



UFSCar
Centro de Ciências Exatas e de Tecnologia
Departamento de Química

Trabalho de Conclusão de Curso

Introdução ao Planejamento de Experimentos (*Design of Experiments, DoE*): Conceitos Básicos e Aplicações

Aluno: Felipe Ferri Hilário, RA 640077

Orientador: Edenir Rodrigues Pereira-Filho



FUNDAÇÃO UNIVERSIDADE FEDERAL DE SÃO CARLOS

DEPARTAMENTO DE QUÍMICA - DQ/CCET

Rod. Washington Luís km 235 - SP-310, s/n - Bairro Monjolinho, São Carlos/SP, CEP 13565-905

Telefone: (16) 33518206 - <http://www.ufscar.br>

DP-TCC-FA nº 10/2021/DQ/CCET

Graduação: Defesa Pública de Trabalho de Conclusão de Curso

Folha Aprovação (GDP-TCC-FA)

FOLHA DE APROVAÇÃO

FELIPE FERRI HILÁRIO

INTRODUÇÃO AO PLANEJAMENTO DE EXPERIMENTOS (DESIGN OF EXPERIMENTS, DOE): CONCEITOS BÁSICOS E APLICAÇÕES

Trabalho de Conclusão de Curso

Universidade Federal de São Carlos – Campus São Carlos

São Carlos, 30 de junho de 2021

ASSINATURAS E CIÊNCIAS

Cargo/Função	Nome Completo
Orientador	Prof. Dr. Edenir Rodrigues Pereira Filho
Membro da Banca 1	M.Sc. José Augusto Garcia
Membro da Banca 2	Dra. Raquel Cardoso Machado



Documento assinado eletronicamente por **Caio Marcio Paranhos da Silva**, Professor(a), em 05/07/2021, às 11:17, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site <https://sei.ufscar.br/autenticacao>, informando o código verificador **0435548** e o código CRC **8C071BEB**.

Referência: Caso responda a este documento, indicar expressamente o Processo nº 23112.012967/2021-13

SEI nº 0435548

Modelo de Documento: Grad: Defesa TCC: Folha Aprovação, versão de 02/Agosto/2019

Lista de abreviações:

DOE – Design of Experiments

LA – Low Acil (baixo acil)

PVP – Polyvinil Pirrolidone (poli vinil pirrolidona)

HA - High Acil (alto acil)

ED – Efficiency Dissolution (eficiência de dissolução)

ANOVA – Análises de variância

SQ – Soma quadrática

SQReg – Soma quadrática da regressão

SQRes – Soma quadrática dos resíduos

SQT – Soma quadrática total

SQEP – Soma quadrática do erro puro

SQFaj – Soma quadrática da falta de ajuste

GL – Graus de liberdade

MQ – Média quadrática

MQReg – Média quadrática da regressão

MQRes – Média quadrática do resíduo

MQT – Média quadrática total

MQEP – Média quadrática do erro puro

MQFaj – Média quadrática da falta de ajuste

IC – Intervalo de confiança

Sumário:

1. Introdução

1.1 Objetivos

1.2 Comparação entre análises univariada e multivariadas

1.3 Aplicações e Vantagens do DoE

2. Construção de um modelo de regressão com um exemplo didático

2.1. Obtenção de um Modelo de Regressão

2.2 Análises de Variância

2.3 Intervalos de Confiança

2.4 Análise do Modelo Obtido

3. Aplicação de design de experimentos em química forense

4. Aplicação de DoE em Planejamento de Misturas

5. Conclusão

1. Introdução.

1.1 Objetivos

Design de experimentos é uma técnica de planejamento experimental e tratamento de dados que tem como objetivo uma visualização adequada de como as variáveis de um experimento interferem na resposta. Através dessa técnica é possível prever qual deve ser a melhor condição experimental para um determinado experimento.

Esse trabalho tem como objetivo introduzir o conceito de desenho experimental, bem como descrever como a técnica é aplicada. Desde a confecção de um desenho experimental até o tratamento de dados e interpretação dos resultados obtidos.

Inicialmente, será discorrido sobre os conceitos teóricos da técnica e os cálculos necessários para um tratamento de dados correto. Por fim, serão mostrados exemplos da literatura. O primeiro exemplo trata de uma aplicação em química forense e, o segundo exemplo trata de uma aplicação em um estudo ambiental de poluição de solos por Chumbo devido ao descarte de material eletrônico.

Ao final da leitura, o leitor deve ser capaz de aplicar os conceitos aqui expostos e interpretar os resultados obtidos. Dessa forma o, o trabalho aqui exposto deve funcionar como um tutorial para interessados no assunto.

1.2 Comparação entre Análises Univariadas e Multivariadas:

Diversos experimentos falham em seu propósito devido à falta de planejamento adequado e, nesses casos mesmo o tratamento estatístico mais rigoroso não é capaz de compensar um planejamento inadequado. Por planejamento deve-se entender a maneira como são avaliadas as variáveis que interferem nos resultados de um determinado experimento [1].

As otimizações feitas através de análise univariada, isto é, analisando uma variável por vez, ignoram um ponto essencial: não se sabe se as variáveis analisadas são de fato independentes. Analisando a Figura 1 fica claro como a abordagem univariada é insuficiente para encontrar a melhor condição para o experimento de forma eficaz. Nota-se que nenhum dos 19 testes propostos na

Figura 1a foram capazes de encontrar a melhor condição de análise. Por outro lado, uma abordagem multivariada (Figura 1b) permite uma visualização melhor do gráfico de contorno, entregando resultados consideráveis com menos experimentos. Se considerarmos um planejamento completo, com a obtenção de um modelo de regressão, é possível prever como a resposta se comporta em todo o domínio experimental, considerando um determinado nível de confiabilidade.

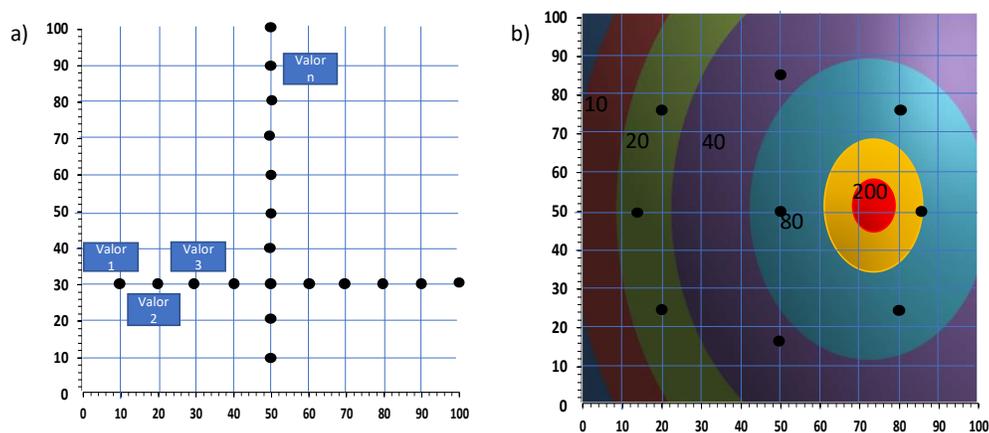


Figura 1: a) abordagem univariada de avaliação de experimentos e b) abordagem multivariada de otimização de experimentos e gráfico de contorno.

1.3 Aplicações e Vantagens do DoE

Este trabalho introduz o conceito de planejamento de experimentos (*Design of Experiments*, DoE) e otimização, apresentando propostas para planejamentos experimentais e procedimentos, visando obter a maior quantidade possível de informação com o menor número de experimentos.

Antes de elaborar um DoE, é crucial que se tenha em mente o propósito exato do experimento a ser avaliado. Isso pode parecer óbvio, entretanto, estabelecer uma resposta adequada para o experimento estudado pode ser um desafio. Levando em conta que diversos fatores podem afetar o resultado de um experimento e que o mesmo experimento pode ser avaliado de diversas formas, fica evidente a importância de um planejamento adequado. Este

trabalho apresenta diversas propostas e exemplos práticos de como preparar um DoE e a melhor maneira de avaliar os resultados obtidos [1].

Leardi [2] considerou que para a preparação de um DoE, cinco etapas devem ser seguidas:

(1) Definir o objetivo do experimento. Isso pode parecer totalmente óbvio, entretanto muitos experimentalistas iniciam suas atividades sem saber ao certo qual o resultado esperar ou mesmo qual o resultado desejado. Isso é uma consequência da ideia de que depois de obter resultados, sejam eles quais forem, você será capaz de tirar algum proveito deles.

(2) Detectar todos os fatores que podem influenciar nos resultados. Uma atenção particular deve ser prestada as palavras “podem Influenciar”. Isso significa que não é correto considerar um número pré-definido de fatores levando em conta que todos com certeza vão influenciar a resposta ou respostas. Em muitos casos não se sabe se o fator de fato influencia significativamente a resposta avaliada, então um dos objetivos do planejamento deve ser esclarecer se esse fator de fato possui influência sobre os resultados.

(3) Planejar, enfim, a série de experimentos a ser realizada. Uma vez que as variáveis que podem influenciar no experimento foram definidas e o domínio experimental estabelecido, bem como as respostas que se deseja avaliar e o objetivo dos experimentos, planejar uma série de experimentos requer apenas alguns minutos.

(4) Realizar os experimentos. No pensamento tradicional essa é a etapa mais importante, entretanto na perspectiva do DoE essa é apenas uma etapa que não pode ser evitada para a obtenção dos resultados que serão utilizados para obtenção de um modelo de regressão.

(5) Analisar as informações obtidas com os experimentos. Essa etapa transforma dados em informação e leva a uma conclusão lógica, deve ser entendida como o ponto chave do processo.

Em muitos casos, apenas um planejamento não é capaz de prover uma solução completa para o problema avaliado. Assim, a informação obtida na etapa 5 é utilizada para reformular o problema: remover os fatores que não influenciam na resposta avaliada, redefinir o domínio experimental e modificar o modelo formulado.

Tendo em mente que a possibilidade de ter que formular um segundo planejamento está sempre presente, é prudente não investir todos os recursos disponíveis na primeira parcela de experimentos.

2. Construção de um Modelo de Regressão com um Exemplo Didático

Ferraz et. al [3] fez uso de um planejamento fatorial para otimizar a formulação de pílulas contendo teofilina. Os fatores analisados foram: composição do agente ligante (“*binding agent*”) e composição do diluente. Os autores analisaram 3 diferentes agentes ligantes e 3 diluentes. A resposta analisada foi a eficiência de dissolução (ED), onde o desejado era obter um alto valor. Os detalhes técnicos acerca da formulação e confecção das pílulas não serão abordados nesse trabalho, bem como os cálculos feitos para obtenção da eficiência de dissolução e as técnicas de análise utilizadas. O planejamento realizado pelos autores é exposto na Tabela 1, nota-se que o planejamento realizado foi qualitativo, isto é, foram testados 3 diferentes tipos de ligantes: LA (goma de baixo acil), PVP (poli vinil pirrolidona) e HA (goma de alto acil), e 3 diferentes solventes: Starch 1500, Mannitol e Maltodextrin. O número de experimentos necessários em um planejamento fatorial pode ser obtido pela equação:

$$E = N^v \quad (1)$$

Na Equação 1 acima, E é o número de experimentos, N a quantidade de níveis em que cada variável é avaliada e v o número de variáveis. No exemplo em questão, tem-se um planejamento fatorial 3^2 , duas variáveis analisadas em 3 níveis, dessa forma o total de experimentos realizados foi 9.

Tabela 1: DoE qualitativo preparado pelos autores do exemplo 1.

Formulação	Ligante		Diluyente		ED
	Real	Codificado	Real	Codificado	
1	LA	-1	Starch 1500	-1	0,5923
2	LA	-1	Mannitol	0	0,7743
3	LA	-1	Maltodextrin	1	0,6418
4	PVP	0	Starch 1500	-1	0,8367
5	PVP	0	Mannitol	0	0,8302
6	PVP	0	Maltodextrin	1	0,8237
7	HA	1	Starch 1500	-1	0,8275
8	HA	1	Mannitol	0	0,7557
9	HA	1	Maltodextrin	1	0,7052

2.1 Obtenção de um Modelo de Regressão

Pereira Filho et. al [4] descreveram com clareza o procedimento necessário para a obtenção de um modelo de regressão e o mesmo procedimento será utilizado neste trabalho. Em paralelo, diversos programas computacionais são dedicados ao tratamento estatístico de dados para essa finalidade, como Minitab e Statistics. Normalmente, o principal objetivo quando se trata de um DoE é a obtenção de um modelo de regressão. Por modelo de regressão, deve-se entender uma equação matemática capaz de relacionar as variáveis com a resposta avaliada. As equações abaixo exemplificam alguns modelos que podem ser obtidos no exemplo em questão.

$$y = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2 \quad (2)$$

$$y = b_0 + b_1X_1 + b_2X_2 + b_{11}X_1^2 + b_{22}X_2^2 + b_{12}X_1X_2 \quad (3)$$

Na Equação 2, observa-se um modelo de regressão linear, y corresponde a resposta, X_1 a primeira variável e X_2 a segunda variável, no exemplo em questão essas variáveis são ligantes e diluentes. Observa-se que cada variável está acompanhada de um coeficiente ($b_1, b_{11}, b_{12}...$), estes são ditos coeficientes de regressão, b_0 corresponde a uma constante. Além das variáveis X_1 e X_2 , nota-se também a interação entre as duas variáveis, X_1X_2 , a influência de cada termo da Equação 2 sobre a resposta avaliada depende do valor do coeficiente que acompanha esse termo.

Aqui será feito o cálculo de um modelo quadrático (Equação 3). O primeiro processo de um planejamento fatorial é a codificação das variáveis e a obtenção de uma matriz \mathbf{X} . A matriz \mathbf{X} dá origem aos coeficientes de regressão a serem calculados e apresenta n linhas e p colunas, as linhas correspondem a cada experimento realizado e as colunas aos coeficientes de regressão a serem calculados.

A coluna referente a constante deve ter valor 1 em todas as linhas e as colunas referentes aos dois primeiros coeficientes, b_1 e b_2 , são obtidos a partir dos níveis codificados das variáveis X_1 e X_2 mostrados na Tabela 1, o coeficiente b_{12} é obtido pela multiplicação da coluna correspondente ao coeficiente b_1 pela coluna correspondente ao coeficiente b_2 , o coeficiente b_{11} e b_{22} são obtidos elevando ao quadrado cada linha das colunas referentes aos coeficientes b_1 e b_2 . A Tabela 2 mostra em detalhes a matriz \mathbf{X} e o resultado obtido em cada experimento no exemplo estudado. O valor dos coeficientes pode ser obtido através da seguinte equação:

$$b = (X^t X)^{-1} X^t y \quad (4)$$

Na equação (4), X corresponde a matriz X , X^t a matriz X transposta, y as respostas obtidas em cada experimento e b a um vetor com os coeficientes. No vetor b cada linha corresponde a um coeficiente de regressão e estes estão organizados na mesma ordem em que foram colocados na matriz X . A Tabela 3 mostra a matriz correspondente a $(X^tX)^{-1}$, a diagonal principal dessa matriz traz a variância relativa a cada coeficiente calculado. O modelo calculado por meio da Equação 4 para o exemplo em questão é igual a:

$$b = 0,8628 + 0,0467X_1 - 0,0143X_2 - 0,1141X_{11} - 0,0489X_{22} - 0,0429X_{12} \quad (5)$$

Neste momento é importante verificar se todos os coeficientes obtidos são significativos ou não dentro de um determinado nível de significância. Para tanto é conveniente comparar os valores previstos com os calculados. A Equação 6 mostra como os resultados previstos podem ser calculados:

$$\hat{y} = Xb \quad (6)$$

em que \hat{y} é um vetor 9:1 contendo os valores previstos para cada experimento realizado, X é a matriz 9:6 exposta na Tabela 2 e b é o vetor 6:1 calculado através da Equação 4. A Tabela 4 mostra os resultados obtidos na prática e os resultados previstos pelo modelo em cada experimento. O ideal é que esses resultados sejam mais similares possível.

Tabela 2: Matriz X e matriz y

Matriz X						Resultados (matriz y)
b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	
1	-1	-1	1	1	1	0,5923
1	-1	0	1	0	0	0,7743
1	-1	1	1	1	-1	0,6418
1	0	-1	0	1	0	0,8367
1	0	0	0	0	0	0,8302
1	0	1	0	1	0	0,8237
1	1	-1	1	1	-1	0,8275
1	1	0	1	0	0	0,7557
1	1	1	1	1	1	0,7052

Tabela 3: Matriz obtida pela equação $(X^tX)^{-1}$ com as variâncias obtidas em destaque

Matrix resultante da operação $(X^tX)^{-1}$					
0,556	0,000	0,000	-0,333	-0,333	0,000
0,000	0,167	0,000	0,000	0,000	0,000
0,000	0,000	0,167	0,000	0,000	0,000
-0,333	0,000	0,000	0,500	0,000	0,000
-0,333	0,000	0,000	0,000	0,500	0,000
0,000	0,000	0,000	0,000	0,000	0,250

Tabela 4: Resultados previstos e resultados obtidos experimentalmente.

Experimento	Resultados	
	Experimentais	Previstos
1	0,592	0,625
2	0,774	0,702
3	0,642	0,682
4	0,837	0,828
5	0,830	0,863
6	0,824	0,800
7	0,828	0,804
8	0,756	0,795
9	0,705	0,689

2.2 Tabela ANOVA

Para validar o modelo é necessário realizar as análises de variância (tabela ANOVA). Nesta tabela são relacionados diversos parâmetros do modelo proposto. O primeiro a ser verificado é a soma quadrática (SQ) devida aos coeficientes de regressão (SQReg). Neste caso, é calculado o

somatório do quadrado da diferença entre os valores previstos e a média dos valores obtidos experimentalmente. No exemplo em questão a SQReg calculada foi 0,0525 e possui um número de graus de liberdade igual a quantidade de coeficientes calculados menos 1 ($p-1$). A situação desejada é que a SQReg seja mais alta o possível.

O próximo valor a ser calculado é a soma quadrática referente aos resíduos (SQRes). Esse valor estima a quantidade de informação que o modelo calculado não foi capaz de modelar. A SQRes é a somatória do quadrado da diferença entre os valores experimentais e previstos. No exemplo em questão o valor obtido para a SQRes foi 0,0120. O número de graus de liberdade da SQRes deve ser igual ao número de experimentos menos o número de coeficientes calculados ($n-p$), no caso em questão $9-6=3$. A soma quadrática total (SQT) é a soma das duas somas quadráticas anteriores e, de forma análoga, o número de graus de liberdade da SQT é igual à soma dos graus de liberdade da SQReg e SQRes ($3+5=8$).

Quando o experimento apresenta réplicas é possível calcular a soma quadrática do erro puro (SQEP). Esse valor é calculado pela somatória do quadrado da diferença entre cada réplica e a média das réplicas. A SQEP é importante, pois as réplicas possuem como objetivo fornecer uma ideia de como é o erro intrínseco aos experimentos realizados, ou seja, a magnitude do erro puro ou erro aleatório cometido durante a realização dos experimentos. O número de graus de liberdade da SQEP é igual a somatória do número de réplicas em cada condição experimental menos 1. Assim, da SQRes deve-se subtrair a SQEP, o valor obtido mostra a incapacidade do modelo em se ajustar aos pontos experimentais obtidos e é a soma quadrática referente a falta de ajuste (SQfaj). O número de graus de liberdade da SQfaj é igual ao número de experimentos independentes, m , (réplicas não são experimentos independentes) menos o número de coeficientes calculados.

No exemplo em questão, não foram realizadas réplicas, logo não é possível obter o valor referente ao erro puro e, conseqüentemente, a SQfaj adquire o mesmo valor da SQRes, assim como o número de graus de liberdade referente a esse valor. A Tabela 5 mostra de forma resumida as fórmulas para os cálculos das SQ's e seus respectivos graus de liberdade.

Tabela 5: Fórmulas para a obtenção das somas quadráticas.

Parâmetro	Soma quadrática (SQ)	Graus de Liberdade (GL)	Legenda
Regressão (SQR)	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	p-1	P= número de coeficientes
Resíduo (SQRes)	$\sum_{i=1}^n (y_i - \hat{y})^2$	n-p	n= número de experimentos
Total (SQT)	SQReg + SQRes	n-1	m= número de experimentos independentes
Erro puro	$\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	n-m	\hat{y}_i = resposta prevista
Falta de ajuste	SQRes-SQEP	m-p	\bar{y} = média das respostas experimentais

Para cada soma quadrática está associado um grau de liberdade, dividindo cada soma quadrática pelo seu grau de liberdade associado, obtém-se as médias quadráticas dos parâmetros do modelo calculado. As médias quadráticas (MQ) podem ser utilizadas para verificar a qualidade do modelo obtido. Para isso, é necessário comparar os valores de variância obtidos (MQ) por meio de um teste F. Dividindo a MQReg (média quadrática da regressão) pela MQRes (média quadrática do resíduo) obtemos o valor F da regressão, usado para validação estatística do modelo.

Como é desejável que a MQReg seja maior que a MQRes o valor de F obtido com essa divisão deve ser algumas vezes maior que aquele tabelado para os respectivos graus de liberdade. Um valor de F obtido maior que o valor de F tabelado significa uma maior probabilidade de hipótese nula. Por hipótese nula deve-se entender que as duas variâncias não são comparáveis estatisticamente e, dessa forma, a MQReg difere-se da MQRes.

Outro teste F que deve ser realizado é aquele proveniente da falta de ajuste. Neste caso, o ideal é que a hipótese nula não seja verdadeira e, dessa forma a MQFaj deve ser estatisticamente comparável com a MQEP. O valor F da falta de ajuste é obtido dividindo-se a MQfaj pela MQEP, este deve ser menor que 1 e menor que o valor tabelado de F.

A Tabela 6 mostra as fórmulas para o cálculo das médias quadrática e dos testes F, a Tabela 7 mostra o resultado das análises de variância para o exemplo em questão. Devido à falta de réplicas (experimentos com a mesmas condições experimentais) não foi possível calcular a MQEP.

Tabela 6: Fórmulas para a obtenção das médias quadrática (MQ's)

Parâmetro	Média quadrática (MQ)	Teste F	F tabelado
Regressão (MQReg)	$\frac{SQR}{p-1}$	$\frac{MQR}{MQRes}$	$F_{(p-1),(n-p)}$
Resíduo (MQRes)	$\frac{SQRes}{n-p}$		
Total (MQT)	$\frac{SQT}{n-1}$		
Erro puro (MQEP)	$\frac{SQEP}{n-m}$		
Falta de ajuste (MQfaj)	$\frac{SQfaj}{m-p}$	$\frac{MQfaj}{MQEP}$	$F_{(m-p),(n-m)}$

Tabela 7: Análises de variância obtidas para o exemplo 1.

Análise de Variância					
	Soma quadrática (SQ)	Graus de Liberdade (GL)	Média quadrática	Teste F	F tabelado
Regressão	0,052	5,000	0,010	2,631	9,013
Resíduo	0,012	3,000	0,004		
Total	0,064	8,000	0,008		
Erro Puro	0	0	0		
Falta de Ajuste	0,012	3,000	0,004	-	-

2.3 Intervalos de Confiança

Uma vez realizadas as análises de variância, deve-se obter um intervalo de confiança para os coeficientes calculados. O intervalo de confiança de um determinado coeficiente compreende os valores que esse coeficiente deve assumir, com uma confiabilidade estatística pré-determinada. Os intervalos de confiança dos coeficientes são calculados a partir do valor de t tabelado, considerando uma determinada confiabilidade e um número de graus de liberdade. Para o exemplo ilustrado neste trabalho será utilizado a MQRes. A Equação 7 mostra o cálculo dos intervalos de confiança [4].

$$IC = t_{conf,(GLRes-1)} \sqrt{MQRes (diag(X^t X)^{-1})} \quad (7)$$

Na equação acima, IC corresponde a uma matriz 6:1 onde as linhas são o intervalo de confiança de cada coeficiente, t é o valor de t tabelado para uma determinada confiabilidade e com um número de graus de liberdade igual ao número de graus de liberdade da média quadrática utilizada menos um. A Tabela 8 mostra os coeficientes obtidos e os intervalos de confiança com 95%, 90%, 85% e 80% de confiabilidade para o valor de t.

Tabela 8: Intervalos de confiança.

Coef	Valor	Intervalo de Confiança (IC)			
		95% de Confiabilidade	90% de Confiabilidade	85% de Confiabilidade	80% de Confiabilidade
B0	0,863	0,137	0,089	0,065	0,050
B1	0,047	0,075	0,049	0,036	0,027
B2	-0,014	0,075	0,049	0,036	0,027
B11	-0,114	0,130	0,084	0,062	0,047
B22	-0,049	0,130	0,084	0,062	0,047
B12	-0,043	0,092	0,060	0,044	0,033

É desejável que os coeficientes obtidos tenham valores maiores que o intervalo de confiança associado, quando um coeficiente tem valor menor que o intervalo de confiança associado a ele concluímos que esse coeficiente é insignificante, isto é, não deve ser levado em conta para a análise do modelo obtido.

Para a realização dos cálculos descritos nesta seção, foi preparada uma rotina de cálculos nos programas Matlab e Octave. Os gráficos de contorno e de superfície foram obtidos com o Excel e, os cálculos de derivada parcial para a obtenção do valor máximo das equações obtidas foram realizados também com o Excel.

2.4 Análise do modelo obtido com o exemplo 1 e considerações

A equação abaixo mostra o modelo obtido para o exemplo em questão considerando um intervalo de confiança com 80% de confiabilidade:

$$ED = 0,863_{\pm 0,050} + 0,047_{\pm 0,027}X_1 - 0,114_{\pm 0,047}X_1X_1 - 0,049_{\pm 0,047}X_2X_2 - 0,043_{\pm 0,033}X_1X_2 \quad (8)$$

Na equação (8) ED corresponde a eficiência de dissolução, X_1 corresponde ao ligante e X_2 ao diluente. A equação representa uma superfície, chamada comumente de superfície de resposta. Analisando a superfície obtida é possível prever como as variáveis analisadas interferem na resposta. Os autores do exemplo desejavam obter a maior eficiência de dissolução (ED) possível, os valores de X_1 e X_2 onde a resposta é máxima podem ser obtidos através das derivadas parciais ou analisando o gráfico de superfície obtido.

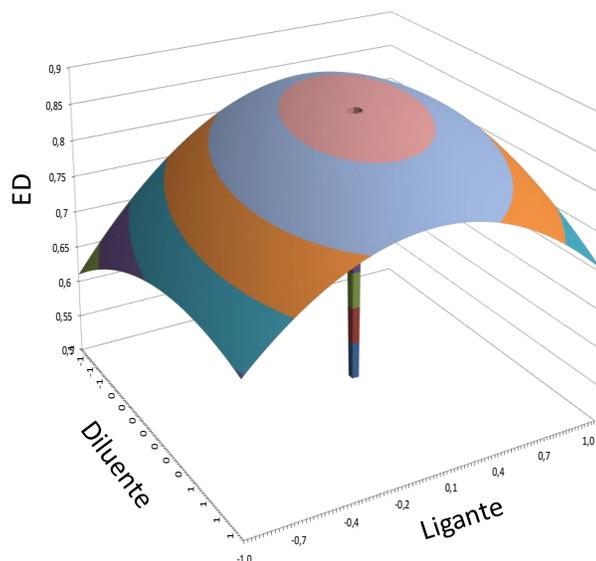


Figura 2: Superfície de resposta obtida para o modelo de regressão do exemplo 1.

A Figura 2a mostra a superfície de resposta obtida. Nota-se que os valores codificados de X_1 e X_2 que apresentam uma maior eficiência de dissolução são 0 e 0, respectivamente. O diluente com valor codificado 0 corresponde ao Mannitol e o ligante com valor codificado igual a 0 corresponde ao PVP. Através da análise do modelo obtido com um planejamento fatorial qualitativo concluiu-se, então, o uso de Mannitol como diluente e de PVP como ligante para a confecção de pílulas apresenta uma eficiência de dissolução por volta de 0,86.

3. Aplicação de Design de experimentos em química forense

F. F. Hilário et al. [5] utilizaram o design de experimentos para otimizar os parâmetros instrumentais da LIBS (Laser Induced Breakdown Spectroscopy). O objetivo do trabalho foi obter diferenciação visual entre dois traços escritos com canetas de mesma cor em uma imagem hiperespectral. Na etapa de otimização instrumental, foram analisadas interseções escritas em forma de cruz, onde cada traço foi feito com uma caneta de cor diferente. O objetivo era obter os parâmetros instrumentais onde o sinal obtido para um traço fosse o mais diferente o possível do sinal obtido em outro traço.

Através da LIBS é possível estabelecer padrões de disparo sobre uma superfície de interesse e, em seguida, o sinal obtido em cada ponto de disparo do laser funciona como um pixel de uma imagem. Dessa forma construir um mapa químico da amostra de interesse.

Para obter as imagens hiperespectrais, os autores fizeram uso de uma técnica de decomposição matricial chamada análise de componentes principais. Essa técnica baseia-se na decomposição de uma matriz em dois vetores, os scores e os loadings. Os pontos de disparo que adquiriam um valor de score negativo eram coloridos de azul e, os pontos de disparo que adquiriam valor de score positivo eram coloridos de amarelo. A Figura 2 mostra de forma explicativa o procedimento adotado pelos autores para a obtenção dos mapas de scores e otimização do resultado estudado.

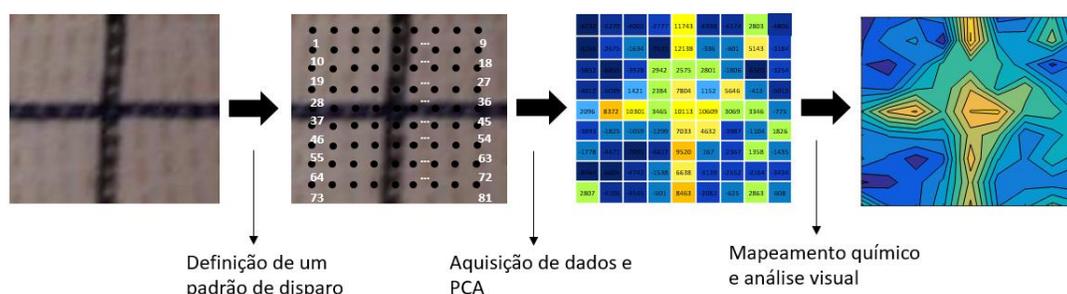


Figura 3: Esquema ilustrativo do trabalho desenvolvido por Hilário F. F. et al.

Inicialmente, foi realizado um planejamento fatorial composto central (planejamento 1), onde as variáveis avaliadas foram *tempo de delay* e *energia do pulso do laser*. Entretanto, os autores não obtiveram respostas significativas para o cálculo de um modelo de regressão, então realizou-se um novo planejamento composto central (planejamento 2). Os autores utilizaram as condições experimentais que apresentaram a melhor resposta no primeiro planejamento como ponto central do segundo planejamento e restringiram a variação das variáveis *tempo de delay* e *energia do pulso do laser*. Além disso, foi adicionada uma terceira variável, *número de pulsos por ponto de disparo*. O planejamento realizado pode ser visto na Tabela 9 e a Figura 4 ilustra de forma clara como planejamento 1 foi aproveitado para a elaboração do planejamento 2.

Tabela 9: Planejamentos elaborados pelos autores no exemplo 2.

Experimentos	Planejamento 1				Planejamento 2				Número de pulsos por ponto de disparo	
	Energia do Laser (mj)		Tempo de Delay (μ s)		Energia do Laser (mj)		Tempo de Delay (μ s)			
	Cod.	Real	Cod.	Real	Cod.	Real	Cod.	Real	Cod.	Real
1	-1	22	-1	0,36	1	92	1	1,78	1	3
2	1	78	-1	0,36	1	92	-1	1,5	1	3
3	-1	22	1	1,64	-1	64	1	1,78	1	3
4	1	78	1	1,64	-1	64	-1	1,5	1	3
5	0	50	0	1	1	92	1	1,78	-1	5
6	0	50	0	1	1	92	-1	1,5	-1	5
7	0	50	0	1	-1	64	1	1,78	-1	5
8	-1,41	10	0	1	-1	64	-1	1,5	-1	5
9	1,41	90	0	1	0	78	0	1,64	0	3
10	0	50	-1,41	0,2	0	78	0	1,64	0	3
11	0	50	1,41	1,9	0	78	0	1,64	0	3
12	-	-	-	-	0	78	0	1,64	-1	1
13	-	-	-	-	-1,41	58	0	1,64	0	3
14	-	-	-	-	1,41	98	0	1,64	0	3
15	-	-	-	-	0	78	-1,41	1,44	0	3
16	-	-	-	-	0	78	1,41	1,84	0	3
17	-	-	-	-	0	78	0	1,64	1	5

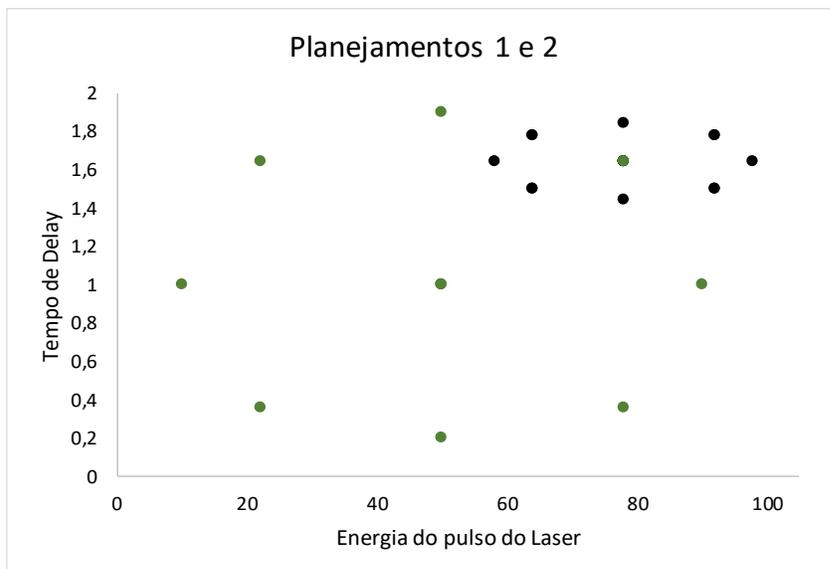


Figura 4: Tempo de Delay vs Energia do Pulso do Laser, com o planejamento 1 em verde e o planejamento 2 em preto. É possível notar como o planejamento 1 foi aproveitado para a confecção do planejamento 2.

As repostas avaliadas no segundo planejamento foram as diferenças entre a média dos sinais obtidos em cada traço e entre os traços e o papel. Dessa forma, foram estudadas 3 repostas ao total: sinal da caneta azul menos sinal da caneta preta, sinal da caneta preta menos sinal do papel e sinal da caneta azul menos o sinal do papel. Assim, o desejável era que a diferença entre as médias fosse o maior possível.

Para avaliar as 3 repostas de forma simultânea, os autores fizeram uso de uma técnica de normalização chamada desejabilidade. As equações 9 e 10 mostram os cálculos necessários para a desejabilidade:

$$d = \left(\frac{(M_y - L)}{(H - L)} \right)^n \quad (9)$$

$$D_g = \sqrt[n]{d_1 d_2 \dots d_n} \quad (10)$$

Na equação 9, d corresponde a desejabilidade individual obtida em cada condição experimental, M_y corresponde ao valor a ser normalizado, L ao menor valor obtido para a resposta

avaliada e H ao maior valor. O n corresponde a um fator de importância, se a resposta apresentar regressão melhor ou for mais importante que as demais esse valor pode ser 2 ou 3, no trabalho em questão todas as respostas apresentaram regressão e importância semelhantes, então esse fator foi 1 para todas as respostas. Dg corresponde à média geométrica das respostas obtidas em uma determinada condição experimental e é chamado desejabilidade global. O intuito da desejabilidade é condensar diversas respostas em apenas um valor, chamado de desejabilidade global (Dg)

A Tabela 10 mostra as respostas obtidas, os valores normalizados de cada resposta e a desejabilidade geral obtida em cada condição experimental do planejamento 2. Essa técnica permite avaliar as 3 respostas de forma simultânea, obtendo a condição experimental que favorece igualmente todas as respostas.

Tabela 10: respostas obtidas em cada experimento realizado no planejamento 2, as respectivas desejabilidades de cada resposta e, por fim, a desejabilidade global.

Experimento	Azul-Preto	Azul-Papel	Preto-Papel	d1	d2	d3	Dg
1	3179	8042	11221	0,55	1,00	1,00	0,82
2	5045	451	5496	0,94	0,00	0,43	0,00
3	3014	474	3488	0,52	0,00	0,23	0,07
4	1677	7042	8719	0,24	0,87	0,75	0,54
5	3480	2826	6306	0,61	0,31	0,51	0,46
6	514	5014	5528	0,00	0,60	0,43	0,00
7	3127	1953	1174	0,54	0,20	0,00	0,00
8	648	3607	2959	0,03	0,42	0,18	0,13
9	5352	5396	10748	1,00	0,65	0,95	0,85
10	4183	6297	10481	0,76	0,77	0,93	0,81
11	4020	4966	8985	0,72	0,59	0,78	0,69
12	1342	4718	6060	0,17	0,56	0,49	0,36
13	2065	7380	5315	0,32	0,91	0,41	0,49
14	1611	6968	5358	0,23	0,86	0,42	0,43
15	1459	3744	5202	0,20	0,43	0,40	0,32
16	2151	5981	8131	0,34	0,73	0,69	0,55
17	4868	2399	7267	0,90	0,26	0,61	0,52

A desejabilidade global obtida em cada condição experimental foi utilizada para os cálculos do modelo de regressão. O procedimento para a obtenção do modelo de regressão e validação dele através das análises de variância foi realizado conforme descrito no exemplo anterior. A equação 11 mostra o modelo de regressão obtido, levando em conta apenas os coeficientes significantes, para a resposta D_g , com o intervalo de confiança de cada coeficiente em subscripto. A Tabela 11 as análises de variância calculadas para o modelo.

$$D_g = 0.75_{(\pm 0.15)} + 0.10x_{3(\pm 0.09)} - 0.13x_{1(\pm 0.12)}^2 - 0.14x_{2(\pm 0.12)}^2 - 0.25x_{3(\pm 0.16)}^2 + 0.23x_1x_2_{(\pm 0.11)} \quad (11)$$

Na equação acima D_g corresponde a desejabilidade global, X_1 a energia do pulso do laser, X_2 ao tempo de delay e X_3 ao número de pulsos. Assim como no exemplo anterior, a equação obtida descreve uma superfície onde o ponto de máximo permite obter o melhor valor de desejabilidade global e, conseqüentemente, a melhor diferenciação entre dois traços em um mapa hiperespectral. A Figura 5 mostra a relação da desejabilidade global com o tempo de delay e energia do pulso do laser.

Tabela 11: Tabela ANOVA

Análises de Variância					
	Soma Quadrática (SQ)	Graus de Liberdade (GL)	Média Quadrática (MQ)	Teste F	F tabelado
Reg.	1,157	5,000	0,231	11,546	3,2
Res.	0,220	11,000	0,020	0,000	
Tot.	1,377	16,000	0,086	0,000	
EP.	0,014	2,000	0,007	0,000	
Faj.	0,207	9,000	0,023	3,354	19,4

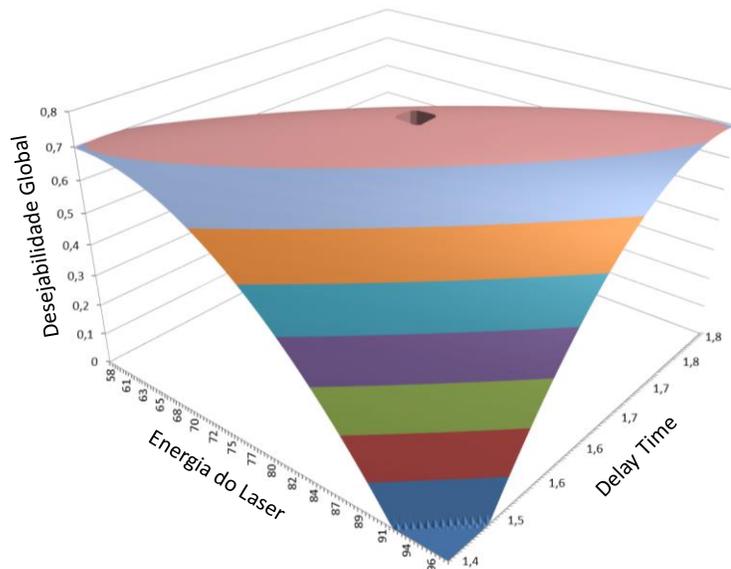


Figura 5: Superfície de resposta obtida para a desejabilidade global.

No exemplo anterior, por ser um planejamento qualitativo com apenas duas variáveis, apenas o gráfico da superfície de resposta foi suficiente para obter a melhor condição experimental. Entretanto, no exemplo atual o planejamento foi realizado de forma quantitativa e três variáveis foram investigadas. A melhor condição experimental, nesse caso, pode ser obtida derivando a equação 11 e igualando o resultado a 0. A função solver do Excel foi utilizada para esse propósito, permitindo obter o valor das variáveis X_1 , X_2 e X_3 onde D_g atinge o valor máximo. A tabela 12 mostra a melhor condição experimental obtida e o valor previsto para D_g .

Dessa forma, através do desenho experimental, foi possível prever a melhor condição experimental para a obtenção de um mapa químico que deve entregar uma diferenciação visual significativa entre a interseção de traços escritos a mão em um papel de cheque. O DoE foi fundamental para a obtenção de um resultado bom e para uma compreensão clara de como as variáveis estudadas interferiam no resultado.

Tabela 12: Desejabilidade global prevista em nível alto, médio e baixo.

Variável	Nível alto		Nível medio		Nível baixo	
	cod	real	cod	real	cod	real
V1	0	78	0	78	0	78
V2	0	1,64	0	1,64	0	1,64
V3	1	5	0,20	3	0,00	3
Dg prev.	1,02		0,76		0,59	

4. Planejamento de Misturas

O planejamento de misturas é bastante empregado em otimizações nas quais a relação entre as variáveis é fixa. Nesse tipo de planejamento, os níveis das variáveis não podem ser alterados de forma independente e sempre é mantida uma relação entre elas. FF Hilário et al. [6] Utilizou exemplos da literatura para apresentar ao público o conceito de planejamento de misturas.

Uma das aplicações apresentadas pelos autores foi o uso de planejamento de misturas para visualizar a relação da composição de solos com a absorção de chumbo presente em resíduos eletrônicos (placas de circuito impresso), realizado por Damasceno et al. [6]. No trabalho em questão, variou-se a proporção entre os 3 principais componentes de amostras de solo: matéria orgânica, óxido de ferro e caulinita. E observou-se como diferentes formulações de solo absorviam o chumbo presente em resíduos eletrônicos, logo, as respostas estudadas foram: concentração de chumbo nas frações orgânica (F2) e residual (F3) do solo.

A Figura 6 e a Tabela 13 mostram o planejamento elaborado por Damasceno et al., assim como as concentrações de chumbo obtidas nas diferentes frações de solo em cada condição experimental.

Tabela 13: Planejamento fatorial realizado pelos autores no exemplo 3 e os resultados obtidos em cada experimento.

Experimento	Matéria Orgânica		Óxido de Ferro		Caulinita		Respostas	
	cod	real(g.)	cod	real(g.)	cod	real(g.)	F2	F3
1	1	80	0	10	0	10	1931	2799
2	0	10	1	80	0	10	2034	2818
3	0	10	0	10	1	80	138	692
4	0,5	45	0,5	45	0	10	1521	2146
5	0	10	0,5	45	0,5	45	1081	1691
6	0,5	45	0	10	0,5	45	1012	1007
7	0,333	33,3	0,333	33,3	0,333	33,3	1058	1309
8	0,667	56,67	0,167	21,66	0,167	21,66	1765	1592
9	0,167	21,66	0,667	56,67	0,167	21,66	1578	1750
10	0,167	21,66	0,167	21,66	0,667	56,67	777	840

Segundo os autores, foram realizadas 3 réplicas em cada nível experimental, entretanto, são mostrados apenas o valor médio e desvio padrão obtido. A obtenção de um modelo de regressão pode ser realizada como descrito no exemplo 1 deste trabalho, os autores do tutorial também descrevem um procedimento para o tratamento de dados com o programa livre Octave e Excel. A equação 13 descreve o modelo de regressão calculado pelos autores:

$$[Pb] = b_1V_1 + b_2V_2 + b_3V_3 + b_{12}V_1V_2 + b_{13}V_1V_3 + b_{23}V_2V_3 + b_{123}V_1V_2V_3 \quad (12)$$

Na equação (12), $[Pb]$ é a concentração de chumbo na fração estudada, V_1 é a quantidade de matéria orgânica no solo, V_2 é a quantidade de óxido de ferro e V_3 é a quantidade de caulinita, b_1 , b_2 , b_3 , b_{12} , b_{13} , b_{23} e b_{123} são os coeficientes de regressão.

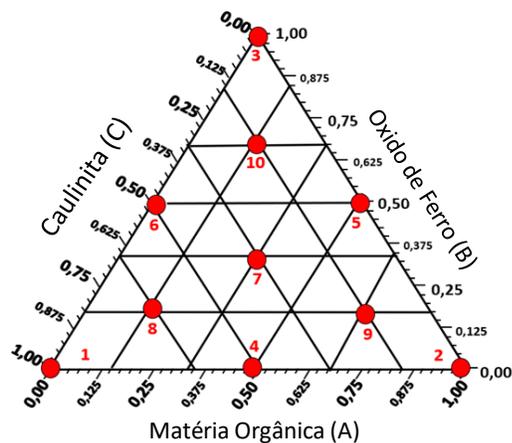


Figura6: Planejamento de mistura preparado pelos autores no exemplo 3

O modelo de regressão obtido para as respostas F2 e F3 são expostos nas equações 13 e 14, respectivamente, levando em conta apenas os coeficientes significantes com 95% de confiabilidade:

$$F2 = 1874_{\pm 328}V_1 + 1903_{\pm 328}V_2 \quad (13)$$

$$F3 = 2804_{\pm 396}V_1 + 2726_{\pm 535}V_2 + 695_{\pm 353}V_3 - 3133_{\pm 1831}V_1V_2 - 3340_{\pm 1831}V_1V_3 \quad (14)$$

As análises de variância para as respostas F2 e F3 podem ser vistas nas tabelas 13 e 14, respectivamente. Nota-se que, devido à falta de réplicas, as somas quadráticas do erro puro não foram calculadas para as respostas estudadas. A ausência da soma quadrática do erro puro não compromete o modelo obtido, tendo em vista que a obtenção dos intervalos de confiança dos coeficientes calculados pode ser obtida através da média quadrática dos resíduos, como de fato foi feito.

Tabela 14: Tabela ANOVA obtida para a resposta F2

Análises de Variância					
	Soma quadrática (SQ)	Graus de Liberdade (GL)	Média quadrática (MQ)	F calculado	F tabelado
Reg.	$3,58 \cdot 10^6$	1	$3,58 \cdot 10^6$	97,87	5,32
Res.	$2,98 \cdot 10^5$	8	$3,73 \cdot 10^4$		
Tot.	$3,88 \cdot 10^6$	9	$4,31 \cdot 10^5$		
EP	0	0	0		
Faj.	$2,98 \cdot 10^5$	8	$3,73 \cdot 10^4$		

Tabela 15: Tabela ANOVA obtida para a resposta F3

Análises de Variância					
	Soma quadrática (SQ)	Graus de Liberdade (GL)	Média quadrática (MQ)	F calculado	F tabelado
Reg.	$4,94 \cdot 10^6$	4	$1,23 \cdot 10^6$	56,03	5,19
Res.	$1,10 \cdot 10^5$	5	$2,20 \cdot 10^4$		
Tot.	$5,05 \cdot 10^6$	9	$5,61 \cdot 10^5$		
EP	0	0	0		
Faj.	$1,10 \cdot 10^5$	5	$2,20 \cdot 10^4$		

De maneira análoga aos exemplos anteriores, as equações obtidas possibilitam a obtenção de um gráfico de superfície, entretanto, a visualização da influência da composição do solo na absorção de chumbo pode ser observada através de um gráfico de contorno. Hilário et al. disponibilizaram um template preparado no Microsoft Excel onde é possível visualizar o gráfico de contorno de modelos de regressão para misturas de 3 componentes. O gráfico de contorno obtido com o template citado para as respostas F2 e F3 são expostos na Figura 7.

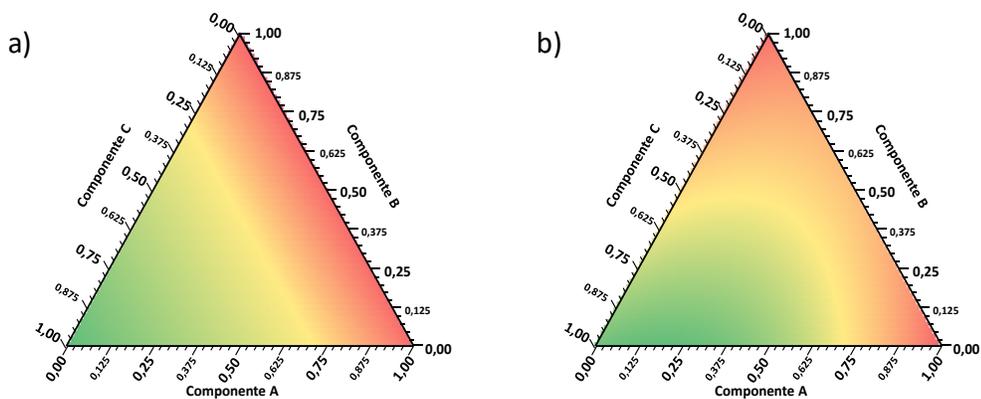


Figura 7: a) Superfície de contorno obtida para a resposta F2, b) Superfície de contorno obtida para a resposta F3

A melhor condição de trabalho foi calculada através do recurso solver do Excel e os resultados obtidos são exibidos na Tabela 16. Observa-se que a concentração de óxidos de ferro possui grande contribuição para a absorção de chumbo na frações orgânica (F2) e residual (F3) do solo.

Tabela 16: Composições de solo que entregam a maior concentração de chumbo em cada fração do solo estudada.

	F2	F3
V(a) %	100	0
V(b) %	0	100
V(c) %	0	0
[Pb] prevista (mg/kg)	2804	1903

4. Conclusão

No trabalho desenvolvido, apresentou-se o conceito de Design of Experiments (DoE). Inicialmente foi exposta uma aplicação em formulações farmacêuticas e a mesma aplicação foi usada para mostrar o tratamento de dados que deve ser realizado após a confecção de um DoE e realização dos experimentos planejados. Mostrou-se os cálculos necessários para obter um modelo de regressão, realizar as análises de variância e obter os intervalos de confiança dos coeficientes calculados.

O segundo exemplo mostra a aplicação de DoE em química forense, onde um planejamento de duas etapas foi utilizado para otimizar os parâmetros instrumentais da técnica LIBS. O modelo de regressão obtido no trabalho foi calculado conforme descrito no primeiro exemplo, bem como as análises de variância e os intervalos de confiança.

Por fim, mostrou-se a aplicação de DoE para a avaliação de como a composição de uma mistura de 3 componentes interfere em uma resposta. O exemplo utilizado estudou como diferentes composições de solo interferiam na absorção de chumbo devido ao descarte de lixo eletrônico.

O conceito de Design de experimentos foi exposto, bem como aplicações práticas presentes na literatura, fornecendo ao leitor ferramentas úteis para a confecção de um DoE e tratamento de dados.

Referências:

1- James Miller, Jane C. Miller. "Statistics and Chemometrics for Analytical Chemistry", ed. Pearson Education Limited, 2010, UK.

2- R. Leardi, "Experimental Design in Chemistry: a Tutorial", *Analytical Chimica Acta* 622 (2019), pag. 161-172. DOI: 10.1016/j.aca.2009.06.015

3- E. J. Barbosa, H. G. Ferraz, "Gellan gum and polyvinylpyrrolidone (PVP) as binding agents in extrusion/spheronization pellet formulations", *Acta Phamr.* 69 (2019), pag. 99-109. DOI:10.2478/acph-2009-0007

4- E. R. Pereira-Filho, "Planejamento Fatorial em Química: Maximizando a Obtenção de Resultados", ed. EDUFSCAR, 2020, São Carlos-SP, Brasil.

5- F.F. Hilário, M. L. de Mello, E. R. Pereira-Filho, "Forensic analysis of hand-written documents using laser-induced breakdown spectroscopy (LIBS) and chemometrics", *Anal. Methods*, 2021,13, 232-241. DOI: 10.1039/D0AY02089C

6- F. F. Hilário, J. P. Castro, T. E. Barros e E. R. Pereira-Filho, "Planejamento de Misturas e Visualização da Região Ótima com Planilhas do Excel: Um Tutorial", *Quim. Nova*, Vol. XY, No. 00, 1-8, 200. DOI: 10.21577/0100-4042.20170750