,18E-03	1,37E-02	1,83E+00	9,737064E-01	9,481042E-01
3	8,24E-02	4,12E-04	2,03E-02	1,85E-03	1,15E-02	1,41E+00	9,884055E-01	9,769453E-01
4	3,36E-03	1,68E-05	4,10E-03	3,74E-04	2,37E-03	3,06E-01	9,989823E-01	9,979656E-01
5	2,38E-03	1,19E-05	3,45E-03	3,15E-04	1,74E-03	2,34E-01	9,994860E-01	9,989723E-01
6	1,42E-03	7,11E-06	2,67E-03	2,43E-04	1,45E-03	1,92E-01	9,997227E-01	9,994455E-01
7	9,96E-04	4,98E-06	2,23E-03	2,04E-04	1,24E-03	1,64E-01	9,997424E-01	9,994849E-01
8	1,28E-03	6,39E-06	2,53E-03	2,31E-04	1,33E-03	1,77E-01	9,997040E-01	9,994081E-01
9	9,96E-04	4,98E-06	2,23E-03	2,04E-04	1,23E-03	1,61E-01	9,997428E-01	9,994857E-01
10	9,97E-04	4,98E-06	2,23E-03	2,04E-04	1,24E-03	1,64E-01	9,997559E-01	9,995118E-01
11	9,99E-04	4,99E-06	2,23E-03	2,04E-04	1,17E-03	1,57E-01	9,997519E-01	9,995039E-01
12	9,92E-04	4,96E-06	2,23E-03	2,03E-04	1,22E-03	1,64E-01	9,997527E-01	9,995055E-01
13	9,94E-04	4,97E-06	2,23E-03	2,04E-04	1,20E-03	1,70E-01	9,997565E-01	9,995131E-01
14	9,97E-04	4,98E-06	2,23E-03	2,04E-04	1,16E-03	1,55E-01	9,997566E-01	9,995133E-01
15	9,87E-04	4,94E-06	2,22E-03	2,03E-04	1,17E-03	1,57E-01	9,997564E-01	9,995128E-01
16	9,94E-04	4,97E-06	2,23E-03	2,03E-04	1,21E-03	1,65E-01	9,997566E-01	9,995133E-01
17	9,68E-04	4,84E-06	2,20E-03	2,01E-04	1,18E-03	1,65E-01	9,997585E-01	9,995171E-01
18	9,98E-04	4,99E-06	2,23E-03	2,04E-04	1,13E-03	1,65E-01	9,997620E-01	9,995241E-01
19	9,96E-04	4,98E-06	2,23E-03	2,04E-04	1,21E-03	1,75E-01	9,997576E-01	9,995153E-01
20	9,99E-04	4,99E-06	2,23E-03	2,04E-04	1,17E-03	1,60E-01	9,997593E-01	9,995187E-01

Entre as 20 topologias avaliadas para o protótipo V, a topologia com 18 neurônios na camada escondida foi a que apresentou maior número de simulações, com o conjunto de treinamento, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9997620 e R_g^2 = 0,9995241). Nas figuras 87 e 88, estão ilustrados, os comportamentos das funções de erro de treinamento e do coeficiente de determinação global em função do número de neurônios escondidos.

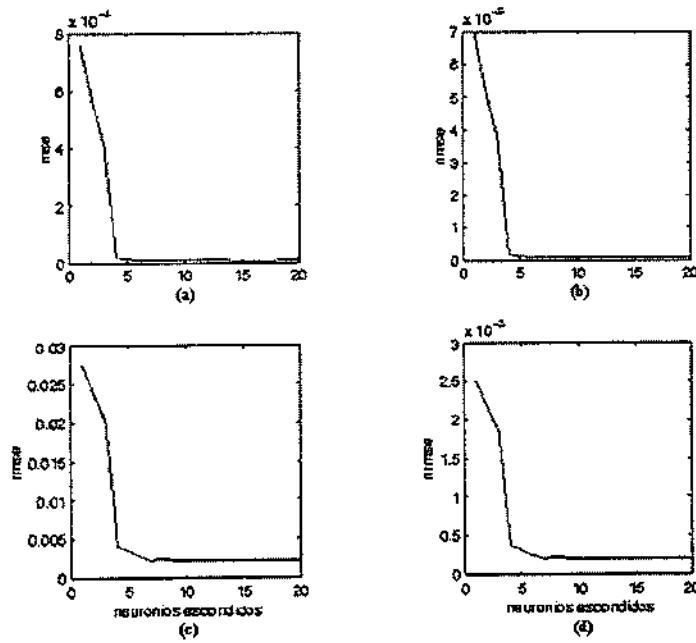


Figura 87. Protótipo V - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nrmse$ usando conjunto de treinamento em função do número de neurônios escondidos.

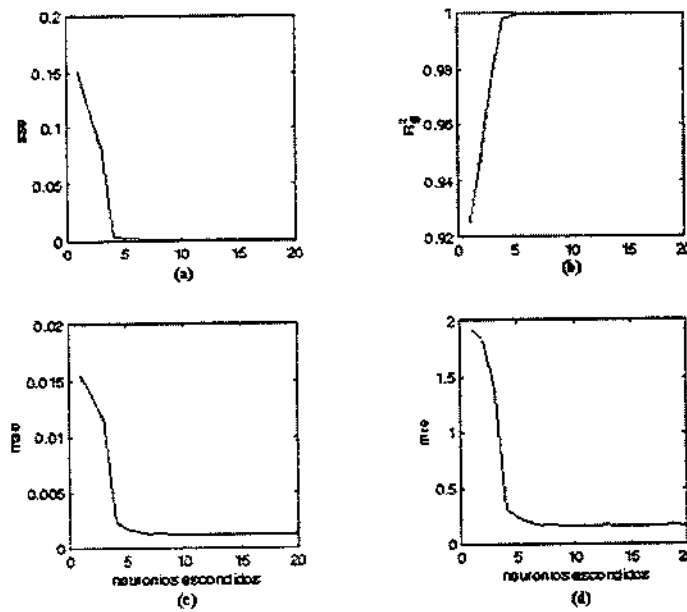


Figura 88. Protótipo V - Índices de avaliação do desempenho (a) sse (b) Rg^2 (c) mae e (d) mre usando conjunto de treinamento em função do número de neurônios escondidos.

Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação estão apresentados na tabela 54.

Tabela 54. Protótipo IV - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	3,48E-01	1,74E-03	4,17E-02	3,74E-03	2,40E-02	3,88E+00	9,531006E-01	9,084007E-01
2	2,95E-01	1,47E-03	3,84E-02	3,44E-03	2,18E-02	3,67E+00	9,621816E-01	9,257934E-01
3	2,67E-01	1,34E-03	3,66E-02	3,28E-03	2,03E-02	3,43E+00	9,776035E-01	9,557087E-01
4	1,68E-01	8,41E-04	2,90E-02	2,59E-03	1,33E-02	2,65E+00	9,875182E-01	9,751923E-01
5	1,66E-01	8,32E-04	2,88E-02	2,59E-03	1,27E-02	2,57E+00	9,884780E-01	9,770887E-01
6	1,84E-01	9,21E-04	3,04E-02	2,73E-03	1,45E-02	2,80E+00	9,889720E-01	9,780656E-01
7	1,68E-01	8,39E-04	2,90E-02	2,60E-03	1,31E-02	2,62E+00	9,889446E-01	9,780114E-01
8	1,73E-01	8,65E-04	2,94E-02	2,64E-03	1,37E-02	2,69E+00	9,890693E-01	9,782581E-01
9	1,79E-01	8,96E-04	2,99E-02	2,69E-03	1,41E-02	2,76E+00	9,892708E-01	9,786567E-01
10	1,80E-01	9,02E-04	3,00E-02	2,70E-03	1,44E-02	2,79E+00	9,887832E-01	9,776923E-01
11	1,73E-01	8,66E-04	2,94E-02	2,64E-03	1,33E-02	2,66E+00	9,888822E-01	9,778879E-01
12	1,65E-01	8,23E-04	2,87E-02	2,58E-03	1,30E-02	2,61E+00	9,889493E-01	9,780207E-01
13	1,64E-01	8,21E-04	2,87E-02	2,57E-03	1,28E-02	2,55E+00	9,893568E-01	9,788269E-01
14	1,79E-01	8,96E-04	2,99E-02	2,69E-03	1,42E-02	2,76E+00	9,879015E-01	9,759494E-01
15	1,95E-01	9,76E-04	3,12E-02	2,80E-03	1,38E-02	2,73E+00	9,864111E-01	9,730068E-01
16	1,81E-01	9,06E-04	3,01E-02	2,70E-03	1,45E-02	2,78E+00	9,874107E-01	9,749799E-01
17	1,65E-01	8,23E-04	2,87E-02	2,57E-03	1,26E-02	2,55E+00	9,883006E-01	9,767381E-01
18	1,38E-01	6,91E-04	2,63E-02	2,35E-03	1,02E-02	2,20E+00	9,888120E-01	9,777492E-01
19	1,61E-01	8,03E-04	2,83E-02	2,54E-03	1,26E-02	2,54E+00	9,871906E-01	9,745453E-01
20	1,41E-01	7,07E-04	2,66E-02	2,38E-03	1,05E-02	2,26E+00	9,893078E-01	9,787299E-01

Entre as 20 topologias analisadas para o protótipo V, a topologia com 13 neurônios na camada escondida (topologia 2-13-1) foi a que apresentou maior número de simulações, com o conjunto de validação, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9893568 e R_g^2 =0,9788269). Entretanto, a topologia contendo 6 neurônios na camada escondida (topologia 2-6-1) foi selecionada como a melhor por ser mais simples (menor número de neurônios) e apresentar capacidade de generalização e robustez (*medmed_R*=0,9889720 e R_g^2 =0,9780656) praticamente iguais a da rede com 13 neurônios na camada escondida.

Os comportamentos das funções de erro de validação e do coeficiente de determinação global em função do número de neurônios escondidos estão ilustrados nas figuras 89 e 90.

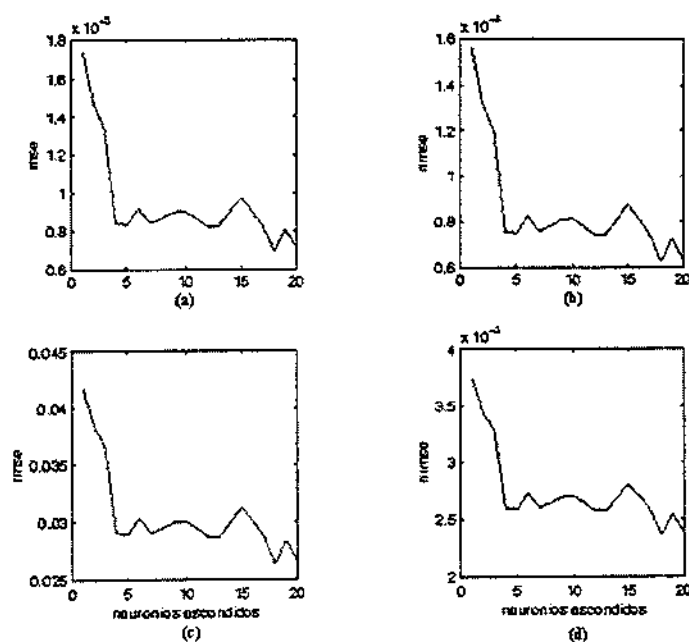


Figura 89. Protótipo V - Índices de avaliação do desempenho (a) *mse* (b) *nmse* (c) *rmse* e (d) *nrmse* usando conjunto de validação em função do número de neurônios escondidos

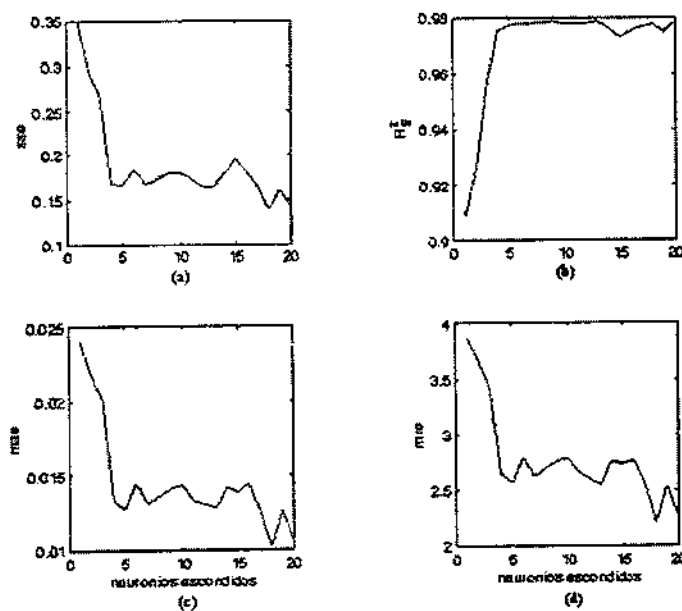


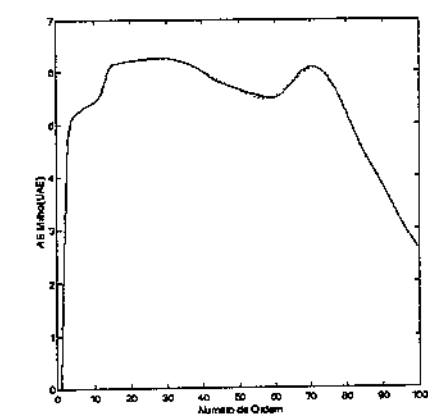
Figura 90. Protótipo V - Índices de avaliação do desempenho (a) *sse* (b) Rg^2 (c) *mae* e (d) *mre* usando conjunto de validação em função do número de neurônios escondidos.

Conforme se pode observar através das figuras 90 e 91, as trajetórias das funções de erro usando o conjunto de validação não são tão suaves quanto às obtidas usando o conjunto de treinamento. As funções de erro declinam rapidamente até o quinto neurônio escondido, quando passam a apresentar um comportamento oscilatório com o aumento do número de neurônios escondidos. A curva do coeficiente de determinação global cresce rapidamente até o sexto neurônio, quando estabiliza em torno de 0,98.

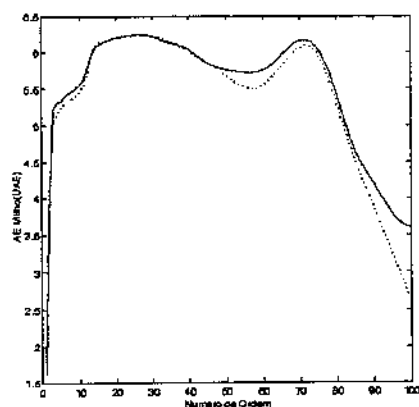
Os valores de atividade de emulsificação para emulsões água-em-óleo de milho, estimados pelo protótipo V com topologia 2-6-1, apresentaram um excelente ajuste aos dados do conjunto de treinamento, conforme pode ser verificado através das figuras 91a e 91c ($R=1$ e $R_g^2=0,9994455$). Os valores de atividade de emulsificação para emulsões água-em-hexadecano, estimados pelo protótipo V com topologia 2-6-1, apresentaram um bom ajuste aos dados do conjunto de validação, conforme pode ser verificado através das figuras 84b e 84d ($R=0,999$ e $R_g^2=0,9780656$).

A variação do número de épocas e dos parâmetros de treinamento (*sse*, *mse*, *ssx*, *np* e *gamk*) do protótipo V em função do número de neurônios escondidos pode ser observada através da tabela 55 e da figura 92.

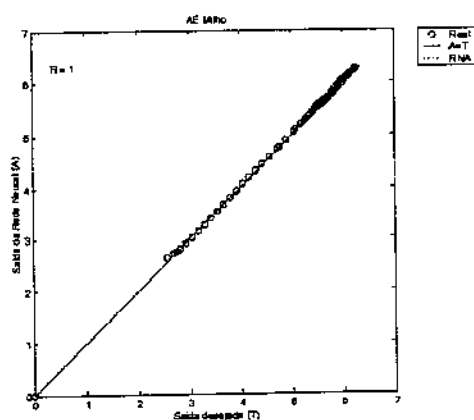
Os parâmetros *sse* e *gamk* começam a se estabilizar a partir do sexto neurônio escondido. Neste caso, a melhor topologia selecionada usando o critério sugerido por Foresee e Hagan (1997) – coincidiu com a determinada usando o critério de seleção adotado no presente trabalho. O parâmetro *ssx*, no entanto cresce continuamente não estabilizando junto com o *sse* e o *gamk* como sugerem Foresee e Hagan (1997).



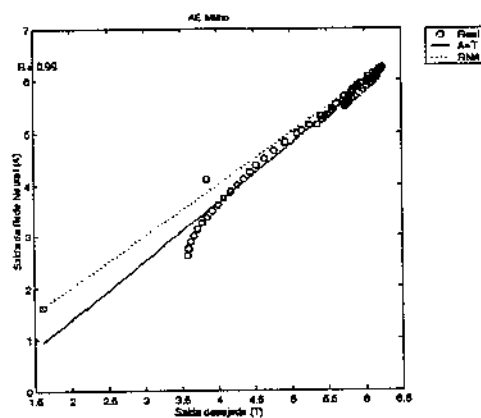
(a)



(b)



(c)



(d)

Figura 91. Protótipo V - Comparação (perfis cinéticos) entre valores de saída experimentais (—) e gerados pela rede (---) para atividade de emulsificação para emulsão água-em-óleo de milho (a) usando dados do conjunto de treinamento (b) usando dados do conjunto de validação. Comparação (ajuste linear) entre valores de saída experimentais e gerados pela rede para concentração de biomassa (c) usando dados do conjunto de treinamento (d) usando dados do conjunto de validação.

Tabela 55. Protótipo V - Parâmetros de treinamento da rede (*época*, *sse*, *mse*, *ssx*, *np* e *gamk*) em função do número de neurônios (*SI*) na camada escondida.

<i>SI</i>	<i>época</i>	<i>sse</i>	<i>mse</i>	<i>ssx</i>	<i>np</i>	<i>gamk</i>
1	30	1,52E-01	7,62E-04	3,72E+03	5	4,96E+00
2	187	1,15E-01	5,73E-04	5,72E+02	9	7,20E+00
3	195	8,24E-02	4,12E-04	2,80E+03	13	9,82E+00
4	591	3,36E-03	1,68E-05	3,50E+03	17	1,47E+01
5	1000	2,38E-03	1,19E-05	1,36E+03	21	1,52E+01
6	763	1,42E-03	7,11E-06	2,08E+03	25	1,92E+01
7	452	9,96E-04	4,98E-06	2,39E+03	29	2,02E+01
8	1000	1,28E-03	6,39E-06	1,57E+03	33	2,00E+01
9	322	9,96E-04	4,98E-06	2,45E+03	37	2,22E+01
10	409	9,97E-04	4,98E-06	3,00E+03	41	2,27E+01
11	301	9,99E-04	4,99E-06	3,76E+03	45	2,28E+01
12	179	9,92E-04	4,96E-06	3,44E+03	49	2,39E+01
13	211	9,94E-04	4,97E-06	4,76E+03	53	2,46E+01
14	208	9,97E-04	4,98E-06	4,93E+03	57	2,62E+01
15	268	9,87E-04	4,94E-06	6,44E+03	61	2,46E+01
16	327	9,94E-04	4,97E-06	7,67E+03	65	2,88E+01
17	488	9,68E-04	4,84E-06	9,91E+03	69	2,77E+01
18	49	9,98E-04	4,99E-06	1,18E+04	73	2,89E+01
19	178	9,96E-04	4,98E-06	1,30E+04	77	2,91E+01
20	138	9,99E-04	4,99E-06	1,43E+04	81	3,12E+01

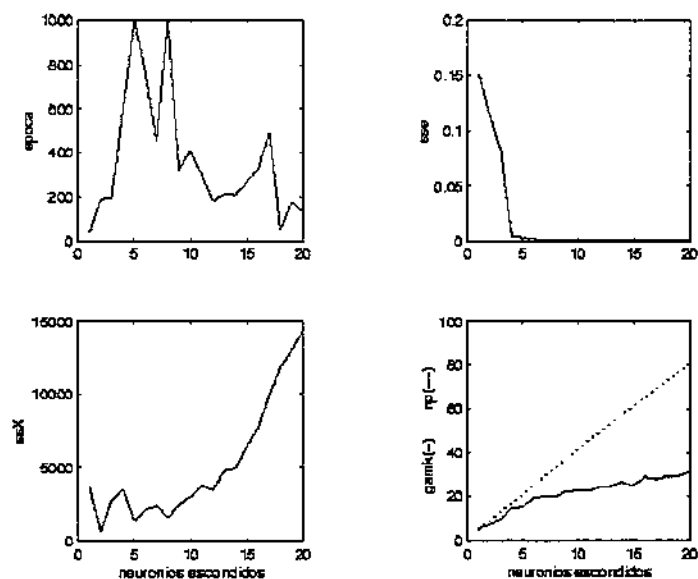


Figura 92. Protótipo V - Época e parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de neurônios escondidos.

A figura 93 apresenta a evolução dos parâmetros de treinamento *sse*, *ssx* e *gamk* em função do número de épocas, para a topologia 2-6-1, selecionada como a melhor do protótipo V.

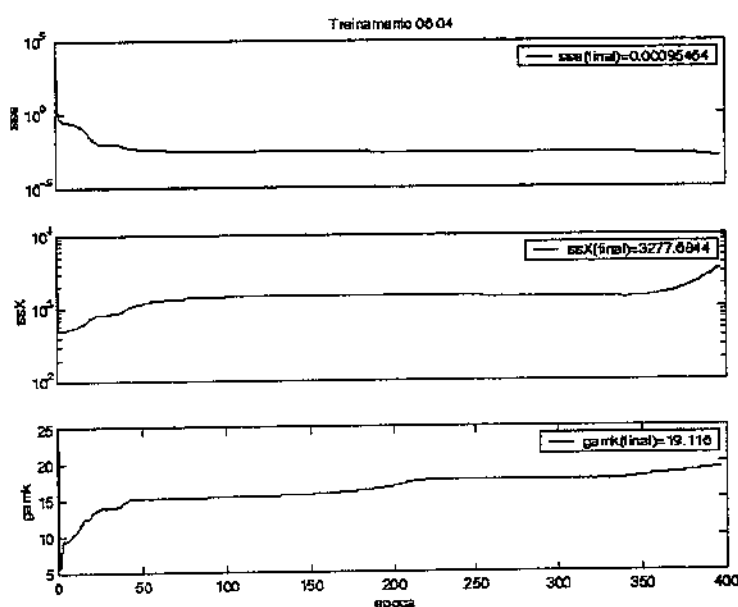


Figura 93. Protótipo V - Parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de épocas.

A topologia 2-6-1 do protótipo V foi capaz de fornecer excelentes estimativas da atividade de emulsificação para emulsões água-em-óleo de milho usando o conjunto de validação. Logo, um *softsensor* neural para estimação de atividade de emulsificação para emulsões água-em-óleo de milho, representado pela equação (32), pode ser implementado *on-line* usando as matrizes de pesos (IW^1 e LW^2) e bias ($b1$ e $b2$) da camada escondida e da camada de saída apresentadas nas tabelas 56 e 57.

$$a = \text{purelin}(LW^2 \logsig(IW^1 p + b1) + b2) \quad (32)$$

onde: a - é a matriz dos padrões de saída do *softsensor* V
purelin - é a função de transferência linear usada na camada de saída
logsig - é a função de transferência sigmoideal usada na camada escondida
 p - é a matriz dos padrões de entrada do *softsensor* V

Tabela 56. Matrizes de pesos e bias da camada escondida do *softsensor* para inferência da atividade de emulsificação para emulsões água-em-óleo-de milho treinado com 100 padrões entrada-saída (Protótipo V com 6 neurônios na camada escondida).

Camada escondida		
Matriz de pesos (IW^1)		bias (b^1)
-1,01E+01	9,23E+00	4,98E-01
-1,92E+01	-1,92E+00	1,52E+01
-3,57E+01	-7,14E-01	1,32E+01
4,13E+00	-8,82E+00	1,82E+00
8,69E+00	1,12E+01	-1,62E+00
-4,27E+00	1,24E+01	-6,66E+00

Tabela 57. Matrizes de pesos e bias da camada de saída do *softsensor* para inferência da atividade de emulsificação para emulsões água-em-óleo-de milho treinado com 100 padrões entrada-saída (Protótipo V com 6 neurônios na camada escondida).

Camada de saída						
Matriz de pesos (LW^2)						bias (b^2)
9,15E+00	2,17E-01	-1,39E-01	5,54E-01	1,41E+01	-6,00E+00	-1,40E+01

5.3.1.6. Protótipo VI – *Softsensor* para inferência de atividade de emulsificação para emulsões água-em-óleo de canola

Medidas *on-line* de pH e de oxigênio dissolvido obtidas a 28° C e 150 rpm foram usadas como entradas para o protótipo VI (Figura 40). A atividade de emulsificação para emulsões água-em-óleo de canola foi selecionada como variável de saída. Vinte diferentes topologias do protótipo VI foram avaliadas usando conjunto de treinamento composto por 100 padrões de entrada-saída. Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando conjunto de treinamento encontram-se na tabela 58.

Tabela 58. Protótipo VI - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de treinamento.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	1,92E-01	9,62E-04	3,10E-02	2,83E-03	1,68E-02	2,15E+00	9,043734E-01	8,178912E-01
2	1,87E-01	9,36E-04	3,06E-02	2,79E-03	1,65E-02	2,15E+00	9,241765E-01	8,541021E-01
3	9,46E-03	4,73E-05	6,88E-03	6,27E-04	4,08E-03	4,99E-01	9,955317E-01	9,910834E-01
4	6,25E-03	3,13E-05	5,59E-03	5,10E-04	3,12E-03	3,87E-01	9,887742E-01	9,776745E-01
5	9,98E-04	4,99E-06	2,23E-03	2,04E-04	1,27E-03	1,58E-01	9,903476E-01	9,807883E-01
6	9,29E-04	4,64E-06	2,15E-03	1,96E-04	1,09E-03	1,39E-01	9,995488E-01	9,990978E-01
7	2,71E-03	1,35E-05	3,68E-03	3,35E-04	1,70E-03	2,16E-01	9,991303E-01	9,982613E-01
8	9,99E-04	4,99E-06	2,23E-03	2,04E-04	1,14E-03	1,46E-01	9,995517E-01	9,991036E-01
9	9,86E-04	4,93E-06	2,22E-03	2,02E-04	1,12E-03	1,42E-01	9,995434E-01	9,990869E-01
10	8,79E-04	4,40E-06	2,10E-03	1,91E-04	1,11E-03	1,46E-01	9,995555E-01	9,991112E-01
11	9,94E-04	4,97E-06	2,23E-03	2,03E-04	1,06E-03	1,35E-01	9,995493E-01	9,990987E-01
12	9,93E-04	4,97E-06	2,23E-03	2,03E-04	1,26E-03	1,56E-01	9,995372E-01	9,990745E-01
13	9,79E-04	4,89E-06	2,21E-03	2,02E-04	1,08E-03	1,37E-01	9,995720E-01	9,991441E-01
14	9,83E-04	4,92E-06	2,22E-03	2,02E-04	1,06E-03	1,34E-01	9,995473E-01	9,990947E-01
15	9,96E-04	4,98E-06	2,23E-03	2,04E-04	1,07E-03	1,38E-01	9,995361E-01	9,990724E-01
16	9,99E-04	4,99E-06	2,23E-03	2,04E-04	1,07E-03	1,36E-01	9,995414E-01	9,990830E-01
17	7,60E-04	3,80E-06	1,95E-03	1,78E-04	1,03E-03	1,36E-01	9,995494E-01	9,990989E-01
18	9,89E-04	4,95E-06	2,22E-03	2,03E-04	1,21E-03	1,60E-01	9,995385E-01	9,990772E-01
19	9,94E-04	4,97E-06	2,23E-03	2,03E-04	1,04E-03	1,34E-01	9,995420E-01	9,990842E-01
20	9,80E-04	4,90E-06	2,21E-03	2,02E-04	1,25E-03	1,51E-01	9,995418E-01	9,990839E-01

Entre as 20 topologias avaliadas para o protótipo VI, a topologia com 13 neurônios na camada escondida foi a que apresentou maior número de simulações, com o conjunto de treinamento, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,995720 e R_g^2 =0,9915441). Nas figuras 94 e 95 encontram-se ilustrados os comportamentos das funções de erro de treinamento e do coeficiente de determinação global em função do número de neurônios escondidos.

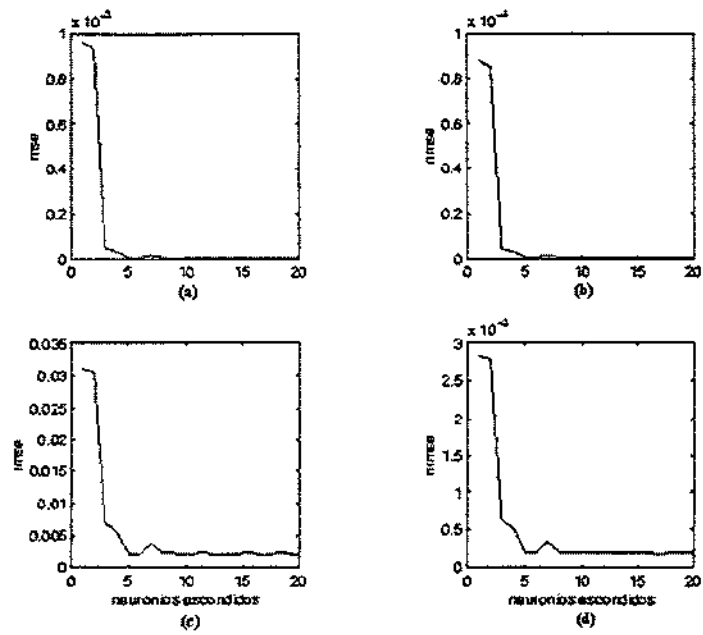


Figura 94. Protótipo VI - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nmrmse$ usando conjunto de treinamento em função do número de neurônios escondidos.

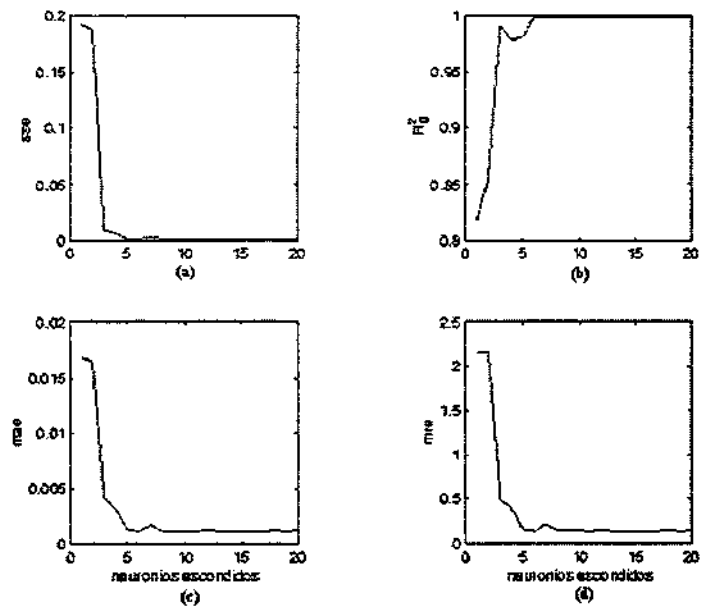


Figura 95. Protótipo VI - Índices de avaliação do desempenho (a) sse (b) Rg^2 (c) mae e (d) mre usando conjunto de treinamento em função do número de neurônios escondidos.

Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação estão apresentados na Tabela 59.

Tabela 59. Protótipo VI - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	4,51E-01	2,26E-03	4,75E-02	4,31E-03	2,45E-02	3,43E+00	8,382581E-01	7,026766E-01
2	4,46E-01	2,23E-03	4,72E-02	4,28E-03	2,44E-02	3,47E+00	8,652169E-01	7,486003E-01
3	2,00E-01	1,00E-03	3,16E-02	2,87E-03	1,66E-02	2,31E+00	9,547534E-01	9,115540E-01
4	1,98E-01	9,89E-04	3,14E-02	2,86E-03	1,58E-02	2,22E+00	9,434974E-01	8,901874E-01
5	1,88E-01	9,39E-04	3,06E-02	2,78E-03	1,57E-02	2,20E+00	9,464270E-01	8,957241E-01
6	1,90E-01	9,48E-04	3,08E-02	2,80E-03	1,57E-02	2,21E+00	9,612229E-01	9,239495E-01
7	2,03E-01	1,01E-03	3,18E-02	2,89E-03	1,57E-02	2,22E+00	9,601442E-01	9,218768E-01
8	2,04E-01	1,02E-03	3,19E-02	2,90E-03	1,55E-02	2,20E+00	9,594503E-01	9,205450E-01
9	1,89E-01	9,47E-04	3,08E-02	2,80E-03	1,59E-02	2,22E+00	9,602707E-01	9,221198E-01
10	1,64E-01	8,20E-04	2,86E-02	2,60E-03	1,39E-02	1,97E+00	9,605736E-01	9,227015E-01
11	1,90E-01	9,50E-04	3,08E-02	2,80E-03	1,48E-02	2,10E+00	9,601675E-01	9,219216E-01
12	1,81E-01	9,03E-04	3,01E-02	2,73E-03	1,48E-02	2,09E+00	9,609314E-01	9,233892E-01
13	1,59E-01	7,96E-04	2,82E-02	2,56E-03	1,38E-02	1,94E+00	9,589974E-01	9,196760E-01
14	2,00E-01	1,00E-03	3,16E-02	2,87E-03	1,51E-02	2,15E+00	9,595941E-01	9,208209E-01
15	1,85E-01	9,25E-04	3,04E-02	2,76E-03	1,48E-02	2,10E+00	9,583502E-01	9,184352E-01
16	1,82E-01	9,11E-04	3,02E-02	2,74E-03	1,47E-02	2,08E+00	9,611973E-01	9,239002E-01
17	1,64E-01	8,19E-04	2,86E-02	2,60E-03	1,42E-02	2,00E+00	9,590747E-01	9,198243E-01
18	1,99E-01	9,95E-04	3,15E-02	2,87E-03	1,61E-02	2,27E+00	9,584017E-01	9,185339E-01
19	1,82E-01	9,12E-04	3,02E-02	2,74E-03	1,46E-02	2,07E+00	9,593201E-01	9,202950E-01
20	1,78E-01	8,92E-04	2,99E-02	2,71E-03	1,44E-02	2,04E+00	9,600146E-01	9,216280E-01

Entre as 20 topologias analisadas para o protótipo VI, a topologia com 6 neurônios na camada escondida (topologia 2-6-1) foi a que apresentou maior número de simulações, com o conjunto de validação, com maiores coeficientes de correlação para as quatro variáveis de saída simultaneamente (*medmed_R*=0,9612229 e R_g^2 = 0,9239495). Entretanto, a topologia contendo 5 neurônios na camada escondida (topologia 2-5-1) também poderia ser selecionada como a melhor por ser mais simples (menor número de neurônios) e apresentar capacidade de generalização e robustez (*medmed_R* = 0,9464270 e R_g^2 = 0,8957241) similar a da rede com 6 neurônios na camada escondida.

Os comportamentos das funções de erro de validação e do coeficiente de determinação global em função do número de neurônios escondidos estão ilustrados nas figuras 96 e 97.

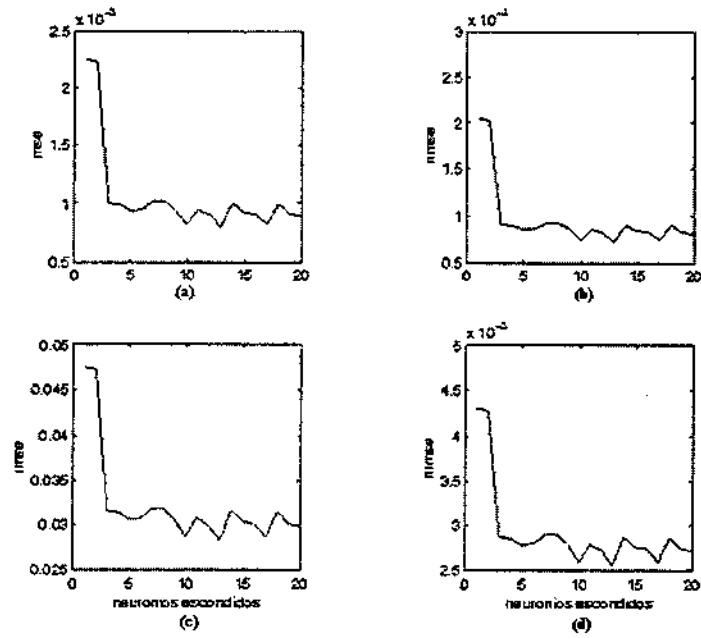


Figura 96. Protótipo VI - Índices de avaliação do desempenho (a) *mse* (b) *nmse* (c) *rmse* e (d) *nrmse* usando conjunto de validação em função do número de neurônios escondidos.

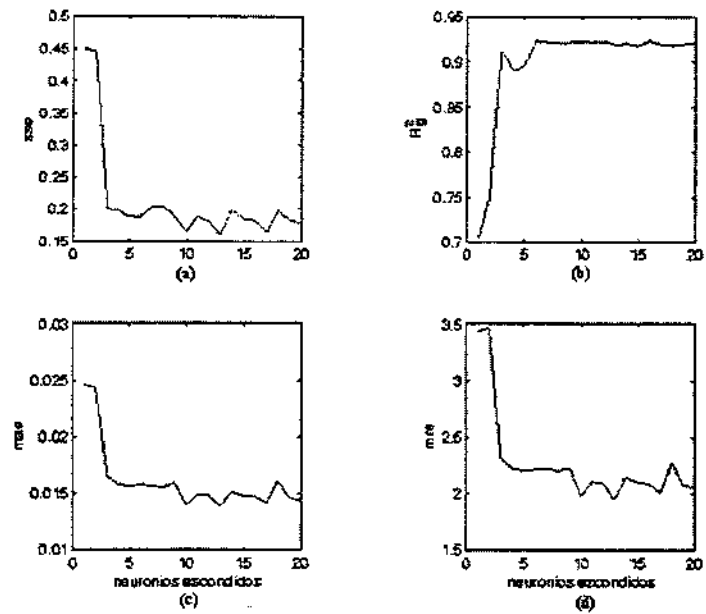


Figura 97. Protótipo VI - Índices de avaliação do desempenho (a) *sse* (b) Rg^2 (c) *mae* e (d) *mre* usando conjunto de validação em função do número de neurônios escondidos.

As trajetórias das funções de erro do protótipo VI, usando o conjunto de validação, não são suaves. As funções de erro declinam rapidamente até o terceiro neurônio escondido, apresentam uma breve estabilidade até o sexto neurônio e passam então a apresentar um comportamento oscilatório com o aumento do número de neurônios escondidos. A curva do coeficiente de determinação global cresce rapidamente até o terceiro neurônio escondido, decresce no quarto - formando um pico no terceiro neurônio - e estabiliza em torno de 0,92 a partir do quinto neurônio escondido..

Os valores de atividade de emulsificação para emulsões água-em-óleo de canola, estimados pelo protótipo VI com topologia 2-6-1, apresentaram um excelente ajuste aos dados do conjunto de treinamento, conforme pode ser verificado através das figuras 98a e 98c ($R=1$ e $R_g^2=0,9990978$). Os valores de atividade de emulsificação para emulsões água-em-hexadecano, estimados pelo protótipo V com topologia 2-6-1, apresentaram um bom ajuste aos dados do conjunto de validação, conforme pode ser verificado através das figuras 98b e 98d ($R=0,964$ e $R_g^2=0,9239495$).

A variação do número de épocas e dos parâmetros de treinamento (sse , mse , ssx , np e $gamk$) do protótipo VI em função do número de neurônios escondidos pode ser observada através da tabela 60 e da figura 99.

Os parâmetros sse e $gamk$ começam a se estabilizar a partir do quinto/sexto neurônio escondido. Neste caso, a melhor topologia selecionada usando o critério sugerido por Foresee e Hagan (1997) – coincidiu com a determinada usando o critério de seleção adotado no presente trabalho. O parâmetro ssx , no entanto, cresce continuamente não estabilizando junto com o sse e o $gamk$, como sugerem Foresee e Hagan (1997).

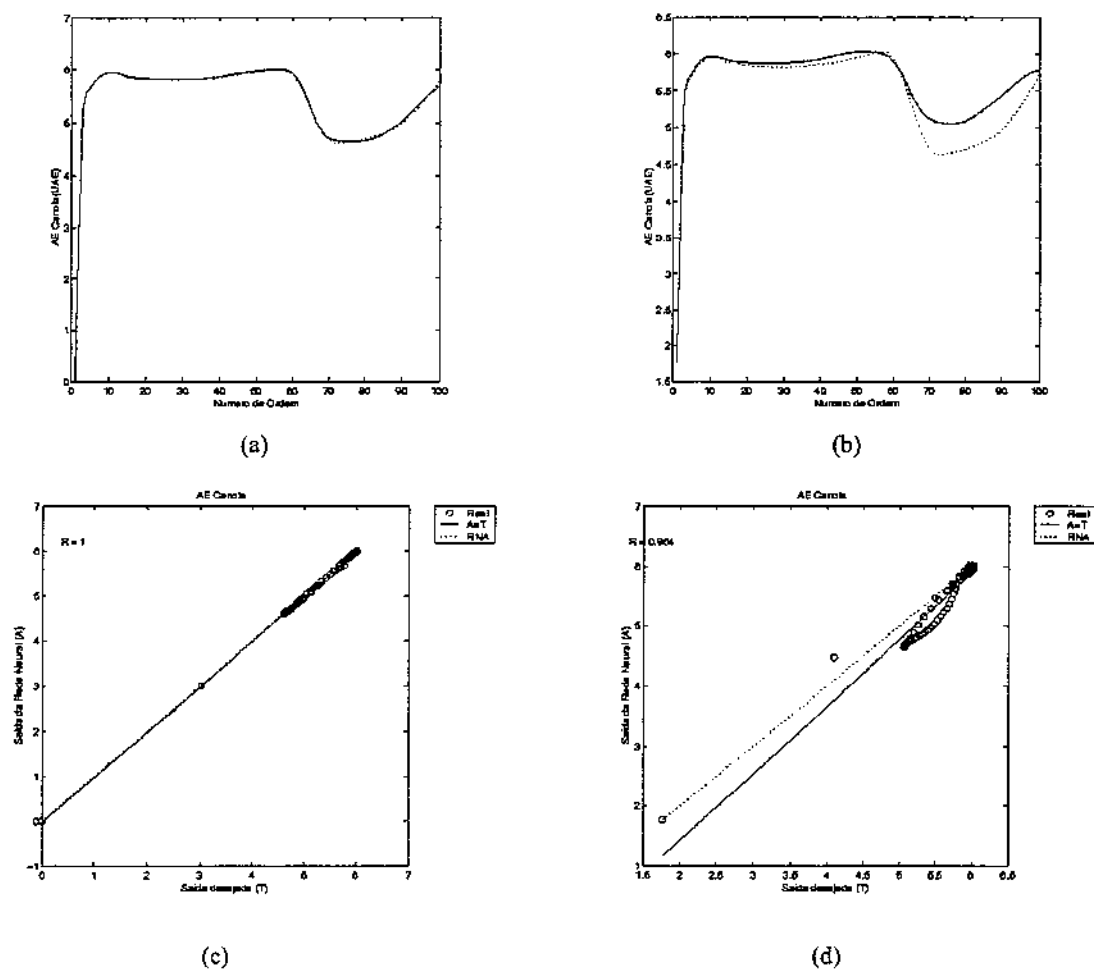


Figura 98. Protótipo VI - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para atividade de emulsificação para emulsão água-em-óleo de canola (a) usando dados do conjunto de treinamento (b) usando dados do conjunto de validação. Comparação (ajuste linear) entre valores de saída experimentais e gerados pela rede para concentração de biomassa (c) usando dados do conjunto de treinamento (d) usando dados do conjunto de validação.

Tabela 60. Protótipo VI - Parâmetros de treinamento da rede (*época*, *sse*, *mse*, *ssx*, *np* e *gamk*) em função do número de neurônios (*SI*) na camada escondida.

<i>SI</i>	<i>época</i>	<i>sse</i>	<i>mse</i>	<i>ssx</i>	<i>np</i>	<i>gamk</i>
1	32	1,92E-01	9,62E-04	8,80E+01	5	4,30E+00
2	75	1,87E-01	9,36E-04	5,35E+01	9	5,87E+00
3	188	9,46E-03	4,73E-05	1,22E+03	13	9,12E+00
4	259	6,25E-03	3,13E-05	9,04E+02	17	1,27E+01
5	205	9,98E-04	4,99E-06	6,45E+03	21	1,65E+01
6	221	9,29E-04	4,64E-06	4,01E+03	25	1,71E+01
7	757	2,71E-03	1,35E-05	2,86E+03	29	1,99E+01
8	252	9,99E-04	4,99E-06	2,78E+03	33	1,98E+01
9	222	9,86E-04	4,93E-06	2,98E+03	37	2,07E+01
10	89	8,79E-04	4,40E-06	3,02E+03	41	2,17E+01
11	217	9,94E-04	4,97E-06	4,71E+03	45	2,11E+01
12	88	9,93E-04	4,97E-06	5,98E+03	49	2,10E+01
13	122	9,79E-04	4,89E-06	4,44E+03	53	2,30E+01
14	297	9,83E-04	4,92E-06	5,92E+03	57	2,79E+01
15	124	9,96E-04	4,98E-06	7,30E+03	61	2,59E+01
16	120	9,99E-04	4,99E-06	1,05E+04	65	2,34E+01
17	64	7,60E-04	3,80E-06	1,11E+04	69	2,88E+01
18	36	9,89E-04	4,95E-06	1,04E+04	73	2,83E+01
19	32	9,94E-04	4,97E-06	1,30E+04	77	2,70E+01
20	24	9,80E-04	4,90E-06	1,36E+04	81	3,02E+01

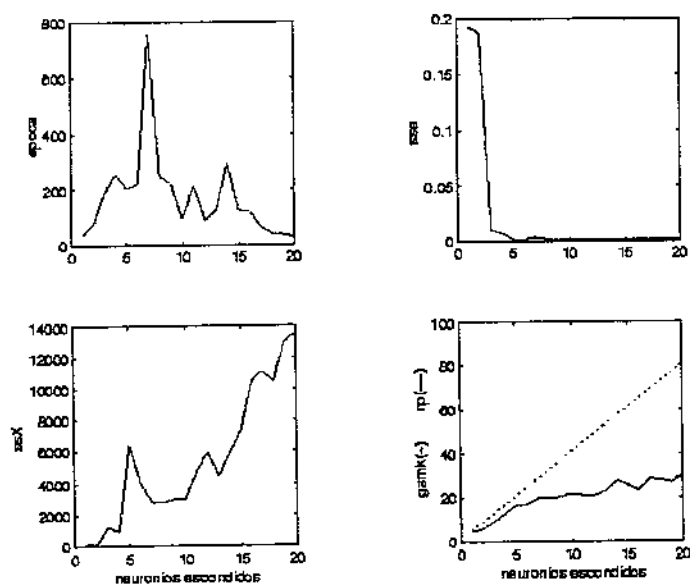


Figura 99. Protótipo VI - Época e parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de neurônios escondidos.

Na figura 93 é apresentada a evolução dos parâmetros de treinamento *sse*, *ssx* e *gamk* em função do número de épocas, para a topologia 2-6-1, selecionada como a melhor topologia do protótipo VI.

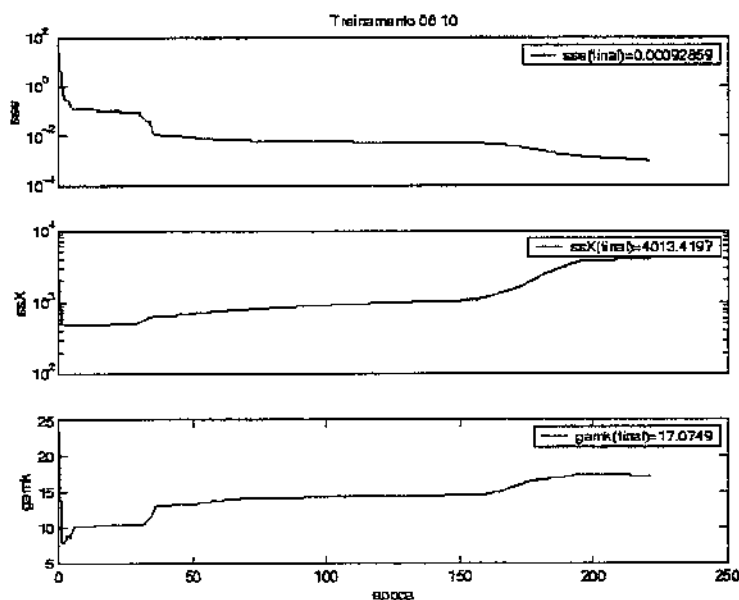


Figura 100. Protótipo VI - Parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de épocas.

A topologia 2-6-1 do protótipo VI foi capaz de fornecer excelentes estimativas da atividade de emulsificação para emulsões água-em-óleo de milho usando o conjunto de validação. Logo, um *softsensor* neural para estimação de atividade de emulsificação para emulsões água-em-óleo de milho, representado pela equação (33), pode ser implementado *on-line* usando as matrizes de pesos (IW^1 e LW^2) e bias ($b1$ e $b2$) da camada escondida e da camada de saída apresentadas nas tabelas 61 e 62.

$$a = \text{purelin}(LW^2 \text{logsig}(IW^1 p + b1) + b2) \quad (33)$$

onde: a - é a matriz dos padrões de saída do *softsensor* VI
 purelin - é a função de transferência linear usada na camada de saída
 logsig - é a função de transferência sigmoidal usada na camada escondida
 p - é a matriz dos padrões de entrada do *softsensor* VI

Tabela 61. Matrizes de pesos e bias da camada escondida do *softsensor* para inferência da atividade de emulsificação para emulsões água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo VI com 6 neurônios na camada escondida).

Camada escondida		
Matriz de pesos (IW^1)		bias (b^1)
-1,39E+01	1,36E+01	2,69E+00
-1,62E+01	1,36E+01	-2,05E+00
-1,65E+01	1,41E+01	4,44E+00
8,92E+00	-4,63E+00	-4,32E-01
2,46E+01	-3,03E+00	-6,55E+00
-2,10E+00	1,19E+01	-6,64E+00

Tabela 62. Matrizes de pesos e bias da camada de saída do *softsensor* para inferência da atividade de emulsificação para emulsões água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo VI com 6 neurônios na camada escondida).

Camada de saída						
Matriz de pesos (LW^2)					bias (b^2)	
7,65E+00	1,21E+01	-2,65E+00	7,10E+00	2,14E+00	-5,93E+00	-8,34E+00

5.3.1.7. Comparação dos desempenhos dos protótipos

A comparação das principais características dos protótipos de *softsensores* selecionados para serem usados no processo de produção de biossufactante por *Candida lipolytica* encontram-se sintetizadas na tabela 63.

Tabela 63. Número de neurônios escondidos, número de épocas, coeficiente de ajuste/generalização e robustez e coeficiente de determinação global dos protótipos de *softsensores* neurais I, II, III, IV, V e VI.

Protótipo	SI	épocas	medmed_R	Rg ²
I	4	803	0,9798977	0,9601996
II (100 padrões)	7	458	0,9831673	0,966179
II (1000 padrões)	7	1000	0,9773317	0,9551771
III	4	234	0,9844187	0,9690801
IV	7	528	0,9975107	0,9950276
V	6	763	0,9889720	0,9780656
VI	6	221	0,9612229	0,9239495

Conforme pode-se observar através da tabela 63 todos os protótipos apresentaram coeficientes de generalização (*medmed_R*) superiores a 0,96 e coeficientes de determinação global (R_g^2) superiores a 0,92. Resultados que atestam o potencial de uso destes protótipos de *softsensores* neurais na estimação e previsão *on-line* de biomassa e atividade de emulsificação no processo de produção de bioemulsificante por *Candida lipolytica*. Melhores resultados ainda poderiam ter sido obtidos se os dados de densidade ótica e concentração de biomassa tivessem sido ajustados utilizando função sigmoidal-Boltzman. Isto entretanto não foi feito para que os dados de todas as variáveis de entrada e saída tivessem um único tratamento. Assim como as curvas de atividade de emulsificação são altamente não lineares e não poderiam ser ajustadas com a função sigmoidal-Boltzman, as curvas de densidade ótica e biomassa tiveram o mesmo tratamento que elas, ou seja eliminação de *outliers* e suavização e ampliação por meio de *spline* cúbica, conforme descrito no item 5.2.2. do presente capítulo.

Os protótipos I e II são constituídos por múltiplas entradas e múltiplas saídas. O protótipo I teve como entradas todas as variáveis - medidas através de sensores *on-line* (pH e oxigênio dissolvido) ou de análises *off-line* (densidade ótica e salinidade do cultivo livre

de células) - que poderiam influenciar as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola. Ambos os protótipos apresentaram coeficientes de generalização ($medmed-R$) e coeficiente de determinação global (R_g^2) superiores a 0,97 demonstrando elevada capacidade preditiva nas condições estudadas.

Estes dois protótipos foram projetados para investigar a influência de informações adicionais do processo - representadas pelas medidas de densidade ótica e salinidade do cultivo livre de células provenientes de análise *off-line* - sobre a capacidade preditiva dos protótipos. Nas condições operacionais estudadas, mais informação não aumentou a capacidade preditiva do protótipo I, provavelmente por conta dos ruídos introduzidos pela medidas *off-line*, mas simplificou a estrutura do mesmo. Apesar do bom resultado do protótipo I, a inclusão da densidade ótica e da salinidade do líquido metabólico livre de células no seu padrão de entrada não é realista, desde que se assume que estas duas variáveis estão disponíveis *on-line*, o que não é o caso no momento, já que o Núcleo de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco não dispõe de sensores *on-line* para elas e a aquisição de novos sensores elevaria o custo de produção de bioemulsificante por *Candida lipolytica*. O protótipo II mesmo tendo como entradas apenas pH e oxigênio dissolvido - variáveis medidas *on-line* - e as mesmas variáveis de saída do protótipo I, apresentou melhor capacidade de generalização que o protótipo I, sugerindo com isso a adequação e suficiência destas duas variáveis para fazer inferência da concentração de biomassa e das atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola.

Com o objetivo de avaliar a influência do número de padrões de entrada-saída sobre a capacidade de generalização, outro teste foi realizado com o protótipo II. O resultado deste teste, que pode ser verificado na tabela 63, mostrou que, neste caso, o aumento do número de padrões entrada-saída não produziu aumento na capacidade preditiva do protótipo II. Quando 100 padrões de entrada-saída foram apresentados ao protótipo II, o coeficiente de determinação global (R_g^2) foi igual a 0,966179. Quando 1000 padrões de entrada-saída foram apresentados ao protótipo II, o coeficiente de determinação global (R_g^2) foi igual a 0,9551771.

Os protótipos III, IV, V e VI foram projetados com duas variáveis de entrada e uma variável de saída, para verificar se produziram melhores resultados que o protótipo II, projetado com duas variáveis de entrada e quatro variáveis de saída. Os protótipos III, IV, V e VI foram projetados com as mesmas variáveis de entrada do protótipo II (pH e oxigênio dissolvido) e com apenas uma das variáveis de saída do protótipo II (biomassa ou atividade de emulsificação para emulsão água-em hexadecano ou atividade de emulsificação para emulsão água-em-óleo de milho ou atividade de emulsificação para emulsão água-em-óleo de canola). Conforme pode-se verificar na tabela 63, com execução do protótipo VI ($mmed_R=0,9612229$; $R_g^2=0,9239495$), todos os outros protótipos (III, IV e V) com duas variáveis de entrada e uma simples variável de saída apresentaram melhor capacidade de generalização ($mmed_R>0,9844187$; $R_g^2>0,9690801$) que o protótipo II com duas variáveis de entrada e quatro variáveis de saída. Entretanto, a diferença entre as capacidades de generalização dos protótipos III, IV, V e VI em relação ao protótipo II não foi significativa. Portanto, em comparação aos protótipos III, IV, V e VI, o protótipo II treinado com 100 padrões de entrada-saída, pelo seu bom desempenho ($mmed_R=0,9831673$; $R_g^2=0,966179$) e sua capacidade de prever mais de uma variável de saída - mostrou ser o protótipo de *softsensor* com maior potencial para ser selecionado e implementado no processo de produção de bioemulsificante por *Candida lipolytica*. A integração em uma plataforma em tempo real dos protótipos de *softsensores* neurais inteiramente especificados, selecionados após simulações usando dados do conjunto de teste (Figura 35 - dados da terceira replicata do ensaio experimental realizado a 28° C e 150 rpm) e dados adicionais de novos ensaios experimentais, conforme descrito a seguir no item 5.3.2., facilitará a eliminação de defeitos latentes e permitirá uma melhor avaliação das previsões dos protótipos de *softsensores* neurais.

Entre as principais vantagens da implementação de um *softsensor* neural para inferência de concentração de biomassa e de atividade de emulsificação no Núcleo de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco podemos citar: a simplicidade e o baixo custo de implementação do *softsensor* neural e a redução do custo global do processo de produção de bioemulsificante por *Candida lipolytica* proveniente do monitoramento, otimização e controle em tempo real das variáveis concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola.

5.3.1.8. Novos critérios para seleção dos melhores *softsensores* neurais

A principal dificuldade na modelagem de sistemas através de redes neurais é a identificação da topologia (números de neurônios na(s) camada(s) escondida(s)) mais apropriada para o sucesso da aplicação. A definição da arquitetura ideal é determinada geralmente através de métodos de tentativa e erro aliados a experiência e sensibilidade do projetista. Algoritmos genéticos também vêm sendo usado na determinação da melhor arquitetura, assim como técnicas que buscam identificar a melhor arquitetura através da diminuição de neurônios redundantes e do compartilhamento de pesos.

No presente trabalho, além dos modelos de *softsensores* identificados acredita-se que também foram obtidos como resultados dois novos critérios (métodos) para seleção da melhor topologia de um determinado protótipo de *softsensor* neural:

- *Coefficientes de ajuste/generalização e robustez ($medmed_R$) e coeficiente de determinação global (R_g^2)*

Para selecionar a melhor topologia para os protótipos de *softsensores* baseado em redes neurais propostos neste trabalho, avaliou-se o maior espaço de arquiteturas possíveis dentro de uma perspectiva prática. Ou seja, procurou-se minimizar o aspecto arbitrário da especificação do número de neurônios, ampliando-se o espaço do número de neurônios da camada escondida. Para cada protótipo de *softsensor*, topologias com número de neurônios escondidos variando de 1 a 20 foram treinadas 10 vezes e avaliadas através de simulações com conjuntos de treinamento e de validação, com apoio do programa **NeuronArTech** desenvolvido durante o presente trabalho. O desempenho de cada protótipo foi avaliado através de várias funções de erro, de técnicas de visualização e especialmente através dos coeficientes $medmed_R$ e R_g^2 , definidos no presente trabalho como medidas da capacidade de generalização/ajuste e robustez. A melhor topologia de um determinado protótipo foi selecionada como sendo aquela com menor número de neurônios escondidos que apresentasse maior coeficiente $medmed_R$ ou R_g^2 .

- *Primeiro pico (máximo) da curva do número de épocas em função do número de neurônios*

A avaliação do acerto ou adequação da melhor topologia selecionada para cada protótipo proposto foi inicialmente ratificada através de comparação visual, em relação as topologias não selecionadas, do ajuste das curvas de saída simuladas usando dados do conjunto de validação. Posteriormente, percebeu-se que a melhor topologia selecionada usando o coeficiente de generalização (*medmed-R*) sempre coincidia com ou era vizinha imediata da topologia após o primeiro pico (máximo) da curva do número de épocas em função do número de neurônios escondidos. Este fato pode ser constatado rápida e facilmente através das tabelas de parâmetros de treinamento da rede (*época, sse, mse, ssx, np e gamk*) em função do número de neurônios (*SI*) na camada escondida dos protótipos I (Tabela 30), II-100 padrões (Tabela 35), II-1000 padrões (Tabela 40), III (Tabela 45), IV (Tabela 50), V (Tabela 55) e VI (Tabela 60). Como estes dois critérios são totalmente diferentes e independentes, um reforça a adequação do outro como método para seleção da topologia ótima de uma determinada rede neural.

5.3.2. Testes dos protótipos de *softsensores* neurais

Dados do conjunto de teste, ou seja, dados da terceira replicata do ensaio realizado a 28°C e 150 rpm (Figura 35), foram usados para avaliar o desempenho (capacidade preditiva) dos protótipos de *softsensores* inteiramente especificados, representados pelas equações 27, 28, 29, 30, 31, 32 e 33. Todos os seis protótipos de *softsensores* apresentaram estimações para biomassa e atividades de emulsificação dentro de uma faixa aceitável de 5% dos valores experimentais e coeficientes de determinação global (R_g^2) superiores a 0,92, atestando as elevadas capacidades de generalização de todos os modelos de redes neurais propostos, treinados e validados no presente trabalho. Ainda assim, existe espaço para melhoramentos dos protótipos, devido às restrições e variações provenientes da quantidade de ensaios realizados e do comportamento não linear e variante no tempo dos dados experimentais. Portanto, novos ensaios experimentais e testes adicionais deverão ser realizados para selecionar confiável e eficientemente o(s) protótipo(s) mais adequado(s) para ser(em) instalado(s) no processo de produção de biossurfactante por *C. lipolytica*.

5.3.3. Ratificação da suficiência das variáveis dos padrões de entrada

A importância do pH, do oxigênio dissolvido, da densidade ótica e da salinidade do líquido metabólico livre de células para o processo de produção de bioemulsificante por *Candida lipolytica*, foi constatada em experimentos realizados durante a ampliação de escala do processo (Capítulo 4). Portanto, estas variáveis foram usadas nos padrões de entrada dos seis protótipos propostos neste trabalho e produziram bons resultados (Itens 5.3.1. e 5.3.2.) na estimação de concentração de biomassa e de atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola, usando dados experimentais obtidos em biorreator, operado em batelada, a 28° C e 150 rpm. Portanto, novos treinamentos e simulações com dados experimentais obtidos em biorreator, operado em batelada, a 29,5 °C e 225 rpm (Tabela 18 – ponto central planejamento 2²/Figuras 33 e 35 – Base de Dados) - foram realizados para confirmar a adequação e a suficiência das referidas variáveis de entrada para fazer inferência da concentração de biomassa e atividade(s) de emulsificação de bioemulsificante produzido por *Candida lipolytica* no processo estudado.

Como foi definido anteriormente (item 5.2.4.), o protótipo I tem como variáveis de entrada pH, concentração de oxigênio dissolvido, densidade ótica e salinidade do líquido metabólico livre de células, enquanto os demais protótipos (II, III, IV, V e VI) têm como variáveis de entrada apenas pH e concentração de oxigênio dissolvido. Os protótipos I e II foram selecionados para confirmar a adequação e a suficiência das variáveis de entrada, pois estimam simultaneamente quatro variáveis de saída (concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola), enquanto os demais protótipos estimam apenas uma variável de saída (concentração de biomassa ou atividade de emulsificação para emulsões água-em-hexadecano ou atividade de emulsificação para emulsões água-em-óleo de milho ou atividade de emulsificação para emulsões água-em-óleo de canola).

A metodologia para treinamento, validação e teste dos protótipos e os critérios de seleção de número ótimo de neurônios escondidos adotados a seguir foram os mesmos descritos anteriormente neste trabalho. Para ratificar a capacidade de predição dos

protótipos I e II, topologias com número de neurônios escondidos variando de 1 a 20 foram treinadas e validadas 10 vezes cada uma. A seleção do número ótimo de neurônios dos protótipos baseou-se nos dois critérios discutidos no item 5.3.1.8.. De acordo com o primeiro destes critérios, a melhor topologia de determinado protótipo será aquela que - na simulação com o conjunto de validação para o menor número de neurônios escondidos possível - apresentar o maior coeficiente de ajuste/generalização e robustez ($medmed_R$) ou equivalentemente o maior coeficiente correlação global (R_g^2). Nos treinamentos dos protótipos, os maiores coeficientes de generalização e robustez, muitas vezes, estão associados a altos números de neurônios escondidos, o que pode indicar sobreajuste da rede. Portanto, a topologia determinada como a melhor, será aquela que apresentar na validação os maiores coeficientes de generalização e robustez e menor erro para o menor número de neurônios escondidos possível. O critério do primeiro pico da curva do número de épocas em função do número de neurônios foi usado para confirmar a escolha do número ótimo de neurônios.

5.3.3.1. Protótipo I – *Softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola

Os resultados dos índices de avaliação de desempenho (sse , mse , $rmse$, $nrmse$, mae , mre , $medmed_R$, R_g^2) em função do número de neurônios (SI) na camada escondida, obtidos usando o conjunto de treinamento estão apresentados na tabela 64.

Entre as 20 topologias testadas para o protótipo I, a topologia com 19 neurônios na camada escondida foi a que apresentou maior número de simulações, com o conjunto de treinamento, com maiores coeficientes de ajuste e robustez ($medmed_R=0,9999745$) e de determinação global ($R_g^2=0,9998490$). As trajetórias das funções de erro mse , $rmse$, $nrmse$, mae e mre são típicas do processo de treinamento de qualquer rede neural, onde existe um rápido declínio do erro com o aumento do número de neurônios escondidos, conforme se pode observar através das figuras 101 e 102. As trajetórias das funções de erro $nrmse$ e sse têm um rápido declínio até o terceiro neurônio e depois apresentam um comportamento oscilatório com o aumento do número de neurônios. A curva do coeficiente de determinação global cresce rapidamente até o terceiro neurônio quando começa a se estabilizar em torno de 0,99.

Tabela 64. Protótipo I - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, Rg^2) em função do número de neurônios(*SI*) na camada escondida,obtidos usando o conjunto de treinamento.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	Rg^2
1	1,12E+03	7,69E-03	8,77E-02	1,26E-03	8,19E-01	5,91E+03	7,444260E-01	5,541701E-01
2	1,93E+02	4,85E-03	6,96E-02	1,00E-03	4,83E-01	5,54E+03	8,770871E-01	7,692818E-01
3	8,82E+01	7,82E-04	2,80E-02	4,03E-04	2,18E-01	5,03E+02	9,584999E-01	9,187221E-01
4	1,22E+02	9,55E-05	9,77E-03	1,41E-04	1,11E-01	2,38E+02	9,955848E-01	9,911890E-01
5	2,69E+02	2,09E-04	1,44E-02	2,08E-04	1,04E-01	1,19E+03	9,966378E-01	9,932869E-01
6	2,09E+02	1,80E-05	4,24E-03	6,11E-05	4,26E-02	2,91E+02	9,980766E-01	9,961569E-01
7	1,86E+02	3,15E-06	1,78E-03	2,56E-05	1,66E-02	9,39E+01	9,989043E-01	9,978097E-01
8	5,54E+02	2,50E-06	1,58E-03	2,28E-05	1,35E-02	5,96E+01	9,996343E-01	9,992688E-01
9	3,49E+02	2,47E-06	1,57E-03	2,26E-05	1,31E-02	6,51E+01	9,998433E-01	9,996865E-01
10	8,42E+02	2,48E-06	1,57E-03	2,27E-05	1,25E-02	5,40E+01	9,999380E-01	9,998760E-01
11	3,71E+02	2,49E-06	1,58E-03	2,27E-05	1,09E-02	5,69E+01	9,999440E-01	9,998881E-01
12	9,00E+02	2,49E-06	1,58E-03	2,27E-05	1,00E-02	6,33E+01	9,999494E-01	9,998987E-01
13	6,28E+02	2,49E-06	1,58E-03	2,27E-05	1,14E-02	6,42E+01	9,999627E-01	9,999254E-01
14	9,61E+02	2,49E-06	1,58E-03	2,27E-05	1,12E-02	7,08E+01	9,999683E-01	9,999366E-01
15	7,78E+02	2,47E-06	1,57E-03	2,26E-05	1,00E-02	8,91E+01	9,999659E-01	9,999318E-01
16	5,44E+02	2,49E-06	1,58E-03	2,27E-05	1,08E-02	5,79E+01	9,999708E-01	9,999417E-01
17	7,53E+02	2,49E-06	1,58E-03	2,27E-05	1,02E-02	4,60E+01	9,999739E-01	9,999478E-01
18	5,46E+02	2,50E-06	1,58E-03	2,28E-05	8,40E-03	7,56E+01	9,999698E-01	9,999396E-01
19	7,57E+02	2,48E-06	1,58E-03	2,27E-05	1,15E-02	9,81E+01	9,999745E-01	9,999490E-01
20	9,03E+02	2,48E-06	1,58E-03	2,27E-05	1,26E-02	5,81E+01	9,999759E-01	9,999517E-01

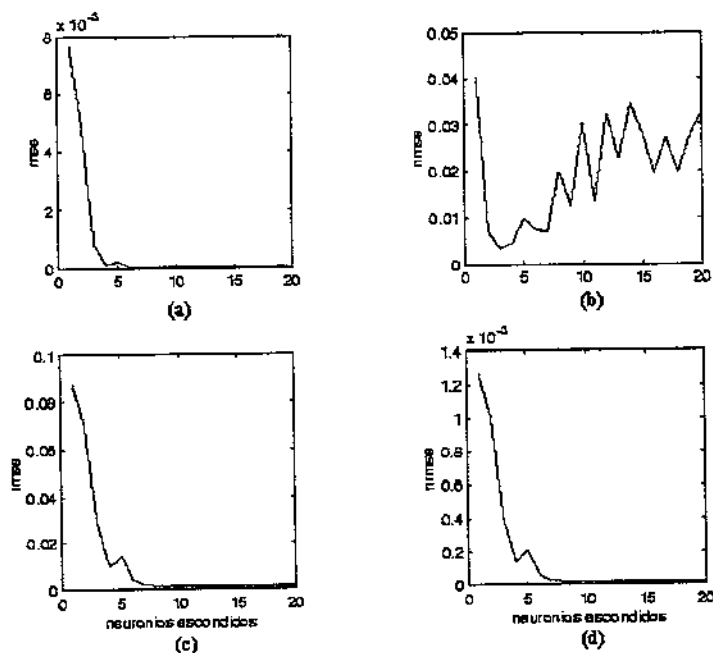


Figura 101. Protótipo I - Índices de avaliação do desempenho (a) *mse* (b) *nmse* (c) *rmse* e (d) *nrmse* usando conjunto de treinamento em função do número de neurônios escondidos.

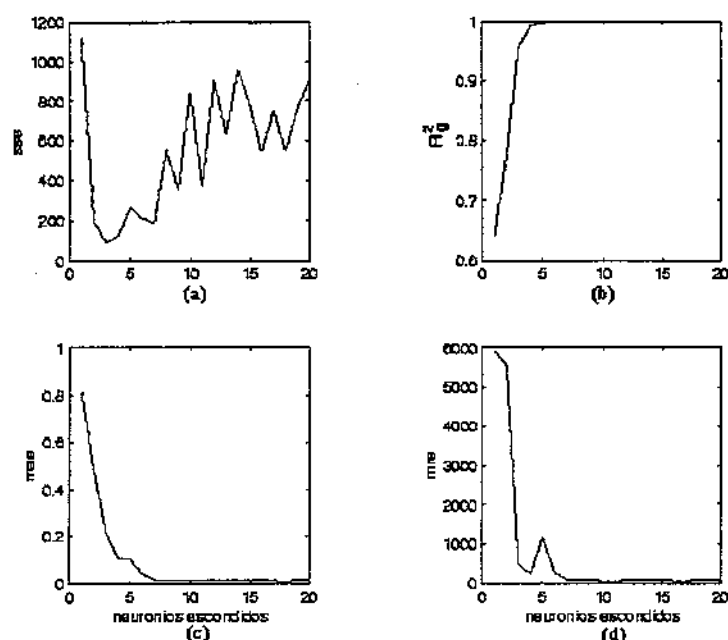


Figura 102. Protótipo I - Índices de avaliação do desempenho (a) sse (b) R_g^2 (c) mae e (d) mre usando conjunto de treinamento em função do número de neurônios escondidos.

Os resultados dos índices de avaliação de desempenho (sse , mse , $rmse$, $nmse$, mae , mre , $medmed_R$, R_g^2) em função do número de neurônios (SI) na camada escondida, obtidos usando o conjunto de validação estão apresentados na tabela 65.

Tabela 65. Protótipo I - Índices de avaliação de desempenho (sse , mse , $rmse$, $nmse$, mae , mre , $medmed_R$, R_g^2) em função do número de neurônios (SI) na camada escondida, obtidos usando o conjunto de validação.

SI	sse	mse	$rmse$	$nmse$	mae	mre	$medmed_R$	R_g^2
1	8,85E+02	2,21E+00	1,49E+00	2,17E-02	8,73E-01	8,38E+03	8,037230E-01	6,459706E-01
2	1,78E+02	4,45E-01	6,67E-01	9,71E-03	5,23E-01	8,32E+03	8,680856E-01	7,535726E-01
3	5,54E+01	1,39E-01	3,72E-01	5,42E-03	2,82E-01	3,08E+02	9,587664E-01	9,192330E-01
4	8,81E+01	2,20E-01	4,69E-01	6,83E-03	2,96E-01	2,06E+02	9,453334E-01	8,936551E-01
5	2,30E+02	5,76E-01	7,59E-01	1,10E-02	4,71E-01	1,79E+04	9,262763E-01	8,579879E-01
6	1,86E+02	4,64E-01	6,81E-01	9,91E-03	4,35E-01	1,26E+03	9,182109E-01	8,431113E-01
7	1,76E+02	4,39E-01	6,63E-01	9,64E-03	4,41E-01	1,40E+03	9,061417E-01	8,210928E-01
8	5,23E+02	1,31E+00	1,14E+00	1,66E-02	5,29E-01	2,22E+03	8,685726E-01	7,544183E-01
9	3,18E+02	7,95E-01	8,92E-01	1,30E-02	4,39E-01	1,39E+03	8,697233E-01	7,564187E-01
10	8,06E+02	2,02E+00	1,42E+00	2,07E-02	6,03E-01	3,69E+03	8,560986E-01	7,329048E-01
11	3,39E+02	8,48E-01	9,21E-01	1,34E-02	4,42E-01	2,37E+03	8,591385E-01	7,381189E-01
12	8,60E+02	2,15E+00	1,47E+00	2,13E-02	6,54E-01	4,14E+03	8,583748E-01	7,368073E-01
13	5,91E+02	1,48E+00	1,22E+00	1,77E-02	5,84E-01	3,71E+03	8,590873E-01	7,380310E-01
14	9,21E+02	2,30E+00	1,52E+00	2,21E-02	6,90E-01	3,54E+03	8,549418E-01	7,309255E-01
15	7,41E+02	1,85E+00	1,36E+00	1,98E-02	5,59E-01	3,43E+03	8,641442E-01	7,467452E-01
16	5,12E+02	1,28E+00	1,13E+00	1,65E-02	5,49E-01	3,64E+03	8,566168E-01	7,337923E-01
17	7,14E+02	1,79E+00	1,34E+00	1,94E-02	6,69E-01	7,87E+03	8,569818E-01	7,344179E-01
18	5,13E+02	1,28E+00	1,13E+00	1,65E-02	5,75E-01	6,02E+03	8,538356E-01	7,290353E-01
19	7,18E+02	1,80E+00	1,34E+00	1,95E-02	5,83E-01	3,58E+03	8,600960E-01	7,397652E-01
20	8,69E+02	2,17E+00	1,47E+00	2,14E-02	6,40E-01	5,96E+03	8,621120E-01	7,432372E-01

Entre as 20 topologias testadas para o protótipo I, a topologia com 3 neurônios na camada escondida (4-3-4) foi a que apresentou maior número de simulações, com o conjunto de validação, com maiores coeficientes de generalização e robustez ($medmed_R=0,9587664$) e de determinação global ($R_g^2=0,9192330$), ou seja, a topologia que apresentou maior robustez e maior capacidade de generalização para as quatro variáveis de saída simultaneamente.

Conforme se pode observar através das figuras 103 e 104, as trajetórias das funções de erro usando o conjunto de validação não são tão suaves quanto às obtidas usando o conjunto de treinamento. Percebe-se que apesar do rápido declínio das funções de erro até o terceiro neurônio escondido, a partir daí elas começam a oscilar, à medida que o número de neurônios na camada escondida aumenta. O coeficiente de determinação global cresce rapidamente até o terceiro neurônio, quando atinge seu valor máximo, então decresce continuamente até começar a se estabilizar em torno de 0,75.

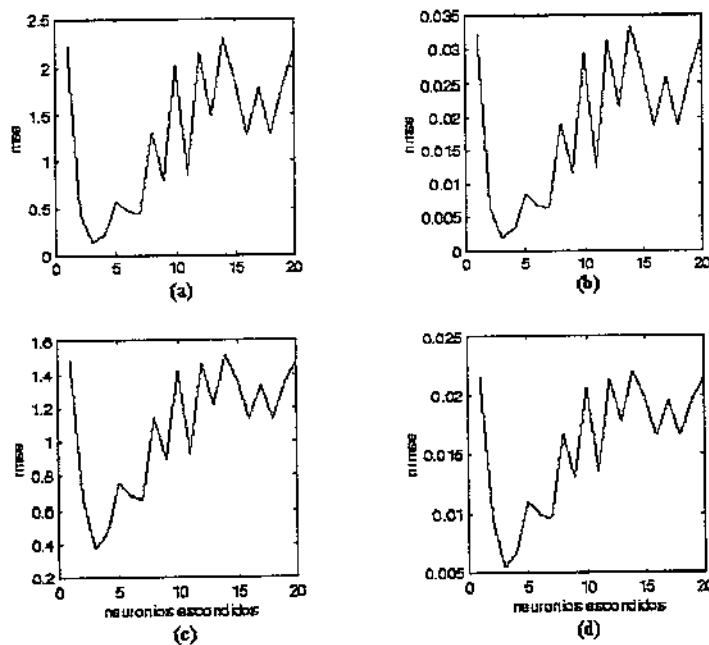


Figura 103. Protótipo I - Índices de avaliação do desempenho (a) *mse* (b) *nmse* (c) *rmse* e (d) *nrmse* usando conjunto de validação em função do número de neurônios escondidos.

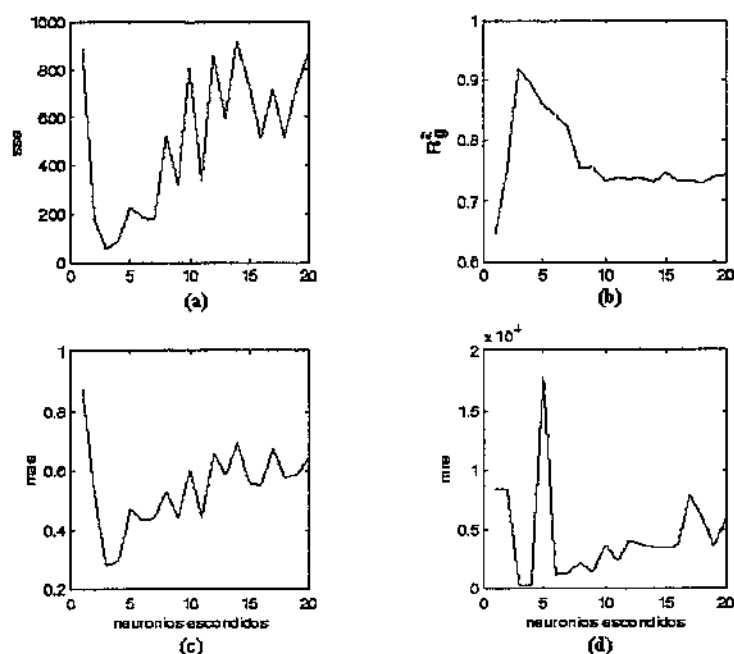


Figura 104. Protótipo I - Índices de avaliação do desempenho (a) sse (b) R_g^2 (c) mae e (d) mre usando conjunto de validação em função do número de neurônios escondidos.

De forma similar ao que ocorreu com este mesmo protótipo usando dados experimentais obtidos a 28° C e 150 rpm, o comportamento oscilatório do erro na simulação com dados do conjunto de validação não é decorrente da escolha indevida das variáveis de entrada, mas sim da qualidade dos dados de entrada, sujeitos a ruídos e distúrbios, já que os valores da densidade ótica e da salinidade do líquido metabólico livre de células foram provenientes de medidas realizadas *off-line*.

O desempenho do protótipo I para a topologia 4-3-4 usando o conjunto de dados de treinamento pode ser avaliado através das figuras 105 e 106 e comparado através das figuras 107 e 108 com o desempenho do protótipo I para a topologia 4x3x4 usando o conjunto de dados de validação. As curvas previstas pelo protótipo I com topologia 4-3-4 para as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola ajustaram-se bem aos dados do conjunto de treinamento (coeficientes de correlação R superiores a 0,95 e R_g^2 igual a 0,9187221), conforme se pode observar através das figuras 105 e 106.

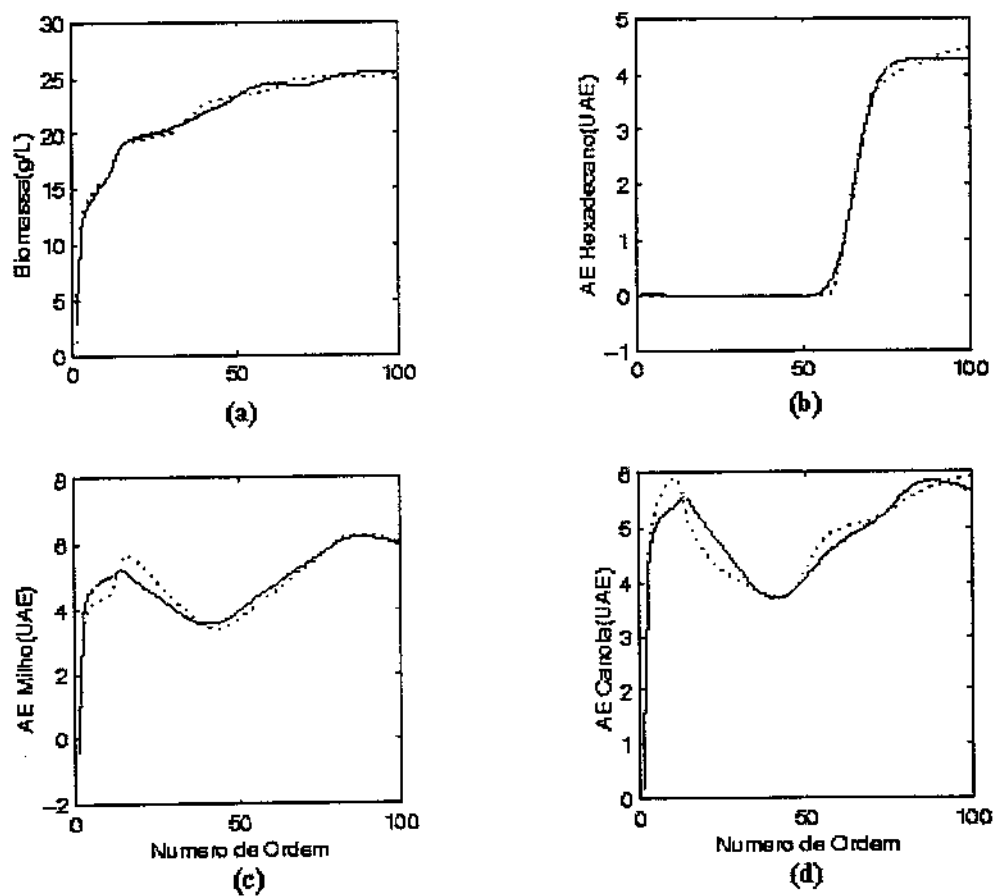


Figura 105. Protótipo I - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de treinamento.

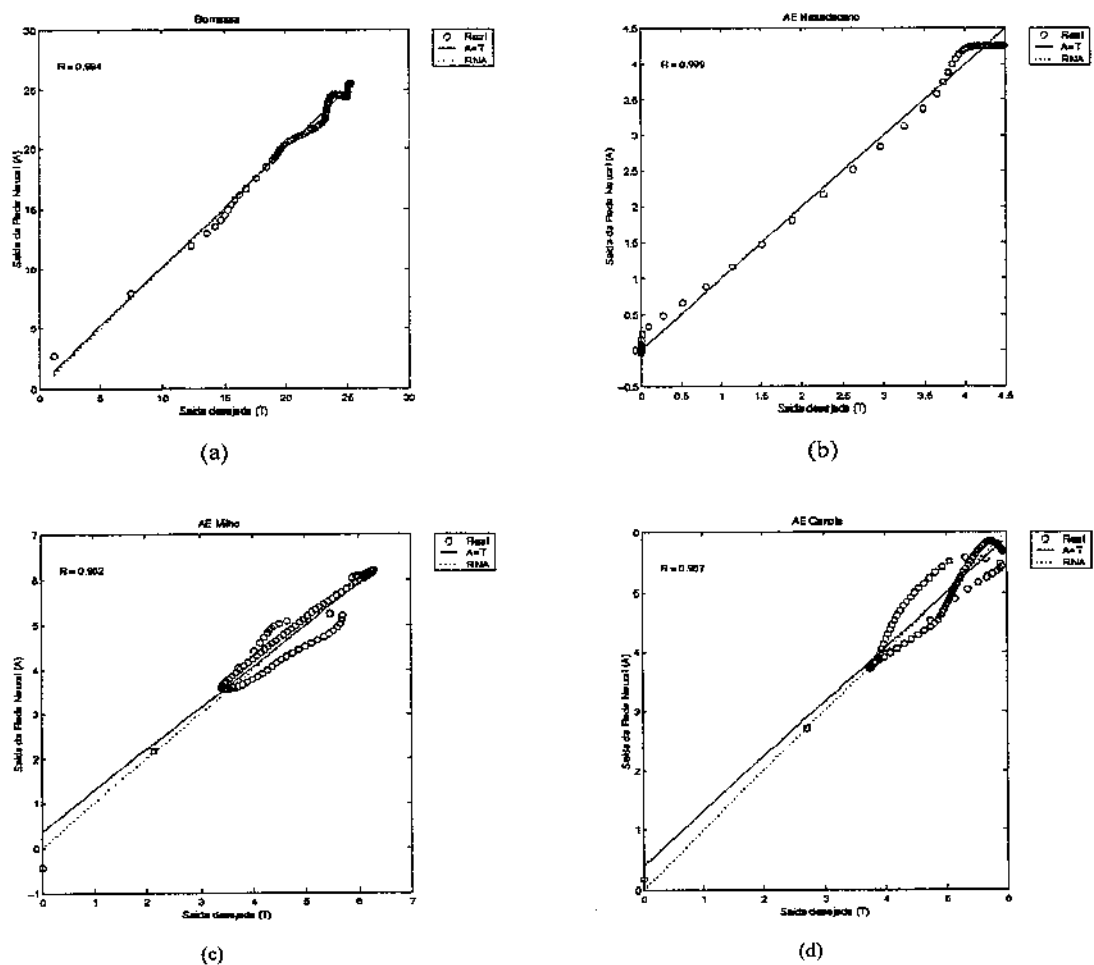


Figura 106. Protótipo I - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-óleo de canola usando dados do conjunto de treinamento.

Conforme se pode observar através das figuras 107 e 108, as curvas previstas pelo protótipo I com topologia 4-3-4 para as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola também apresentaram um ajuste satisfatório aos dados do conjunto de validação (coeficientes de correlação R superiores a 0,90 e R_g^2 igual a 0,9192330).

Este resultado mostra que o protótipo I com topologia 4-3-4 apresenta boa capacidade de generalização e corrobora com resultados anteriores sobre o comportamento oscilatório do erro, neste caso, estar associado à qualidade dos dados de entrada e não a falta de adequação das variáveis de entrada escolhidas.

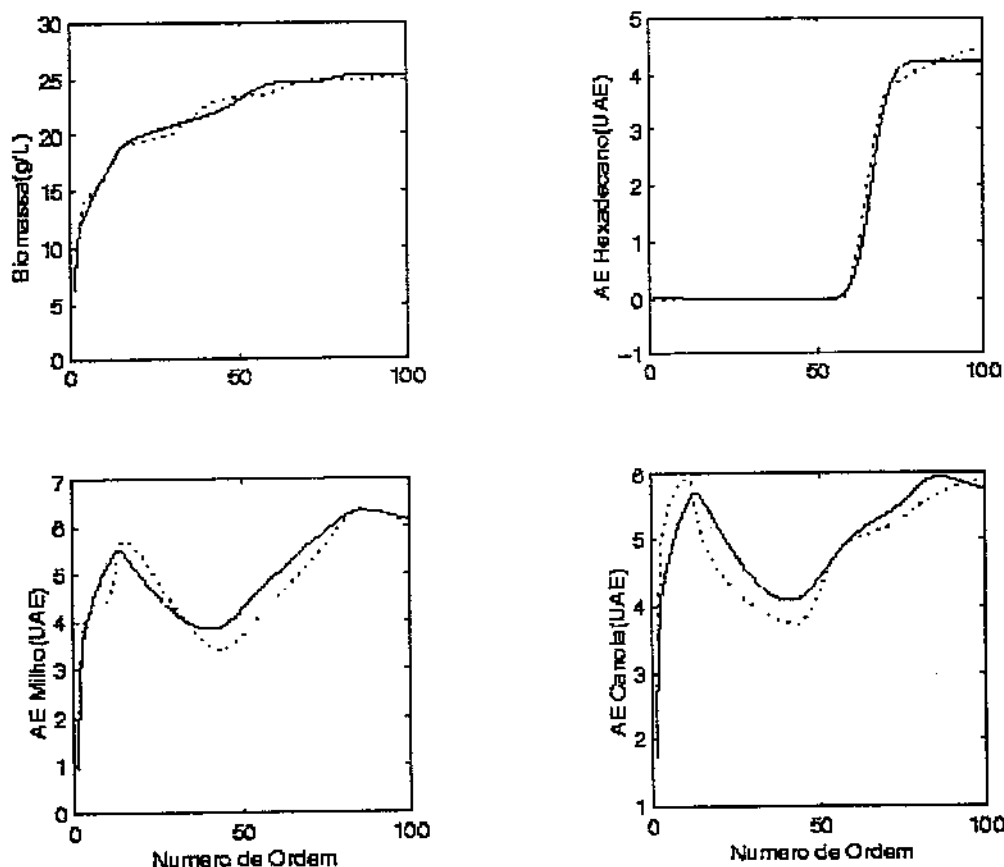
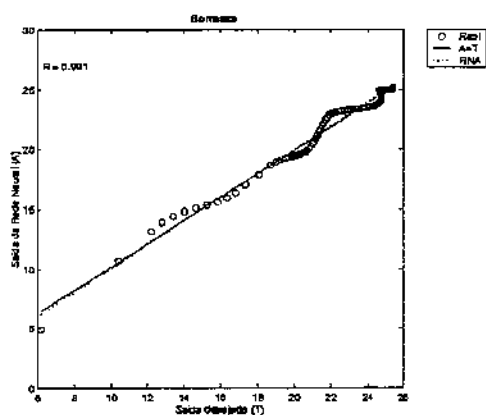
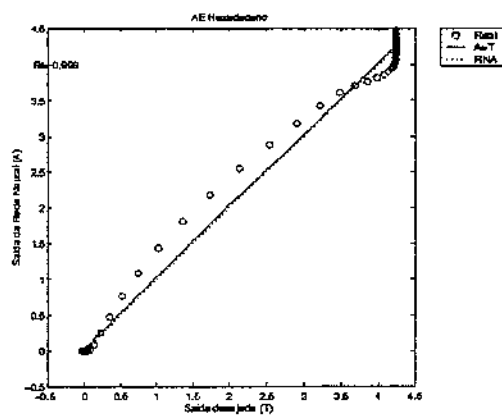


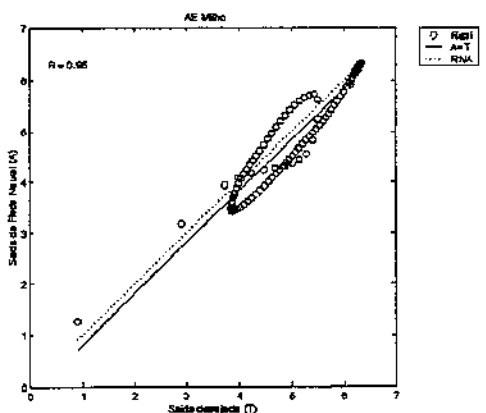
Figura 107. Protótipo I - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de validação.



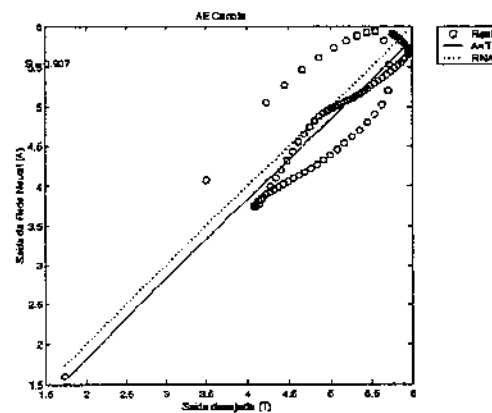
(a)



(b)



(c)



(d)

Figura 108. Protótipo I - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de validação.

A variação - em função do número de neurônios escondidos - do número de épocas e dos parâmetros de treinamento usados no algoritmo de Levenberg-Marquardt associado com regularização bayesiana (*sse*, *mse*, *ssx*, *np* e *gamk*) pode ser analisada através da tabela 66 e da figura 109.

A análise da tabela 66 permite verificar que mais uma vez a melhor topologia selecionada usando o coeficiente de generalização (*medmed-R*) coincide com a topologia selecionada usando o critério do primeiro pico da curva do número de épocas em função do número de neurônios escondidos. O primeiro pico da curva do número de épocas em função do número de neurônios escondidos ocorre para 2 neurônios escondidos. Portanto o número ótimo de neurônios escondidos por este critério é igual a 3, confirmando o resultado obtido pelo critério do coeficiente de generalização (*medmed-R*).

A figura 110 apresenta a evolução dos parâmetros de treinamento *sse*, *ssx* e *gamk* em função do número de épocas, para a topologia 4-3-4, selecionada como a melhor do protótipo I.

Tabela 66. Protótipo I - Parâmetros de treinamento da rede (*época*, *sse*, *mse*, *ssx*, *np* e *gamk*) em função do número de neurônios (*SI*) na camada escondida.

<i>SI</i>	<i>época</i>	<i>sse</i>	<i>mse</i>	<i>ssx</i>	<i>np</i>	<i>gamk</i>
1	9,10E+01	3,08E+00	7,69E-03	3,89E+01	13	11,69035
2	1,75E+02	1,94E+00	4,85E-03	5,59E+01	22	17,82085
3	7,80E+01	3,13E-01	7,82E-04	3,04E+03	31	27,64477
4	3,37E+02	3,82E-02	9,55E-05	1,16E+04	40	37,36389
5	2,94E+02	8,34E-02	2,09E-04	1,04E+03	49	46,02361
6	5,59E+02	7,20E-03	1,80E-05	1,16E+04	58	56,09966
7	1,00E+03	1,26E-03	3,15E-06	7,14E+03	67	66,08719
8	3,35E+02	1,00E-03	2,50E-06	3,86E+03	76	70,25468
9	3,48E+02	9,88E-04	2,47E-06	2,92E+03	85	77,61923
10	3,70E+02	9,92E-04	2,48E-06	2,64E+03	94	82,88351
11	2,96E+02	9,98E-04	2,49E-06	2,64E+03	103	85,07552
12	1,66E+02	9,96E-04	2,49E-06	2,24E+03	112	89,69546
13	3,05E+02	9,96E-04	2,49E-06	2,48E+03	121	88,82228
14	2,65E+02	9,95E-04	2,49E-06	2,36E+03	130	92,93191
15	1,39E+02	9,89E-04	2,47E-06	2,28E+03	139	94,08277
16	2,10E+02	9,97E-04	2,49E-06	2,77E+03	148	93,43056
17	3,46E+02	9,96E-04	2,49E-06	2,34E+03	157	95,24413
18	2,49E+02	9,99E-04	2,50E-06	2,46E+03	166	98,40498
19	9,10E+01	9,94E-04	2,48E-06	2,36E+03	175	90,77003
20	3,36E+02	9,94E-04	2,48E-06	2,46E+03	184	95,05967

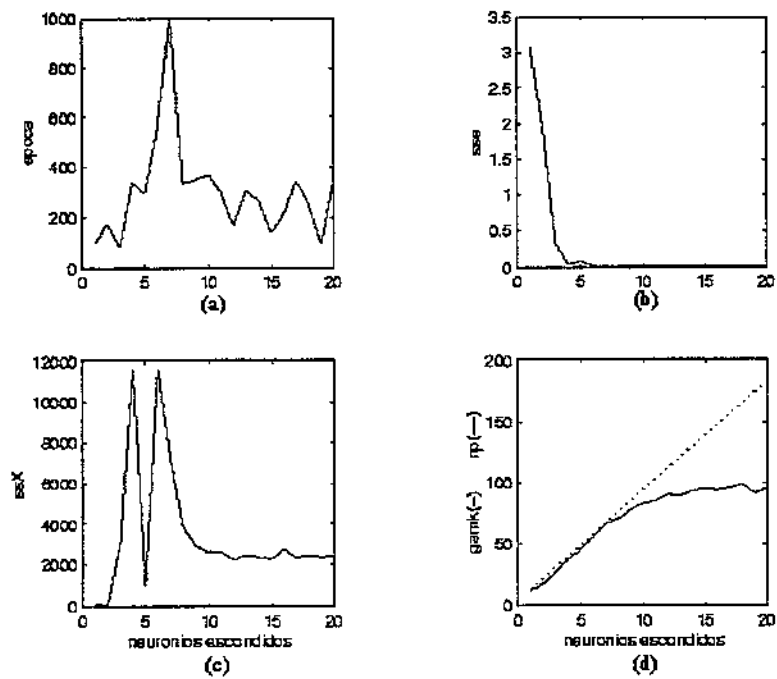


Figura 109. Protótipo I - Época e parâmetros de treinamento (sse , ssx e $gamk$) em função do número de neurônios escondidos.

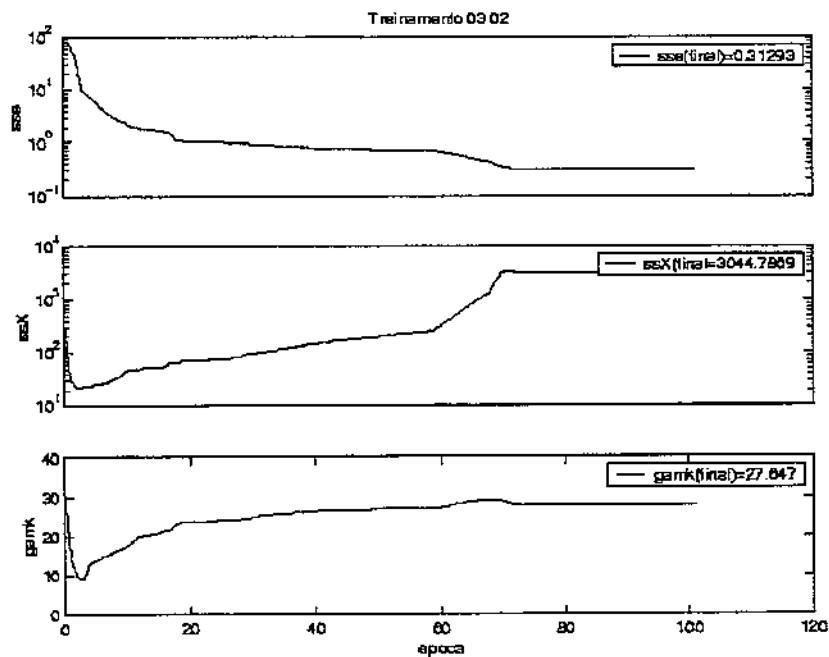


Figura 110. Protótipo I - Parâmetros de treinamento (sse , ssx e $gamk$) em função do número de épocas.

A topologia 4-3-4 do protótipo I foi capaz de fornecer boas estimativas da concentração de biomassa e das atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola, usando o conjunto de validação composto por dados obtidos a 29,5°C e 225 rpm. Logo, um *softsensor* neural completamente especificado para estimação das variáveis de estado, nestas condições operacionais, pode ser implementado *on-line* usando as matrizes de pesos (IW^1 e LW^2) e bias ($b1$ e $b2$) da camada escondida e da camada de saída apresentadas nas tabelas 67 e 68. O *softsensor* neural I pode ser representado pela seguinte equação :

$$a = \text{purelin}(LW^2 \logsig(IW^1 p + b1) + b2) \quad (27)$$

onde: a - é a matriz dos padrões de saída do *softsensor* I
purelin - é a função de transferência linear usada na camada de saída
logsig - é a função de transferência sigmoideal usada na camada escondida
 p - é a matriz dos padrões de entrada do *softsensor* I

Tabela 67. Matrizes de pesos e bias da camada escondida do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo I com 3 neurônios na camada escondida).

Camada escondida				
Matriz de pesos (IW^1)				bias (b^1)
-4,38E+00	-4,74E+00	8,65E+00	-5,67E+00	1,49E+00
1,38E+00	1,62E-01	-2,30E-01	-2,01E+00	-5,02E-01
-5,46E+00	-2,27E+01	-4,16E+00	4,15E+01	-2,39E+01
-4,38E+00	-4,74E+00	8,65E+00	-5,67E+00	1,49E+00

Tabela 68. Matrizes de pesos e bias da camada de saída do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo I com 3 neurônios na camada escondida).

Camada de saída			
Matriz de pesos (LW^2)			bias (b^2)
-2,05E-02	-1,44E+00	-5,08E-02	1,10E+00
-4,68E-03	1,93E-02	7,68E-01	9,46E-02
-8,75E-01	-2,83E+00	3,89E-02	1,92E+00
-9,61E-01	-2,94E+00	-1,89E-02	2,06E+00

5.3.3.2. Protótipo II – *Softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola

Vinte diferentes topologias do protótipo II foram avaliadas usando conjunto de treinamento composto por 100 padrões de entrada-saída. Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando conjunto de treinamento estão apresentados na tabela 69.

Tabela 69. Protótipo II (100 padrões) - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de treinamento.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	1,23E+03	1,49E-02	1,22E-01	1,76E-03	7,59E-01	2,94E+03	7,998074E-01	6,396918E-01
2	4,80E+02	7,87E-03	8,87E-02	1,28E-03	6,16E-01	9,00E+02	8,806695E-01	7,755788E-01
3	2,47E+02	2,90E-03	5,38E-02	7,75E-04	2,56E-01	9,52E+02	9,782055E-01	9,568859E-01
4	1,45E+02	3,23E-04	1,80E-02	2,59E-04	1,37E-01	2,40E+02	9,971191E-01	9,942465E-01
5	1,79E+02	1,41E-04	1,19E-02	1,71E-04	7,32E-02	1,56E+02	9,989235E-01	9,978481E-01
6	2,43E+02	1,91E-04	1,38E-02	1,99E-04	8,86E-02	2,53E+02	9,995776E-01	9,991553E-01
7	9,62E+02	1,44E-04	1,20E-02	1,73E-04	7,04E-02	6,51E+02	9,999432E-01	9,998864E-01
8	8,57E+02	2,08E-05	4,56E-03	6,56E-05	3,11E-02	2,16E+02	9,999557E-01	9,999114E-01
9	2,58E+02	5,14E-06	2,27E-03	3,26E-05	1,39E-02	1,15E+02	9,999615E-01	9,999231E-01
10	5,81E+02	5,01E-06	2,24E-03	3,22E-05	1,21E-02	6,43E+01	9,999605E-01	9,999209E-01
11	8,44E+02	1,03E-05	3,21E-03	4,62E-05	1,64E-02	1,26E+02	9,999637E-01	9,999274E-01
12	5,39E+02	5,00E-06	2,24E-03	3,22E-05	1,03E-02	6,85E+01	9,999651E-01	9,999302E-01
13	3,29E+02	5,39E-06	2,32E-03	3,34E-05	1,12E-02	7,60E+01	9,999647E-01	9,999295E-01
14	6,20E+02	4,99E-06	2,23E-03	3,22E-05	9,92E-03	5,70E+01	9,999685E-01	9,999371E-01
15	4,85E+02	5,00E-06	2,23E-03	3,22E-05	1,01E-02	4,81E+01	9,999727E-01	9,999454E-01
16	5,43E+02	4,99E-06	2,23E-03	3,22E-05	9,92E-03	1,27E+02	9,999704E-01	9,999408E-01
17	5,69E+02	4,99E-06	2,23E-03	3,22E-05	8,11E-03	4,90E+01	9,999711E-01	9,999423E-01
18	5,41E+02	5,00E-06	2,24E-03	3,22E-05	7,87E-03	6,04E+01	9,999711E-01	9,999421E-01
19	3,43E+02	4,99E-06	2,23E-03	3,22E-05	9,15E-03	5,25E+01	9,999734E-01	9,999469E-01
20	4,38E+02	4,98E-06	2,23E-03	3,21E-05	7,23E-03	6,19E+01	9,999732E-01	9,999464E-01

Entre as 20 topologias testadas para o protótipo II, a topologia com 20 neurônios na camada escondida foi a que apresentou maior número de simulações, com o conjunto de treinamento, com maiores coeficientes de ajuste e robustez (*medmed_R*=0,9999732) e de determinação global (R_g^2 =0,9999464), ou seja, a topologia que apresentou maior robustez e maior capacidade de generalização para as quatro variáveis de saída simultaneamente.

Conforme se pode observar através das figuras 111 e 112, as trajetórias das funções de erro *mse*, *rmse*, *nrmse*, *mae* e *mre* são relativamente suaves e apresentam um rápido declínio do erro até o quarto neurônio escondido, e a partir deste ponto começam a se aproximar de zero à medida que o número de neurônios escondidos aumenta. A curva do coeficiente de determinação global apresenta um comportamento oposto (Figura 112b). As trajetórias das funções de erro *nrmse*, *sse* e *mre* têm um

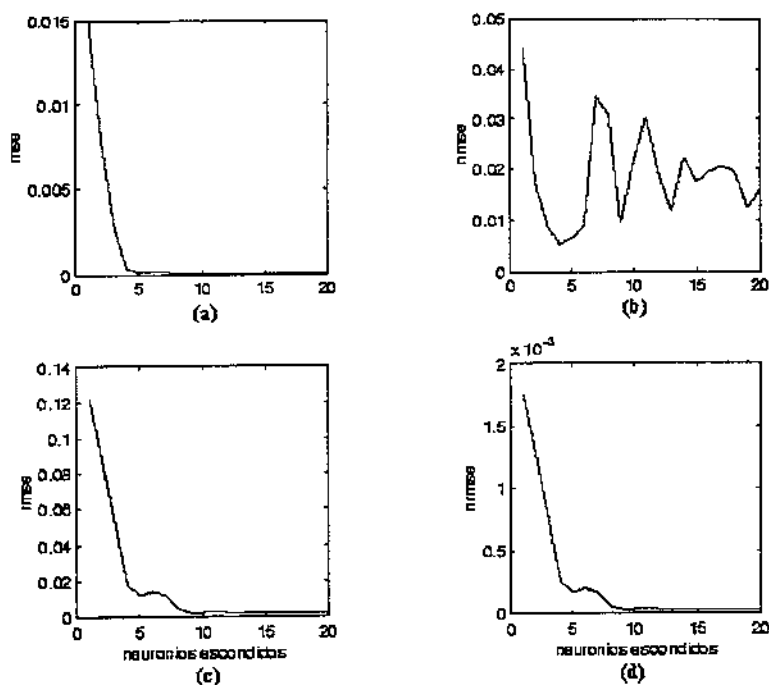


Figura 111. Protótipo II (100 padrões) - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nrmse$ usando conjunto de treinamento em função do número de neurônios escondidos.

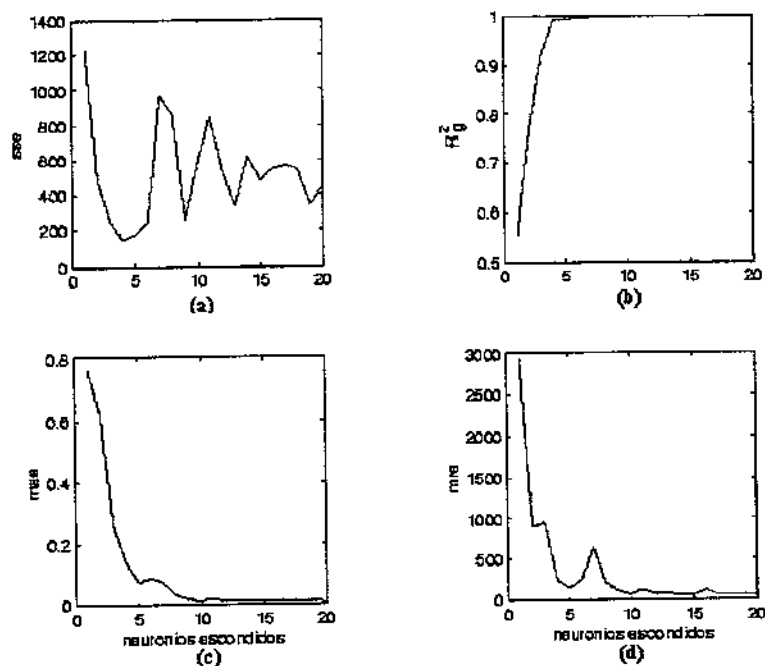


Figura 112. Protótipo II (100 padrões) - Índices de avaliação do desempenho (a) sse (b) R_g^2 (c) mae e (d) mre usando conjunto de treinamento em função do número de neurônios escondidos.

rápido declínio até o quarto neurônio e depois apresentam um comportamento oscilatório com o aumento do número de neurônios.

Os resultados dos índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida, obtidos usando o conjunto de validação estão apresentados na tabela 70.

Tabela 70. Protótipo II (100 padrões) - Índices de avaliação de desempenho (*sse*, *mse*, *rmse*, *nrmse*, *mae*, *mre*, *medmed_R*, R_g^2) em função do número de neurônios (*SI*) na camada escondida (*SI*), obtidos usando o conjunto de validação.

<i>SI</i>	<i>sse</i>	<i>mse</i>	<i>rmse</i>	<i>nrmse</i>	<i>mae</i>	<i>mre</i>	<i>medmed_R</i>	R_g^2
1	9,65E+02	2,41E+00	1,55E+00	2,26E-02	8,00E-01	8,44E+02	7,543844E-01	5,690959E-01
2	4,62E+02	1,15E+00	1,07E+00	1,56E-02	6,90E-01	6,12E+02	8,189966E-01	6,707554E-01
3	1,89E+02	4,73E-01	6,87E-01	1,00E-02	4,16E-01	9,51E+02	8,773578E-01	7,697566E-01
4	1,07E+02	2,68E-01	5,18E-01	7,53E-03	3,52E-01	2,10E+02	9,604929E-01	9,225467E-01
5	1,46E+02	3,64E-01	6,04E-01	8,78E-03	4,22E-01	1,71E+02	9,428628E-01	8,889903E-01
6	2,05E+02	5,12E-01	7,16E-01	1,04E-02	4,63E-01	4,26E+02	9,231573E-01	8,522194E-01
7	9,31E+02	2,33E+00	1,53E+00	2,22E-02	8,91E-01	8,85E+03	8,887173E-01	7,898184E-01
8	8,16E+02	2,04E+00	1,43E+00	2,08E-02	7,41E-01	1,27E+04	8,561230E-01	7,329465E-01
9	2,30E+02	5,75E-01	7,58E-01	1,10E-02	4,73E-01	7,08E+03	8,948393E-01	8,007374E-01
10	5,51E+02	1,38E+00	1,17E+00	1,71E-02	5,71E-01	4,14E+03	8,970824E-01	8,047568E-01
11	8,09E+02	2,02E+00	1,42E+00	2,07E-02	6,82E-01	1,42E+04	9,019734E-01	8,135561E-01
12	5,07E+02	1,27E+00	1,13E+00	1,64E-02	6,11E-01	1,05E+04	9,020806E-01	8,137494E-01
13	2,97E+02	7,42E-01	8,62E-01	1,25E-02	5,67E-01	1,41E+04	8,919529E-01	7,955799E-01
14	5,90E+02	1,48E+00	1,21E+00	1,77E-02	6,82E-01	9,17E+03	8,908365E-01	7,935896E-01
15	4,53E+02	1,13E+00	1,06E+00	1,55E-02	6,20E-01	1,09E+04	8,976165E-01	8,057153E-01
16	5,26E+02	1,31E+00	1,15E+00	1,66E-02	7,41E-01	9,71E+03	8,834321E-01	7,804523E-01
17	5,34E+02	1,34E+00	1,16E+00	1,68E-02	6,14E-01	1,16E+04	8,937943E-01	7,988682E-01
18	5,12E+02	1,28E+00	1,13E+00	1,65E-02	6,15E-01	1,12E+04	9,108894E-01	8,297196E-01
19	3,22E+02	8,04E-01	8,97E-01	1,31E-02	5,46E-01	9,00E+03	8,702824E-01	7,573914E-01
20	4,08E+02	1,02E+00	1,01E+00	1,47E-02	5,91E-01	1,22E+04	8,950904E-01	8,011868E-01

Entre as 20 topologias testadas para o protótipo II, a topologia com 4 neurônios na camada escondida (topologia 2-4-4) foi a que apresentou maior número de simulações, com o conjunto de validação, com maiores coeficientes de generalização e robustez (*medmed_R*=0,9604929) e de determinação global (R_g^2 =0,9225467), ou seja, a topologia que apresentou maior robustez e maior capacidade de generalização para as quatro variáveis de saída simultaneamente.

Através das figuras 113 e 114, percebe-se que as curvas das funções de erro do protótipo II usando o conjunto de validação apresentam um rápido declínio do erro com o aumento dos neurônios escondidos e começam a se estabilizar a partir de 4 neurônios.

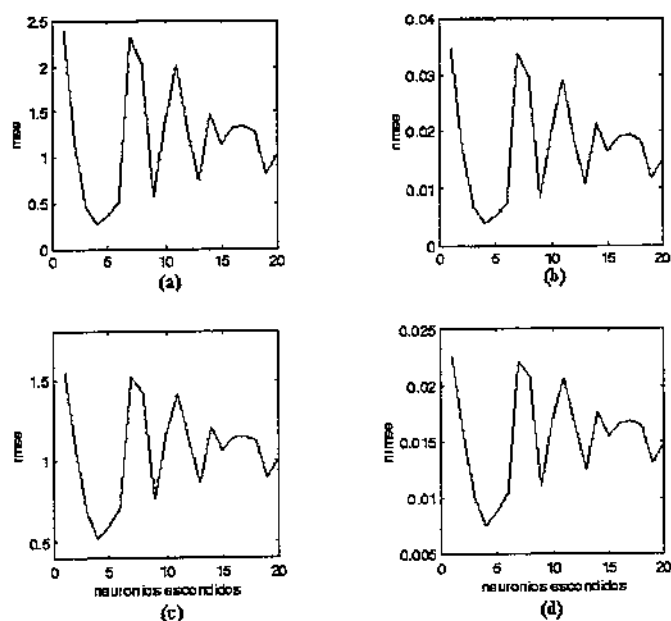


Figura 113. Protótipo II (100 padrões) - Índices de avaliação do desempenho (a) mse (b) $nmse$ (c) $rmse$ e (d) $nmrmse$ usando conjunto de validação em função do número de neurônios escondidos

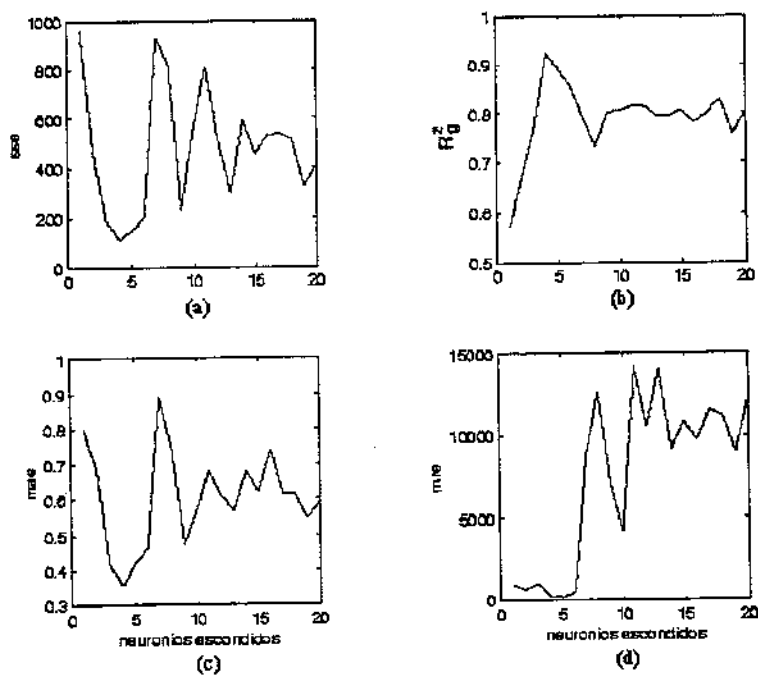


Figura 114. Protótipo II (100 padrões) - Índices de avaliação do desempenho (a) sse (b) R_g^2 (c) mae e (d) mre usando conjunto de validação em função do número de neurônios escondidos.

Neste caso, as trajetórias das funções de erro do protótipo II usando o conjunto de validação são bem menos suaves que às obtidas usando o conjunto de treinamento. Esta é uma situação onde aumento do número de neurônios escondido não melhora a capacidade preditiva da rede, possivelmente devido ao mapeamento de não linearidades de alta ordem, que podem ser consideradas como ruídos.

As figuras 115 e 116 permitem observar que as curvas previstas pelo protótipo II com topologia 2-4-4 para as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola ajustaram-se quase perfeitamente aos dados do conjunto de treinamento (coeficientes de correlação R aproximadamente iguais a 1 e R_g^2 igual a 0,9225467). Conforme se pode observar através das figuras 117 e 118, as curvas previstas pelo protótipo II com topologia 2-4-4 para as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola também apresentaram ajuste satisfatório aos dados do conjunto de validação (coeficientes de correlação R superiores a 0,94 e $R_g^2=0,9225467$).

A comparação dos desempenhos dos dois protótipos (comparação da tabela 65 com a tabela 70, dos gráficos 103 e 104 com os gráficos 113 e 114 e dos gráficos 107 e 108 com os gráficos 117 e 118) permite observar que a remoção das variáveis - densidade ótica e salinidade do líquido metabólico livre de células - do padrão de entrada do protótipo I, resultou em uma pequena diminuição dos erros (*sse*, *mse*, *nmse*, *rmse*, *nrmse*, *mae* e *mre*) e em pequeno aumento dos coeficientes de generalização e robustez. O protótipo I com topologia 4-4-4 demonstrou robustez, pois mesmo trabalhando com duas variáveis com ruído provenientes de análises *off-line* manteve sua capacidade de generalização e apresentou excelente ajuste aos dados do conjunto de validação. Estes resultados ratificam àqueles, obtidos na condição operacional de 28°C e 150 rpm, sobre a adequação das entradas do protótipo I para estimar as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola. Por outro lado, eles também corroboram com os resultados, obtidos na condição operacional de 28°C e 150 rpm, sobre a suficiência das variáveis de entrada pH e oxigênio dissolvido para estimar as variáveis de saída concentração de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola.

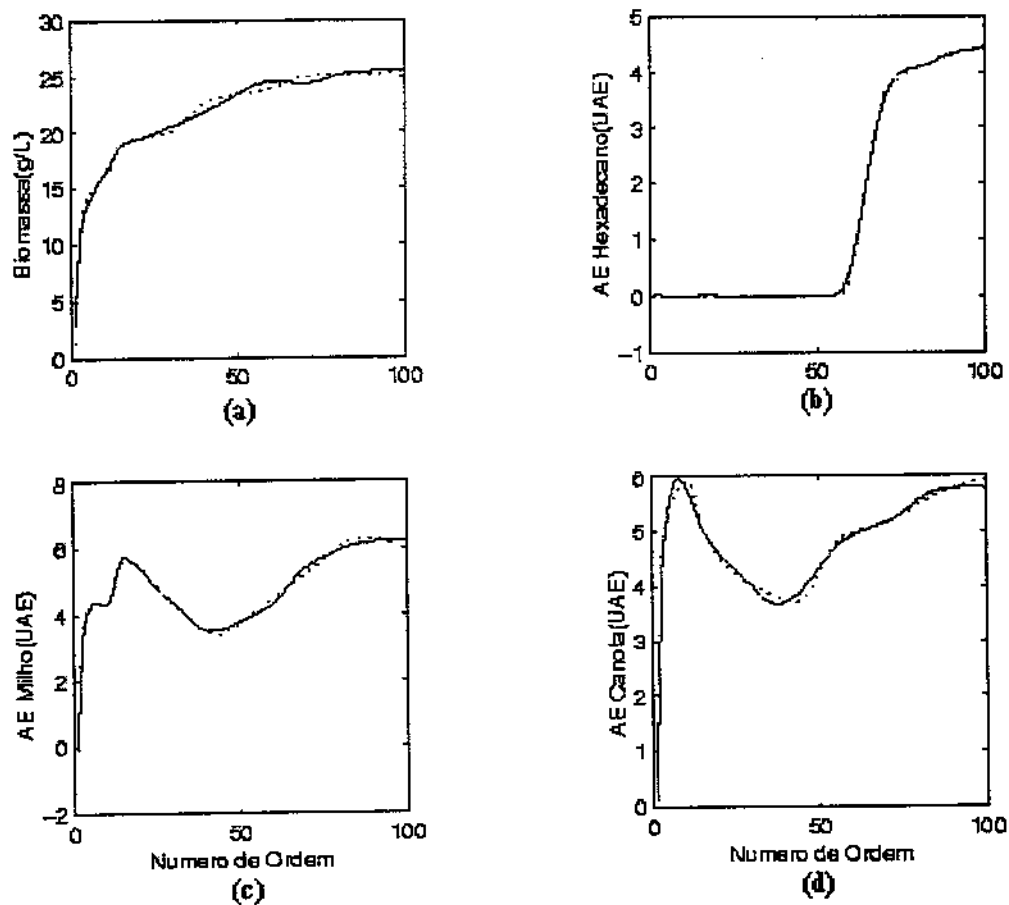
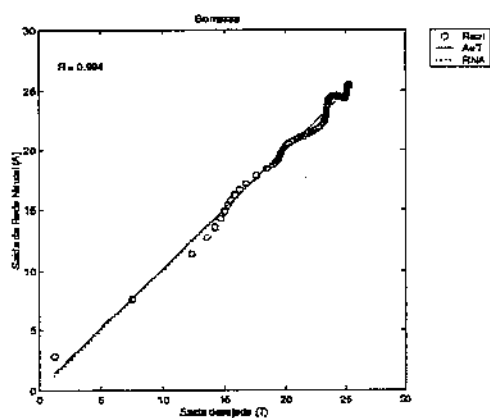
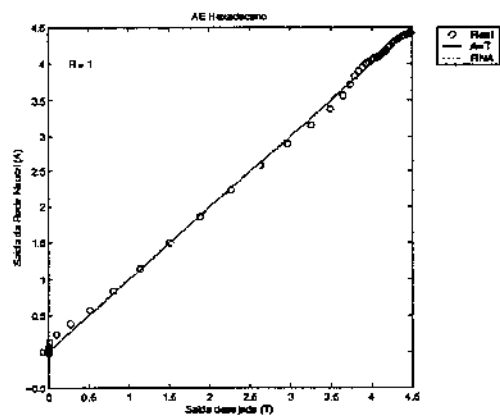


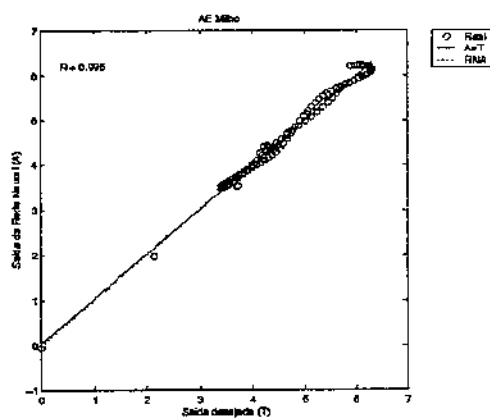
Figura 115. Protótipo II (100 padrões) - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de treinamento.



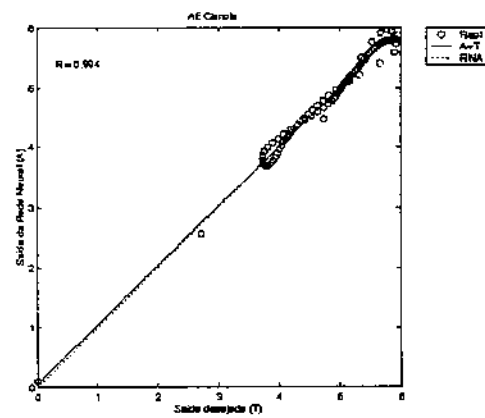
(a)



(b)



(c)



(d)

Figura 116. Protótipo II (100 padrões) - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de treinamento.

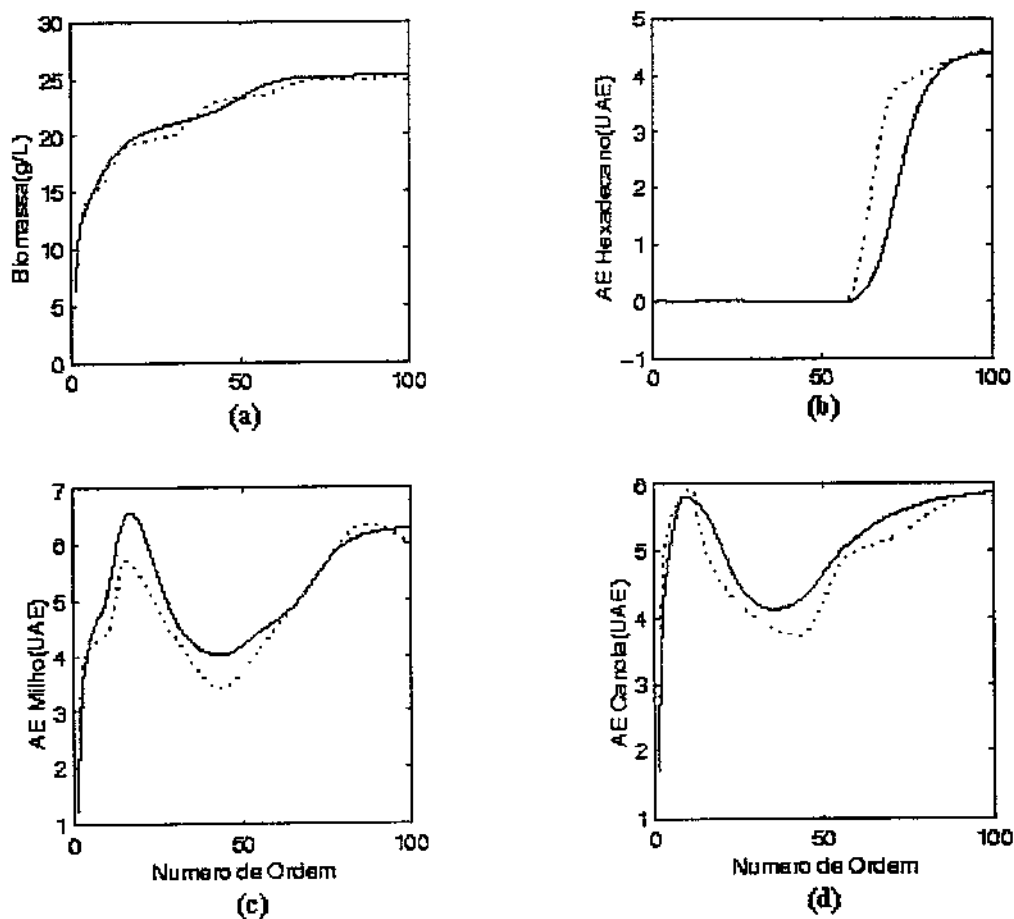
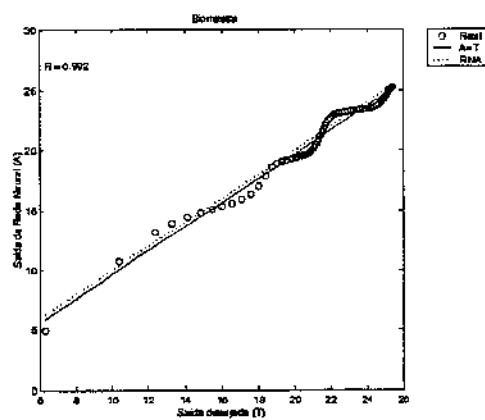
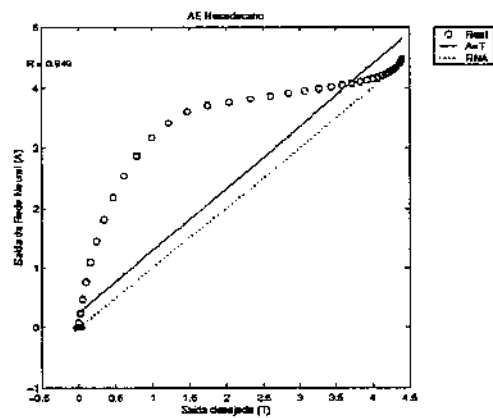


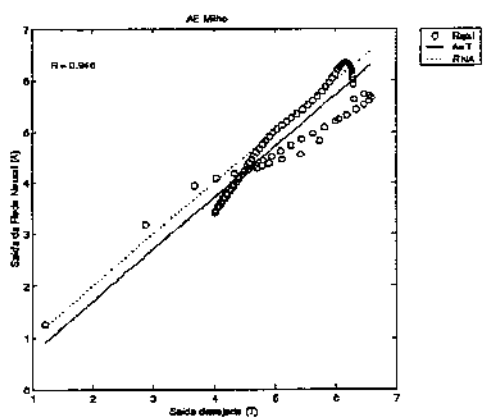
Figura 117. Protótipo II (100 padrões) - Comparação (perfis cinéticos) entre valores de saídas experimentais (—) e gerados pela rede (---) para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em-hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de validação.



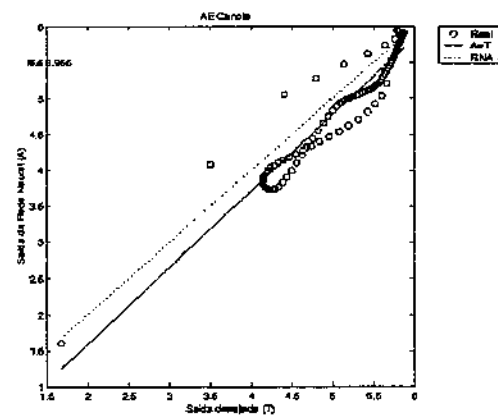
(a)



(b)



(c)



(d)

Figura 118. Protótipo II (100 padrões) - Comparação (ajuste linear) entre valores de saídas experimentais e gerados pela rede para (a) concentração de biomassa e atividades de emulsificação para (b) emulsões água-em- hexadecano (c) emulsões água-em-óleo de milho e (d) emulsões água-em-óleo de canola usando dados do conjunto de de validação.

A variação do número de épocas e dos parâmetros de treinamento (*sse*, *mse*, *ssx*, *np* e *gamk*) com o número de neurônios escondidos (*SI*) pode ser analisada através da tabela 71 e da figura 119.

A topologia 2-2-4 foi identificada como a melhor do protótipo II tanto pelo critério de selecionar a topologia com menor número de neurônios escondidos, que apresentasse o maior coeficiente de determinação global (R_g^2) usando o conjunto de validação, quanto pelo critério de selecionar a topologia com número de neurônios escondidos igual ao número de neurônios escondidos após o primeiro pico da curva do número de épocas em função do número de neurônios escondidos. A identificação da melhor topologia por meio de dois critérios de seleção totalmente diferentes reforça ao mesmo tempo a confiabilidade e a eficiência dos dois critérios empregados e o acerto da escolha do número ótimo de neurônios escondidos.

A figura 120 apresenta a evolução dos parâmetros de treinamento *sse*, *ssx* e *gamk* em função do número de épocas, para a topologia 2-4-4.

Tabela 71. Protótipo II (100 padrões) - Parâmetros de treinamento da rede (*época*, *sse*, *mse*, *ssx*, *np* e *gamk*) em função do número de neurônios (*SI*) na camada escondida.

<i>SI</i>	<i>época</i>	<i>sse</i>	<i>mse</i>	<i>ssx</i>	<i>np</i>	<i>gamk</i>
1	3,80E+01	2,99E+00	1,49E-02	8,48E+03	11	9,98358
2	7,10E+01	1,57E+00	7,87E-03	1,07E+04	18	16,81826
3	2,01E+02	5,80E-01	2,90E-03	1,93E+04	25	23,86746
4	4,75E+02	6,47E-02	3,23E-04	4,51E+04	32	29,99158
5	9,79E+02	2,82E-02	1,41E-04	4,82E+04	39	37,94739
6	5,81E+02	3,82E-02	1,91E-04	2,96E+04	46	42,92316
7	8,51E+02	2,88E-02	1,44E-04	8,20E+03	53	49,25141
8	6,73E+02	4,15E-03	2,08E-05	9,18E+03	60	58,29826
9	1,00E+03	1,03E-03	5,14E-06	4,28E+04	67	64,05911
10	1,00E+03	1,00E-03	5,01E-06	1,65E+04	74	70,40339
11	1,00E+03	2,06E-03	1,03E-05	1,14E+04	81	75,18718
12	3,98E+02	1,00E-03	5,00E-06	1,51E+04	88	83,95290
13	1,00E+03	1,08E-03	5,39E-06	2,34E+04	95	82,86218
14	6,16E+02	9,99E-04	4,99E-06	1,41E+04	102	89,07503
15	2,67E+02	9,99E-04	5,00E-06	1,52E+04	109	95,96265
16	5,84E+02	9,99E-04	4,99E-06	1,97E+04	116	101,54712
17	1,85E+02	9,97E-04	4,99E-06	1,51E+04	123	100,11338
18	3,47E+02	1,00E-03	5,00E-06	1,67E+04	130	104,92740
19	4,28E+02	9,99E-04	4,99E-06	1,50E+04	137	99,96844
20	1,41E+02	9,95E-04	4,98E-06	1,54E+04	144	113,54533

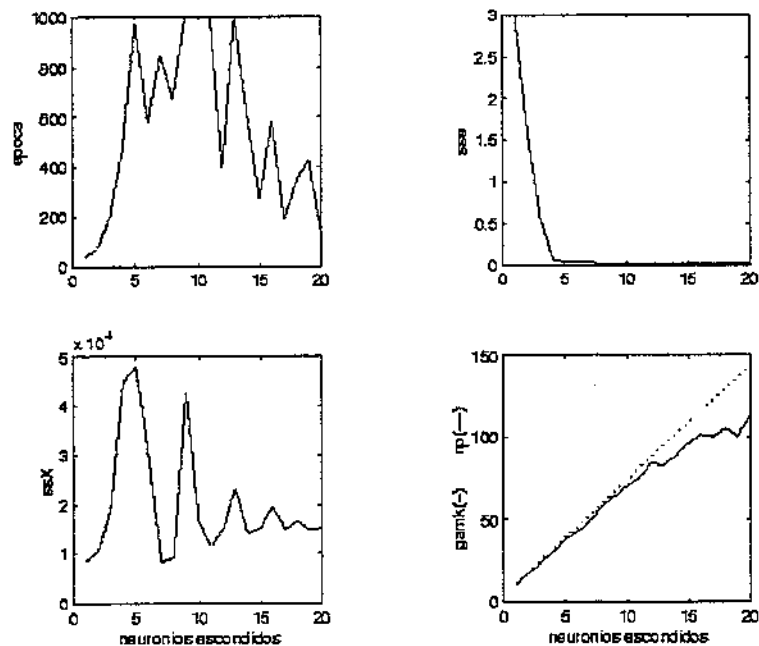


Figura 119. Protótipo II (100 padrões) - Época e parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de neurônios escondidos.

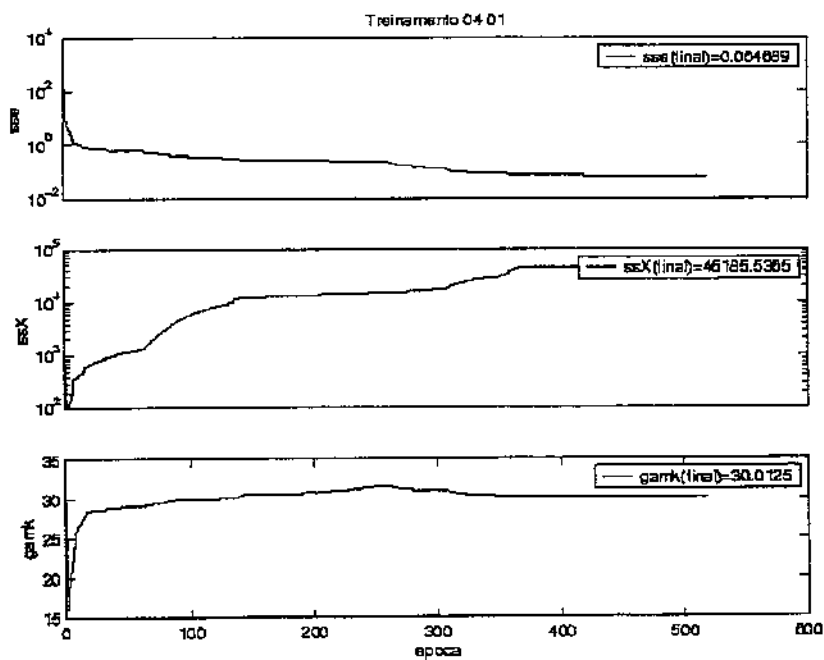


Figura 120. Protótipo II (100 padrões) - Parâmetros de treinamento (*sse*, *ssx* e *gamk*) em função do número de épocas.

A topologia 2-4-4 do protótipo II foi capaz de fornecer estimativas satisfatórias da concentração de biomassa e das atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo de milho e água-em-óleo de canola, usando o conjunto de validação composto por dados obtidos a 29,5°C e 225 rpm. Logo, um *softsensor* neural completamente especificado para estimação das variáveis de estado pode ser implementado *on-line* usando as matrizes de pesos (IW^1 e LW^2) e bias ($b1$ e $b2$) da camada escondida e da camada de saída apresentadas nas tabelas 36 e 37. O *softsensor* neural II pode ser representado pela seguinte equação:

$$a = \text{purelin}(LW^2 \logsig(IW^1 p + b1) + b2) \quad (28)$$

onde: a - é a matriz dos padrões de saída do *softsensor* II

purelin - é a função de transferência linear usada na camada de saída

$\logsig$ - é a função de transferência sigmoidal usada na camada escondida

p - é a matriz dos padrões de entrada do *softsensor* II

Tabela 72. Matrizes de pesos e bias da camada escondida do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo II com 7 neurônios na camada escondida).

Camada escondida		
Matriz de pesos (IW^1)		bias (b^1)
1,42E+01	9,61E+00	-1,06E+01
1,22E+01	1,05E+01	-9,76E+00
-1,65E+01	2,05E+02	-2,34E+01
1,08E-01	3,54E-01	5,89E-01

Tabela 73. Matrizes de pesos e bias da camada de saída do *softsensor* para inferência de biomassa e atividades de emulsificação para emulsões água-em-hexadecano, água-em-óleo-de milho e água-em-óleo de canola treinado com 100 padrões entrada-saída (Protótipo II com 7 neurônios na camada escondida).

Camada de saída				
Matriz de pesos (LW^2)			bias (b^2)	
-8,24E-01	8,35E-01	3,49E-02	-1,02E+01	7,56E+00
-8,58E-02	8,29E-02	-8,03E-01	2,36E-01	7,42E-01
-7,16E+00	7,83E+00	-1,71E-01	-1,66E+01	1,17E+01
-2,65E+00	3,54E+00	4,47E-03	-2,13E+01	1,48E+01

5.4. Conclusão

Os protótipos de *softsensores* baseados em redes neurais artificiais - treinados, validados e testados com dados experimentais do processo de produção de biossurfactante por *Candida lipolytica* - mostraram ser ferramentas eficientes e viáveis para fazer estimativa e previsão em tempo real da concentração de biomassa e das atividades de emulsificação para emulsões água-em hexadecano, água-em-óleo de milho e água-em-óleo de canola, nas condições estudadas. Os dois novos critérios (métodos) propostos e empregados com sucesso neste trabalho, para determinação do número ótimo de neurônios escondidos de uma determinada arquitetura de rede neural foram fundamentais para seleção e bom desempenho dos protótipos de *softsensores* na modelagem do comportamento não linear do processo de produção de biossurfactante por *Candida lipolytica*. Em simulações com os conjuntos de treinamento, validação e teste, compostos por dados obtidos a 28°C e 150 rpm, todos os protótipos apresentaram coeficientes de generalização ($medmed_R$) superiores a 0,96 e coeficientes de determinação global (R_g^2) superiores a 0,92. O protótipo de *softsensor* com topologia 2-7-4 - tendo como entradas pH e oxigênio dissolvido e como saídas concentração de biomassa e as atividades de emulsificação para emulsões água-em hexadecano, água-em-óleo de milho e água-em-óleo de canola - foi o que apresentou em simulações com o conjunto de validação, maior potencial para ser selecionado e implementado em uma plataforma de teste em tempo real no Núcleo de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco. Simulações com conjuntos de treinamento, validação e teste, compostos por dados obtidos a 29,5°C e 225 rpm, ratificaram a adequação e suficiência do pH e da concentração de oxigênio dissolvido para fazer estimativa da concentração de biomassa e das atividades de emulsificação para emulsões água-em hexadecano, água-em-óleo de milho e água-em-óleo de canola. Entretanto, devido às limitações e incertezas provenientes da qualidade e da quantidade dos dados experimentais e do comportamento altamente não linear e variante no tempo, característico dos processos biotecnológicos; seria prudente que a decisão sobre o protótipo a ser instalado no processo real, só fosse tomada de forma definitiva após alguns testes adicionais *off-line*, com dados de novos ensaios experimentais, nas condições estudadas, serem realizados para ratificar e assegurar, mais ainda, a adequação, a capacidade preditiva e a robustez dos protótipos e evitar gastos desnecessários com testes *on-line*.

6. CONCLUSÃO GERAL E PERSPECTIVAS FUTURAS

Técnicas de planejamento experimental, de análise de superfície de respostas e inteligência artificial foram empregadas com sucesso na otimização dos componentes do meio de produção de biossurfactante por *Candida lipolytica*, na ampliação do processo da escala de frascos para escala de biorreator de bancada e no desenvolvimento de protótipos de *softsensores* neurais para estimação e previsão *on-line* de biomassa e atividades de emulsificação para emulsões água-em hexadecano, água-em-óleo de milho e água-em-óleo de canola. Uma metodologia sistemática para melhoramento da eficiência e redução do custo do processo de produção de biossurfactante por *Candida lipolytica* no Núcleo de Pesquisa em Ciências Ambientais da Universidade Católica de Pernambuco foi estabelecida, podendo inclusive ter sua esfera de aplicação estendida a outros processos biotecnológicos.

Como perspectivas de projetos futuros a serem desenvolvidos visando dar prosseguimento a linha de pesquisa do presente trabalho, merecem destaques:

- Otimização em escala de biorreator de bancada dos componentes do meio de produção de bioemulsificante por *Candida lipolytica*;
- Otimização em escala de biorreator de bancada das condições operacionais do processo de produção de biossurfactante por *Candida lipolytica*;
- Ampliação processo de produção de bioemulsificante por *Candida lipolytica* para escala semi-piloto;
- Implementação do *softsensor* desenvolvido no processo de produção de bioemulsificante por *Candida lipolytica*, dentro de um laço de controle em uma plataforma de teste em tempo real;
- Desenvolvimento de *softsensores* utilizando redes neurais recursivas.
- Aplicação a outros processos biotecnológicos das metodologias desenvolvidas no presente trabalho para construção de *softsensores* baseados em redes neurais.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ABURUWAIDA, A.S.; BANAT, I.M.; HADITIRTO, S.; SALEM, A.; KADRI, M. Isolation of biosurfactant-producing bacteria. Product characterization and evaluation. *Acta Biotechnol.*, v.2, p.315-324, 1991.
- ABELÉM, A.J.G. **Redes Neurais Artificiais na Previsão de Séries Temporais**. Rio de Janeiro, 1994. Dissertação (Mestrado em Engenharia Elétrica) - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, 1994.
- ADINARAYANA, K.; ELLAIAH, P. Response surface optimization of the critical medium components for the production of alkaline protease by a newly isolated *bacillus* sp. *J. Pharm. Pharmaceut Sci.*, v.5, p.272-278, 2002. Disponível em: <<http://www.ualberta.ca/~csps>>. Acesso em: 05 de out. de 2004.
- AHMAD, M.N.; HOLLAND, C.R.; McKAY, G. Mass transfer studies in batch fermentation: mixing characteristics. *J. Food Eng.*, v.23, p.145, 1994.
- ALBIOL, J.; CAMPMAYÓ, C.; CASA, C.; POCH, M. Biomass estimation in plant cell cultures: a neural network approach. *Biotechnol. Prog.*, v.11, p.88-92, 1995.
- ARMILLATO, L. **Estudo da produção de ácido cítrico por *Candida lipolytica* NRLL Y 1095**. Campinas, 2000. 83 p. Dissertação (Mestrado em Engenharia de Alimentos) – Faculdade de Engenharia de Alimentos, UNICAMP, 2000.
- ARINO, S.; MARCHAL, R.; VANDECASTEELE, J.P. Production of new extracellular glycolipids by a strain of *Cellulomonas cellulans* (*Oerskovia xanthineolytica*) and their structural characterization. *Can. J. Microbiol.*, v.44, p.238-243, 1998.
- ASSIS, J.A.; MACIEL FILHO, R. Softsensors development for on-line bioreactor state estimation. *Comp. Chem. Engng.*, v.24, p.1099-1103, 2000.
- ASSOCIATION OF ANALYTICAL CHEMISTS. **Official methods of analysis**. Washigton, 1984. 1018p.
- ATLAS, R.M. (Ed.). **Petroleum Microbiology**. New York: Macmillan publishing Co, p.352-353, 1984.
- ATLAS, R.M.; BARTHA, R. Degradation and mineralization of petroleum in sea water: limitation by nitrogen and phosphorus. *Biotechnol. Bioeng.*, v.14, p.309-318, 1972.
- AZZOF, E.M. Reducing error in neural network time series forecasting. *Neural Comp. Appl.*, v.1, p.240-247, 1993.

- BANAT, I.M. Characterization of biosurfactants and their use in pollution removal – state of art (review). **Acta Biotechnol.**, v.15, p.251-267, 1995.
- BANAT, I.M., MAKKAR, R.S.; CAMEOTRA, S.S. Potential commercial applications of microbial surfactants. **Appl. Microbiol. Biotechnol.**, v.53, p.495-508, 2000.
- BARROS NETO, B.; SCARMÍNIO, I.S.; BRUNS, R.E. **Planejamento e Otimização de Experimentos**. Campinas:Editora Unicamp,1996.
- BASTIN, G.; DOCHAIN, D. **On line estimation and adaptive control of bioreactors**. Amsterdam:Elsevier, 1990.
- BELGARDT, K.H., KULMANN, W., MEYER, H.D., SCHÜGERL, K.; THOMA, M. Application of an extended Kalman filter for state estimation of a yeast fermentation. **IEE Proc.**, v.133, n.5, p.226-234, 1986.
- BHAT, N.; McAVOY, T.J. Use of neural nets for dynamic modeling and control of chemical process systems. **Comp. Chem. Engng.**, v.14, n.4/5, p.573-583, 1990.
- BHAT, N.; McAVOY, T.J. Determining model structure for neural models by networks stripping. **Comp. Chem. Engng.**, v.16, n.4, p.271-281, 1992.
- BHATTACHARJEE, J.K. Microorganisms as potential sources of foods. **Adv. Appl. Microbiol.**, v.13, p.139-159, 1970.
- BISHOP, C.M. **Neural Networks for Pattern Recognition**. Oxford: Oxford, University Press,1995
- BRANDÃO, L.V.C. **Produção de biossurfactantes por leveduras do gênero *Candida*, isoladas de solo de mata e de solo poluído com chumbo**. Rio Claro, São Paulo, 102p. Tese (Doutorado em Ciências Biológicas) - Instituto de Biociências, UNESP, 2001.
- BREMERMAN, H.J.; ANDERSON, R.W. **An alternative to back- propagation: a simple rule for synaptic modification for neural net training and memory**. Berkeley: Dept of Mathematics - Univ. of California, 1989. (Internal Report).
- BOGNOLO, G. Biosurfactants as emulsifying agents for hydrocarbons. **Colloid. Surf.**, v.152, n.1-2, p.41-52, 1999.
- BOX, G.E.P.; WETZ, J. **Criteria for judging adequacy of estimation by an approximation response function**. University of Wisconsin. Technical Report, n.9, 1973.
- BOX, G.E.P.; HUNTER, W.G.; HUNTER, J.S. **Statistics for experimenters: an introduction to design, data analysis and model building**. New York: Wiley, 1978.

- BOX, G.E.P.; DRAPPER, N.R. **Empirical model building and response surfaces**. New York: Wiley, 1987.
- CAMEOTRA, S.S.; BOLLAG, J.M. Biosurfactant-enhanced bioremediation of polycyclic aromatic hydrocarbons. **Crit. Rev. Environ. Sci. Technol.**, v.33, n.2, p.111-126, 2003.
- CAMPOS-TAKAKI, G.M. (Coord.) **Eficiência Tecnológica dos sensores e catalisadores (Biológicos e Químicos) na redução da Poluição do Petróleo e Gás Natural**. Recife: Núcleo de Pesquisas em Ciências Ambientais – UNICAP, 1999. (Relatório Técnico, Projeto FNDCT/CTPETRO).
- CARVALHO, G.L. **Efeito dos hidrocarbonetos aromáticos policíclicos, naftaleno e fluoranteno, no crescimento de amostras de *Cunninghamella elegans* e biotransformação do naftaleno**. Dissertação (Mestrado em Biologia de Fungos) – Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Recife, 1999.
- CHAKRABORTY, K. Forecasting the Behavior of Multivariate Time Series Using Neural Networks. **Neural Networks**, v.5, p.961-970, 1992.
- CHEN, H.C. Optimizing the concentration of carbon, nitrogen and phosphorus in a citric acid fermentation with response surface method. **Food Biotechnol.**, v.10, p.13-27, 1996.
- CHÉRUY, A. Software sensors in bioprocess engineering. **J. Biotechnol.**, v.52, p.193-199, 1997.
- CHTOUROU, M.; NAJIM, K.; ROUX, G.; DAHHOU, B. Control of a bioreactor using a neural network. **Bioproc. Eng.**, v.8, p.251-254, 1993.
- CHU, W.B.Z.; CONSTANTINIDES, A. Modeling, optimization and computer control of the cephalosporin C fermentation process. **Biotechnol. Bioeng.**, v.32, p.277, 1988.
- CIBENKO, G. Approximation by superpositions of a sigmoidal function. **Math. Contr. Signal Syst.**, v.2, p.303-314, 1989.
- CIRIGLIANO, M.C.; CARMAN, G.M. Isolation of a bioemulsifier from *Candida lipolytica*. **Appl. Environ. Microbiol.**, v.48, n.4, p.747-750, 1984.
- CIRIGLIANO, M.C.; CARMAN, G.M. Purification and characterization of liposan, a bioemulsifier from *Candida lipolytica*. **Appl. Environ. Microbiol.**, v.50, p.846-850, 1985.
- COOPER, D.G.; GOLDENBERG, B.G. Surface-active agents from two *Bacillus* species. **Appl. Environ. Microbiol.**, v.23, p.224-229, 1987.
- COOPER, D.G.; PADDOCK, D.A. Production of a biosurfactant from *Torulopsis bombicola*. **Appl. Environ. Microbiol.**, v.47, p.173-176, 1984.

- COOPER, D.G; ZAJIC, J.E. Surface-active compounds from microorganisms. **Adv.Appl.Microbiol.**, v.26, p.229-253, 1980.
- CORZO, G.; REVAH, S. Productions and characteristics of the lipase from *Yarrowia lipolytica* 681. **Biores. Technol.**, v.70, p.173-180, 1999.
- COWAN, J.D. A mathematical theory of central nervous activity. London.1967. Thesis (Ph.D.) – University of London, 1967.
- CROLLA, A.; KENNEDY, K.J. Optimization of citric acid production from *Candida lipolytica* Y-1095 using n-paraffin. **J. Biotechnol.**, v.89, p.27-40, 2001.
- CROLLA, A.; KENNEDY, K.J. Fed-batch production of acid citric by *Candida lipolytica* grown on n-paraffins. **J. Biotechnol.**, v.110, p.73-84, 2004.
- CRUZ, A. J.G.; HOKKA, C.O.; GIORDANO, R.C. A simulation of the penicillin G production bioprocess applying neural networks. **Braz. J. Chem. Eng.**, v.14, n.4, 1997. ISSN 0104-6632.
- DARLINGTON, W.A. Aerobic hydrocarbon fermentation. A pratical evaluation. **Biotech. and Bioeng.**, v.11, p.269,1964.
- DEMUTH, H.; BEALE, M. **Neural network toolbox.**: for use with Matlab. User's Guide. Version 4. Release 12. Natick: The Mathworks Inc., 2001.
- DESAI, J.D.; BANAT, I.M. Microbial production of surfactant and their commercial potential. **Microbiol. Molec. Biol. Rev.**, p.47-44, 1997.
- DESAI, K.M.; VAIDYA, B.K.; SINGHAL, R.S.; BHAGWAT, S.S. Use of an artificial neural network in modeling yeast biomass and yeld of β -glucan. **Proc. Biochem.**, xxx:xxx-xxx, 2004. (In press).
- DESHPADE, M.; DANIELS, L. Evaluation of sophorolipid biosurfactant production by *Candida bombicola* using animal fat. **Biores. Tecnol.**, v.54, p.143-150,1995.
- Di MASSIMO, C. ; MONTAGUE, G.A.; WILLIS, M.J.; THAM, M .T . ; MORRIS, A.J. Towards improved penicilin fermentation via artificial neural networks. **Comp. Chem. Engng**, v.16, n.4, p.283-291, 1992.
- DOHERTHY, S.K. ; GOMM, J.B. ; WILLIAMS, D. Experiment design for non-linear system identification using neural network. **Comp. Chem. Engng**, v.21, n.3, p.327-346, 1997.
- DUBOIS, M.; GILLES, K.A.; HAMILTON, K.J.; REBERS, P.A.; SMITH, F. Colorimetric method for determinationof sugar and related substances. **Anal. Chem.**, v.28, p.350-356, 1956.
- FIECHTER, A. Biosurfactants: moving towards industrial application. **Trends Biotechnol.**, v.10, n.9, p.208-217, 1992.

- FORESEE, F.D. HAGAN, M.T. Gauss-Newton Approximation to Bayesian Learning. In: **Proceedings of the 1997 International Joint Conference on Neural Networks**, 1997, p.1930-1935, 1997. Disponível em: <http://hagan.ecen.ceat.okstate.edu/icnn97a.pdf>. Acesso em: 04/02/2003.
- FRATTINI, A.M.; PEDROSA, L.S.; PEREIRA, J.A.F.R. A self tuning for multicomponent batch distillation with soft sensor inference based on a neural network. **Comp. Chem. Engng.**, p.S261-S264, 1999.
- GELINAS, P.; GOULET, J. Morphology of baker's yeast and dissolved oxygen saturation during fed-batch growth. **Lett. Appl. Microbiol.**, v.12, p.164-170, 1991.
- GHOSH, S.; SWAMINATHAN, T. Optimization of process variables for the extractive fermentation of 2-3 butanediol by *Klebsiella oxytoca* in aqueous two-phase system using response surface methodology. **Chem. Biochem. Eng. Q.**, v.17, p.319-325, 2003.
- GLASSEY, J., MONTAGUE, G.A., WARD, A.C.; KARA, B. Enhanced supervision of recombinant *E.Coli* fermentations via artificial neural networks. **Proc. Biochem.**, v.29, p.387-398, 1999.
- GLASSEY, J., MONTAGUE, G.A., WARD, A.C.; KARA, B. Artificial neural network based experimental design for enhancing fermentation development. **Biotech. Bioeng.**, v.44, p.397-405, 1994b.
- GLASSEY, J.; IGNOVA, M.; WARD, A.C.; MONTAGUE, G.A.; MORRIS, A.J. Bioprocess supervision: neural networks and knowledge-based systems. **J. Biotechnol.**, v.52, p.201-205, 1997.
- GONTARSKI, C.A.; RODRIGUES, P.R.; MORI, M.; PRENEM, L.F. Simulation of an industrial waste water treatment plant using artificial neural networks. **Comp. Chem. Engng.**, v.24, p.1719-1723, 2000.
- GRAINDORGE, P.; CHARBONNIER, S.; MAGNIN, J.P.; MAUVY, C.; CHÉRUY, A. A software nsensor of biological activity based on a redox probe for the control of *Thiobacillus ferrooxidans* cultures. **J. Biotech.**, v.35, p.87-96, 1994.
- GUERRA- SANTOS, L.; KÄPPELI, O.; FIECHTER, A. *Pseudomonas aeruginosa* biosurfactant production in continuous culture with glucose as carbon source. **Appl. Environ. Microbiol.**, p.301-305, 1984.
- HAGAN, M.T.; MENHAJ, M. Trainning feedforward networks with the Maquardt algorithm. **IEEE Trans. Neural Net.**, v.5, p.989-993, 1992.
- HAGAN, M.T.; DEMUTH, B.D.; BEALE, M. **Neural network design**. Boston: PWS Publishing Company, 1996.
- HUMPHREY, A.E. A critical review of hydrocarbon fermentations and their industrial utilization. **Biotech. and Bioeng.**, v.9, p.3, 1967.

- HARAYAMA, S. Polycyclic aromatic hydrocarbon bioremediation design. **Curr. Opin. Biotechnol.**, v.8, p.268-273, 1997.
- HENSIRISAK, P. Scale-up the use of a microbubble dispersion to increase oxygen transfer in aerobic fermentation of Baker's Yeast. Blacksburg, 89p. Masters Thesis - Biological System Engineering Faculty, Virginia Polytechnic Institute and State University, 1997.
- HOROWITZ, A.; GUTNICK, D.; ROSENBERG, E. Sequential Growth of Bacteria on Crude Oil. **Appl. Microbiol.**, p.10-19, 1975.
- HUI, Y.H.(Ed.) **Bailey's industrial oil and fat products**. 5^a ed., New York: Wiley, v.2, p.125-158, v.3, p.1-63, 1996.
- JOHNSON, A. The control of fed-batch fermentation process – A survey, **Automatica**, v.23, p.691, 1987.
- JOHANSON, V.; SINGH, M.; SAINI, V.S.; ADHIKARI, D.K.; SISTA, V.; YADAV, N.K. Bioemulsifier production by an oleaginous yeast *Rhodotorula glutinis* lip-30. **Biotechnol. Lett.**, v.14, p.487-490, 1992.
- KAPLAN, N., ZOSIM, Z.; ROSENBERG, E. *Acinetobacter calcoaceticus* BD4 emulsan: reconstitution of emulsifying activity with pure polysaccharide and protein. **Appl. Environ. Microbiol.**, v.53, p.440-446, 1987.
- KARIM, M.N.; RIVERA, S.L. Artificial neural networks in bioprocess state estimation. **Adv. Biochem. Eng.**, p.461-33, 1992a.
- KARIM, M.N.; RIVERA, S.L. Comparison of feed-forward and recurrent neural networks for bioprocess state estimation. **Comp. Chem. Engng.**, v.16(Suppl.), p.S369-S377, 1992b.
- KAROL, E. O.; BARTHOLOMEW, W.H.; SFAT, M.R. Oxygen transfer and agitation in submerged fermentations. **Agric. Food Chem.**, v.1, n.4, p.302, 1953.
- KHURY, A.I.; CORNELL, J.A. **Response surfaces: designs and analyses**. New York: Dekker, 1987.
- KIHLBERG, R. The microbe as a source of food. **Ann. Rev. Microbiol.**, v.26, p.427-466, 1972.
- KIM, S.H.; LIM, E.J.; LEE, S.O.; LEE, J.D.; LEE, T.H. Purification and characterization of biosurfactant from *Norcadia* sp. L-417. **Biotechnol. Appl. Biochem.**, v.31, p.249-253, 2000.
- KITAMOTO, D.; FUZISHIRO, T.; YANAGISHITA, H.; NAKANE, T.; NAKAHARA, T. Production of mannosylerythritol lipids as biosurfactants by resting cells of *Candida antarctica*. **Biotech. Lett.**, v.14, n.4, p.305-310, 1992.

- KENNEDY, M.J.; PRAPULLA, S.G.; THAKUR, M.S. Designing fermentation media: a comparison of neural networks to factorial design. **Biotechnol.Tec.**, v.6, n.4, p.293-298, 1992.
- KROTHAPALLY, M.; PALANKI, S. A neural network strategy for batch process optimization. **Comp.Chem. Engng.**, v.21(Supp), p.S463-S468.
- KROTHAPALLY, M.; PALANKI, S. A neural network strategy for batch process optimization. **Comp.Chem. Engng.**, v.21 (Suppl.), p.S463-S468, 1997.
- LATRILLE, E.; CORRIEU, G.; THIBAUT, J. Neural network models for final process time determination in fermented milk production, **Comp. Chem. Engng.**, v.18, n.11/12, p.1171-1181, 1994.
- LEE, B-K.; KIM, J.K.. Production of *Candida utilis* molasses in different culture types. **Aquacult. Engng.**, v.25, p.11-124, 2001.
- LIN, S.-C. Biosurfactants recent advances. **J. Chem. Tecnol. Biotechnol.**, v.66, p.109-120, 1996.
- LINKO, P.; ZHU, Y.-H. Neural network programming in bioprocess variable estimation and state prediction. **J. Biotechnol.**, 21:253-270 , 1991.
- LINKO, P.; ZHU, Y.-H. Neural network modelling for real-time variable estimation and prediction in the control of glucoamilase fermentation. **Proc. Biochem.**, v.27, p.275-283 , 1992.
- LINKO, S.; RAJALAHTI, T.; ZHU, Y.-H. Neural state estimation and prediction in amino acid fermentation. **Biotechnol. Tech.**, v.9, p.607-612 , 1995.
- LINKO, S.; LUOPA, J.; ZHU, Y.-H. Neural network as software sensors in enzyme production. **J. Biotechnol.**, v.52, p.257-266 , 1997.
- LINKO, S.; ZHU, Y.-H.; LINKO, P. Applying neural network as software sensors for enzyme engineering. **Tibitech.**, v.17, p.155-162 , 1999.
- LIU, Y-C., WANG, F-S., LEE, W-C. On-line monitoring and controlling system for fermentation process. **Biochem. Engng. J.**, v.7, p.17-25, 2001.
- MacKAY, D.J.C. Bayesian Interpolation. **Neural Comput.**, v.4, p.415-447, 1994.
- MARÇAL, M.C.R. **Produção de biopolímeros por *Candida lipolytica* em meios suplementados por óleos vegetais (babaçu, coco, dendê).** Recife, 1991. 147 p. Dissertação (Mestrado em Nutrição) - Centro de Ciências da Saúde, UFPE, 1991.
- MARQUARDT, D., An algorithm for least squares estimation of non-linear parameters. **Joc. Soc. Ind. Appl. Math.**, v.11, p.431-441, 1963

- MARTIN, A.M. GODDARD, S., BEMISTER, P. Production of *Candida utilis* biomass as aquaculture feed. **J. Sci. Food Agric.**, v.61, p.363-370, 1993.
- MONTAGUE, G.A.; MORRIS, A.J.; WARD, A.C. Fermentation, monitoring and control: a perspective. **Biotechnol. Genet. Engng. Rev.**, v.7, p.147-188, 1989.
- MONTAGUE, G.A.; MORRIS, A.J.; THAM, M.T. Enhancing bioprocess operability with generic software sensors. **J. Biotechnol.**, v.25, p.183-201, 1992.
- MONTGOMERY, D.C. **Design and analysis of experiments**. New York: Wiley, 1997.
- MONTGOMERY, D.C.; RUNGER, G.C. **Applied statistics and probability for engineers**. New York: John Wiley, 1994.
- MOU, D.G. AND COONEY, C.L. Growth monitoring and control through computer-aided on-line mass balancing in a fed-batch penicillin fermentation, **Biotechnol. Bioeng.**, v.25, p.255, 1983.
- MULLIGHAN, C.N.; GIBBS, B.F. Factors influencing the economics of biosurfactants. In N.Kosaric (ed.), **Biosurfactants: production, properties, applications**. Marcel Dekker, Inc., New York, N.Y. p.329-371.
- MURALIDHAR, R.V.; CHIRUMAMILA, R.R.; MARCHANT, R.; NIGAM, P.A. response surface approach for the comparison of lipase production by *Candida cylindracea* using two different carbon sources. **Biochem. Engng. J.**, v.9, p.17-23, 2001.
- MUSIAL, I.; RYMOWICZ, W.; CIBIS, E. Optimization of single-cell biomass production by *Yarrowia lipolytica* using response surface methodology and pulse method. **Eletr. J. Polish Agricult. Universities**, v.7, 2004. Disponível em: <http://www.ejpau.media.pl/series/volume7/issue1/biotechnology/art-02.html>.> Acesso em: 10/02/2004.
- MYERS, R.H.; MONTGOMERY, D.C. **Response surface methodology: process and production optimization using designed experiments**. New York: Wiley, 1995.
- NAHAS, E.P.; HENSON, M.A.; SEBORG, D.E. Nonlinear internal model control strategy for neural network models. **Comp. Chem. Engng.**, v.16, n.12, p.1039-1057, 1992.
- NAJJAR, Y.M.; BASHEER, I.A.; HAJMEER, M.N. Computational neural network for predictive microbiology. **Method. Int. J. Food Microbiol.**, v.34, p.27-49, 1997.
- NASCIMENTO, A.E.; CAMPOS-TAKAKI, G.M. Effects of sodium dodecyl sulfate on lipase of *Candida lipolytica*. **Appl. Biochem. Biotechnol.**, v.49, p.93-99, 1994.
- NASCIMENTO, C.A.O.N.; GIUDICI, R.; GUARDANI, R. Neural network based approach for optimization of industrial chemical process. **Comp. Chem. Engng.**, v.24, p.2303-2314, 2000.

- NAVON-VENEZIA, S.; ZOSIM, Z.; GOTTLIEB, A.; LEGMANN, R.; CARMELI, S.; RON, E.Z.; ROSENBERG, E. Alasan, a new bioemulsifier from *acinetobacter radioresistens*. **Appl. Environ. Microbiol.**, v.61, p.3240-3244, 1995.
- MARTIN, A.M.; GODDARD, S.; BEMISTER, P. Production of *Candida utilis* biomass as aquaculture feed. **J. Sci. Food Agric.**, v.61, p.363-370, 1993.
- PALEJWALA, S.; DESAI, J.D. Production of an extra cellular emulsifier by a gram negative bacterium. **Biotechnol. Lett.**, v.2, p.115-118, 1989.
- PAREILLEUX, A. Hydrocarbon assimilation by *Candida lipolytica*: formation of a biosurfactant: effects on respiratory activity and growth. **Eur. J. Appl. Microbiol. Biotechnol.**, v.8, p.91-101, 1979.
- PLACKET, R.L.; BURMAN, J.P. The design of optimum multifactorial experiments. **Biometrika**, v.33, p.305-325, 1946.
- PODDIG, T.. Short_Term Forecasting of the USD/DM-Exchange rate. In: **Neural Network in Capital Markets**, 1993, Londres. Anais... Londres, 1993. p.1.
- PONS, M-N. (Ed.) **Bioprocess Monitoring and Control. Hanser series in Biotechnologie**. Munich, Viena, New York, Barcelona: Hanser Publishers, Nancy: Marie-Noele Pons, Laboratoires des Science du Génie Chimique, 1992. 365 p.
- PURI, S.; BEG, Q.K.; GUPTA, R. Optimization of alkaline protease production from *Bacillus sp.* by response surface methodology. **Curr. Microbiol.**, v.44, p.286-290, 2002.
- RANE, K.D.; SIMS, K.A. Citric acid production by *Yarrowia lipolytica*: effects of nitrogen and biomass concentration on yield and productivity. **Biotech.Lett.**, v.18, p.1139-1144, 1996.
- RAO, P.V.; JAYARAMAN, K.; LAKSHMANA, C.M. Production of lipase by *Candida rugosa* in solid-state fermentation. 2: Medium optimization and effect of aeration. **Proc. Biochem.**, v.28, p.391-395, 1993.
- RAPP, P.; BOCK, H.; WRAY, V.; WAGNER, F. Formation, isolation and characterization of trehalose dimycolates from *Rhodococcus erythropolis* grown on n-alkanes. **J.Gen. Microbiol.**, v.115, p.491-503, 1979.
- RAO, J.L.U.M.; SATYANARAYANA, T. Statistical optimization of a high maltose-forming hyperthermostable *Geobacillus thermoleovorans* using response surface methodology. **J. Appl. Microbiol.**, v.95, p.712-718, 2003.
- REFENES, A.N. Currency exchange rate prediction and neural network design strateies. **Neural Comp. Appl. J.**, v.1, n.1, p.46-58, 1992.
- REISFELD, A.; ROSENBERG, E.; GUTNICK, D. Microbial degradation of crude oil: factors affecting the dispersion in sea water by mixed and pure culture. **Appl. Microbiol.**, v.24, p.363-368, 1972.

- REVAH, S.; LEBEAULT, J.M. Accelerated production of blue cheese flavors by fermentation on granular curds with lipase addition. *Le Lait*, v.69, p.281-289, 1989.
- RIPLEY, B.D. **Pattern Recognition and Neural Networks**. Cambridge: Cambridge University Press, 1996.
- ROBERT, M.; MERCADÉ, E.; ANDRÉS, C.; ESPUNY, M.J.; MANRESA, M.A.; GUINEA, J. Optimización de la producción de biotensoactivos por *Pseudomonas aureginosa* 44T1. *Grasas y Aceites*, v.42, n.1, p1-7, 1991.
- ROBBINS, J.W.; TAYLO, K.B. Optimization of *Echerichia coli* growth by controlled addition of glucose, *Biotechnol. Bioeng.*, v.34, p.1289, 1989.
- ROCHA, J.A. Estudos preliminares da produção de bioemulsificantes através de *Candida lipolytica* utilizando como fonte de carbono óleo de babaçu. In: **Encontro Nacional de Microbiologia Ambiental**, 1996, Fortaleza. Resumo... Fortaleza, 1996. p.5.
- RODRIGUES, J.N. **Reestruturação por Mistura e Interesterificação da Gordura do Leite com Óleo de Milho**. São Paulo, 2002. 132p. Tese (Doutorado em Tecnologia Bioquímico-Farmacêutica) - Faculdade de Ciências Farmacêuticas, USP, 2002.
- ROSE, A.H.; HARRISON, J.S. Yeast technology. In: **The yeasts**, London: Academic Press, 1970, v.3, 590p.
- ROSENBERG, E.; ENGLANDER, E.; HOROWITZ, A.; GUTNICK, D. In: **Impact of the use of microorganisms on aquatic environment**. USA: Bourquin, A.W. *et al.*, p.157, 1975. (EPA Report 660-3-75-001).
- ROSENBERG, E.; ZUCKERBERG, A.; RUBINOVITZ, C.; GUTNICK, D. Emulsifier of *Arthrobacter* RAG-1: Isolation and emulsifying properties. **Appl. Environ. Microbiol.**, v.37, n.3, p.402-408, 1979.
- ROSENBERG, E. Microbial surfactants. **Crit. Rev. Biotechnol.**, v.3, p.109-132, 1986.
- ROSENBERG, E.; RON, E.Z. High and low-molecular-mass microbial surfactants. **Appl. Microbiol. Biotechnol.**, v.52, p.154-162, 1999.
- RUFINO, R.D.; SARUBBO, L.A.; CAMPOS-TAKAKI, G.M. Produção de biossurfactante por *Candida lipolytica* utilizando substratos de baixo custo. In: **XV Simpósio Nacional de Bioprocessos**, 2005, Recife. Anais ... Recife, 2005.
- RUMELHART, D.E.; McCLELLAND, J.D. **Parallel Distributed Processing: Exploration in the Microstructure of Cognition**, v.1, Cambridge: MIT, 1986.
- RUMSEY, G.L. Recent Advances in nutrition of salmonids. **Salmonoid**, v.2, n.4, p.14-17, 1978.

- SAN, K.Y.; STEPHANOPOULOS, G. Optmization of fed-batch penicilin fermentation: a case of singular optimal control with state constraints. **Biotechnol. Bioeng.**, v.34, p.72, 1989.
- SARUBBO, L.A. **Otimização da produção de biopolímeros por *Candida lipolytica* utilizando óleo de babaçu como fonte de carbono.** Recife, 1997. 108 p. Dissertação (Mestrado em Nutrição) - Centro de Ciências da Saúde, UFPE, 1997.
- SARUBBO, L.A.; PORTO, A.L.F.; CAMPOS-TAKAKI, G.M. The use of babassu oil as substrate to produce bioemulsifiers by *Candida hypolitica*. **Can.J.Microbiol.**, v.45, p.423-426, 1999.
- SARUBBO, L.A.; MARÇAL, M.C.; NEVES, M.L.C.; SILVA, M.P.C.; PORTO, A.L.F.; CAMPOS-TAKAKI, G.M. Bioemulsifier production in batch culture using glucose as carbon source by *Candida lipolitica*. **App. Biochem. BioTec.**, v.95, p.1-9, 2001.
- SAVKOVIC-STEVANOVIC, J. Neural networks for process analysis and optimization: modeling and aplications. **Comp.Chem.Engng.**, v.18, n.11/12, p.1149-1155, 1994.
- SHENE, C.; DIEZ, C.; BRAVO, S. Neural Network for the prediction of state of *Zymomonas mobilis* CP4 batch fermentations. **Comp.Chem. Engng.**, v.23, p.1097-1108, 1999.
- SHIMIZU, K. A. Tutorial review on bioprocess systems engineering. **Comp.Chem. Engng.**, v.20, p.6/7, p.915-941, 1996.
- SILVA, T.A.L.; VILAR, D.S.; SARRUBO, L.A.; MESSAIS, A.S.; CAMPOS-TAKAKI, G.M. Crescimento e detecção de atividade lipásica e de surfactantes produzidos por amostras de *Candida* em derivados de petróleo. In: **I Encontro de Pesquisa em Ciências Ambientais**, 2001, Recife. Resumo... Recife, 2001. p.41.
- SINGER, M.E. Microbial biosurfactants. **Microbes Oil Recovery**, v.1, p.19-38, 1985.
- SINGH, M.; SAINI, V.S.; ADHIKARI, D.K.; DESAI, J.D.; SISTA, V.R. Production of bioemulsifier by a SCP-producing straing of *Candida tropicalis* during hydrocarbon fermentation. **Biotechnol. Lett.**, v.14, p.487-490, 1992.
- STEPHANOPOULOS, G.; STEPHANOPOULOS, G. Artificial intelligence in development and design of biochemical process, **Trends in Biotechnol.**, p.241-249, 1986.
- STEPHANOPOULOS, G.; SAN, K.Y. Studies on on-line bioreactor identification. I. Theory, **Biotechnol. Bioeng.**, v.26, p. 101-115, 1990.
- STRICKLAND, J.D.H.; PARSONS, T.R. **A manual of sea water analysis.** Bulletin fisheries research ord of Canada, 1965. 125p.

- STROHM, J., DALE, R.F.; PEPPLER, H.J. Polarographic measurement of dissolved oxygen in yeast fermentations. **Appl. Microb.**, v.7, p.235, 1959.
- SYU, M.-J.; TSAO, G.T. Neural Network Modeling of batch cell growth pattern. **Biotechnol. Bioeng.**, v.42, p.376-380, 1993.
- TAYLOR, B.; DARRAH, M.; MOATS, C. Verification and validation of neural networks: a sampling of research in progress. In: **Proceedings SPIE 2003 – The International Society for Optical Engineering**, San Diego, CA. Proceedings ... San Diego, 2003. v.6 5103-4, p.1-9, 2003.
- TANG, Z. Time series forecasting using neural network vs. Box Jenkins methodology In: **Artificial Neural Network: Forecasting Time Series**. Los Alamitos, CA: IEEE Computer Society Press, p.20-27, 1994.
- THAM, M.T.; MONTAGUE, A.; MORRIS, A.J.; LANT, P.A. Softsensors for process estimation and inferential control. **J. Proc. Cont.**, v.1, p.3-14, 1991.
- THANH, N.C.; SIMARD, R.E. Biological treatment of wastewater by yeasts. **J.WPCF**, v.45, n. 4, p.674-680, 1971.
- THIBAUT, J.; VAN BREUSEGEM, V.; CHÉRUY, A. On-line prediction of fermentation variables using neural networks. **Biotechnol. Bioeng.**, v.26, p.1041-1048, 1990.
- THOLUDUR, A.; RAMIREZ, W.F. Optimization of fed-batch bioreactor using neural network parameter functions models. **Biotechnol.Prog.**, v.12, p.302-309, 1996.
- THOLUDUR, A.; RAMIREZ, W.F.R.; McMILLAN, J.D. Interpolated parameter functions for neural network models. **Comp.Chem.Engng.**, v.24, p.2545-2553, 2000.
- VAN BREUSEGEM, V.; THIBAUT, J.; CHÉRUY, A. Adaptive neural models for on-line prediction in fermentation. **Can. J. Chem. Eng.**, v.69, p.481-487, 1991.
- VANCE-HARROP, M.H. ; SARUBBO, L.A.; CARNEIRO DA CUNHA, M.G.; BUARQUE-GUSMÃO, N.; CAMPOS-TAKAKI, G.M. Produção de biossurfactante em meio de cultura de baixo custo suplementado com óleo de milho por *Candida lipolytica*. **Revista Symposium**, v.2, p.23-27, 1999.
- VANCE-HARROP, M.H. **Influência de fontes de carbono D-glicose e óleo de babaçu no crescimento de *Candida lipolytica* e na produção de biossurfactantes**. Recife, 2000. 72p. Dissertação (Mestrado em Biologia de Fungos) - Centro de Ciências Biológicas, UFPE, 2000.
- VANCE-HARROP, M.H.; BUARQUE-GUSMÃO, N.; CAMPOS-TAKAKI, G.M. Influência da Água do Mar na Produção de Surfactantes por *Candida lipolytica*. In: **VII Encontro Nacional de Microbiologia Ambiental**, 2000, Recife. Resumo... Recife, 2000. p.141.

- VANCE-HARROP, M.H.; CAMPOS-TAKAKI, G.M. Degradação e Remoção de Hidrocarbonetos Aromáticos Policíclicos através do Uso de Surfactantes Sintético e produzidos por *Candida lipolytica*. In: **2ª Mostra UNICAP de Pesquisa, Pós-Graduação e Extensão**, 2001, Recife. Resumo... Recife, 2001. p.92.
- VANCE-HARROP, M.H.; BUARQUE-GUSMÃO, N.; CAMPOS-TAKAKI, G.M. New bioemulsifiers produced by *Candida lipolytica* using D-glucose and babassu oil as carbon sources. **Braz. J. of Microbiol.**, v.34, p.120-123, 2003.
- VANCE-HARROP, M.H. **Potencial biotecnológico de *Candida lipolytica* na produção de biossurfactantes, nos processos de remoção e bioabsorção de pireno (derivado do petróleo)**. Recife, 2004. Tese (Doutorado em Biologia de Fungos) - Centro de Ciências Biológicas, UFPE, 2004.
- VASILACHE, A.; DAHHOU; ROUX, G.; OMA, G. Classification of fermentation process models using recurrent neural networks. **Int. J. of Sys. Sci.**, v.32, n.9, p.1139-1153, 2001.
- VILLAR, D.S.; SILVA, T.A.L.; SARUBBO, L.A.; MESSIAS, A.S.; CAMPOS-TAKAKI, G.M. Detecção de surfactantes com atividade de emulsificação produzidos por leveduras do gênero *Candida*. In: **XXI Congresso Brasileiro de Microbiologia**, 2001, Foz de Iguaçu. Resumo... Foz de Iguaçu, 2001. p.352.
- VRDOLJAK, M.M.; MARINKOVIC, G.; PAVUSEK, I.; JOHANIDES, V. Capability for degradation of crude oil hydrocarbons by sea-water microbial association of yeasts and bacteria from kvarner bay. **Periodicum Biologorum**, v.94, p.169-178, 1992.
- WANG, Y.; LUOPA, J.; RAJALAHTI, T.; LINKO, S. Strategies for the production of lipase by *Candida rugosa*; neural estimation of biomass and lipase activity. **Biotechnol. Tech.**, v.9, n.10, p.741-746, 1995.
- WASSERMAN, P.D. **Advanced methods in neural computing**. New York: Van Nostrand Reinhold, 1993.
- WILLIS, M.J.; MONTAGUE, G.A.; DI MASSIMO, C.; THAM, M.T.; MORRIS, A.J. Artificial neural networks in process estimation and control. **Automatica**, v.26, n.6, p.1181-1187, 1992.
- YANO, T.; KOBAYASHI, T.; SHIMIZU, S. Control system of dissolved oxygen concentration employing a microcomputer, **J. Ferment. Technol.**, v.59, p.295, 1981.
- YET-POLE, I.; WU, W.-T.; LIU, Y.-C. Neural network modelling for on-line state-estimation in fed-batch culture of L-lysine production. **Chem. Engng. J.**, v.61, p.35-40, 1996.

- ZAJIC, J.E.; GUIGNARD, H; GERSON, D.F. Properties and biodegradation of a bioemulsifiers from *Corynebacterium hydrocarboclaustus*. **Biotechnol. Bioeng.**, v.19, p.1303-1320, 1977.
- ZHANG, Q.; REID, J.F.; LITCHFIELD, J.B.; REN,J.; CHANG, S.-W. A prototype neural network supervised control system for *Bacillus thuringiensis* fermentations. **Biotech. Bioeng.**, v.43, p.483-489, 1994.
- ZHOU, Q.H.; KOSARIC, N. Utilization of canola oil and lactose to produce biosurfactant with *Candida bombicola*. **JAACS**, v.72, n.1, p.117-121, 1995.
- ZHU, Y.-H.; RAJALAHTI, T.; LINKO, S. Application of neural network to lysine production. Chem. Engng. J., v.62, p.207-214, 1996.
- ZINJARDE, S.; CHINNATHAMBI, S.; LACHKE, A.H.; PANT, A. Isolation of an emulsifier from *Yarrowia lipolytica* NCIM 3589 using a modified mini isoelectric focusing unit. **Lett. Environ. Microbiol.**, v.37, p. 414-420, 1979.

ANEXOS

ANEXO A - Produção Científica: Trabalhos apresentados em eventos, submetidos e aceitos para publicação e/ou publicados em periódicos, durante o desenvolvimento desta tese de doutorado.

PRODUÇÃO CIENTÍFICA

1. TRABALHOS APRESENTADOS EM EVENTOS NACIONAIS

- **XXIII Congresso Brasileiro de Microbiologia, Florianópolis, 17 a 20 de Novembro de 2003**
 - *Biosurfactant production by Candida lipolytica using experimental design.* Albuquerque, C.D.C.; Sarubbo, L.A, Filleti,A.M.F.; Campos-Takaki, G.M.
- **I Semana de Integração da Universidade Católica de Pernambuco, Recife, 2003**
 - *Desenvolvimento de Softsensor Neural para Inferência de Biomassa e de Atividade de Emulsificação de Biossurfactante produzido por Candida lipolytica.* Albuquerque, C.D.C.; Filleti,A.M.F.; Campos-Takaki, G.M.
- **II Semana de Integração da Universidade Católica de Pernambuco – Setembro de 2004- Recife Pernambuco**
 - *Biosurfactant Production by Candida lipolytica _ Study of the Culture Medium Components using Factorial Design.* Albuquerque, C.D.C.; Vance_Harrop, M.H., Sarubbo, L.A.; Filleti,A.M.F.; Campos-Takaki, G.M.
- **IV Congresso Brasileiro de Micologia - 17 a 20 de Outubro de 2004 -Ouro Preto - Minas Gerais.**
 - *Estabilidade de Emulsões formadas entre n-hexadecano e cultivos livres de células de Candida lipolytica UCP 988.* Albuquerque, C.D.C.; Filleti,A.M.F.; Campos-Takaki, G.M.
- **IX Encontro Nacional de Microbiologia Ambiental – 16 a 19 de novembro de 2004 - Curitiba –Paraná**
 - *Estabilidade de Emulsões formadas por Biossurfactantes produzidos por Candida lipolytica UCP 988:* Albuquerque, C.D.C., Fileti, A.M.F. e Campos-Takaki, G.M.

- *Produção de Biosurfactante por Candida lipolytica* UCP 988. Albuquerque, C.D.C.; Filleti, A.M.F.; Campos-Takaki, G.M.
- **XXIII CONGRESSO BRASILEIRO DE MICROBIOLOGIA – 22 a 25 de novembro de 2005 – Santos – São Paulo.**
 - Otimização de produção de bioemulsificante por *Candida lipolytica* usando metodologia de superfície de resposta
 - Ampliação de escala de processo de produção de bioemulsificante por *Candida lipolytica*
- **IV ENCONTRO DE PESQUISA DA FACULDADE DE ENGENHARIA QUÍMICA - Novembro de 2005 – Campinas – São Paulo.**
 - Ampliação de escala e estudo dos efeitos da agitação e da temperatura no processo de produção de bioemulsificante por *Candida lipolytica*
 - Neural software sensors: Application to biosurfactant production process.

2. TRABALHOS APRESENTADOS EM EVENTOS INTERNACIONAIS

- **BioMicroWorld 2005 - Internacional Conference on Environmental, Industrial and Applied Microbiology - Badajoz Spain - March 2005.**
 - “Neural Network Based Software Sensors: Application to biosurfactant Production by *Candida lipolytica* “. Albuquerque, C.D.C., Campos-Takaki, G.M. and Fileti, A.M.F

3. ARTIGOS ACEITOS PARA PUBLICAÇÃO INTERNACIONAL

- **Artigo completo aceito para publicação no livro "Recent Research Developments in Multidisciplinary Applied Microbiology (2005). Understanding and Exploiting Microbes and Their Interactions. Biological, Physical, Chemical and Engineering Aspects", a ser publicado e distribuído internacionalmente pela Wiley-VCH.**

- “Neural Network Based Software Sensors: Application to biosurfactant Production by *Candida lipolytica* “. Albuquerque, C.D.C., Campos-Takaki, G.M. and Fileti, A.M.F..
- **Artigo completo aceito para publicação no periódico Canadian Journal of Microbiology em Janeiro de 2005**
 - *Optimizing the medium components in bioemulsifiers production by Candida lipolytica using response surface method.*

4. ARTIGO PUBLICADO

- Artigo completo publicado no periódico Revista Symposium. ISSN 0039-7695 v.1, p.109-105,2005.
 - *Biosurfactant Production by Candida lipolytica _ Study of the Culture Médium Components using Factorial Design.* Albuquerque, C.D.C.; Vance_Harrop, M.H., Sarubbo, L.A.; Filleti,A.M.F.; Campos-Takaki, G.M.

ANEXO B - Carta de aceitação do artigo "Neural Network Based Software Sensors: Application to Biosurfactant Production by *Candida lipolytica*" para publicação no livro "Recent Research Developments in Multidisciplinary Applied Microbiology (2005). Understanding and Exploiting Microbes and Their Interactions. Biological, Physical, Chemical and Engineering Aspects" a ser publicado e distribuído internacionalmente pela Willey-VCH.

Dear Dr G. M. Campos-Takaki,

I am pleased to inform that your paper entitled "**NEURAL NETWORK BASED SOFTWARE SENSORS: APPLICATION TO BIOSURFACTANT PRODUCTION BY *Candida lipolytica***", related to a work presented during the last BioMicroWorld-2005 Conference on Environmental, Industrial and Applied Microbiology, held in March 2005 in Badajoz (Spain), and submitted as a full-length manuscript, has been found suitable and accepted for publication in the book "*Recent Research Developments in Multidisciplinary Applied Microbiology (2005). Understanding and Exploiting Microbes and Their Interactions. Biological, Physical, Chemical and Engineering Aspects*", to be published and internationally distributed by Wiley-VCH. A copy of the book will be sent to you by post as soon as it has been released (expected by the end 2005).

Please note that if you intend to make any minor amendment, you may do it, but we should receive your final paper no later than October 17th 2005.

A formal acceptance letter will follow this email notification in short. You will be also contacted soon as regard as to the copyright transfer to Wiley that the publisher requires to publish your article.

Thank you very much for your contribution to the book.

Best regards,

A. Mendez-Vilas
Editor
FORMATEX Research Centre
C/Zurbaran, 1, 2ª Planta, Local 1
06001 Badajoz
SPAIN

ANEXO C - Artigo "Neural Network Based Software Sensors: Application to Biosurfactant Production by *Candida lipolytica*" a ser publicado e distribuído internacionalmente pela Willey-VCH. no livro "Recent Research Developments in Multidisciplinary Applied Microbiology (2005). Understanding and Exploiting Microbes and Their Interactions. Biological, Physical, Chemical and Engineering Aspects."

NEURAL NETWORK BASED SOFTWARE SENSORS: APPLICATION TO BIOSURFACTANT PRODUCTION BY *Candida lipolytica*

C.D.C. Albuquerque^{1,3}, G. M. Campos-Takaki^{2*}, A. M. F. Fileti³

¹Núcleo de Pesquisas em Ciências Ambientais-NPCIAMB, Departamento de Estatística e Informática-UNICAP; ²Núcleo de Pesquisas em Ciências Ambientais-NPCIAMB, Departamento de Química -UNICAP; ³Departamento de Engenharia de Processos Químicos, Faculdade de , UNICAMP.

Abstract - This paper showed that on-line estimation and multi-step ahead prediction of emulsification activity and biomass concentration could be satisfactorily carried out employing well-trained feedforward backpropagation neural networks with one hidden layer. The results showed that neural 'software sensors' supplied for biomass concentration and emulsification activity on-line estimation and prediction within an acceptable variation of 5% of the experimental values. Coefficients of determination higher than 0.90 indicated excellent agreement of the neural network models with experimental test values, obtained for biomass concentration and emulsification activity.

Keywords: neural network, software sensor, biosurfactant, emulsification activity, biomass, *Candida lipolytica*

1 Introduction

The lack of reliable sensors for on-line measurements of primary process variables constitutes a great problem for bioprocess control and automation purposes [1]. To overcome this obstacle, numerous 'software sensors' have been proposed in the literature. 'Software sensors' are mathematical algorithms, which provide reliable real time estimation of unmeasured variables by using their correlation with available data. Well-designed software sensors are cheap and precious tools, able to work in parallel with, and eventually replace, a real sensor when it is affected by a fault or taken off for maintenance [2]. Among conventional techniques used for development of 'software sensors', artificial neural networks have showed to be a powerful tool for modelling and control of complex bioprocess. The present work deals with the development of neural network based software sensors for real time estimation and prediction of biomass concentration and emulsification activity in a biosurfactant production process by *Candida lipolytica*. Bioemulsifiers have received increasing attention in recent years because of their role in the growth of microorganisms on water-insoluble hydrophobic materials such as hydrocarbons and because of their commercial potential in the cosmetics, food and agriculture industries [3]. However, the most promising applications are cleaning of oil-contaminated recovery, recovery of crude oil from sludge, and bioremediation of sites contaminated with hydrocarbons, heavy metals, and other pollutants [4]. The choice of inexpensive raw materials is important to the overall economics of the process because they account for fifty percent of the final product cost [5]. The yeast *Candida lipolytica* has been used with success, inclusive in this work, in the biopolymers production with emulsification activity, using culture medium of low cost, containing corn oil and sea water [6].

2 Materials and Methods

2.1 Microorganism: *Candida lipolytica* UCP 988 were maintained at 4°C, on yeast mold agar (YMA) slants containing (w/v): yeast extract (0.3%), malt extract (0.3%), D-glucose (1%), tryptone (0.5%), and agar (1.5%). The pH was adjusted to 5.0 with HCl.

* e-mail: takaki@unicap.br

2.2 Biosurfactant production: Seed medium - SWDW-PASUG-2 – was composed of Sea Water (50.0% v/v), Distillated Water (50.0% v/v), Potassium Phosphate (2.628% w/v), Ammonium Sulphate (2.130% w/v), Urea (0.544% w/v), D-Glucose (5% w/v). The initial pH of the production media was adjusted to 5.3 with NaOH 40%. The inoculum for bioemulsifier production was prepared in four Erlenmeyer flasks with capacity of 500 mL containing 100 mL of SWDW-PASUG-2 medium. Then this suspension was incubated at 28° C for 48 h at 150 rpm, and the culture having approximately 10^8 cells/mL, was used to inoculate the bioreactor at 10% v/v. Cultivations were conducted in a 5L bioreactor (BioFlo2000, New Brunswick) containing 4 L of SWDW-PASUG-2 medium: Sea Water (0.5% v/v), Distillated Water (0.5% v/v), Potassium Phosphate (2.628% w/v), Ammonium Sulphate (2.130% w/v), Urea (0.544% w/v), Corn Oil (5% v/v). The initial pH of the production media was adjusted to 5.3 with NaOH 40%. Dissolved oxygen and pH of the media were not controlled during the experiments.

2.3 Analysis: Samples were taken every 0h, 4h, 18h and 24 h in the first day and every 24h during seven days, and biomass concentration (given as dry weight), emulsification activity, optical density and cell-free filtrate salinity were assayed off-line. Real time information was obtained by monitoring the temperature, agitation rate, pH and dissolved oxygen. Emulsification activity was evaluated according to the method described by Cirigliano and Carman [7]. Cell-free filtrates were prepared for each culture, and the emulsification activity for water-in-hexadecane emulsions, water-in-corn oil emulsions and water-in-canola oil were determined.

2.4 Training Data and Training Procedure: The data set required to train and test the neural software sensors was obtained from biosurfactant production experiments carried out using corn oil and sea water based media in a 5L bioreactor, under different temperature and agitation conditions. A 2^2 full-factorial design with three replicates at the center point was employed. Five experimental data set were obtained at 28°C and 150 rpm; 28°C and 300 rpm; 31 °C and 150 rpm; 31 °C and 300 rpm and 29.5 °C and 225 rpm [8]. Data was recorded from the available on-line sensors and from off-line sample analysis performed every 0h, 4h, 18h and 24 h in the first day and every 24h during seven days. Several neural network topologies with one hidden layer were tested in the biosurfactant production process for estimation and prediction of emulsification activity and biomass concentration [9]. The input process variables included pH, temperature, agitation rate, dissolved oxygen, optic density, salinity of the cell-free filtrate and also biomass concentration and emulsification activity at times (t-1) and (t-2), among other variables. The measurements patterns used to train and test the network were normalized between 0.1 and 0.9, because these values were found to improve convergence speed. The training and test sets were smoothed and expanded by interpolation using a piecewise smoothing cubic spline. Initially, all data of the learning set were presented to the network and a directed mode learning procedure was used. The procedure was performed ten times with different initial random weights, for each architecture, and the more accurate model was chosen. After the connection weight matrix determined in the step 1 was used to perform semi-directed learning with a prediction horizon of 1h or 2h and a moving window of 1h [10]. The training algorithm used was the Levenberg-Marquardt based backpropagation algorithm, in conjunction with Bayesian regularization. The sigmoid and linear functions

were used, respectively, as activation functions in hidden and output layers. The root mean square error (RMSE) and the determination coefficient (R^2) were used to compare model performances.

3 Results and discussion

The contribution that neural networks can make in biosurfactant production scientific research was investigated. This investigation was concerned with evaluation of various modeling strategies, based on neural networks, when only small experimental data set was disponible. The various neural models were trained and tested with experimental process data. The noisy off-line data was first smoothed using piece smoothing cubic spline. The smoothing of relatively sparse off-line data was necessary not only to eliminate the noise related to the measurement errors but also to expand, by interpolation, the data set used for neural model training. Neural network were constructed using Neural Network Toolbox 4.01/ Matlab 6.01 (Mathworks). The performance of neural networks were evaluated based on the root mean square error (RMSE) and the determination coefficient (R^2), using normalized variable between 0.1 and 0.9. After the training was completed, the networks were tested by presenting experimental data sets which were not used during training. The proposed networks performance was verified on a number of test cases. Also the correct operation was analysed in the presence of measurement noise and disturbance. A comparison of operations speeds of all networks was made. To evaluate the performance for several architectures, various neural networks prototypes were introduced and various parameters were tried for each neural network. By using biosurfactant production as the example case, the effect of the number of hidden neuron on the goodness of fit in state variable estimation was investigated for each output variable (emulsification activity and biomass concentration) by multivariate regression [9]. Neural networks of 10-M-2 topology, with pH, temperature, agitation rate, dissolved oxygen, optic density, salinity of the cell-free filtrate and also biomass concentration and emulsification activity at times (t-1) and (t-2) forming the input vector, and emulsification activity and biomass, at time t, forming the output vector were employed. Each network was trained separately according with the methods section. The results of testing with two example experiments, evaluated statistically on the basis of the coefficient of determination (R^2) are shown in Fig 1. In all examined cases the optimum number of hidden neuron was 7, with an obvious increase in the calculation time and overfitting when too many hidden neurons were used. The smallest number of hidden neurons, which gave a satisfactory result, was chosen in order to minimize the calculation time [9]. In the present work, neural network of a 10-7-1 topology was successfully used for estimation of emulsification activity and biomass concentration, separately. A relatively good fit to the experimental off-line analyzed data was clearly evident, with $R^2 = 0.95$ for emulsification activity and $R^2 = 0.92$ for biomass concentration. The results of the testing procedure for the simultaneous prediction both of emulsification activity and biomass concentration two time-steps ahead were good. The testing procedure with a well trained neural network of 10-8-2 topology performed very nicely, especially in the prediction of the emulsification activity. Highly satisfactory coefficients of determination, $R^2 = 0.94$ for emulsification activity and $R^2 = 0.90$ for biomass, were obtained.. The results also showed that, in general, for neural models, the RMS values are slightly lower (1.9%-3.1%) and nearly independent of the horizon prediction.

This preliminary results showed that neural 'software sensors' supplied for biomass concentration and emulsification activity on-line estimation and prediction within an acceptable variation of 5% of the experimental values. Coefficients of determination higher than 0.90 indicated excellent agreement of the neural network models with experimental test values, obtained for biomass concentration and emulsification activity.

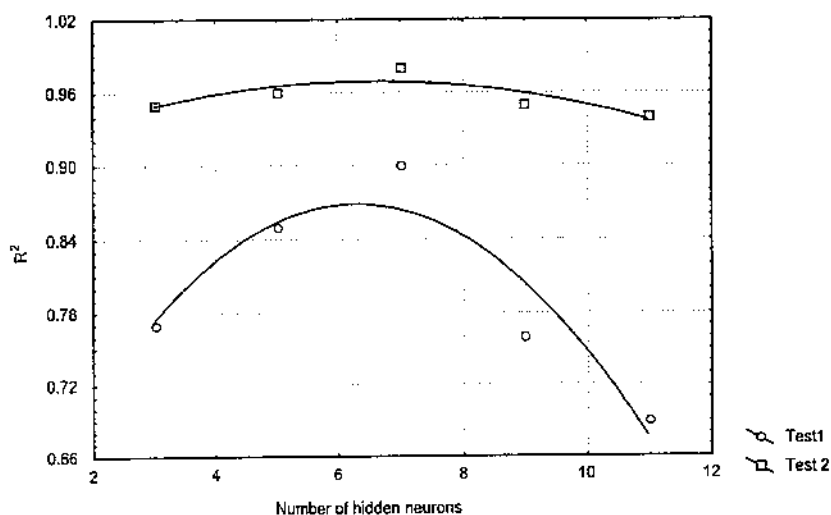


Figure 1. Estimation of biosurfactant activity with neural networks of varying number of hidden neurons, tested with two example cases.

Acknowledgments

This work was carried out with financial support by UNICAP, UNICAMP, FINEP/CT-PETRO, CNPq and CNPq/CT-PETRO.

References

- [1] J.Thibault, V.V.Breusegem and A.Chéury. *Biotechnology and Bioengineering*, **36**, 1041 (1990).
- [2] L.Fortuna, A. Rizzo, M.Sinatra and M.G. Xibilia. *Control Engineering Practice*, **11**,1491 (2003).
- [3] E. Rosenberg. *Critical Reviews Biotechnology*. **3**, 109 (1986).
- [4] J.D. Desai and I.M. Banat. *Microbiol. Mol. Bio. Rev.*. **61**(1),47, (1997).
- [5] Q-H Zhou and N. Kosaric. *JAOCS*, **72**, 67 (1995).
- [6] M.H. Vance-Harrop, L.A.Sarubbo, M.G. Carneiro da Cunha, N.Gusmão, G.M. Campos-Takaki. *Revista Symposium*, **2**,23,(1999).
- [7] M.C.Cirigliano and G.M.Carman. *Applied and Environmental Microbiology*. **74**7, (1984).
- [8] M.N. Karim and S.I. Rivera. *Advances in Biochemical Engineering Biotechnology*, **46**,1(1992).
- [9] S. Linko, J.Luopa and Y.-H.Zhu. *Journal of Biotechnology* , **52**, 257 (1997).
- [10] E.Latrilie, G.Corrieu and J. Thibault. *Computers Chemical. Engineering*, **11/12**, 1171 (1994).

ANEXO D - Carta de aceitação do artigo “Optimizing the medium components in bioemulsifiers production by *Candida lipolytica* using response surface method” pelo periódico Canadian Journal of Microbiology (CJM).

FILE: A04-053, 04-008121

Dr. G. Campos-Takaki
Departamento de Química
Núcleo de Pesquisas em Ciências Ambientais
Universidade Católica de Pernambuco-UNICAP
Rua Nunes Machado, 42.
Bloco J Térreo, Boa Vista
Recife, PE
50050-590, Brazil

Dear Dr. Campos-Takaki:

Re: Optimizing the medium components in bioemulsifiers production by *Candida lipolytica* using response surface method

We are pleased to inform you that your manuscript is acceptable for publication in the Canadian Journal of Microbiology.

Your contribution will appear in the Applied Microbiology & Biotechnology section of the first available issue.

Galley proofs will be mailed to you directly from the printer (approximately 12 weeks from the date of acceptance).

Corresponding authors must proofread and provide approval for their galley proofs within 48 hours of receipt.

If you wish to see your manuscript in the next available issue, please return your revised paper to us as soon as possible. Those authors planning to be away for any length of time should designate a backup to proofread and approve galley proofs. Galley proofs not approved within a minimum time period will be published as is. CJM is now publishing papers as they are ready on the web, in advance of the paper copy. Once you have approved your galley proofs, you can expect to find your paper published on the web-site within 6-8 weeks at <http://cjm.nrc.ca>.

If any changes are to occur to your contact details from this point forward please contact our Publication Officer,
Ms. Seiden, in order to keep the line of communication flowing during the publication process.

Also, if you have any questions regarding publication or proofs, please contact our Publication Officer:

Jacqueline Costigan PHONE: (613) 993-9038
Publications Office FAX: (613) 952-7656
Research Press EMAIL: Jacqueline.Costigan@nrc-cnrc.gc.ca
National Research Council of Canada
Ottawa, Ontario K1A 0R6

If your paper includes biomolecular interactions, between any combination of protein, DNA, RNA, ligand, or molecular complex, please identify this to the editorial office at cjm.journal@usask.ca.

The Journal is participating in a pilot project with BIND (Biomolecular Interaction Network Database) to make biomolecular interactions available to the scientific community. If you identify your paper as containing molecular interactions, the Journal will provide your paper directly to BIND to facilitate the inclusion of your research in this rapidly growing, freely available database at www.blueprint.org.

We thank you again for submitting your research for publication in the Canadian Journal of Microbiology.

Yours sincerely,

Drs. J.J. Germida and H.G. Deneer
Editors

bt

cc: Dr. C. Greer

ANEXO E - Artigo “Optimizing the medium components in bioemulsifiers production by *Candida lipolytica* using response surface method” a ser publicado no periódico Canadian Journal of Microbiology (CJM) até junho de 2006.

Optimizing the medium components in bioemulsifiers production by *Candida lipolytica* with a response surface method

C.D.C. Albuquerque^{1, 4}, A.M.F. Filetti², G.M. Campos-Takaki^{3, 4}

¹Departamento de Estatística e Informática, Universidade Católica de Pernambuco, Rua Nunes Machado, 42, Bloco J, Térreo, Boa Vista, 50050-590, Recife, PE, Brasil.

²Departamento de Engenharia de Sistemas de Processos Químicos, Faculdade de Engenharia Química, Universidade Estadual de Campinas, Cidade Universitária “Zeferino Vaz”, Caixa Postal 6066, 13081-970, Campinas, SP, Brasil.

³Departamento de Química, ⁴Núcleo de Pesquisa em Ciências Ambientais, Universidade Católica de Pernambuco, Rua Nunes Machado, 42, Bloco J, Térreo, Boa Vista, 50050-590, Recife, PE, Brasil.

¹Corresponding Author: (e-mail: takaki@unicap.br).

Abstract

A response surface methodology was used to study the bioemulsifier production by *Candida lipolytica*. A 2^4 full experimental design was previously carried out to investigate the effects and interactions of the concentrations of corn oil, urea, ammonium sulfate and potassium dihydrogen orthophosphate on the emulsification activity (EA) of the bioemulsifier produced by *Candida lipolytica*. The best EA value (3.727 unity of emulsification activity (UEA)) was obtained with medium composed of 0.4 g of urea, 1.1 g of ammonium sulfate, 2.04 g of potassium dihydrogen orthophosphate, 5 mL of corn oil, 50 mL of distilled water and 50 mL of seawater. A curvature check was performed and revealed lack of fit of the linear approximation. The proximity of the optimum point was evident and also was the need for quadratic model and second-order designs that incorporate the effect of the curvature. Medium constituents were then optimized for the EA using a three-factor central composite design and response surface methodology. The second order model showed statistical significance and predictive ability. It was found that the maximum EA produced was 4.415 UEA and the optimum levels of urea, ammonium sulfate and potassium dihydrogen orthophosphate were, respectively, 0.544 % (m/v), 2.131 % (m/v) and 2.628 % (m/v).

Key words: emulsification activity, factorial design, central composite design, optimization, biosurfactant.

Abstrait

Une méthodologie de surface de réponse a été employée pour étudier la production de bioémulsifier par *Candida lipolytica*. Des 24 pleines conceptions expérimentales précédente ont été effectuées pour étudier les effets et les interactions des concentrations d'huile de maïs, d'urée, de sulfate d'ammonium et d'orthophosphate de dihydrogène de potassium sur l'activité d'émulsification du biosurfactant produit par *Candida lipolytica*. La meilleure valeur d'activité d'émulsification (3.727 UEA) a été obtenue avec le milieu composé par 0,4 g d'urée, de 1,1 g de sulfate d'ammonium, de 2,04 g de phosphate de potassium, de 5 ml d'huile de maïs, de 50 ml d'eau distillée et de 50 ml d'eau de mer. Le contrôle de la courbure a été effectué et manque indiqué d'ajustement de l'approximation linéaire, indiquant la proximité du point optimum et du besoin du modèle quadratique et des conceptions de second ordre qui incorporent l'effet de la courbure. Des constituants moyens ont été alors optimisés pour l'activité d'émulsification en utilisant la conception composée centrale de trois-facteur et la méthodologie extérieure de réponse. Le modèle du second degré a montré la signification statistique et les capacités prédictives. On l'a constaté que l'activité maximum d'émulsification produite était 4,415 UEA et les niveaux optima de l'urée, sulfate d'ammonium et l'orthophosphate de dihydrogène de potassium étaient respectivement, 0,544 % (m/v), 2,131 % (m/v) et 2,628 % (m/v).

Mots Clés: Activité d'émulsification, Conception factorielle, Conception composée centrale, Optimisation, biosurfactant.

Introduction

Bioemulsifiers are amphipathic molecules that can be divided into (i) low-molecular-weight compounds such as glycolipids and phospholipids, which lower the interfacial tension between hydrophobic liquids and water and thus reduce the energy required to form emulsions, and (ii) polymers which stabilize emulsions (Rosenberg 1986).

Surfactants are surface-active compounds capable of reducing surface and interfacial tension at the interfaces between liquids, solids and gases, thereby allowing them to mix or disperse readily as emulsion in water or others liquids. The enormous market demand for surfactants is currently met by numerous synthetic, mainly petroleum-based, chemicals surfactants. These compounds are usually toxic to the ambient and non-biodegradable. They may accumulate and their production, processes and by-products can be environmentally hazardous. Tightening environmental regulations and increasing awareness for the need to protect the ecosystem have effectively resulted in an increasing interest in biosurfactants as possible alternatives to chemicals surfactants. Biosurfactants are amphiphilic compounds of microbial origin with considerable potential in commercial applications within various industries. They have advantages over their chemical counterparts in biodegradability and effectiveness at extreme temperature or pH and in have lower toxicity. Biosurfactant are beginning to acquire a status as potential performance-effective molecules in various fields. At present are mainly used in studies on enhanced oil recovery and hydrocarbon bioremediation. The solubilization and emulsification of toxic chemicals by biosurfactants have also been reported (Banat et al., 2000).

There are several examples, especially in the earlier studies, of biosurfactants that both lowered surface tension and stabilized emulsions (Cooper and Goldenberg 1987).

Biosurfactant activities can be determined by measuring the changes in superficial and interfacial tensions, stabilization or destabilization of emulsion and hydrophilic-lipophilic balance (HBL). An emulsion is formed when one liquid is dispersed as microscope droplets in another liquid continuous phase. Biosurfactants may stabilize (emulsifiers) or destabilize (deemulsifiers) the emulsion. The emulsification activity (EA) is assayed by the ability of the surfactant to generate turbidity due to suspended hydrocarbons, in an aqueous assay system. Generally, biosurfactant are produced during microorganisms growth on hydrocarbons which are usually expensive and therefore increase the overall process costs. However, other cheaper, water soluble substrates, as

glucose and ethanol, are sometimes used. In the search for cheaper raw material for biosurfactant production, industrial effluents have recently shown good promise (Desai and Banat 1997).

The yeast *Candida lipolytica* has been used with success in the biopolymers production with EA, using culture medium of low cost, containing typical vegetable oils of the North and Northeast area of Brazil (Marçal 1991, Rocha 1996, Sarubbo et al. 1999, Vance-Harrop et al. 1999, Vance-Harrop 2000, Sarubbo et al. 2001, Vance-Harrop et al. 2003).

Many bioemulsifiers and their production processes have been patented, but so far only a few have been commercialized. The fermentation process holds key to improving the overall process economics in bioemulsifier production. Recent developments in the area of optimization of fermentations conditions have resulted in a significant increase in production yields, making them more commercially attractive (Desai and Banat 1997).

To reduce the production costs of the bioemulsifier in relation to the synthetic surfactants, improvements of culture and production medium are carried out usually through 'one-factor-at-a time' methodology. The disadvantage of this optimization technique is that it does not include the interaction effects among the variables employed and it does not describe the net effect of the various medium constituents on the EA. On the other hand, full-factorial design that includes all possible combinations for each of the factors is a powerful tool for understanding complex processes. Optimization through factorial design and response surface methodology (RSM) is a common practice in biotechnology for the optimization of media components and culture conditions (Rao et al. 1993; Chen 1996; Crola and Kennedy 2001; Muralidhar et al. 2001; Adinarayana and Ellaiah 2002; Puri et al. 2002; Rao and Satyanarayana; 2003). Statistical optimization not only allows quick screening of a large experimental domain, but also reflects the role of each of the components. Basically this optimization process involves three major steps: performing the statistically designed experiments, estimating the coefficients in a mathematical model, and predicting the response and checking the adequacy of the model (Rao and Satyanarayana, 2003).

In the present study, initially, a 2^4 full factorial design was carried out to investigate the effects and interactions of the components and concentrations of the corn oil and urea, ammonium sulfate and potassium dihydrogen orthophosphate, on the EA of the bioemulsifier produced by *Candida lipolytica*. Thereafter, a three factor composite design

and a response surface methodology were used to locate the optimum levels of urea, ammonium sulfate and potassium dihydrogen orthophosphate and to further investigate the interactions among these factors during the bioemulsifier production by *Candida lipolytica* on corn oil.

Materials and methods

Yeast Culture

The yeast *Candida lipolytica* UCP 065 was obtained from the Culture Collection of the Núcleo de Pesquisas em Ciências Ambientais, Catholic University of Pernambuco, maintained at 4°C, on yeast mold agar (YMA) monthly.

Inoculum preparation

The inoculum for bioemulsifier production was prepared by aseptically transferring several loopfulls from a fresh culture of *Candida lipolytica*, grown on agar plates at 28° C for 48 h, to 250 mL Erlenmeyer flask, containing 50 mL of yeast malt broth (YMB), and incubated during 48 hours at 28° C on a rotatory shaker at 150 rpm.

EA Assay

The EA of each cell-free filtrate of the 127 h cultures was evaluated according to the method described by Cirigliano and Carman (1984). The method defines one unit of emulsification activity (UEA) as that amount of emulsifier that resulted in an absorbance of 1.0 at 540nm.

Optimization of the Bioemulsifier Production Medium Constituents

A 2⁴ full factorial design was previously carried out to verify the effects and interactions of the corn oil, urea, ammonium sulfate and potassium dihydrogen orthophosphate concentrations on the EA of bioemulsifiers produced by *Candida lipolytica*. Afterwards, a three factor central composite design was employed for optimizing the medium constituents. EA was considered as the response variable. Corn oil, urea, ammonium sulfate and potassium dihydrogen orthophosphate concentrations were considered as independent variables. The optimum values of the selected variables were obtained by solving the regression equation and also by analyzing the response surface contour plots. Variance analysis, regressions coefficients determination and three-dimensional graphics were performed using Statistica® software version 6.0 (Statsoft.Inc, USA).

The 2⁴ Full Factorial Design

In this design, a set of 20 experiments, with 4 replicates at the central points, was performed. The range and levels of the components (factors or independent variables) under study are given in Table 1. The values at the lower level were based on the results obtained by VANCE-HARROP (2000) in preliminary experiments. According to the 2⁴ full factorial design matrix (Table 3) the EA was studied at various combinations of the medium constituents using corn oil as carbon source. The components of each medium were dissolved in 50 mL of sea water and 50 mL of distilled water and the pH was adjusted to 5.3. One-milliliter aliquots of YMB suspension, containing 10⁷ cells/mL of *Candida lipolytica*, were used to inoculate each of twenty 250 mL Erlenmeyer flasks, containing 100 mL of sterile production medium. The flasks were incubated on a rotatory shaker at 150 rpm, for 127 hours at 28° C.

Three Factor Central Composite Design

In this second design, a set of 17 experiments, with 3 replicates at the central point was performed. The range and levels of the components under study were adapted and are given in Table 2. According to the three factor central composite design matrix (Table 4), the EA was studied at various combinations of the medium constituents using corn oil as carbon source. The components of each medium were dissolved in 50 mL of sea water and 50 mL of distilled water and the pH was adjusted to 5.3. One-milliliter aliquots of YMB suspension, containing 10⁷ cells/mL of *Candida lipolytica*, were used to inoculate each of seventeen 250 mL Erlenmeyer flasks, containing 100 mL of sterile production medium. The flasks were incubated on a rotatory shaker at 150 rpm, for 127 hours at 28° C.

Results and Discussion

The results of the 2⁴ experimental design and the three factor central composite design showed the ability of *Candida lipolytica* to grow on corn oil and to produce an inducible extracellular EA capable of stabilizing water-in-oil emulsion, under various media in batch cultures. Experimental setups are outlined, respectively, in Tables 3 and 4. The last columns show the uncoded values of the response variable EA, expressed UEA. The independent variables - corn oil concentration, urea concentration, ammonium sulfate concentration and potassium dihydrogen orthophosphate concentration - are coded according to the Equation [1]:

$$X_i = \frac{x_i - x_0}{\Delta x} \quad [1]$$

where X_i , x_i and x_0 are the coded value, uncoded value and the value at the center point respectively of the i -th independent variable and Δx is the step value change.

The 2⁴ full factorial design

Data presented in Table 3 indicate that high EA values were obtained under all test conditions studied in factorial design, mainly in nitrogen-phosphorus rich medium. These results are in accordance with literature (Resfield et al., 1972, Palejwala and Desai 1989, Vance-Harrop et al. 1999, Vance-Harrop 2000, Vance-Harrop et al. 2003). The best EA value (3.727 UEA) was obtained with medium composed of 0.4 g of urea, 1.1 g of sulfate of ammonia, 2.04 g of potassium dihydrogen orthophosphate, 5 mL of corn oil, 50 mL of distilled water and 50 mL of sea water (Table 3 - run number 15). Figure 1 illustrates the Pareto Chart, with 95% confidence level, for effect estimates, in absolute values. The Pareto chart was used for identifying which estimated effects are most important. On the chart, it can be seen that potassium dihydrogen orthophosphate concentration, ammonium sulfate concentration and corn oil concentration had a significant influence on the EA. On the other hand, there was no significant influence of the urea concentration on the EA in the range studied. The analysis also showed that four of the interactions, i.e., urea with potassium dihydrogen orthophosphate, urea with ammonium sulfate, corn oil with ammonium sulfate and corn oil with potassium dihydrogen orthophosphate, had significant effect on the EA. Increasing the potassium dihydrogen orthophosphate concentration, the ammonium sulfate concentration and the interaction of the urea with the phosphate showed positive significant effects on the EA. However, increasing the corn oil concentration and the interaction between urea and ammonium sulfate and the interaction between corn oil and ammonium sulfate produced negative significant effects. Curvature check was carried out and revealed lack of fit of the linear approximation. Figure 1 also shows that curvature effect crosses the 95% confidence level, indicating the optimum point proximity and the need of quadratic model and second-order designs that incorporate the effect of the curvature.

Three factors central composite design

Factor levels used for this design (Table 2) were chosen based on the results of the previous 2^4 full factorial design (Table 3). The concentration of corn oil was fixed in 5% (v/v), due to the negative significant effect produced in the 2^4 full factorial design. Medium containing corn oil in 5% (v/v) showed to be a better medium for bioemulsifier production by *Candida lipolytica*. The independent variables, ammonium sulfate concentration and potassium dihydrogen orthophosphate concentration, which presented positive significant effect, were chosen for study in the second phase of the optimization process. Although, in the 2^4 full factorial designs, urea concentration was not significant, this factor was maintained in the three factor central design, for a new study. The lower level of the 2^3 composite central design (Table 2) was chosen based on run 15 (Table 3). The results for the three factors central composite design are presented in Table 4. The experiments of the present design showed that high EA values were obtained under all test conditions studied and that the highest EA values were found at central level conditions (zero level – runs numbers 15, 16 e 17). The average EA at zero level was 4.346 UEA, 262.77 % higher than the EA of the standard bioemulsifier production process. The average EA of the experiments in the three factor central composite design is 3.632 UEA, 51,08 % higher than the average result (2.404 UEA) of the 2^4 full factorial design. These results are also very favorable when compared with those in which *Candida lipolytica* was cultivated with different carbon sources and medium components. (Cirigliano and Carman 1984, Marçal 1991, Rocha 1996, Sarubbo et al. 1999, Vance-Harrop et al. 1999, Vance-Harrop 2000, Sarubbo et al. 2001, Vance-Harrop et al. 2003).

The experimental data were then fitted to a quadratic model. The following regression equation for EA (Y) was obtained:

$$\begin{aligned} Y = & 4.347191 + 0.106366x_1 + \\ & + 0.076504x_2 - 0.295195x_3 - \\ & - 0.331140x_1^2 - 0.230730x_2^2 - \\ & - 0.327958x_3^2 - 0.284500x_1x_2 + \\ & + 0.256000x_1x_3 + 0.11450x_2x_3 \end{aligned} \quad [2]$$

where x_1 , x_2 and x_3 represent, respectively, coded values of urea concentration, ammonium sulfate concentration and potassium dihydrogen orthophosphate concentration.

The significance of each coefficient of the regression equation was determined by Student's t -test and p -values, which are listed in Table 5. The larger the magnitude of the t -value and the smaller the p -value, the more significant is the corresponding coefficient. This implies that the quadratic main effects of urea concentration, ammonium sulfate concentration and potassium dihydrogen phosphate concentration are more significant than their respective first order effects. Only first order effect of potassium dihydrogen phosphate was significant. The ammonium sulfate with potassium dihydrogen orthophosphate interaction effect was not significant.

The analysis of variance of the second order model shows that the model is highly significant (Table 6), as is evident from Fisher F test, where the calculated F value ($F_{\text{model}} = 17.39$) is much greater than the tabular F value ($F_{\text{tab}}(0.05, 9, 7) = 3.68$). The value of the determination coefficient ($R^2 = 0.95719$) indicates that only 4.281% of the total variation are not explained by the model, demonstrating the goodness of fit of the model. The value of the adjusted determination coefficient ($\text{Adj. } R^2 = 0.90214$) - that corrects the R^2 value for the sample size and for the number of terms in the model - is also very high, which also indicates a high significance of the model. The high value of signifies An excellent correlation between the experimental and predicted values was obtained once the correlation coefficient ($R = 0.97836$) is close to 1.

By means of the equation [2], the optimum coded values for urea concentration, ammonium concentration and potassium dihydrogen orthophosphate, were found to be respectively -0.052 , 0.085 and -0.456 . The corresponding optimum uncoded values were calculated to be 0.544% (m/v), 2.131% (m/v) and 2.628% (m/v), respectively. The maximum EA predicted in these optimum conditions was 4.415 UEA. The experimental runs indicated as best results an average EA of about 4.346 UEA under zero level conditions of the central composite design carried out (urea 0.7% (m/v), ammonium phosphate 2.05% (m/v) and potassium dihydrogen orthophosphate 3.12% (m/v)). This confirms the closeness of the model to the experimental results.

Graphical study also corroborated the high significance of regression model. Using three parameter models, as described by Equation [2], often difficult to visualize the interrelationships that exist between the various parameters. In order to understand the interactions of the medium components and the optimum concentration of each component

required for maximum EA, the three-dimensional response surface curves and bi-dimensional contour curves were then plotted.

Figure 2 shows the response surface and the contour plots obtained as a function of urea concentration vs. ammonium sulfate concentration, while potassium dihydrogen orthophosphate was maintained at zero level. An increase in the EA was observed when urea and ammonium sulfate concentrations moved towards the zero level conditions.

Figure 3 shows the response surface and contour plots obtained as a function of urea concentration vs. potassium dihydrogen phosphate concentration, while ammonium sulfate concentration was maintained at zero level. An increase in the EA was also observed when urea and potassium dihydrogen phosphate concentrations moved towards the zero level conditions.

Figure 4 shows the response surface and contour plots obtained as a function of ammonium sulfate concentration vs. potassium dihydrogen phosphate concentration, while urea concentration was maintained at zero level. Again, an increase in the EA was observed when ammonium sulfate and potassium dihydrogen phosphate concentrations moved towards the zero level conditions.

Predicted and observed values of EA are shown on Figure 5. The predicted values were calculated using the quadratic model developed (Equation 2). The residual analysis showed that the EA model has a normal distribution (Figure 6) and constant variance (Figure 7).

To validate the mathematical model, the EA was determined under different experimental conditions. Three experiments were conducted on predicted optimum conditions of the medium components. The experimental maximum EA perfectly agreed with predicted maximum EA (Table 6). The difference between the model prediction (4.415 UEA) and the experimental results (4.435 UEA) was less than 0.5 %. This validates the second order model (Equation [2]) used in the present work since the difference can be explained by experimental errors.

Conclusions

Statistical optimization of medium components for the production of bioemulsifier by *Candida lipolytica* was performed using experimental design and surface response methodologies. A mathematical model to represent the EA, as a function of urea, ammonium sulfate e potassium dihydrogen orthophosphate concentrations, was proposed.

The model did not show lack-of-fit and the determination coefficient indicated that the model could explain 95.72% of the variability in the EA. According to the model, the urea, ammonium sulfate e potassium dihydrogen orthophosphate concentrations necessary to attain the maximum EA (4.415 UEA) were respectively 0.544 % (m/v), 2.131 % (m/v) and 2.628 % (m/v). Validation of the model was accomplished by experiments carried out on optimized medium conditions. The results showed that the EA value predicted by the model was similar to that obtained experimentally (difference < 0.5%). The optimization enhanced the EA of the standard bioemulsifier production process in 262.8%. Response surface methodology demonstrated to be a powerful and efficient tool for media components optimization and therefore it has been used in the process scale-up to optimize the BioFlo 2000 fermentor operations conditions.

Acknowledgements

The authors are thankful to Dra. M.H. Vance-Harrop, Dra. L.A. Sarubbo and Dra.L.O.Franco for their suggestions. This work was carried out with financial support by the Catholic University of Pernambuco - UNICAP, FINEP/CT-PETRO, CNPq and CNPq /CT-PETRO.

References

- Adianarayana, K., and Ellaiah, P. 2002. Response surface optimization of the critical medium components for the production of alkaline protease by a newly isolated *bacillus* sp. J.Pharm. Pharmaceut Sci 5:272-278. Available from www.ualberta.ca/~csps. Accessed 12 October 2004.
- Banat, I.M., Makkar, R.S., and Cameotra, S.S. 2000. Potential commercial applications of microbial surfactants. Appl. Microbiol. and Biotechnol. 53 :495-508.
- Cirigliano, C.M., and Carman, G.M. 1984. Isolation of a bioemulsifier from *Candida lipolytica*. Appl. Environ. Microbiol. 48:747-750
- Chen, H.C. 1996. Optimizing the concentration of carbon, nitrogen and phosphorus in a citric acid fermentation with response surface method. Food Biotechnology. 10:13-27.
- Cooper, D.G., and Goldenberg, B.G. 1987. Surface Active Agents from Two *Bacillus* Species. Appl. Environ. Microbiol. 23:224-229.

- Crolla, A., and Kennedy, K.J. 2001. Optimization of citric acid production from *Candida lipolytica* Y-1095 using n-paraffin. *Journal of Biotechnology*. **89**:27-40.
- Desai, J.D., and Banat, I.M. 1997. Microbial production of surfactant and their commercial potential. *Microbiology and Molecular Biology Reviews*. **1**:47-64.
- Marçal, M.C.R. 1991. Produção de Biopolímeros por *Candida lipolytica* em meios suplementados por óleos vegetais (babaçu, coco, dendê). *Master Thesis* Universidade Federal de Pernambuco, Recife, Brasil.
- Muralidhar, R.V., Chirumamila, R.R., Marchant, R., and Nigam, P. 2001. A response surface approach for the comparison of lipase production by *Candida cylindracea* using two different carbon sources. *Biochemical Engineering Journal*. **9**:17-23.
- Palejwala, S., and Desai, J.D. 1989. Production of an extra cellular emulsifier by a gram negative bacterium. *Biotechnol. Lett.* **2**:115-118.
- Puri, S., Beg, Q.K., and Gupta, R. 2002. Optimization of alkaline protease production from *Bacillus sp.* by response surface methodology. *Current Microbiology*, **44**:286-290.
- Rao, P.V., Jayaraman, K., and Lakshmana, C.M. 1993. Production of lipase by *Candida rugosa* in solid-state fermentation. 2: Medium optimization and effect of aeration. *Process Biochemistry* **28**:391-395
- Rao, J.L.U.M., and Satyanarayana, T. 2003. Statistical optimization of a high maltose-forming hyperthermostable *Geobacillus thermoleovorans* using response surface methodology. *J. Appl. Microbiol.*, **95**:712-718.
- Reisfield, A., Rosenberg, E., and Gutnick, D. 1972. Microbial degradation of crude oil: factors affecting the dispersion in sea water by mixed and pure culture. *Appl. Microbiol.* **24**:363-368.
- Rocha, J.A. 1996. Estudos preliminares da produção de bioemulsificantes através de *Candida lipolytica* utilizando óleo de babaçu como fonte de carbono. In: Encontro Nacional de Microbiologia Ambiental –ENAMA, 5, 1996, Resumo, Fortaleza.
- Rosenberg, E. 1986. Microbial surfactants. *Crit. Rev. Biotechnol.* **3**:109-132.
- Sarubbo, L.A., Porto, A.L.F., and Takaki, G.M.C. 1999. The use of babassu oil as substrate to produce bioemulsifiers by *Candida lipolytica*. *Can. J. Microbiol.* **45**:423-426.
- Sarubbo, L.A., Marçal, M.C., Neves, M.L.C., Silva, M.P.C., Porto, A.L.F., and Takaki, G.M.C. 2001. Bioemulsifier Production in Batch Culture Using Glucose as Carbon Source by *Candida lipolytica*. *App. Biochem. BioTec.*, **95**:1-9.
- Vance-Harrop, M.H., Sarubbo, L.A., Carneiro da Cunha, M.G., Buarque Gusmão, N., and Campos Takaki, G.M. 1999. Produção de Biossurfactante em meio de cultura de baixo custo suplementado com óleo de milho por *Candida lipolytica*. *Revista Symposium*, **2**:23-27.

Vance-Harrop, M.H. 2000. Influência das fontes de carbono, D-glicose e óleo de babaçu no crescimento de *Candida lipolytica* e na produção de biosurfactantes. *Master Thesis*. Universidade Federal de Pernambuco, Recife, Brasil.

Vance-Harrop, M.H., Gusmão, N.B., and Campos-Takaki, G.M. 2003. New emulsifiers produced by *Candida lipolytica* using D-Glucose and Babassu oil as carbon source. *Brazilian Journal of Microbiology*. 34:120-123.

Table 1. Levels and experimental values of the 2⁴ full factorial design variables.

Independent Variable	Level		
	-1	0	+1
Corn Oil (% v/v)	5.00	7.50	10.00
(NH ₂) ₂ CO (% w/v)	0.10	0.25	0.40
(NH ₄) ₂ SO ₄ (% w/v)	0.10	0.60	1.1
KH ₂ PO ₄ (% w/v)	0.68	1.36	2.04

Table 2. Levels and experimental values of the 2^3 central composite design variables. Corn oil is fixed in 5%(v/v).

Independent Variable	Level				
	-1.68	-1	0	+1	+1.68
(NH ₂) ₂ CO (% w/v)	0.196	0.4	0.7	1.0	1.204
(NH ₄) ₂ SO ₄ (% w/v)	0.454	1.1	2.05	3.0	3.646
KH ₂ PO ₄ (% w/v)	1.3056	2.04	3.12	4.2	4.934

Table 3. The 2⁴ full factorial design results for emulsification activity.

Run n°	Corn Oil	(NH ₂) ₂ CO	(NH ₄) ₂ SO ₄	KH ₂ PO ₄	EA ^a
1	-1	-1	-1	-1	2.161
2	+1	-1	-1	-1	2.239
3	-1	+1	-1	-1	1.897
4	+1	+1	-1	-1	1.475
5	-1	-1	+1	-1	3.691
6	+1	-1	+1	-1	2.107
7	-1	+1	+1	-1	2.167
8	+1	+1	+1	-1	2.218
9	-1	-1	-1	+1	1.739
10	+1	-1	-1	+1	2.273
11	-1	+1	-1	+1	3.182
12	+1	+1	-1	+1	3.053
13	-1	-1	+1	+1	3.614
14	+1	-1	+1	+1	3.223
15	-1	+1	+1	+1	3.727
16	+1	+1	+1	+1	3.607
17	0	0	0	0	1.337
18	0	0	0	0	1.429
19	0	0	0	0	1.341
20	0	0	0	0	1.598

^a Emulsification activity (EA) obtained in each run were measured in Unity of Activity of the Emulsification (UAE).

Table 4. Three-factor central composite design results for emulsification activity.

Run n°	(NH ₂) ₂ CO	(NH ₄) ₂ SO ₄	KH ₂ PO ₄	EA ^a
1	-1	-1	-1	3.830
2	+1	-1	-1	3.806
3	-1	+1	-1	3.982
4	+1	+1	-1	3.340
5	-1	-1	+1	2.250
6	+1	-1	+1	3.770
7	-1	+1	+1	3.380
8	+1	+1	+1	3.242
9	-1.68	0	0	3.202
10	+1.68	0	0	3.640
11	0	-1.68	0	3.480
12	0	+1.68	0	3.930
13	0	0	-1.68	3.940
14	0	0	+1.68	2.920
15	0	0	0	4.436
16	0	0	0	4.408
17	0	0	0	4.194

^a Emulsification activity (EA) obtained in each run were measured in Unity of Activity of the Emulsification (UAE).

Table 5. Results of the regression analysis using the three factor composite design. Least square fit and parameter estimation

Factor	Coefficient	t-value	p-value
Mean/Inter.	4.347191*	56.99062*	0.000308
(NH ₂) ₂ CO (Q)	-0.331140*	-8.39890*	0.013882
KH ₂ PO ₄ (Q)	-0.327958*	-8.31819*	0.014147
[3]KH ₂ PO ₄ (L)	-0.295195*	-8.24076*	0.014408
1L by 2L	-0.284500*	-6.07870*	0.026012
1L by 3L	0.256000*	5.46976*	0.031837
(NH ₄) ₂ SO ₄ (Q)	-0.230730*	-5.85216*	0.027979
2L by 3L	0.114500	2.44644	0.134245
[1](NH ₂) ₂ CO (L)	0.106366	2.96935	0.097167
[2](NH ₄) ₂ SO ₄ (L)	0.076504	2.13572	0.166224

* Significant 95% confidence level term

L – linear term

Q- quadratic term

Table 6. Analysis of Variance (ANOVA) for the model regression representing emulsification activity

Source	SS	df	MS	F
Regression	4.675787	9	0.519532	17.388446
Residue	0.209143	7	0.029878	
Lack of fit	0.174095	5	0.034819	1.98693
Pure error	0.035048	2	0.017524	
Total	4.884930	16		

Table 7. Validation of response surface model. Observed experimental average emulsification activity obtained under predicted optimum conditions versus predicted maximum emulsification activity

Replicate	Emulsification Activity
1	4.38467
2	4.498999
3	4.421625
Average Experimental	4.435098
Predicted	4.414905

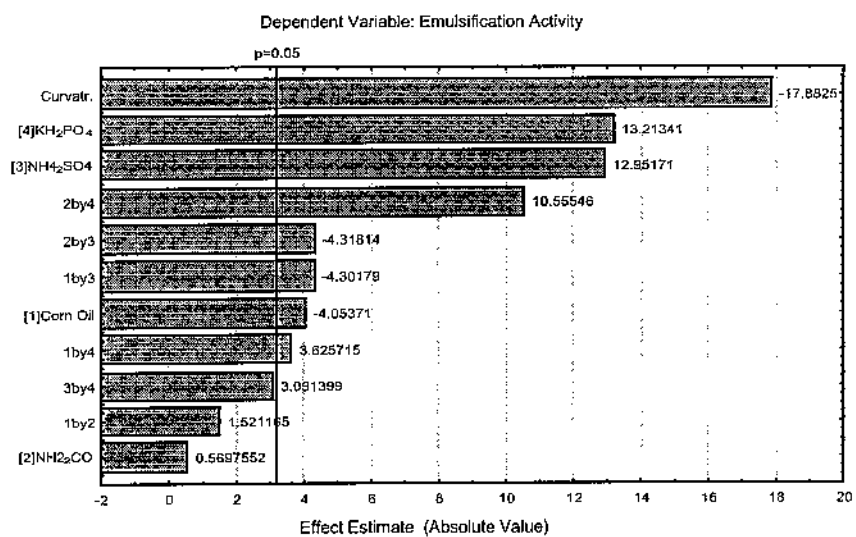


Figure 1.

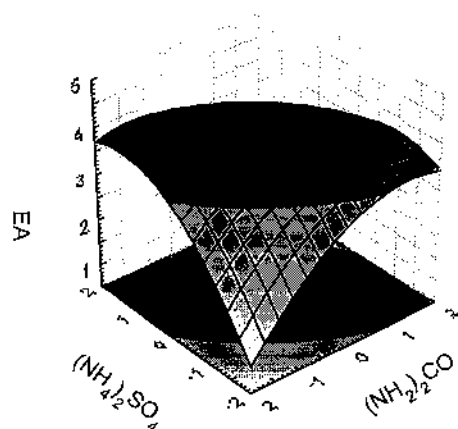


Figure 2.

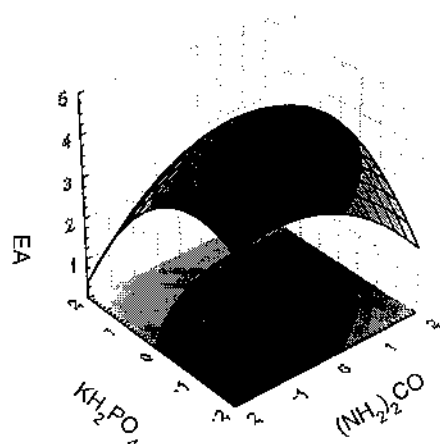


Figure 3.

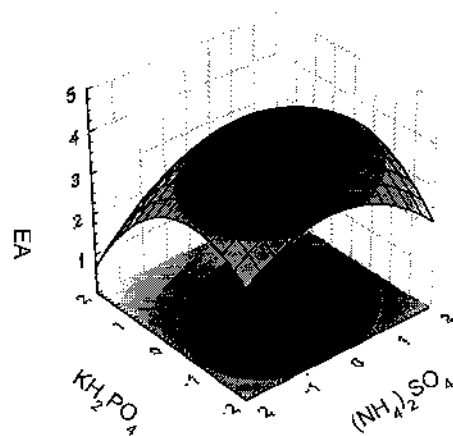


Figure 4.

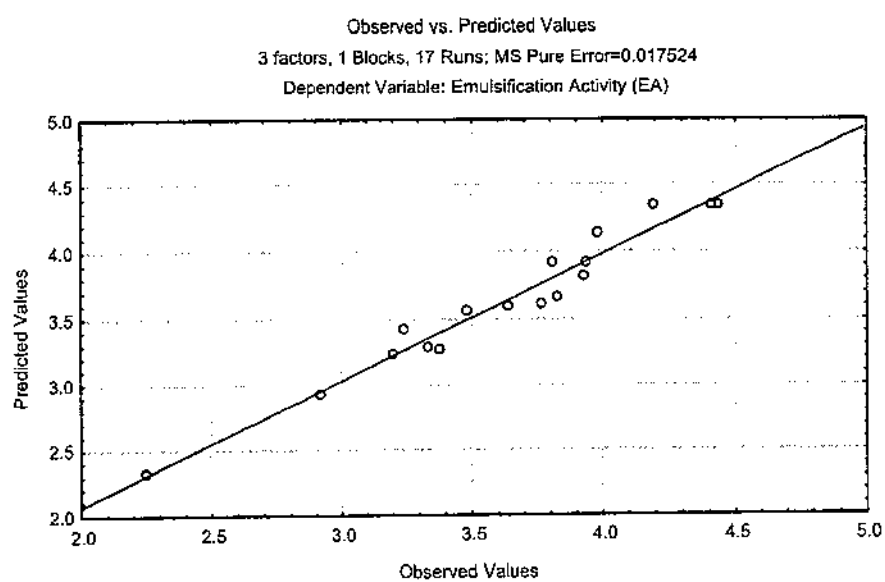


Figure 5.

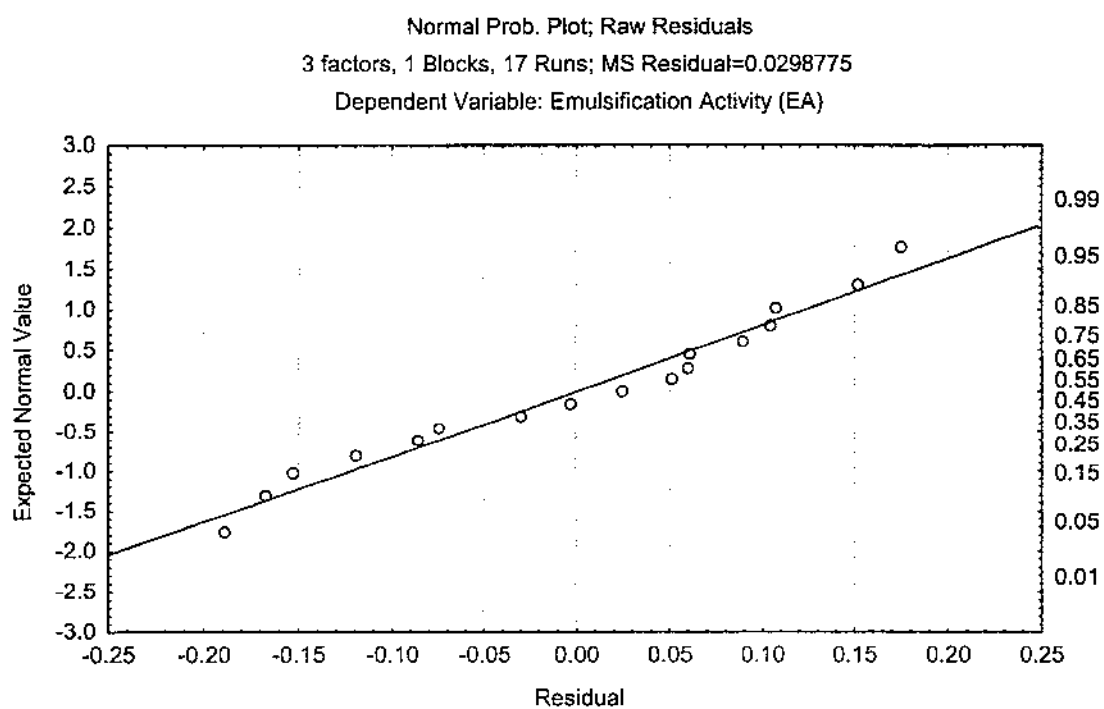


Figure 6.

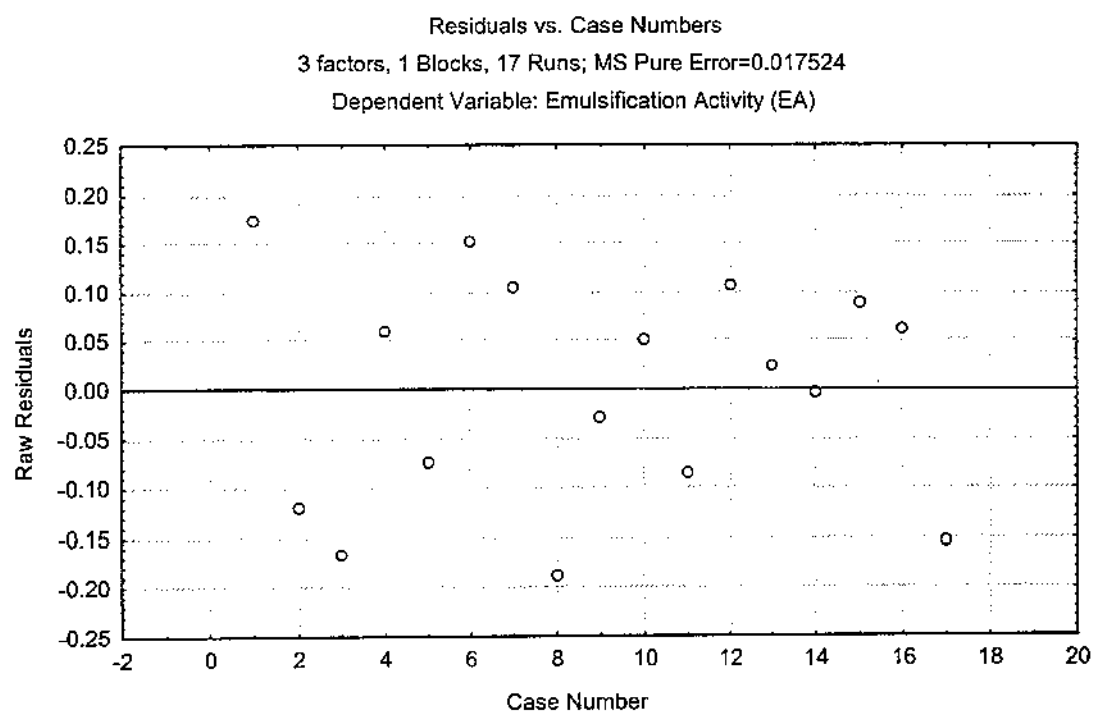


Figure 7.

Captions Figures

Figure 1. Pareto Chart of standardized effects for for the 2^3 full factorial design. The point at which the effect estimates were statically significant (at $p=0.05$) is indicated by the vertical solid line.

Figure 2. Emulsification activity (EA) response surface showing interaction between urea and ammonium sulfate concentrations.

Figure 3. Emulsification activity (EA) response surface showing interaction between and potassium dihydrogen orthophosphate concentrations.

Figure 4. Emulsification activity (EA) response surface showing interaction between ammonium sulfate and potassium dihydrogen orthophosphate concentrations.

Figure 5. Predicted versus observed values plot for emulsification activity (EA).

Figure 6. Normal probability plot for the residual from the emulsification activity (EA) model.

Figure 7. Residual versus case numbers plot for emulsification activity (EA).

ANEXO F - Artigo “Biosurfactant Production by *Candida lipolytica* - Study of the Culture Medium Components using Factorial Design” publicado no periódico Revista Symposium, ISSN 0039-7695, v.1, p.109-115, 2005.

Biosurfactant production by *Candida lipolytica* – study of the culture medium components using factorial design

Albuquerque, C.D.C.¹ ; Vance-Harrop, M.H.²; Sarubbo, L.A.³, Fileti, A.M.F.⁴; Campos-Takaki, G.M.⁵

^{1,3,5}Catholic University of Pernambuco, Recife, Pernambuco, Brazil;

²Agriculture Ministry, Recife, Pernambuco, Brazil;

⁴Campinas State University, Campinas, São Paulo, Brazil.

Abstract - A 2⁴ full factorial design was carried out to investigate the interactions of the effects and concentrations of the corn oil, urea, ammonium sulphate and potassium dihydrogen orthophosphate on the emulsification activity of the biosurfactant produced by *Candida lipolytica*. The results indicated that the concentrations of potassium dihydrogen orthophosphate and ammonium sulphate, in their highest levels, and the corn oil concentration, in the lowest level, were significant and favored the increase of emulsification activity. The research also showed that the interaction of the urea with the phosphate had positive significant effect on emulsification activity increase. The interaction between urea and ammonium sulphate and the interaction between corn oil and ammonium sulphate showed negative significant effect, observed through emulsification activity reduction. The best emulsification activity value (3.727 UEA) was obtained with medium composed by 0.4 g of urea, 1.1 g of sulfate of ammonia, 2.04 g of potassium phosphate, 5 mL of corn oil, 50 mL of distilled water and 50 mL of sea water. **Key-words:** Emulsification Activity, Corn Oil, Surfactant.

PRODUÇÃO DE BIOSURFACTANTE POR *Candida lipolytica* – ESTUDO DE COMPONENTES DE MEIO DE CULTURA USANDO PLANEJAMENTO EXPERIMENTAL

Resumo – Um planejamento fatorial completo 2⁴ foi realizado para investigar os efeitos e interações das concentrações de óleo de milho, uréia, sulfato de amônio e fosfato monobásico de potássio sobre a atividade de emulsificação do biosurfactante produzido por *Candida lipolytica*, com. Os resultados indicaram que as concentrações de fosfato monobásico de potássio e de sulfato de amônio, em seus níveis mais altos, e a concentração de óleo de milho em seu nível mais baixo, foram significativas favorecendo ao aumento da atividade de emulsificação. O estudo também mostrou que a interação da uréia com o fosfato, exerceu um efeito positivo significativo sobre o aumento da atividade de emulsificação. As interações da uréia com a amônia e do óleo de milho com a amônia apresentaram efeitos negativos significativos, observados através da redução da atividade de emulsificação. A melhor atividade de emulsificação (3,727 UAE) foi obtida com um meio composto por 0,4 g de uréia, 1,1 g de sulfato de amônio, 2,04 g de fosfato de potássio, 5 mL de óleo de milho, 50 mL de água destilada e 50 mL de água do mar. **Palavras-chaves:** Atividade de Emulsificação, Óleo de Milho, Surfactante.

Introduction

The interest in industrial biotechnology and its importance opens up challenging possibilities of research in this area. Biosurfactant are valuable microbial amphiphilic molecules with effective surface-activity and biological properties applicable to several industries and processes. Microbes synthesize them, especially during growth on water-immiscible substrates, providing an alternative to chemically prepared conventional surfactants. Because of their structural diversity, low toxicity, and biodegradability, these molecules could be widely used in cosmetic, pharmaceutical, and food process as emulsifiers, humectants, preservatives and detergents. Moreover, they are ecologically safe and can be applied in bioremediation and waste treatments. They can be produced from various substrates, mainly renewable resources such as vegetable oils, distillery and dairy wastes, which are economical but have not been reported in detail (CAMEOTRA, S.S and MAKKAR, R.S., 1998; MAKKAR and CAMEOTRA, S.S,2002)

The yeast *Candida lipolytica* has been used with success in the biopolymers production with emulsification activity, using culture medium of low cost, containing typical vegetable oils of the North and Northeast area of Brazil and sea water (MARÇAL,1991; ROCHA,1996; SARUBBO et al.,1999; VANCE-HARROP et al.,1999; SARUBBO et al.,2001). To reduce the production costs of the biosurfactant in relation to the synthetic surfactants, improvements of culture and production medium are carried out usually through 'one-factor-at-a time' techniques or through factorial design. The 'one-factor-at-a time' techniques are time consuming, they require a large number of experiments to be carried out and they don't consider the effects of interactions between the variables of the process. In the present study, to overcome this difficulty, a 2⁴ full factorial design were carried out to investigate the effects and interactions of the components and concentrations of the corn oil and urea, ammonium sulphate and potassium dihydrogen orthophosphate, on the emulsification activity of the biosurfactant produced by *Candida lipolytica*.

Material and Methods

Yeast culture

The yeast *Candida lipolytica* IA1050 was obtained from the Culture Collection bank of the Núcleo de Pesquisas em Ciências Ambientais, Catholic University of Pernambuco. Yeast colonies were maintained at 4°C, on yeast mold agar (YMA), slants containing 3 g/L yeast extract, 3 g/L malt extract (), 10 g/L D-glucose, 5 g/L tryptone, and 15 g/L agar. The pH was adjusted to 5.0 with HCl. Transfers using aseptic techniques, were made monthly to fresh agar plates to insure cell viability.

Inoculum preparation

The inoculum for biosurfactant production was prepared by aseptically transferring several loop-fulls from a fresh culture of *Candida lipolytica*, grown on agar plates at 28° C for 48 h, into a 250 mL Erlenmeyer flask, containing 50 mL of yeast malt broth (YMB) medium (identical composition of the YMA medium, without agar). The YMB suspension was incubated for 48 hours at 28° C on a rotatory shaker running at 150 rpm.

Shake flasks experiments and standard factorial design

The composition of the production medium varied according to 2⁴ factorial design. A set of 20 experiments, including 4 central points, was performed. The range and levels of the components (factors or independent variables) under study are given in Table 1. The components of each medium were dissolved in 50 mL of sea water and 50 mL of distilled water and the pH was adjusted to 5.3 with HCl. One milliliter aliquots of YMB suspension, containing 10⁷ cells/mL of *Candida lipolytica* IA1055, were used to inoculate each of twenty 250 mL Erlenmeyer flasks, containing 100 mL of sterile production medium. The flasks were incubated on a rotatory shaker at 150 rpm, for 127 hours at 28° C.

Media sterilization

All media used throughout this study were sterilized by autoclaving at 121°C for 20 minutes.

Emulsification activity assay

The emulsification activities of the cell-free filtrate of the 127 h cultures were evaluated according to the method described by CIRIGLIANO and CARMAN (1), that it defines one unit of emulsification activity (UEA) as that amount of emulsifier that effected an emulsion with an absorbance at 540 nm of 1.0.

Table 1 - Levels of the components of the media of the full factorial design 2⁴

INDEPENDENT VARIABLE	LEVEL		
	-1	0	1
Corn Oil (% w/v)	5,0	7,5	10,0
(NH ₂) ₂ CO (% w/w)	0,1	0,25	0,4
(NH ₄) ₂ SO ₄ (% w/w)	0,1	0,6	1,1
KH ₂ PO ₄ (% w/w)	0,68	1,36	2,04

3. Results and Discussion

Our results showed the ability of *Candida lipolytica* to grow on corn oil and produce an inducible extracellular emulsification activity capable of stabilizing oil-in-water emulsion, under various media in batch cultures.

Data presented in Table 2 indicate that high emulsification activity values were obtained under all test conditions studied in factorial design, mainly in nitrogen-phosphorus rich medium. These results are in accordance with literature (REISFELD et al., 1972; VANCE HARROP et al., 1999), that relate the absolute dependence of the source of nitrogen and of the phosphate buffer in the biosurfactant production.

The best emulsification activity value (3.727 UEA) was obtained with medium composed by 0.4 g of urea, 1.1 g of sulfate of ammonia, 2.04 g of potassium phosphate

Vance Harrop (2000) reported an emulsification activity of 2,555 UEA for a culture of *Candida lipolytica* grown on medium containing 0.1 g of urea, 0.1 g of sulfate of ammonia, 0.68 g of potassium phosphate, 5 mL of corn oil, 50 mL of distilled water and 50 mL of sea water, that it was used as initial reference (-1 level) for our experiments. The first experiment of the 2⁴ full factorial design (Table 3) using the same conditions, produced an emulsification activity of 2,161 UEA.

The results also showed the ability of *Candida lipolytica* to produce biosurfactants when grown at high salinity conditions (sea water medium with 36% salinity). Sodium salt is a major component of sea water and frequently breaks the emulsion between oil and water in the practical process (JUNG et al., 1995). The stability of our surfactant to salinity suggests that it is a good candidate for use in marine environments and some industries related to emulsion. Vance Harrop et al. (1999) also obtained a similar result in sea water medium with 13% salinity.

Table 2 - Results of the Complete Factorial Design 2⁴.

Medium	Corn Oil	(NH ₂) ₂ CO	(NH ₄) ₂ SO ₄	KH ₂ PO ₄	Units Emulsification Activity
1	-1	-1	-1	-1	2,161
2	1	-1	-1	-1	2,239
3	-1	1	-1	-1	1,897
4	1	1	-1	-1	1,475
5	-1	-1	1	-1	3,691
6	1	-1	1	-1	2,107
7	-1	1	1	-1	2,167
8	1	1	1	-1	2,218
9	-1	-1	-1	1	1,739
10	1	-1	-1	1	2,273
11	-1	1	-1	1	3,182
12	1	1	-1	1	3,053
13	-1	-1	1	1	3,614
14	1	-1	1	1	3,223
15	-1	1	1	1	3,727
16	1	1	1	1	3,607
17	0	0	0	0	1,337
18	0	0	0	0	1,429
19	0	0	0	0	1,341
20	0	0	0	0	1,598

Figure 1 illustrate the Pareto Chart, with 95% confidence level, for estimates of the effects, in absolute value. On the chart, it can be seen that potassium dihydrogen orthophosphate concentration, amonium sulphate concentration and corn oil concentration

had a significant influence on the emulsification activity. However, in the experimental range there is no significant influence of the urea concentration on the emulsification activity.

The analysis also showed that four of the interactions, i.e., urea with potassium phosphate, urea with ammonium sulphate, corn oil with ammonium sulphate and corn oil with potassium phosphate, had significant effect on the emulsification activity. The increases of the potassium phosphate concentration, ammonium sulphate concentration and the interaction of the urea with the phosphate showed positive significant effects on the emulsification activity. However, the increase of the corn oil concentration and the interactions of the urea with the ammonium sulphate and of the corn oil with the ammonium sulphate produced negative significant effects.

Curvature check also was carried out and revealed lack of fit of the linear approximation. Figure 1 shows that curvature effect crosses the 95% confidence level, indicating the proximity of the optimum point and the need of quadratic model and second-order designs that incorporate the effect of the curvature. The results were analysed using the Statistica® software version 6.0 (Statsoft.Inc, USA)

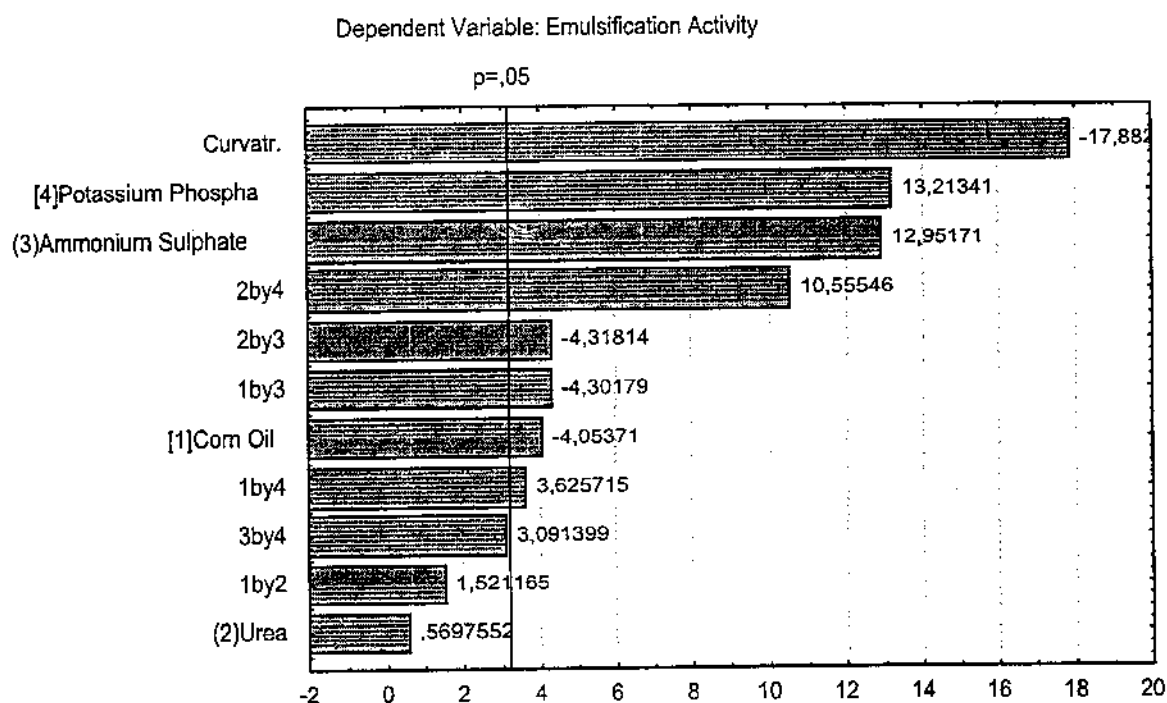


Figure 1 - Pareto Chart of Standardized Effects (Pure error =0.0149511).

New factorial designs in progress will allow the optimization of the process through adjustments of second order polynomials models and neural network based models relating the emulsification activity response variable with the analyzed independent variables. These studies are of interest in order to optimize biosurfactant production by *Candida lipolytica* and therefore reduce production costs.

4. Conclusion

A culture medium for surfactant production by *Candida lipolytica* was studied using 2^4 full factorial design to evaluate the effects and interactions of corn oil concentration, urea concentration, ammonium phosphate concentration and potassium dihydrogen orthophosphate on the emulsification activity response. Based on the experimental results, the following conclusions can be reached:

- the concentrations of potassium dihydrogen orthophosphate and ammonium sulphate, in their highest levels, and the corn oil concentration, in the lowest level, were significant and favored the increase of emulsification activity;
- the interaction of the urea with the phosphate had positive significant effect on emulsification activity increase;
- the interaction between urea and ammonium sulphate and the interaction between corn oil and ammonium sulphate showed negative significant effect, observed through emulsification activity reduction;
- the potassium dihydrogen orthophosphate concentration, ammonium sulphate concentration and corn oil concentration had a significant influence on the emulsification activity increase;
- the best emulsification activity obtained was 3.727 UAE, with medium composed by 0.4 g of urea, 1.1 g of sulfate of ammonia, 2.04 g of potassium phosphate, 5 mL of corn oil, 50 mL of distilled water and 50 mL of sea water;
- the stability of our surfactant to salinity suggests that it is a good candidate for use in marine environments and some industries related to emulsion;

- the need of the process optimization through adjustments of second order polynomials models and/or neural network based models relating the emulsification activity response variable with the analyzed independent variables.

5. Acknowledgements

This work was carried out with financial support by the Catholic University of Pernambuco-UNICAP, FINEP/CT-PETRO, CNPq and CNPq/CT-PETRO.

5. References

1. CAMEOTRA S.S. & MAKAR, R.S.. Synthesis of biosurfactants in extreme conditions. *Applied Microbiology Biotechnology*, v. 50, n. 5, p. 520-529, 1998.
2. CIRIGLIANO, C.M. & CARMAN, G.M.. Isolation of a Bioemulsifier from *Candida lipolytica*. *Applied and Environmental Microbiology*, v. 48, n. 4, p. 747-750, 1984.
3. JUNG,H.K., LEE, J.B., YIM,G.B.,KIM,K.K.. Properties of microbial surfactants S-acid. *J.Biotechnol.Bioeng*, v.19, p.71-77,1995.
4. MAKAR, R.S. & CAMEOTRA S.S.. An update on the use of unconventional substrates for biosurfactants production and their new applications. *Applied Microbiology Biotechnology*, v. 58, n. 4, p. 428-434, 2002.
5. MARÇAL, M.C.R. Produção de Biopolímeros por *Candida lipolytica* em Meios Suplementados por Óleos Vegetais (Babaçu, Coco, Dendê). Recife, 147 p.. Tese (Mestrado em Nutrição) - Centro de Ciências da Saúde, UFPE, 1991.
6. REISFELD, A.; ROSEMBERG, E.& GUTINICK, D.. Microbial degradation of crude oil: factors affecting the dispersion in sea water by mixed and pure cultures. *Applied Microbiology*, v. 24, n. 30, p. 363-368, 1972.
7. ROCHA, J.A., VANCE-HARROP, M.H., CAMPOS-TAKAKI, G.M., GUSMÃO,N.B. Estudos preliminares da produção de bioemulsificantes através de *Candida lipolytica*

utilizando como fonte de carbono óleo de babaçu. In: V Encontro Nacional de Microbiologia Ambiental, 1996, Fortaleza. *Resumo*. 1996.

8. SARUBBO, L.A.; PORTO, A.L.F. & TAKAKI, G.M.C.. The use of babassu oil as substrate to produce bioemulsifiers by *Candida lipolytica*. *Canadian Journal of Microbiology*. v. 45, p. 423-426, 1999.
9. SARUBBO, L.A.; MARÇAL, M.C.; NEVES, M.L.C.; SILVA, M.P.C.; PORTO, A.L.F.; TAKAKI, G.M.C. Bioemulsifier Production in Batch Culture Using Glucose as Carbon Source by *Candida lipolytica*. *Applied Biochemistry and Biotechnology*., v. 95, n. 1, p. 1-9, 2001.
10. VANCE-HARROP, M.H.; SARUBBO, L.A.; CARNEIRO DA CUNHA, M.G.; BUARQUE GUSMÃO, N.; CAMPOS TAKAKI, G.M. Produção de Biossurfactante em meio de cultura de baixo custo suplementado com óleo de milho por *Candida lipolytica*. *Revista Symposium*, v. 2, p. 23-27. 1999.
11. VANCE-HARROP, M.H. Influência das Fontes de Carbono D-Glicose e Óleo de Babaçu no Crescimento de *Candida lipolytica* e na Produção de Biossurfactantes. Recife, 72 p.. Tese (Mestrado em Biologia de Fungos) - Centro de Ciências Biológicas, UFPE, 2000.

Endereço para correspondência: Profª Clarissa Daisy da Costa Albuquerque.
E-mail:cdaisy@unicap.br