

7 Síntese e Conclusões

Nesta tese foram reunidas as técnicas estatísticas necessárias ao planejamento e análise de experimentos com mistura com ou sem variáveis de processo e foi apresentada e detalhada uma metodologia de seleção de modelos em EMP, com uma aplicação real.

Na pesquisa de seleção de modelos foram utilizados os dados do experimento de um misto químico do mecanismo de retardo para ignição de um motor foguete. O misto químico consiste de uma mistura de três componentes. Além das proporções dos componentes da mistura, são consideradas duas variáveis de processo. O objetivo do estudo foi investigar as proporções dos componentes da mistura e os níveis das variáveis de processo que colocam o valor esperado do tempo de retardo (resposta) o mais próximo possível do valor alvo e, ao mesmo tempo, minimizam o tamanho do intervalo de previsão de uma futura resposta.

São muito comuns em experimentos com mistura as restrições nas proporções dos componentes, tornando a região experimental resultante uma sub-região da região original. A análise de dados em regiões muito restritas de mistura pode ser bastante complexa pela própria natureza das variáveis de mistura, podendo ocorrer multicolinearidade entre alguns dos termos do modelo considerado. Visando reduzir o efeito desse tipo de problema, optou-se pelo uso de pseudocomponentes em vez de componentes reais.

Foi apresentada uma grande variedade de modelos combinados para representar os problemas do tipo mistura-processo. Na literatura pesquisada, foi observada a utilização das combinações aditivas e multiplicativas dos diferentes tipos e ordens de modelos de mistura com os modelos para as variáveis de processo. Para o caso do misto químico, as combinações aditivas e multiplicativas de termos de variáveis de mistura e de processo simultaneamente proporcionaram uma redução no *PRESS* e no *MSE*, quando comparados com os resultados obtidos com os modelos aditivo e multiplicativo de Scheffé. Já a variância de uma futura

resposta no ponto do modelo aditivo-multiplicativo de Scheffé foi menor do que a obtida com o modelo aditivo e ficou praticamente no mesmo patamar da obtida com o modelo multiplicativo. Cabe registrar que em todas as referências de EMP pesquisadas não foi observada a utilização de combinações aditivas e multiplicativas de termos de modelo de mistura e de processo simultaneamente.

A utilização de critérios baseados na teoria da informação constituiu uma evolução da pesquisa relacionada com seleção de modelos. Nesta etapa da pesquisa foram também utilizados os dados do EMP do misto químico.

A colinearidade pode tornar instáveis e bastante inflados os estimadores dos coeficientes do modelo. Com isso, certos termos do modelo podem não ser significativos na presença de alguns termos e ser significativos na presença de outros termos. Neste contexto, a seleção *stepwise*, *forward* e *backward* pode ocasionar seleção arbitrária de variáveis que pertencem ao modelo (Harrell, 2001). Uma alternativa foi considerar todas as combinações possíveis de termos do modelo completo e de número de parâmetros e utilizar critérios de seleção de modelos baseados na teoria da informação.

Dentre os diversos critérios baseados na teoria da informação, os critérios AIC_c , AIC_u e HQ_c são os mais indicados de acordo com estudos de simulação de McQuarrie & Tsai (1998). Foi, portanto, elaborada uma rotina em Matlab[®] para calcular o AIC_c , AIC_u e HQ_c para todos os modelos candidatos e selecionar os três modelos que apresentarem respectivamente o menor valor de AIC_c , AIC_u e HQ_c . Este procedimento foi realizado com a utilização, primeiramente, de combinações aditivas e, posteriormente, de combinações multiplicativas de termos de variáveis de mistura e de processo. Ainda nesta etapa da pesquisa, foi analisada a utilização de combinações aditivas e multiplicativas de termos de variáveis de mistura e de processo simultaneamente, obtendo-se resultados melhores do que os obtidos anteriormente. Concluiu-se que os modelos obtidos pelos três critérios de informação não proporcionaram uma melhoria nas estatísticas $PRESS$ e MSE , nem na variância de uma futura resposta no ponto, quando comparados com os resultados obtidos pelo modelo aditivo-multiplicativo de Scheffé selecionado pelo método *backward*.

Numa etapa seguinte da pesquisa, foi elaborada uma rotina Matlab[®] para o cálculo e armazenamento dos valores de AIC_c e seleção de modelos indiferentes. Diante dos resultados obtidos, concluiu-se que a utilização do critério de

indiferença de AIC_c pode levar a escolha de um modelo com menor $PRESS$, MSE ou variância de uma futura resposta no ponto do que o obtido pelo critério AIC_c mínimo.

Finalmente, foi apresentada e detalhada uma metodologia de seleção de modelos em EMP. Na primeira etapa da metodologia, ajustou-se um Modelo Base utilizando um método tradicional passo-a-passo. Na etapa seguinte, foi obtido um modelo melhor, levando em consideração, além dos termos do modelo base, todos os termos equivalentes aos termos do modelo base. Nesta etapa, foi utilizado o critério de indiferença de AIC_c para a seleção dos termos do modelo. Conclui-se que a implementação da segunda etapa da metodologia resultou na obtenção de modelos melhores que o Modelo Base e também melhores do que os modelos obtidos anteriormente.

Com o modelo desenvolvido foram determinadas as proporções ótimas dos componentes da mistura e os níveis ótimos das variáveis de processo. Cabe registrar que a modificação de processo proposta com relação à granulometria da mistura e a introdução do orifício no mecanismo de retardo resultam na redução da variância de uma nova resposta no ponto de interesse.

Uma oportunidade de melhoria do método de seleção de modelos em EMP proposto consiste na diminuição do esforço computacional para o cálculo e armazenamento dos AIC_c de todos os possíveis modelos. Deve-se, portanto, explorar e prever o comportamento do primeiro termo do AIC_c mínimo, $n \ln(\hat{\sigma}_p^2)$, que depende do modelo considerado, com o intuito de saber previamente o número de parâmetros (p) correspondente ao menor valor de AIC_c dentre todos os modelos possíveis.

Como desdobramento futuro da pesquisa relacionada com seleção de modelos pode-se propor a investigação de uma metodologia com uma estratégia de duas etapas, sendo que na segunda etapa haja a introdução total ou parcial de um critério de seleção baseado em alguma medida de diagnóstico de colinearidade, visando a obtenção de modelos com baixos níveis de colinearidade.

Sejam $\lambda_1, \lambda_2, \dots, \lambda_p$ os autovalores de $\mathbf{X}'\mathbf{X}$ e seja λ_{\min} e λ_{\max} o menor e o maior dos autovalores. Pequenos valores de λ_{\min} ($< 0,001$) indicam a presença de colinearidade forte (Khuri, 2005).

Uma importante medida de diagnóstico de colinearidade é o *condition number*, κ , que é definido por (Khuri, 2005):

$$\kappa = \left[\frac{\lambda_{\max}}{\lambda_{\min}} \right]^{1/2}$$

Grandes valores de κ (> 40) indicam colinearidade severa (Khuri, 2005).

Variance Inflation Factors (VIFs) são os elementos da diagonal da matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Grandes valores de VIFs (> 10) indicam que a colinearidade presente é suficiente para resultar numa estimativa pobre dos parâmetros do modelo. (Khuri, 2005)