

5. ESTÁGIO II: PROJETO EXPERIMENTAL

5.1 Adequabilidade do sistema de medição

Nesta tese foram controlados e avaliados os principais fatores que influenciam a exatidão dos resultados de um sistema de medição: o modo de operação, as condições ambientais e a calibração. Nas próximas seções se encontram as informações a respeito do controle e avaliação destes fatores.

5.1.1. Projeto Experimental

Para assegurar a rastreabilidade do sistema de medição, o projeto experimental considerou as três fases Al_2O_3 , MgO e NiO, de reconhecida qualidade metrológica, padrões de referência internacional, que obedeceram as seguintes especificações:

- i) material de referência certificado pela Bundesanstalt für Materialforschung und Prüfung (BAM) com a pureza estabelecida:
 - Al_2O_3 – ZRM2 - (99,76±0,02)%
 - MgO – ZRM-RS 6M – (99,76±0,02)%
 - NiO – ZRM-RS5 - (99,99±0,06)%
- ii) cristalino para minimizar os efeitos de alargamento de picos;
- iii) pó com reduzida capacidade de aglutinação para minimizar os problemas com a micro-absorção;
- iv) reduzida hidroafinidade;
- v) estável durante o período da pesquisa;

Como informação suplementar à fornecida pelo certificado de referência do material foi conduzido o estudo de granulometria de cada fase cujos resultados encontram-se no Anexo B.

A norma QS 9000 *apud* Costa *et al.* (2004) sugere que sejam feitas dez medições por três operadores.

Essa não é uma regra rígida e a decisão do número de amostras foi tomada com base no interesse de avaliar o erro referente à estimação do percentual de massa de cada fase pelo método de Rietveld, w_t , em relação ao *valor verdadeiro*

convencional, w_{nt} , para diferentes conjuntos de níveis, aqui denominados de grupos (Oliveira e Queiroz, 2000).

Foram levados em consideração os custos de análise de cada amostra, a disponibilidade de operadores treinados e condições ambientais que favorecessem as análises.

De acordo com o exposto, a determinação dos níveis de cada fase ou fração foi baseada no artigo de Madens *et al.*(2001) que utiliza um modelo de mistura e considera poucos experimentos (Barros *et al.*, 2001; Nascimento *et al.*, 1999) a partir da variação dos percentuais de massa das fases, denominados de níveis.

O modelo cúbico completo para uma mistura de três componentes é dado pela equação:

$$\hat{w} = b_0 + \sum_{i=1}^3 b_i w_i + \sum_{i \leq j}^3 \sum_{j=1}^3 b_{ij} w_i w_j + \sum_{i \leq j}^3 \sum_{j \leq k}^3 \sum_{k=1}^3 b_{ijk} w_i w_j w_k \quad (5.0)$$

que expressa como os valores da variável de interesse \hat{w} mudam com os valores de três fatores experimentais que estão sendo investigados.

O menor número possível de experimentos com o modelo acima são dez experimentos incluindo as misturas de dois componentes; uma modificação deste, o modelo cúbico especial, fornece um total de sete experimentos. Por alguma razão não mencionada, os autores do artigo mencionados anteriormente obtiveram oito experimentos.

A utilização de poucos experimentos só se justifica na etapa de avaliação de estudo de robustez do método para se verificar se o método é afetado por pequenas variações nas condições de operação (INMETRO, 2000a; Nijhuis *et al.*, 2001).

Esta estratégia se baseia no fato do método, nessa etapa, já estar otimizado, de modo que poucos fatores ou interações poderão ser significativos. Isto é denominado de efeito *esparsidade* e somente em tais casos, um experimento pode ser projetado com um número mínimo de observações.

O que é muito interessante no projeto experimental de Madsen *et al.* (2001) é o fato de eles utilizarem percentuais de massa de diferentes magnitudes.

Aproveitando esta idéia, estabeleceu-se nesta tese a criação grupos de amostras que diferem entre si pela forma de variação dos percentuais de massa entre as fases: alta, média e baixa, além do grupo constituído por amostras de apenas uma fase, mas foram consideradas todas as permutações dos níveis para cada fase formando os grupos A, B, C e D (Quadro 5.1 e Figura 5.1).

O Grupo A foi o grupo com a mais alta variabilidade do percentual de massas entre as amostras.

O Grupo B foi o grupo com média variabilidade e o Grupo C foi o grupo com a mais baixa variabilidade entre as amostras. O Grupo D foi constituído por apenas uma fase e foi criado objetivando subsidiar as análises das características dos picos.

Quadro 5.1. Projeto experimental com as frações de massa do Al_2O_3 , MgO e NiO

GRUPO	Fração de massa $w_i\%$		
	Al_2O_3	MgO	NiO
A	1	4	95
	1	95	4
	4	1	95
	4	95	1
	95	1	4
	95	4	1
B	15	30	55
	15	55	30
	30	15	55
	30	55	15
	55	15	30
	55	30	15
C	32	33	35
	32	35	33
	33	32	35
	33	35	32
	35	32	33
	35	33	32
D	100	0	0
	0	100	0
	0	0	100

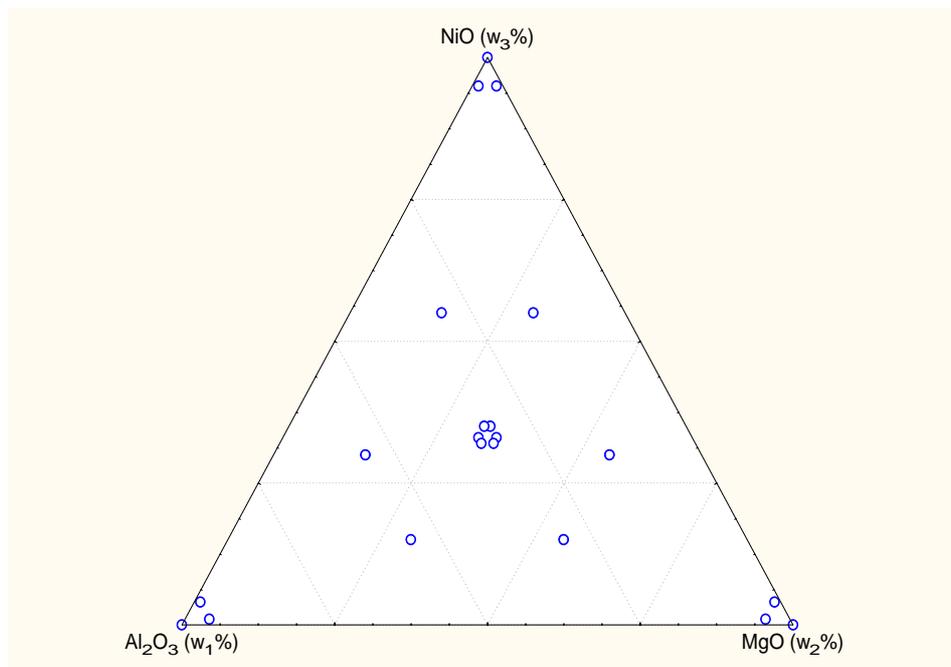


Figura 5.1. Diagrama de composição do padrão ternário.

Para cada padrão ternário estabelecido pelo projeto experimental foram preparadas e lidas três amostras independentes de um grama cada (1g), perfazendo um total de 63 amostras (Tabela B1 no Anexo B)

Das amostras do projeto experimental (Quadro 5.1), quatro experimentos foram sorteados aleatoriamente para o estudo com outra granulometria. Para tal, as fases individuais foram peneiradas em peneira de 400 mesh (0,037mm) certificada pela Rede Brasileira de Calibração e Ensaio, RBC, e posteriormente pesadas e misturadas.

Pela restrição de custo financeiro, apenas cinco amostras foram levadas à microscopia de varredura de elétrons, com temperatura ambiente 19°C e umidade relativa do ar 56%.

As imagens geradas pelo detector de elétrons secundários foram digitalizadas no Microscópio Eletrônico de Varredura, DSM 960 da Zeiss (Figura B4 a B8 no Anexo B).

De acordo com o projeto experimental, cada fração de fase foi obtida pela pesagem individual do material em uma balança analítica Modelo Bel Engineering Mark 210 A (Capacidade: 100mg a 210g; divisão: 0,0001g; tara: 210g; reprodutibilidade: 0,1 mg e linearidade: $\pm 0,3$ mg).

As frações de fases foram misturadas em um béquer com 20 mL de éter de petróleo colocado em um agitador magnético por 30 minutos. Em seguida o béquer foi levado à estufa e lá permaneceu por 30 minutos para evaporação do solvente.

Liu e Kuo (1996) mencionam como principais fontes de incerteza na quantificação de fase a preparação da amostra e o tempo de contagem. Pôde-se observar no Estudo N°2 do Projeto Piloto a alteração do fator de escala do CaCO_3 com diferentes tempos de contagem e tamanhos de passos.

Outro estudo, o Estudo N°3, demonstrou que há indícios que a umidade relativa do ar pode afetar de forma significativa o fator de escala das fases na amostra. Com base nesse estudo e pelas condições de armazenamento da amostra e condições ambientais de medição (Chui *et al.*, 2001; EURACHEM, 1998), depois da estufa, as amostras foram acondicionadas adequadamente em um dessecador para evitar o efeito da umidade e, no mesmo período, nas três primeiras semanas de setembro de 2004, foram feitas as leituras das amostras no difratômetro com as condições ambientais monitoradas.

5.1.2. Leitura

O Projeto Piloto do Estágio I, além de subsidiar a escolha do tempo de contagem, tamanho do passo e grau de polinômio para ajustar a radiação de fundo, subsidiou também a escolha do porta-amostra. No Estudo N°1, foram avaliadas duas formas de preparação da amostra para leitura comuns à rotina do Laboratório de Difração de Raios X do Departamento de Ciência de Materiais e Metalurgia.

Na primeira, a amostra foi espalhada em um porta-amostra raso ($50\mu\text{m}$ de profundidade) com um monocristal de safira, com adição de álcool isopropílico. Na segunda, forma esta adotada nesta tese, o material foi espalhado em um porta-amostra de aço inoxidável de 2 mm de profundidade acordo com o procedimento de Klug e Alexander (1974), com o intuito de minimizar o efeito da orientação preferencial e tentar garantir a passagem dos feixes primários por toda a amostra, mas sem alcançar o porta-amostra.

Para avaliação das incertezas inerentes a esse processo, nove amostras do projeto experimental foram sorteadas aleatoriamente para serem preparadas para leitura por outro operador. A avaliação foi feita utilizando a metodologia descrita na seção 4.3.

5.1.3. Calibração e leitura no difratômetro

As amostras foram lidas no difratômetro D-5000 do Laboratório de Difração de Raios X, do Departamento de Ciência de Materiais e Metalurgia na Pontifícia Universidade Católica do Rio de Janeiro, com as seguintes características:

- Tensão e corrente: 40kV; 30mA
- Geometria Bragg-Brentano;
- Comprimento de onda(λ): $\text{CuK}\alpha$ - monocromador: grafite
- Passo: $0,05^\circ$ (2θ)
- Fenda de espalhamento: $0,6^\circ$
- Fenda antiespalhamento: $0,6^\circ$
- Fenda do detector: 0,2mm
- Tempo: 4s
- Faixa angular: 22 a 90° (2θ)

O equipamento foi calibrado mecanicamente de acordo com as especificações do fabricante e suas configurações para a utilização do método de ajuste por parâmetros fundamentais foram obtidas por calibração utilizando o padrão de referência LaB_6 .

5.2. Avaliação do ajuste do padrão de difração

Foi visto no Capítulo 3 que o modelo estrutural adaptado por Rietveld (The Rietveld Method, Young, R. A., editor, 1995) inclui vários tipos de parâmetros, entre os quais: parâmetros da estrutura cristalina, parâmetros do perfil das reflexões, parâmetros globais, parâmetros de intensidade.

Esses parâmetros permitem calcular, através de um algoritmo, um padrão difratométrico adequado às fases que se pretende estudar, o qual é comparado com o difratograma observado e a diferença entre ambos é minimizada fazendo variar os parâmetros no modelo estabelecido, utilizando um processo de minimização baseado no princípio dos mínimos quadrados.

O programa computacional e os parâmetros de refinamento do método são especificados a seguir.

5.2.1. Programa computacional e parâmetros de refinamento

O programa computacional utilizado nesta tese é o programa computacional TOPAS (Bruker AXS GmbH, Alemanha) e os parâmetros de refinamento utilizados foram:

Parâmetro de radiação de fundo:

1/x

Coeficientes de Tchebycheff de 5º grau

Parâmetro de correção:

Erro do zero (2θ)

Deslocamento da amostra

Absorção (1/cm)

Parâmetro de estruturais de fase:

Parâmetros da rede cristalina do Al_2O_3

Parâmetros da rede cristalina do MgO

Parâmetros da rede cristalina do NiO

Fator de Escala, S_r

Tamanho do cristalito (nm)

Micro-deformação(%)

Coeficiente de absorção linear (1/cm)

Após avaliação visual dos difratogramas entrou no refinamento o parâmetro de orientação preferencial (2.16) na direção (0 0 2) do MgO .

5.2.2. Método de ajuste de pico

Neste trabalho o ajuste dos picos foi feito por parâmetros fundamentais por que permitem definir a forma dos picos e pelo único conhecimento das características instrumentais.

5.2.3. Análise Quantitativa de fase

A fração de massa w_r das três fases foi calculada diretamente do fator de escala S_t (De La Torre *et al.*, 2000; Esteve *et al.*, 2000; Bergese *et al.* 2003) através da equação (2.8):

$$w_t = \frac{S_t(ZMV)_t}{\sum_{j=1}^3 S_j(ZMV)_j} \quad (2.8)$$

Onde:

(ZM) é o conteúdo de massa da célula unitária, obtido como o produto de Z^1 unidades em uma única unidade da célula;

M é a massa da fórmula química.

V é o volume da célula unitária.

O exemplo 5.1 a seguir exibe esta forma de cálculo utilizando a equação (2.9).

Exemplo 5.1 A amostra C62 foi constituída por 32,980% de Al_2O_3 , 35,020% de MgO e 32,000 de NiO, apresentou como parâmetros refinados:

Tabela 5.1. Parâmetros refinados pelo método de Rietveld para a amostra C62.

Parâmetro	Al_2O_3	MgO	NiO
ZM_r	611,768	161,217	224,069
V_r	254,675	74,630	54,616
S_r	$1,739 \times 10^{-4}$	$2,269 \times 10^{-3}$	$1,778 \times 10^{-3}$

O cálculo do percentual de massa do Al_2O_3 , e analogamente para as outras fases, é feito como se segue:

$$w_{Al_2O_3} = \frac{155802,015 \times 1,739 \times 10^{-4}}{158802,015 \times 1,739 \times 10^{-4} + 12031,625 \times 2,269 \times 10^{-3} + 12237,752 \times 1,778 \times 10^{-3}}$$

$$w_{Al_2O_3} = 0,3557 \text{ e } \%w_{Al_2O_3} = 35,57$$

5.2.4.

Critério de ajuste e incertezas

Os indicadores utilizados nesta tese para avaliar a qualidade do ajuste foram R_{WP} e S_{Gof} , definidos na seção 2.26, e para avaliar a qualidade do modelo estrutural refinado, o indicador R_{Bragg} (equação 2.24).

A raiz quadrada dos elementos da diagonal principal da matriz de covariância \mathbf{x} (equação 2.25) é usada como estimativa das incertezas mínimas dos parâmetros correspondentes.

¹ Os valores de Z são 6 moléculas na célula unitária do Al_2O_3 (massa molecular 101,960); 4 moléculas na célula unitária do MgO (massa molecular 40,310) e 3 moléculas na célula unitária de NiO (massa molecular 74,710).

5.3. Repetitividade e reprodutibilidade

As principais medidas de variabilidade desse sistema de medição foram o desvio-padrão associado à *repetitividade* e o desvio-padrão associado à *reprodutibilidade* (ISO 5725-3:1994;EURACHEM, 1998).

O desvio-padrão associado à *repetitividade* foi utilizado para indicar a variabilidade observada entre resultados de três medições de um mesmo padrão ternário, sob condição de repetitividade, para os grupos A, B e C.

A repetitividade desse sistema de medição foi avaliada a partir do modelo de repetitividade e reprodutibilidade, R&R indicado por Costa *et al.* (2004). Este modelo também subsidiou a avaliação da variabilidade entres os grupos.

Para três medições a repetitividade do sistema foi quantificada a partir da estimativa do desvio-padrão, σ_{repe} , pela equação:

$$\hat{\sigma}_{repe} = \frac{\overline{\overline{R}}}{d_2} \quad (5.1)$$

A repetitividade do instrumento foi quantificada por:

$$6\hat{\sigma}_{repe} \quad (5.2)$$

o que corresponde à largura da faixa que contém 99,730% dos resultados sob condições de repetitividade, quando o erro de medição segue uma distribuição normal.

A variabilidade entre os grupos foi quantificada a partir da estimativa de σ_G pela equação:

$$\hat{\sigma}_G = \sqrt{\left(\frac{R_{\overline{S}}}{d_2}\right)^2 - \frac{(\hat{\sigma}_{repe})^2}{nr}} \quad (5.3)$$

com $R_{\overline{S}} = \overline{S}_{max} - \overline{S}_{min}$ e S é o fator de escala.

Para a estimativa de $\hat{\sigma}_{repro}^2$ obteve-se valores negativos dentro do radical. Neste caso a ISO *apud* Kuttatharmmakul *et al.*(1991), considera esta estimativa igual a zero.

A reprodutibilidade do instrumento foi quantificada por:

$$6\hat{\sigma}_{repro} \quad (5.4)$$

A estimativa da capacidade do sistema de medição, quantificada por R&R, foi obtida conforme segue:

$$R \& R = 6\hat{\sigma}_{\text{med}} \quad (5.5)$$

$$\text{com } \hat{\sigma}_{\text{med}} = \sqrt{\hat{\sigma}_{\text{repe}}^2 + \hat{\sigma}_{\text{repro}}^2}$$

o que corresponde à largura da faixa que contém 99,730% dos resultados se o erro de medição seguir uma distribuição normal.

Para avaliar se o sistema de medição foi adequado, comparou-se sua capacidade, expressa pelo índice R&R, com a variabilidade do processo, através da razão:

$$\%R \& R = 100 \frac{R \& R}{6\hat{\sigma}_{\text{total}}} \quad (5.6)$$

$$\text{com a estimativa da variância total } \hat{\sigma}_{\text{total}} = \sqrt{\frac{\sum_{i=1}^g \sum_{j=1}^r \sum_{l=1}^n (S_{ijl} - \bar{S})^2}{gn - 1}}$$

e o número de grupos $g=3$, o número de amostras $r=6$ e o número de replicações $n=3$.

Os valores obtidos para %R&R estão entre 10 e 30% e foram considerados adequados devido aos fatores custo e condições de aplicação do método que foram previstas no projeto experimental.

Esta forma de avaliação também foi utilizada para avaliar a variabilidade do efeito de preparação da amostra por outro operador.

5.4. Estatísticas dos parâmetros de refinamento

5.4.1. Estimador de tendência central

Os parâmetros refinados pelo método de Rietveld apresentaram grande variabilidade entre amostras e não se ajustaram às distribuições teóricas conhecidas, por isso foram utilizados os estimadores M de Tukey para sintetização das informações desses parâmetros.

Os estimadores M são estimadores robustos de tendência central que aplicam uma ponderação aos dados de acordo com as distâncias destes ao centro da

distribuição. Assim, os dados mais próximos do centro da distribuição recebem os maiores pesos. Quando os dados pertencem a uma distribuição simétrica com caudas longas, ou quando os dados têm valores extremos, os estimadores M fornecem melhores estimativas do que as obtidas por outro estimador de tendência central como a mediana ou a média (Hampel *et al.*, 1986).

Os quatro estimadores M existentes são: estimador M de Huber, estimador de onda de Andrews, estimador M de Hampel e estimador bponderado de Tukey. Estes diferem pelos pesos que atribuem aos dados.

O estimador M que foi utilizado nesta tese é o estimador bponderado de Tukey, que atribui peso zero aos dados padronizados mais distantes da constante 4,685 (Norusis, 2002).

5.4.2. Avaliação de dados discrepantes (*outliers*)

A identificação dos dados discrepantes (*outliers*) foi feita a partir do exame dos resíduos padronizados:

$$d_{jk} = \frac{e_{jk}}{\sqrt{QM_r}} \quad (5.7)$$

onde:

e_{jk} é o k-ésimo resíduo da j-ésima concentração

QM_r é o quadrado médio do resíduo

Se os resíduos seguem aproximadamente uma distribuição normal, então, ao serem padronizados, devem se aproximar de uma distribuição normal com média zero e variância um, $d_{jk} \sim N(0,1)$. Dessa forma, aproximadamente 68,27% dos resíduos padronizados devem pertencer ao intervalo de ± 1 ; 95,45% ao intervalo de ± 2 e 99,73% ao intervalo de ± 3 . Assim, resíduos maiores em módulo que três desvios-padrão, são considerados *dados discrepantes* em potencial.

5.5. Correção do efeito de micro-absorção aplicados ao fator de escala

Os valores para a correção do efeito de micro-absorção foram aplicados aos fatores de escala obtidos no refinamento de acordo com a equação (3.4):

$$w_t = \frac{S_t M_t V_t}{\sum_j (S_j M_j V_j / \tau_j) \tau_t} \quad (3.4)$$

5.6. Avaliação do erro relativo referente ao percentual de fase

Nesta tese foi utilizada a definição de erro relativo r_{wt} para avaliar o erro referente à estimação do percentual de massa de cada fase pelo método de Rietveld, w_t , em relação ao valor verdadeiro convencional w_{nt} .

$$r_{wt} = \frac{W_t - W_{nt}}{W_{nt}} \quad (5.8)$$

A propagação de erros considera que:

1. O erro relativo da soma é a soma dos erros relativos de cada parcela, ponderados pela participação da parcela no total da soma.
2. O erro relativo da subtração é a soma do erro relativo do minuendo com o erro relativo do subtraendo, ponderados pela participação de cada um no resultado da subtração.
3. O erro relativo do produto é a soma dos erros relativos dos fatores.
4. O erro relativo da divisão é a soma dos erros relativos do dividendo e do divisor.

5.7. Avaliação do efeito da redução do tamanho da partícula na quantificação de fase

A Tabela 5.2 mostra o esquema da análise de variância que foi conduzida para avaliar a redução do diâmetro da partícula, por peneiramento, a partir do fator de escala e conseqüentemente da quantificação de fase.

Foi considerado o modelo estatístico de efeitos fixos identificado pela equação (5.9):

$$S_{ti} = \mu_t + \tau_{ti} + \beta_{ij} + \varepsilon_{tij} \quad (5.9)$$

onde:

S_{ti} é a observação do fator de escala da fase t , $t=1,2,3$, da amostra sujeita ou não ao peneiramento, $i=1,2$.

μ_t é a média geral das observações do fator de escala da fase t , $t=1,2,3$;

τ_{ti} é o efeito fixo do peneiramento no fator de escala da fase t , $t=1,2,3$ e $i=1,2$;

β_{ij} é o efeito fixo de controle representada pelo grupo $j=1,2,3$ e fase $t=1,2,3$;

ε_{tij} é o efeito aleatório de erros independentes associados às observações do fator de escala, tal que $\varepsilon_{tij} \sim N(0, \sigma^2)$, $t=1,2,3$, $i=1,2$ e $j=1,2,3$.

Tabela 5.2. Esquema da análise de variância para o fator de escala da fase.

Causa de variação	Graus de liberdade	Soma de quadrados (SQ)	Quadrado médio (QM)	F
Grupo	2	SQ_1	$QM_1 = SQ_1/2$	QM_2 / QM_3
Tratamento	1	SQ_2	$QM_2 = SQ_2$	
Resíduo	2	SQ_3	$QM_3 = SQ_3/2$	
Total	5	SQ_{Total}		

De acordo com a ISO e IUPAC, o valor convencional para o nível de significância α deve ser 5% (Currie, 1999). Assim, este será o nível de significância para todas as análises estatísticas desta tese.

Foram utilizados os testes de Anderson-Darling, modificação do teste de Kolmogorov-Smirnov (Shapiro, 1990), para testar a normalidade, e o teste de Bartlett para testar a pressuposição de homogeneidade de variância do modelo (4.9) considerando o nível de significância $\alpha=5\%$ quando o primeiro teste é não significativo (atestando uma distribuição normal) e o teste de Levene caso contrário.

5.8.

Avaliação da variabilidade do fator de escala

Apesar de não constituir o objeto principal do estudo, a análise do conjunto de dados obtidos do planejamento experimental, pelas técnicas multivariadas pode fornecer informações pertinentes a sistemas similares acerca da estrutura do processo de medição.

Os métodos multivariados podem ser aplicados a inúmeras situações práticas cujos objetivos sejam mensurar os efeitos ou relacionamentos entre múltiplas variáveis. Esta abordagem já é muito utilizada em outros sistemas de medição como na Ressonância Magnética Nuclear (Malinowski, 1991).

Na análise de pós por difração de raios X, esses métodos multivariados já estão sendo utilizados na identificação das fases com muitas amostras de padrões de testes,

como nas análises farmacêutica e na industrial (Barr *et al.*, 2004), aplicações estas bem diferentes da que foi utilizada nesta tese.

Considerando-se a inter-relação entre a investigação científica e a análise multivariada, uma vez que esta se aplica naturalmente aos objetivos da primeira, pode-se definir, então, como funções da Análise Multivariada, os mesmos objetivos da investigação científica, tal como propõe Johnson e Wichern (1982):

1. redução de dados com vistas a simplificação estrutural do problema, sem perda de informações (ou minimizando destas perdas);
2. criação de agrupamentos de indivíduos e variáveis;
3. investigação da dependência entre variáveis;
4. predição do comportamento de variáveis em função de outras variáveis; e formulação e teste de hipóteses.

Segundo Hair *et al.* (1998), a análise multivariada pode auxiliar a formulação de questões relativamente complexas de forma específica, possibilitando com isso a condução de pesquisas teoricamente significativas e a avaliação do efeito das variações paramétricas naturalmente encontradas no contexto em que normalmente ocorrem, garantindo a visualização da correlação entre elas, preservando e separando seus efeitos sem causar isolamento de casos ou variáveis.

5.8.1 Análise Fatorial

Dentre as diversas técnicas existentes de análise multivariada, a Análise Fatorial é uma das que permite realizar redução ou sumarização de dados.

A Análise Fatorial, desenvolvida como uma técnica multivariada criada por Hotelling (1933), considera todas as variáveis simultaneamente, assumindo que cada uma delas está relacionada com as demais, a fim de se estudar a inter-relação existente entre as mesmas.

Em termos de pré-requisitos, para aplicação da técnica de Análise Fatorial são necessárias algumas propriedades restritivas, como uma matriz de covariância que tenha uma estrutura específica de covariância e que apresente propriedades inerentes a grandes amostras e estabilidade, que possibilitem a construção de testes de hipóteses e intervalos de confiança.

Ainda existem outros requisitos que devem ser satisfeitos, tais como:

- (a) presença de correlações superiores a 0,30, em módulo, para as variáveis que situam no mesmo fator;
- (b) presença de baixas correlações parciais ou na matriz antiimagem;

(c) teste de Esfericidade de Barlett significativo, o que significa que a matriz de correlações tem correlações significativas no mínimo em algumas das variáveis, e

(d) medida de adequação de dados ao modelo de Kaiser-Meyer-Olkin, KMO, acima de 0,500 para cada variável ou para matriz completa;

Nesta tese foi utilizada a Análise Fatorial para identificar a estrutura de variação dos fatores de escala, proporção de massas e erro relativos a esta proporção. O método utilizado na Análise de Fatores foi o de Componentes Principais, com rotação ortogonal dos eixos via *Varimax*, onde o objetivo é encontrar uma rotação dos fatores que maximize a variância da matriz de pesos, para simplificar a interpretação dos fatores.

Esta análise é baseada em um modelo explícito, em que as variáveis são expressas como uma função linear de um número reduzido de fatores latentes. Na Análise de Componentes Principais deseja-se determinar uma transformação linear das variáveis originais, sendo elas correlacionadas entre si, resultando em novas variáveis chamadas componentes principais, que possuem as seguintes características:

1. São descorrelacionadas;
2. A variação total é igual à variação das variáveis originais;

1. $\text{Var}(c_1) \geq \text{Var}(c_2) \geq \dots \geq \text{Var}(c_p)$, onde $\text{Var}(c_1)$ é a maior possível.

A principal aplicação da Análise de Componentes Principais é a redução de dimensionalidade através da eliminação dos componentes com menor variância. A variância de cada componente principal é igual ao autovalor que lhe corresponde. O primeiro componente é o de maior variância e os outros aparecem em ordem decrescente de variância.

As variáveis da base de dados são descritas pelo modelo:

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon} \quad (5.10)$$

onde:

\mathbf{X} é o vetor das variáveis;

$\boldsymbol{\mu}$ é o vetor do centróide;

\mathbf{L} é matriz de carga dos fatores;

\mathbf{F} é o vetor de fatores;

$\boldsymbol{\varepsilon}$ é o vetor de resíduos do modelo.

Para este modelo assume-se que:

$$E(\mathbf{F}) = \mathbf{0} \quad \text{e} \quad \text{Cov}(\mathbf{F}) = \mathbf{I}$$

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} \quad \text{e} \quad \text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}, \quad \text{onde } \boldsymbol{\Psi} \text{ é uma matriz diagonal.}$$

\mathbf{F} e $\boldsymbol{\varepsilon}$ são independentes.

É freqüente traçar um gráfico chamado de gráfico de declive (*Scree Plot*) que representa a porcentagem da variância explicada pelos componentes encontrados. O seu exame permite que se identifique visualmente o ponto em que os autovalores se tornaram pequenos ou em que a porcentagem da variância explicada é pequena em relação ao ônus de se trabalhar com mais um componente.

A variância de cada variável X_i pode ser dividida em duas parcelas: aquela devida a fatores comuns e aquela devido a fatores individuais, de acordo com o modelo que se segue,

$$\text{Var}(X_i) = s_{ii} = \lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{iq}^2 + \Psi_i \quad (5.11)$$

A variância devido aos fatores comuns é denominada de comunalidade. A comunalidade total é dada por:

$$H = h_1^2 + h_2^2 + \dots + h_p^2 = \sum_{i=1}^p \sum_{j=1}^q \lambda_{ij}^2 \quad (5.12)$$

Assim, a matriz de covariância de \mathbf{X} será a soma da matriz cheia $\mathbf{L}\mathbf{L}^T$ e da matriz diagonal $\boldsymbol{\Psi}$, isto é:

$$\text{Cov}(\mathbf{X}) = \Sigma = \mathbf{L}\mathbf{L}^T + \Psi \quad (5.13)$$

onde:

$\mathbf{L}_{p \times m}$ é a matriz de carga dos fatores, constituída dos coeficientes l_{ij} chamados de carga da i -ésima variável no j -ésimo fator.

$$\Psi_{p \times p} \text{ é a matriz de } \text{Cov}(\epsilon) = E[\epsilon\epsilon^T], \text{ com } \Psi = \begin{bmatrix} \psi_1 & 0 \dots & 0 \\ 0 & \psi_2 \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 \dots & \psi_p \end{bmatrix}$$

Logo, no modelo de Análise Fatorial, os fatores comuns são responsáveis pela estrutura de covariância, enquanto os fatores individuais só influenciam a diagonal da matriz de covariância de \mathbf{X} . Os fatores comuns são responsáveis pela estrutura de dependência entre as variáveis.

Um recurso para melhorar a interpretabilidade \mathbf{L} é realizar uma rotação conjunta dos eixos. Utilizou-se o método de rotação ortogonal *Varimax*, o qual gera uma matriz \mathbf{G} dos fatores que maximizam a variância do quadrado das colunas de $\mathbf{L}^T = \mathbf{G}\mathbf{L}$.

Pode-se provar que as estimativas de máxima verossimilhança das comunalidades são:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots + \hat{l}_{im}^2 \quad \text{para } i=1,2,\dots,p.$$

com a proporção da variância amostral total para o j -ésimo fator dada por:

$$\frac{\hat{l}_{11}^2 + \hat{l}_{12}^2 + \dots + \hat{l}_{im}^2}{s_{11} + s_{22} + \dots + s_{pp}} \quad (5.14)$$

A proporção padronizada da variância amostral total para o j -ésimo fator é dada por:

$$\frac{\hat{l}_{11}^2 + \hat{l}_{12}^2 + \dots + \hat{l}_{im}^2}{p} \quad (5.15)$$

O algoritmo básico do *Varimax* define, para $j=1,\dots,m$ e $i=1,\dots,p$:

$$\tilde{l}_{ij}^* = \frac{\hat{l}_{ij}^*}{\hat{h}_i} \quad (5.16)$$

como os coeficientes finais de rotação escalonados pela raiz quadrada da comunalidade. O procedimento utiliza uma transformação ortogonal que faz com que

$$V = \frac{1}{p} \sum_{j=1}^m \left[\sum_{i=1}^p \tilde{l}_{ij}^{*4} - \left(\sum_{i=1}^p \tilde{l}_{ij}^{*2} \right)^2 / p \right] \quad (5.17)$$

seja a maior possível.

De forma sintética, pode-se dizer que a análise fatorial possui quatro grandes etapas para a sua realização:

1. Cálculo da matriz de correlação (ou de covariância conforme o caso) das variáveis que irão compor os fatores; nesta etapa é verificada a adequação da aplicação da técnica à base de dados em questão;
2. Extração dos fatores e seleção dos mais significativos que representarão os dados; nesta etapa é possível verificar quão bem o modelo representa os dados;
3. Aplicação de rotação, nos fatores mais significativos, para facilitar o entendimento dos mesmos;
4. Armazenamento dos valores das cargas fatoriais, que podem ser utilizados em outras análises.

A análise fatorial é usada nesta tese na avaliação da estrutura do processo de medição, mais especificamente na avaliação da variabilidade do fator de escala (Seção 6.3.2.1)