

2.

METODOLOGIA PROPOSTA E SEUS FUNDAMENTOS TEÓRICOS

2.1. ESPECIFICAÇÃO DA METODOLOGIA

A metodologia de caracterização de recursos minerais, apresentada neste trabalho, decorre em vários passos e combina os métodos factoriais de análise de dados com diferentes técnicas geoestatísticas de estimação (krigagem ordinária, co-krigagem ordinária, krigagem com deriva externa e co-krigagem com deriva externa) e simulação estocástica (Simulação Sequencial Gaussiana).

Este procedimento metodológico, esquematizado na Figura 2.1, tem início com a preparação de toda a informação amostrada para integrar uma única base de trabalho, capaz de assegurar a mesma representatividade e a georreferenciação a todos os suportes de amostragem.

A tentativa de utilizar toda a informação, para maximizar o conhecimento de partida, conduz ao passo seguinte, que consiste na crítica, análise estatística e codificação de toda a informação amostral. No presente caso de estudo, atendendo ao facto da amostragem possuir duas proveniências distintas (sondagens carotadas e em destrutivo), à existência de variáveis

2. Metodologia Proposta e Seus Fundamentos Teóricos

2.1. Especificação da metodologia

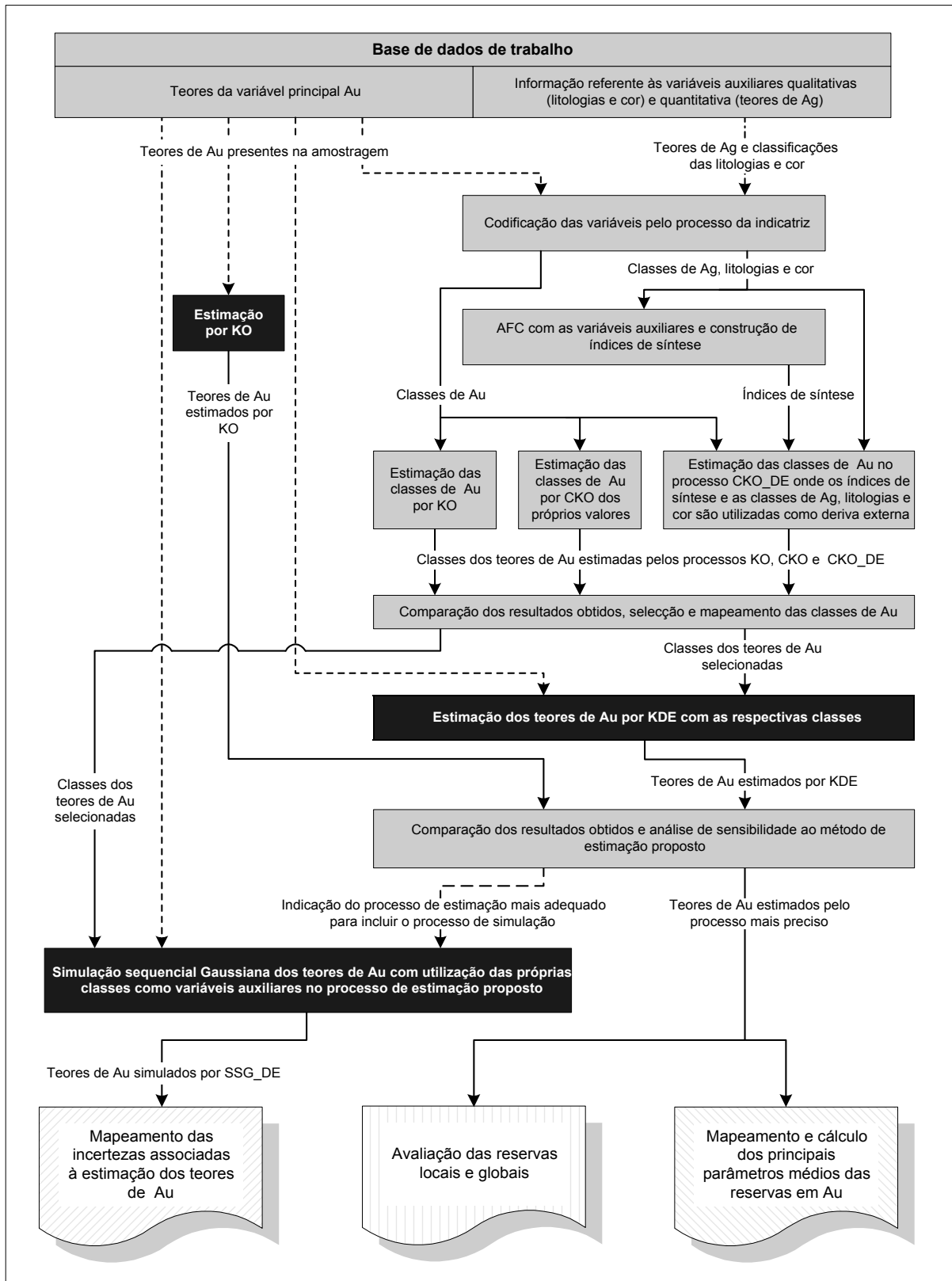


Fig. 2.1 – Fluxograma da metodologia.

apenas observadas em cada uma destas proveniências e à circunstância da baixa recuperação da amostragem (pelo menos no caso das sondagens carotadas) poder contribuir para alguma diferenciação, torna necessárias abordagens particulares para cada um dos conjuntos amostrais e para o conjunto resultante da sua união. Trata-se de uma etapa particularmente importante, que deve ser acompanhada por especialistas, dos domínios aos quais a informação reporta, com o objectivo de tirar desta o melhor partido, em consequência de codificações mais coerentes com a fenomenologia subjacente à amostragem.

Após estes passos preliminares, efectuam-se interpretações pelo método de Análise Factorial das Correspondências (AFC), para verificar as relações existentes entre as variáveis amostradas, e construir índices de síntese das variáveis auxiliares, baseados nas respectivas coordenadas no espaço factorial das suas projecções.

Em termos meramente de estimação espacial de uma variável caracterizada por uma função de distribuição fortemente assimétrica positiva, que normalmente implica o distanciamento da fracção económica da mediana (como é habitual na caracterização de recursos minerais), levanta de imediato a questão: que método utilizar? Como já foi referido, neste trabalho, pretende-se incluir a informação auxiliar na estimação da variável principal. No entanto, a utilização directa das variáveis auxiliares na estimação dos teores de Au, não acrescentou nenhuma precisão, o que levou à decisão de numa primeira etapa utilizar as variáveis auxiliares apenas na estimação de novas variáveis, ditas secundárias, que traduzem as probabilidades dos teores de Au pertencerem a uma determinada classe (foram utilizadas três classes: Au1, Au2 e Au3, que correspondem aos teores: baixos, médios e altos). Para validar as vantagens da utilização das variáveis auxiliares (Ag, litologias, cor e índices de síntese), na estimação das classes de Au, cujo resultado é portanto apenas uma estimação ‘grosseira’ dos respectivos teores, procedeu-se ao cálculo e cotejamento dos resultados obtidos por processos de estimação capazes de utilizar de forma diferenciada a informação auxiliar:

- cokrigagem ordinária (CKO);
- krigagem com deriva externa (KDE);
- CKO com deriva externa (CKO_DE);

com o processo de estimação por krigagem ordinária (KO).

Este cotejamento permitiu numa segunda etapa, estimar a variável principal na sua forma quantitativa, por diferentes processos onde intervieram as classes seleccionadas na primeira.

A metodologia de estimação foi concluída com uma análise de sensibilidade, que permitiu seleccionar o processo de estimação dos teores de Au por KDE com as respectivas classes, como o mais preciso.

Os processos de estimação geoestatística estão claramente vocacionados para caracterizar as distribuições espaciais dos valores mais prováveis de uma variável, podendo verificar-se que a sua utilização na avaliação das fracções extremas da distribuição, conduzem a menores dispersões dos valores estimados, face aos valores experimentais, o que pode provocar desvios significativos de previsão/estimação. Esta inaptidão dos processos de estimação para produzir imagens dos possíveis comportamentos extremos dos domínios em análise conduz à necessidade de recorrer a processos estocásticos de simulação geoestatística (ou mais simplesmente simulação estocástica), para permitir aceder: aos valores extremos em cada ponto; aos comportamentos extremos simultâneos do conjunto de pontos que constitui o domínio; e à incerteza espacial que lhes está associada (só possível através de várias realizações da distribuição conjunta no domínio da variável regionalizada). Assim, para permitir reproduzir mais fielmente o histograma, variância e variabilidade espacial dos dados amostrais e aceder à incerteza do valor de cada ponto (dado pela função de probabilidade), efectuaram-se simulações estocásticas dos teores de Au, pelo processo de simulação sequencial Gaussiana (SSG), com incorporação do processo de estimação seleccionado.

2.2. CONSTRUÇÃO DE ÍNDICES DE SÍNTESE COM BASE NA ANÁLISE FACTORIAL DAS CORRESPONDÊNCIAS

Nas suas aplicações correntes, os métodos ditos factoriais possuem um desígnio puramente descritivo, e permitem realizar estudos globais através de representações gráficas simples que pretendem caracterizar a estrutura da nuvem representativa de N indivíduos no espaço R^P , tradutor do sistema de relações existente entre linhas e colunas do quadro de

partida. Para isso, recorrem à sua projecção num conjunto de rectas ortogonais que passam pelo centro de gravidade da nuvem, onde a primeira é responsável pela maior variação possível da nuvem, a segunda pela maior variação possível restante, e assim por diante, até toda a variação da nuvem ser explicada. Estas rectas, coincidentes com as maiores direcções de alongamento da nuvem, são designadas por eixos factoriais, possuindo valores de explicação da nuvem, ou valores próprios, iguais à relação entre a variância das projecções em cada uma e a variância total.

Esta representação pode projectar os indivíduos e propriedades num espaço de dimensões reduzidas, o que permite expor graficamente num mesmo plano factorial os indivíduos e/ou propriedades intervenientes e deste modo, clarificar as respectivas similitudes e oposições. Assim, para estudar as correlações entre variáveis, basta verificar se as respectivas variações são idênticas nos diferentes indivíduos, bem como a proximidade entre indivíduos, para concluir que são semelhantes (PEREIRA, 1990). São estas características que tornam estes métodos ideais para detectar possíveis estruturas multivariadas nos dados de partida, e indiciar análises complementares.

À semelhança dos processos geoestatísticos e de outras técnicas muito exigentes a nível de cálculo, estes métodos iniciados no princípio do século XX viram o seu aperfeiçoamento e aplicabilidade adiados, até ao desenvolvimento e proliferação das tecnologias informáticas. Com efeito, após os trabalhos pioneiros de Pearson, 1901, Spearman, 1904 e Hotteling, 1933, foi apenas nas décadas de 60 e 70 que se assistiu a um forte desenvolvimento, divulgação e aplicação destas técnicas nos mais diversos domínios, alargando-se claramente o seu âmbito.

2.2.1. ANÁLISE FACTORIAL DAS CORRESPONDÊNCIAS

Com o objectivo de pesquisar as associações entre variáveis em tabelas de contingência, J. P. Benzécri (BENZÉCRI, 1973) formulou uma metodologia, a Análise Factorial das Correspondências (AFC), que assenta no procedimento dos métodos factoriais e é indicada para descrever conjuntos de dados qualitativos, os quais mantêm o significado após a classificação em categorias ou tabelas de descrição lógica. A diversidade de variáveis

a tratar (sejam elas nominais, ordinais ou quantitativas transformadas em qualitativas) exige a codificação da informação de partida numa tabela Q de dimensões $N \times P$ (N linhas por P colunas), onde, na intersecção da linha i (pertencente ao conjunto I de cardinal N) com a coluna j (pertencente ao conjunto J de cardinal P), se encontra um código numérico $K(i, j)$ que relaciona, de algum modo, a linha i com a coluna j (PEREIRA *et al.*, 1988). Esta tabela é designada por matriz disjuntiva completa, sendo com base nela que é calculada uma matriz com estrutura especial, designada por matriz Burt, que serve de base à AFC. A matriz de Burt é quadrada, de dimensão igual ao número total de modalidades de todas as variáveis, e simétrica. Resulta do produto $Q^T Q$, e é constituída por submatrizes em número igual ao quadrado do número total de modalidades. As submatrizes da diagonal principal são matrizes diagonais e representam a frequência dos indivíduos em cada modalidade, e as restantes submatrizes representam a frequência de indivíduos que co-ocorrem no cruzamento das diferentes modalidades.

A AFC pretende explicitar a estrutura de quadros de dados do tipo referido anteriormente, tornando exequível a interpretação da variabilidade de uma população com recurso a um pequeno número de factores independentes. A utilização desta técnica permite apresentar, de forma mais inteligível, a estrutura principal dos dados, através da sua projecção numa dimensão mais reduzida, com uma perda mínima de informação. Assim, o primeiro eixo factorial é encontrado através de uma regressão linear múltipla, ou seja, é a recta em R^P que minimiza a soma dos quadrados das distâncias para cada ponto, onde, para igualizar os valores das propriedades possuidoras de ordens de grandeza diferentes, se ponderam as distâncias pelo inverso das respectivas massas.

2.2.2. CONSTRUÇÃO DE ÍNDICES DE SÍNTESE

Para além do seu desígnio puramente descritivo, que permite expor, de modo mais ou menos claro, as oposições e similitudes entre os objectos em análise, a AFC pode ter uma aplicação mais abrangente do que a análise factorial clássica. De facto os valores das coordenadas das amostras em cada eixo factorial principal são bons representantes da condensação da informação de partida, com uma representatividade igual à soma das

inércias transportadas pelos eixos considerados. É esta potencialidade, que é específica da AFC (princípio baricêntrico), que torna possível a sua utilização como gerador de condensação da informação de partida em índices de síntese, com interesse para articular variáveis de natureza diferente num valor quantitativo, passível de utilização como variável auxiliar dos processos de estimção.

2.3. ESTIMAÇÃO GEOESTATÍSTICA

Uma das mudanças mais significativas na abordagem dos fenómenos relacionados com a quantificação dos recursos naturais, ocorrida no último quartel do séc. XX, prende-se com a conceptualização e generalização da utilização de modelos geomatemáticos, para estimar uma população com base numa amostragem. Esta abordagem, baseada na quantificação da contiguidade espacial da variável em estudo, é realizada no pressuposto de terem existido fenómenos aglutinantes, no domínio em análise, que originaram distribuições de variáveis, cujos valores estão de alguma forma dependentes das posições espaciais ocupadas. Esta noção de dependência espacial levou Krige (KRIGE, 1951), a concluir que as concentrações de ouro só eram explicáveis, se a distância entre observações fosse tomada em conta. Foi baseado nestas conclusões empíricas que Matheron (MATHERON, 1963) desenvolveu a teoria das variáveis regionalizadas (VR), que concilia este aspecto estruturado com uma característica aleatória, tradutora da variação irregular resultante da múltipla variedade de factores dos processos genéticos. É este duplo aspecto, aparentemente contraditório, que permite considerar o valor $z(x_i)$ de uma VR $Z(x)$, no ponto x_i , como uma realização particular da variável aleatória $Z(x_i)$ (PEREIRA, 1979).

Para atingir os seus objectivos (construção de modelos da realidade geológica através de técnicas de estimção), a teoria das VR toma a informação fragmentária de partida disponível como a realização de uma função aleatória que passa a representar o conjunto de variáveis aleatórias localizadas nos pontos amostrais.

A caracterização de uma função aleatória é efectuada pela sua lei de distribuição. Assim, com uma só realização de cada variável aleatória torna-se impossível determinar qualquer parâmetro estatístico, o que coloca a questão de como inferir as respectivas leis de

distribuio ou at apenas alguns dos seus primeiros momentos. Para tornar o modelo exequível, e reduzir os requisitos de que dependem,  necessrio recorrer  imposio de algumas hipteses restritivas adicionais de homogeneidade espacial. Nas aplicaes geoestatísticas,  usual considerar como suficientes as hipteses restritivas de: estacionaridade de 2ª ordem; e hiptese intrínseca, que postulam apenas a estacionaridade dos dois primeiros momentos (da variável e dos seus incrementos).

O conceito de estacionaridade est fortemente ligado  escala de trabalho, uma vez que existem fenômenos que apresentam características no estacionrias a grande escala, mas que verificam essa hiptese localmente, ou seja, existe uma quasi-estacionaridade (SOUSA *et al.*, 1990). Uma funo aleatria diz-se estacionria de 2ª ordem quando os dois primeiros momentos existem e no dependem da localizao espacial. O que, no caso da mdia, corresponde a dizer que  constante em todas as variáveis aleatrias, ou seja,  representativa de todo o domínio em estudo, permitindo realizar a estimaco em locais com valores desconhecidos:

$$E\{Z(x)\} = m(x) = m \quad \forall x \in \text{ao domínio}$$

E no caso da covariância espacial $[C(h)]$ equivale a dizer que depende apenas do vector h que liga os pares de pontos do domínio e  independente da localizao:

$$\text{Cov}\{Z(x+h), Z(x)\} = E\{Z(x+h).Z(x)\} - m^2 = C(h)$$

No entanto, existem alguns fenômenos naturais que apresentam deriva, mesmo localmente, no permitindo a sua classificao como funes aleatrias estacionrias, mas que pertencem  classe das chamadas funes aleatrias intrínsecas. Assim, e uma vez que a estimaco estatística no exige a estacionaridade de 2ª ordem, basta que as funes aleatrias satisfaam as condies exigidas pela classe das funes aleatrias intrínsecas (imposio da estacionaridade dos dois primeiros momentos dos acréscimos espaciais), para que sejam consideradas adequadas, ou seja:

$$\varepsilon(h) = Z(x+h) - Z(x) \quad \forall x \in \text{ao domínio}$$

conduz às duas seguintes condições de estacionaridade:

$$E\{\varepsilon(h)\} = 0 \quad \text{e} \quad E\{\varepsilon(h)^2\} = 2\gamma(h)$$

o que permite definir o variograma da variável $Z(x)$ a partir do 2º momento dos acréscimos

$$\gamma(h) = \frac{1}{2} E\{[Z(x+h) - Z(x)]^2\}$$

2.3.1. KRIGAGEM ORDINÁRIA

O primeiro passo da metodologia geoestatística de estimação consiste na análise estrutural das variáveis amostradas (PEREIRA, 1981), cujo objectivo é caracterizar e quantificar o modo como as VR se dispersam no espaço. A função matemática do variograma $[\gamma(h)]$, dada pela variância da semidiferença dos valores da variável em pontos separados pelo vector h , permite substituir a distância euclidiana, h , pela distância $\gamma(h)$, específica da estrutura espacial do atributo no domínio em estudo, e serve de base ao desenvolvimento e aplicação da generalidade dos processos de estimação.

Uma vez que a VR caracteriza um dado fenómeno que não é normalmente conhecido em todos os pontos do seu domínio, mas somente nos pontos amostrados, torna-se necessário calcular a função do variograma experimental através da seguinte expressão:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2, \quad \forall i \in \text{ao domínio}$$

em que $N(h)$ é o número de pares de pontos separados pelo vector h .

Após o cálculo do variograma experimental, analisam-se as características estruturais mais importantes do fenómeno em estudo e ajusta-se uma função matemática - modelo teórico - que, nas aplicações geoestatísticas relacionadas com as Ciências da Terra, é

frequentemente um modelo esférico. Na maioria dos casos, o comportamento de um variograma varia com a direco em que é calculado, dizendo-se ento que estamos em presena de fenómenos anisótipos, com direces preferenciais de variabilidade, designando-se por direco principal a que apresenta maior continuidade.

Nas Ciéncias da Terra, para estimar valores da VR num suporte desconhecido, considerado homogéneo para a variável a estimar (LUÍS *et al.*, 1998^b), é comum o recurso a um conjunto de processos genericamente denominados por processos de krigagem, como homenagem aos trabalhos desenvolvidos por Krige, que aplicou ponderadores que traduziam as relaes espaciais da variável amostrada. Deste conjunto, o processo mais utilizado é a krigagem ordinária que recorre aos ponderadores lineares de krigagem (GOOVAERTS, 1997) e (SOARES, 2000) para combinar linearmente a informao existente e garantir dois critérios de qualidade: não-enviesamento e minimizao da variância de estimao.

O cálculo do estimador de KO $[z(x_0)]_{KO}^*$ para um suporte x_0 traduz-se numa combinao linear das n amostras experimentais $z(x_\alpha)$, vizinhas do suporte a estimar, onde é garantido o não-enviesamento e a minimizao da variância de estimao, do modo traduzido pela seguinte equao:

$$[z(x_0)]_{KO}^* = \sum_{\alpha=1}^n \lambda_\alpha^{KO} z(x_\alpha)$$

onde: $[Z(x_0)]_{KO}^*$ - representa o valor estimado por KO para a VR $Z(x)$ no suporte x_0 ;

$z(x_\alpha)$ - são os valores da variável nos pontos conhecidos x_α , que intervêm na estimao do suporte x_0 ;

λ_α^{KO} - são os ponderadores a afectar a cada um dos valores $z(x_\alpha)$;

n - é o número de pontos (x_α) que intervêm na estimao do suporte.

Dentro do formalismo geoestatístico, este estimador é obtido pela imposio dos dois critérios de qualidade já referidos:

- i) o estimador Z_v^* deve ser não-enviesado, ou seja, a média dos valores estimados num domínio deve ser igual à média dos valores experimentais:

$$E\{\varepsilon(x_0)\} = E\{[z(x_0)]_{KO}^*\} - E\{Z(x_0)\} = 0$$

Admitindo a estacionaridade da média dentro dos limites do raio de busca, a condição de não-enviesamento do estimador implica que os ponderadores obedeçam à seguinte condição:

$$E\{\varepsilon(x_0)\} = E\left\{\sum_{\alpha=1}^n \lambda_{\alpha}^{KO} z(x_{\alpha})\right\} - E\{z(x_0)\} = \left[\sum_{\alpha=1}^n \lambda_{\alpha}^{KO} - 1\right] m = 0 \Rightarrow \sum_{\alpha=1}^n \lambda_{\alpha}^{KO} = 1$$

- ii) o critério anterior pode ser respeitado e, no entanto, o processo de estimação produzir valores com uma grande dispersão, o que implica uma perda de qualidade face a outro onde essa dispersão seja menor. Atendendo a este facto, existe a necessidade de assegurar a minimização da variância de estimação, o que corresponde a:

$$\begin{aligned} \min Var\{\varepsilon\} &= Var\{[z(x_0)]_{KO}^* - z(x_0)\} = Var\left\{\sum_{\alpha=1}^n \lambda_{\alpha}^{KO} z(x_{\alpha}) - z(x_0)\right\} \\ &= E\left\{\left[\sum_{\alpha=1}^n \lambda_{\alpha}^{KO} z(x_{\alpha}) - z(x_0)\right]^2\right\} \\ &= E\left\{\sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{KO} \lambda_{\beta}^{KO} Z(x_{\alpha}) Z(x_{\beta})\right\} + E\{[Z(x_0)]^2\} - 2E\left\{\sum_{\alpha=1}^n \lambda_{\alpha}^{KO} z(x_{\alpha}) z(x_0)\right\} \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{KO} \lambda_{\beta}^{KO} E\{Z(x_{\alpha}) Z(x_{\beta})\} + E\{[Z(x_0)]^2\} - 2 \sum_{\alpha=1}^n \lambda_{\alpha}^{KO} E\{z(x_{\alpha}) z(x_0)\} \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{KO} \lambda_{\beta}^{KO} C(x_{\alpha}, x_{\beta}) + C(0) - 2 \sum_{\alpha=1}^n \lambda_{\alpha}^{KO} C(x_{\alpha}, x_0) \end{aligned}$$

A minimização efectuada pelo formalismo de Lagrange, através da adição de um

termo nulo, para assegurar que as solues encontradas honram os valores amostrais e o constringimento de no-enviesamento:

$$\begin{aligned} Var\{\varepsilon\} &= Var\{\varepsilon\} + 2\mu \left(\sum_{\alpha=1}^n \lambda_{\alpha}^{KO} - 1 \right) \\ &= C(0) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^{KO} \lambda_{\beta}^{KO} C(x_{\alpha}, x_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha}^{KO} C(x_{\alpha}, x_0) + 2\mu \left(\sum_{\alpha=1}^n \lambda_{\alpha}^{KO} - 1 \right) \end{aligned}$$

Pelo procedimento clssico de igualar a zero as derivadas parciais em ordem a λ_{α} e μ , resulta o seguinte sistema de equaes:

$$\begin{cases} \sum_{\alpha=1}^n \lambda_{\alpha}^{KO} C(x_{\alpha}, x_{\beta}) + \mu = C(x_{\beta}, x_0), & \beta = 1, \dots, n \\ \sum_{\alpha=1}^n \lambda_{\alpha}^{KO} = 1 \end{cases}$$

2.3.2. UTILIZAO DE VARIVEIS AUXILIARES

2.3.2.1. Anlise variogrfica multivariada

Quando existe dependncia espacial entre diferentes VR correlacionadas entre si, pode recorrer-se à variografia multivariada para produzir um quadro de correlaes espaciais entre o nmero total de variveis (N_{vt}), sendo a dependncia espacial entre duas variveis $Z_i(x)$ e $Z_j(x)$ medida pela covarincia cruzada e o variograma cruzado (MARIANO, 1995):

$$C_{ij}(h) = E \left\{ \left[Z_i(x) - m_i \right] \left[Z_j(x+h) - m_j \right] \right\}, \quad \forall i, j$$

$$\gamma_{ij}(h) = \frac{1}{2} E \left\{ \left[Z_i(x) - Z_i(x+h) \right] \left[Z_j(x) - Z_j(x+h) \right] \right\}, \quad \forall i, j$$

onde se pode verificar que $\gamma_{ij}(h) = \gamma_{ji}(h)$, o que já não acontece com a $C_{ij}(h)$ que pode ser diferente de $C_{ji}(h)$ - ou seja, a covariância é uma função não-simétrica relativamente a h .

O grande inconveniente da geoestatística multivariada prende-se com a necessidade de modelar conjuntamente $[N_{vt} \cdot (N_{vt} - 1) / 2]$ variogramas cruzados e N_{vt} variogramas simples. Para simplificar esta tarefa, utilizam-se os mesmos tipos de funções teóricas e amplitudes na definição de cada estrutura dos variogramas de todas as VR, uma vez que também no caso multivariado, as estruturas imbricadas dos variogramas cruzados têm, imperativamente, de assegurar para cada passo h a positividade da variância das combinações lineares existentes, sendo necessário, portanto, recorrer à modelação por combinação linear de modelos básicos definidos positivos (SOUSA, 1988). O modelo linear de co-regionalização das $[N_{vt} \cdot N_{vt}]$ covariâncias simples e cruzadas pode ser definido por :

$$C_{ij}(h) = \sum_{l=0}^L b_{ij}^l C_l(h), \quad \forall i, j$$

$$\gamma_{ij}(h) = \sum_{l=0}^L b_{ij}^l \gamma_l(h), \quad \forall i, j$$

onde b_{ij}^l corresponde aos patamares das estruturas básicas $C_l(h)$.

Existem algumas regras práticas de ajuda à modelização da co-regionalização de sistemas multivariados por modelos lineares, que podem ser vistas em GOOVAERTS, 1997 e SOARES, 2000.

2.3.2.2. Cokrigagem ordinária

Quando se efectua a amostragem de uma grandeza num determinado domínio, é frequente amostrar também informação sobre algumas grandezas auxiliares, que podem ser utilizadas para melhorar os processos de estimação. Este ponto assume particular importância quando a informação principal é escassa e/ou a sua contiguidade espacial é débil. A interdependência entre os valores de um conjunto de VR, presentes num

determinado domínio, originou o desenvolvimento das chamadas Variáveis Corregionalizadas (VC) e levou ao aparecimento da geoestatística multivariada, formulada por Matheron (SOUSA, 1988) com o objectivo, à semelhança da geoestatística univariada, de encontrar os melhores estimadores e respectivos erros de estimaco, tendo por base as contiguidades espaciais de cada e entre as várias VR.

Para integrar toda esta informaco no processo de estimaco de VC, pode utilizar-se o processo de estimaco multivariada de CKO. Este processo, é um interpolador exacto, ou seja, honra todos os valores amostrais da VR principal, advindo as suas principais vantagens comparativamente à estimaco univariada, da capacidade de incorporar, para além das correlaces espaciais existentes entre as amostras da VR principal, as existentes entre as amostras de cada VR auxiliar, e entre todas as amostras de todas as VR intervenientes no processo. Mantém-se também a coerência entre a estimaco da soma dos termos e a estimaco de cada um em particular.

O processo de estimaco recorre à VR principal, conhecida em N_1 pontos amostrados $\{Z_1(x_{\alpha_1}), \alpha_1 = 1, \dots, N_1\}$ e às N_S variáveis auxiliares conhecidas nos N_i pontos amostrados, $\{Z_i(x_{\alpha_j}), \alpha_j = 1, \dots, N_i, i = 2, \dots, N_{S+1}\}$ para produzir um estimador linear, onde o número total de variáveis envolvidas é igual ao número de variáveis auxiliares mais uma ($N_{vt} = N_S + 1$). No caso de existir apenas uma VR auxiliar (Z_2), este processo pode ser descrito pela seguinte combinaco linear dos respectivos valores mais próximos, delimitados pelo raio de busca:

$$[z_1(x_0)]_{CKO}^* = \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} z_1(x_{\alpha_1}) + \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} z_2(x_{\alpha_2})$$

Tratando-se de um estimador que se quer não-enviesado e com variância de estimaco mínima, torna-se necessário assegurar a condico de não-enviezamento:

$$E\{[z_1(x_0)]_{CKO}^* - z_1(x_0)\} = 0 \Rightarrow$$

$$E\left\{\left[\sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} z_1(x_{\alpha_1}) + \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} z_2(x_{\alpha_2})\right] - z_1(x_0)\right\} = 0$$

Dada a estacionaridade das médias locais das duas variáveis $E\{[Z_i(x_0)]\} = E\{Z_i(x_0)\} = m_i$

$$\left[\sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} - 1 \right] m_1 + \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} m_2 = 0$$

é necessário utilizar os seguintes constrangimentos nos ponderadores de CKO:

$$\sum_{\alpha_1}^{N_1} \lambda_{\alpha_1}^{CKO} = 1 \quad e \quad \sum_{\alpha_2}^{N_2} \lambda_{\alpha_2}^{CKO} = 0$$

Relativamente à condição de impor que a dispersão do erro em torno da respectiva média (nula) seja mínima:

$$\begin{aligned} Var\{\varepsilon\} &= Var\left\{ [z_1(x_0)]_{CKO}^* - z_1(x_0) \right\} = Var\left\{ \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} z_1(x_{\alpha_1}) + \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} z_2(x_{\alpha_2}) - z_1(x_0) \right\} \\ &= E\left\{ \left[\sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} z_1(x_{\alpha_1}) + \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} z_2(x_{\alpha_2}) - z_1(x_0) \right]^2 \right\} \\ &= \sum_{\alpha_1=1}^{N_1} \sum_{\beta_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} \lambda_{\beta_1}^{CKO} C_{Z_1}(x_{\alpha_1}, x_{\beta_1}) + \sum_{\alpha_2=1}^{N_2} \sum_{\beta_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} \lambda_{\beta_2}^{CKO} C_{Z_2}(x_{\alpha_2}, x_{\beta_2}) + C_{Z_1}(x_0, x_0) \\ &\quad + 2 \sum_{\alpha_1=1}^{N_1} \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_1}^{CKO} \lambda_{\alpha_2}^{CKO} C_{Z_1 Z_2}(x_{\alpha_1}, x_{\alpha_2}) - 2 \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} C_{Z_1}(x_{\alpha_1}, x_0) - 2 \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} C_{Z_1 Z_2}(x_{\alpha_2}, x_0) \end{aligned}$$

A minimização efectuada pelo formalismo de Lagrange, através da adição de dois termos nulos, assegura, que as soluções encontradas honram os valores amostrais e os constrangimentos de não-enviesamento:

$$Var\{\varepsilon\} = Var\{\varepsilon\} + 2\mu_1 \left(\sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} - 1 \right) + 2\mu_2 \left(\sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} \right)$$

que pelo procedimento clássico de igualar a zero as derivadas parciais em ordem a λ_{α_1} , λ_{α_2} , μ_1 e μ_2 , conduz ao seguinte sistema de equações:

$$\left\{ \begin{array}{l} \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} C_{Z_1}(x_{\alpha_1}, x_{\beta_1}) + \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} C_{Z_1 Z_2}(x_{\alpha_2}, x_{\beta_1}) + \mu_1 = C_{Z_1}(x_{\beta_1}, x_0), \quad \beta_1 = 1, \dots, N_1 \\ \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} C_{Z_1 Z_2}(x_{\alpha_1}, x_{\beta_2}) + \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} C_{Z_2}(x_{\alpha_2}, x_{\beta_2}) + \mu_2 = C_{Z_1 Z_2}(x_{\beta_2}, x_0), \quad \beta_2 = 1, \dots, N_2 \\ \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO} = 1 \\ \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO} = 0 \end{array} \right.$$

A generalização do estimador de CKO para um conjunto de (N_s) VR auxiliares pode ser representada pelo seguinte sistema de equações:

$$\left\{ \begin{array}{l} \sum_{j=1}^{N_{VT}} \sum_{\beta_j=1}^{n_j} \lambda_{\beta_j}^{CKO} C_{ij}(x_{\alpha_i} - x_{\beta_j}) + \mu_i^{CKO} = C_{i1}(x_{\alpha_i} - x_0) \quad \text{com } \alpha_i = 1, \dots, n_i \quad i = 1, \dots, N_{VT} \\ \sum_{\beta_i=1}^{n_i} \lambda_{\beta_i}^{CKO} = \delta_{i1} \quad i = 1, \dots, N_s \quad \text{e onde } \delta_{i1} = 1 \quad \text{para } i = 1, \dots, N_{VT} \quad \text{e } \delta_{i1} = 0 \quad \forall i \neq 1 \end{array} \right.$$

2.3.2.3. Krigagem com deriva externa

Os fenômenos físicos intervenientes na diferenciação de algumas variáveis características de um recurso natural, originam por vezes VR com esperanças não estacionárias, mesmo em áreas dentro dos limites do raio de procura. Em alguns destes casos, a estacionaridade da esperança pode ser encarada como função linear de VR

auxiliares, presentes no domínio em estudo. Isto pode ser exemplificado, para o caso de apenas uma VR auxiliar, pelas relações encontradas entre médias locais de pluviometria e as altitudes topográficas (GOOVAERTS, 1999).

Na KDE, a tendência é modelada como função linear de variáveis auxiliares externas $Y(x)$ (MARÉCHAL, 1984), (GOOVAERTS, 1997) e (SOARES, 2000)

$$m(x) = a_0(x) + a_1(x) y(x)$$

onde os coeficientes a_0 e a_1 são estimados conjuntamente com $Z(x_\alpha)$ e a deriva $m(x_0)$ assume o valor da variável $Y(x_0)$, sendo o estimador em x_0 definido à custa de n amostras vizinhas por:

$$[Z_{KDE}(x_0)]^* = \sum_{\alpha=1}^n \lambda_{\alpha}^{KDE} Z(x_{\alpha})$$

A condição de não-envieçamento é assegurada pela igualdade a zero da diferença entre valores reais e estimados:

$$E\{[Z_{KDE}(x_0)]^* - Z(x_0)\} = 0$$

$$E\left\{\sum_{\alpha=1}^n \lambda_{\alpha}^{KDE} Z(x_{\alpha}) - Z(x_0)\right\} = \sum_{\alpha=1}^n \lambda_{\alpha}^{KDE} m(x_{\alpha}) - m(x_0)$$

$$\sum_{\alpha=1}^n \lambda_{\alpha}^{KDE} [a_0 + a_1 Y(x_{\alpha})] - [a_0 + a_1 Y(x_0)] = 0$$

de onde resulta mais uma condição de universalidade, por cada uma das P variáveis auxiliares da deriva externa, a acrescentar à condição de universalidade da KO, através dos seguintes constrangimentos dos ponderadores λ_i^{KDE} :

$$\sum_{\alpha=1}^N \lambda_{\alpha}^{KDE} Y(x_{\alpha}) = Y(x_0)$$

$$\sum_{\alpha=1}^N \lambda_{\alpha}^{KDE} = 1$$

Para tornar mais inteligível o processo KDE, pode recorrer-se à decomposição na soma de um valor médio e num resíduo:

$$Z(x) = m(x) + R(x), \text{ em que } R(x) \text{ tem média nula e covariância } C_r(h)$$

Neste caso, a condição de optimabilidade, a exemplo do referido anteriormente para a KO, pode ser escrita em função da minimização da variância pelo formalismo de Lagrange, o que significa igualar a zero as $(n + 1 + P)$ derivadas parciais de:

$$E \left\{ \left[[R(x_0)]^* - R(x_0) \right]^2 \right\} + 2\mu_0 \left[\sum_{\alpha} \lambda_{\alpha}^{KDE} - 1 \right] + 2\mu_1 \left[\sum_{\alpha} \lambda_{\alpha}^{KDE} Y(x_{\alpha}) - Y(x_0) \right]$$

onde os ponderadores λ_i^{KDE} são obtidos do sistema de $(n + 1 + P)$ equações lineares:

$$\left\{ \begin{array}{l} \sum_{\alpha=1}^N \lambda_{\alpha}^{KDE} C_R(x_{\alpha}, x_{\beta}) + \mu_0 + \sum_{k=1}^P \mu_k Y(x_{\beta}) = C_R(x_{\beta}, x_0), \quad \beta = 1, \dots, N \\ \sum_{\alpha=1}^N \lambda_{\alpha}^{KDE} = 1 \\ \sum_{\alpha=1}^N \lambda_{\alpha}^{KDE} Y_p(x_{\alpha}) = Y_p(x_0) \quad \text{com } P = 1, \dots, \text{número de variáveis auxiliares} \end{array} \right.$$

A KDE utiliza a informação das VR auxiliares $Y(x)$, previamente amostradas e/ou estimadas, em todos os suportes de amostragem e de estimação da VR principal ($Z(x)$), como variáveis externas auxiliares para realizar um modelo de tendência. Isto implica a necessidade de existir um conhecimento do fenómeno físico, com um grau de confiança

suficientemente forte, que permita estabelecer uma confiana na relao entre a variável principal e as variáveis auxiliares.

2.3.2.4. Cokrigagem ordinária com deriva externa

Nos casos em que seja justificada a utilizao da CKO em suportes onde sejam conhecidas variáveis auxiliares externas, pode recorrer-se à utilizao da CKO com deriva externa (CKO_DE) que é traduzida pelo seguinte sistema de equaes

$$\left\{ \begin{array}{l} \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO-DE} C_{Z_1}(x_{\alpha_1}, x_{\beta_1}) + \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO-DE} C_{Z_1 Z_2}(x_{\alpha_2}, x_{\beta_1}) + \mu_1 + \sum_{k=1}^P \mu_{k+2} Y_k(x_{\beta_1}) = C_{Z_1}(x_{\beta_1}, x_0) \\ \text{com } \beta_1 = 1, \dots, N_1 \\ \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO-DE} C_{Z_1 Z_2}(x_{\alpha_1}, x_{\beta_2}) + \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO-DE} C_{Z_2}(x_{\alpha_2}, x_{\beta_2}) + \mu_2 + \sum_{k=1}^P \mu_{k+2} Y_k(x_{\beta_2}) = C_{Z_1 Z_2}(x_{\beta_2}, x_0) \\ \text{com } \beta_2 = 1, \dots, N_2 \\ \sum_{\alpha_1=1}^{N_1} \lambda_{\alpha_1}^{CKO-DE} = 1 \\ \sum_{\alpha_2=1}^{N_2} \lambda_{\alpha_2}^{CKO-DE} = 0 \\ \sum_{\alpha=1}^N \lambda_{\alpha}^{CKO-DE} Y_p(x_{\alpha}) = Y_p(x_0) \quad \text{com } P = 1, \dots, \text{número de variáveis} \end{array} \right.$$

2.3.3. FORMALISMO DA INDICATRIZ

A visualizao da morfologia de um qualquer atributo característico de um sistema geológico é uma operao geralmente impossível de realizar de modo directo, obrigando a recorrer a modelos estocásticos de definio indirecta da forma, ou seja, a morfologia é

obtida à custa do conjunto de pontos onde ocorrem as maiores probabilidades de pertença ao corpo. Neste campo, a geoestatística pode ser uma excelente ferramenta, visto ser um interpolador óptimo, que incorpora medidas das contiguidades médias do corpo (covariância ou variogramas), para assegurar que a probabilidade estimada de pertença de um qualquer ponto ao corpo depende do grau de pertença das amostras vizinhas. Uma das grandes vantagens deste tipo de modelos é a possibilidade de localizar facilmente as zonas de maior incerteza da forma, conseguida através de pequenas variações dos valores de corte das probabilidades de pertença (SOARES, 1989).

A utilizaço de modelos deste tipo obriga à construco de uma variável do tipo indicatriz, dicotómica, resultante da classificaço dos dados amostrais em uns e zeros. Assim, dada uma classe X e o seu complementar X^c , no domínio $A = X \cup X^c$, onde existe um conjunto de N amostras x_i , localizadas em A com $i = 1, \dots, N$, a variável indicatriz $I(x_i)$ é expressa por:

$$I(x_i) = \begin{cases} 1 & \text{se } x_i \in X \\ 0 & \text{se } x_i \in X^c \end{cases}$$

onde X é uma fase de A e, X^c o seu complementar.

A fase X pode designar uma litologia, um tipo de minério ou qualquer outra categoria de uma variável qualitativa ou intervalo de valores de uma variável quantitativa. Se a variável indicatriz for aplicada a uma área A , subdividida não em duas mas em p fases, de modo que $k(x_i) = \text{prob} \{ x_i \in X_I \}$ com $I = 1, \dots, p$ e $A = \{X_1 \cup X_2 \cup \dots \cup X_p\}$, então a estrutura multifásica no ponto x_i passa a ser representada por um vector de valores de indicatriz $k_1(x_i), k_2(x_i), \dots, k_p(x_i)$, em que cada um é definido por (ALMEIDA, 1992):

$$K_I(x_i) = \begin{cases} 1 & \text{se } x_i \in X_I \\ 0 & \text{se } x_i \notin X_I \end{cases}$$

onde X_I ($I = 1, \dots, p$) representa cada uma das fases de A .

A variável indicatriz tem o significado de uma probabilidade de pertença de cada ponto x_i a cada fase X_I , podendo ser considerada como uma realização de uma função aleatória $I(x)$, com uma lei de distribuição binomial, caracterizada pelos seguintes momentos de primeira e segunda ordem:

$$m_I = E\{I(x)\} \quad \text{e} \quad Var_I = E\{(I(x) - m_I)^2\} = m_I(1 - m_I)$$

A covariância é interpretada como a probabilidade de duas amostras (x_i e $x_i + h$) separadas pelo vector h pertencerem a X . Também à semelhança de qualquer outra VR, quando se pretende estimar a probabilidade de pertença à classe X , pode utilizar-se uma combinação linear do conjunto de valores de indicatriz das amostras experimentais vizinhas, permitindo assim a aplicação dos processos de estimção geoestatísticos, segundo o formalismo da indicatriz.

2.3.3.1. Classificação de estruturas multifásicas e identificação de zonas de incerteza

A codificação pelo formalismo da indicatriz das variáveis presentes num domínio em estudo, conduz a um quadro onde os valores assumidos por cada fase, em cada ponto amostral, pertencem ao intervalo $[0, 1]$ e traduzem a probabilidade de pertença do ponto à fase considerada. No caso particular de uma situação em que as fases de uma variável codificada assumem sempre os valores zero ou um, pode dizer-se que existe uma “miscibilidade” entre as amostras.

A utilização de processos geoestatísticos na estimção de probabilidades de pertença a uma fase num dado domínio levanta a questão da classificação de cada valor estimado em cada suporte, uma vez que raramente o respectivo valor estimado será igual a um. Este facto não é devido aos métodos de estimção utilizados, nem depende dos modelos variográficos encontrados, mas decorre da própria continuidade das VR e das atenuações dos valores extremos, consequentes da utilização de qualquer processo de estimção. Os valores estimados, tal como as amostras, podem ser interpretados como a probabilidade de pertença de cada ponto, a uma das fases em que foi subdividida a variável.

Para transformar os valores estimados para a probabilidade de cada classe em valores que reproduzam o quadro multifásico de partida, utiliza-se um algoritmo de classificação dinâmica, proposto por (SOARES, 1992; 1998), que reproduz as proporções globais das classes e atribui a cada posição (suporte estimado) a categoria com maior probabilidade de ocorrência, através dos seguintes passos:

1. ordenação decrescente dos valores estimados para cada classe;
2. cálculo do número de valores pertencentes a cada classe (n_f) da seguinte forma:
 $n_f = p_i \times n_e$, onde: p_i é a proporção da classe i ; n_e é o número de suportes estimados;
3. escolha e transformação, em cada linha, do valor mais elevado das classes estimadas, sendo em seguida esta linha retirada. O algoritmo prossegue até à atribuição do número de amostras por classe.

A classificação de estruturas multifásicas conduz a valores de corte da variável indicatriz, que atribuem diferentes classes a valores estimados muito próximos. Esta situação pode conduzir a classificações menos boas, uma vez que existe uma zona de valores com idênticas probabilidades de pertença a quaisquer duas classes (LUÍS *et al.*, 1991). Para obviar esta situação, é comum mapear os valores contidos nestas zonas de maior incerteza, para visualizar a sua distribuição.

2.3.4. TESTES DE VALIDAÇÃO CRUZADA

Os modelos deterministas calculam para cada “suporte” não-amostrado uma solução única $z^*(x_\alpha)$, supostamente exacta, que assumem como valor desconhecido $z(x_\alpha)$, sendo incapazes de calcular o erro de estimação [$z^*(x_\alpha) - z(x_\alpha)$]. Esta incapacidade conduz à assunção de tomar como “verdadeiros” os valores por si calculados, que passam a ser admitidos como representantes da grandeza em análise, em todas as subsequentes utilizações.

A incapacidade de qualquer método em contabilizar o erro de estimação, cometido na estimação de um atributo num “suporte”, pode ser parcialmente ultrapassada, quando estes são pontuais e coincidentes com a amostragem, através da utilização do chamado Teste do

Ponto Fictício (TPF) ou “cross validation”, que consiste na eliminação sucessiva do valor de uma dada amostra $z(x_\alpha)$, do conjunto de amostragem, procedendo-se em seguida à estimação do seu valor $z^*(x_\alpha)$, à custa dos valores das amostras vizinhas. O conhecimento do par de valores real-estimado, na totalidade da localização das amostras, $\alpha = 1, \dots, n$, permite efectuar o cálculo de alguns estatísticos, de que se salientam o desvio quadrático médio (*DQM*) que calcula a média do quadrado das diferenças entre os valores reais $z(x_\alpha)$ e os estimados $z^*(x_\alpha)$ e o erro relativo médio (*ERM*):

$$DQM = \frac{1}{n} \sum_{\alpha=1}^n [z(x_\alpha) - z^*(x_\alpha)]^2 \qquad ERM = \frac{1}{n} \sum_{\alpha=1}^n \frac{[z(x_\alpha) - z^*(x_\alpha)]}{z(x_\alpha)}$$

sendo n o número de amostras estimadas. A escolha destes critérios, para aferir a qualidade dos dados estimados, deve-se ao facto de serem bons indicadores da precisão e não-enviesamento das estimações, pretendendo-se que o *DQM* seja mínimo, e que o *ERM* tenda para zero.

O facto do valor de estimação para uma VR poder atingir valores mínimos relativamente baixos e entrar como quociente na equação do *ERM* pode afectar negativamente o valor final deste teste, caso existam algumas amostras com teores baixos que estejam rodeadas de valores elevados. Para contrariar esta possibilidade, calcula-se o erro relativo médio (*ERM_{Z médio}*) onde no quociente entra a média dos valores amostrais da VR ($\bar{z}(x)$), em substituição do valor pontual.

$$ERM_{(Z \text{ médio})} = \frac{1}{n} \sum_{\alpha=1}^n \frac{[z(x_\alpha) - z^*(x_\alpha)]}{\bar{z}(x)}$$

Na estimação multivariada, interessa também assegurar a coerência entre os valores estimados e as restantes variáveis. A construção de tabelas de contingência, após a divisão em classes das variáveis quantitativas, entre pares de variáveis e o cálculo do respectivo χ^2 , permite aferir aquela ocorrência.

2.4. SIMULAÇÃO GEOESTATÍSTICA

2.4.1. INTRODUÇÃO

A estimação geoestatística interpreta a distribuição espacial de uma variável aleatória $z(x)$, como uma realização da função aleatória $Z(x)$, caracterizada por funções de distribuição e de covariância condicionadas aos valores amostrais. Já os processos de simulação estocástica foram idealizados para tentar reproduzir de modo probabilístico a variabilidade dos atributos em estudo. Para isso, sorteiam N_s diferentes realizações da função aleatória $Z(x)$ nos suportes não amostrados e reproduzem, em cada um, características estatísticas como: a média, a função de distribuição e os variogramas encontrados na amostragem.

Por vezes, a densidade local de amostras efectuadas no jazigo pode ser mais elevada nas zonas de teores mais ricos, o que pode condicionar a representatividade da média global e conseqüentemente as médias e variâncias de estimação, principalmente quando é utilizado o processo de krigagem simples. Para obviar a esta situação pode reter-se apenas as amostras próximas de uma rede regular, caso a densidade de amostragem o permita, ou diminuir a influência amostral nas zonas mais amostradas (ISAACS *et al.*, 1989).

Os processos de simulação condicionais ou não condicionais (de acordo com a imposição de honrar os valores da VR nos pontos coincidentes com a amostragem) devem reproduzir o comportamento espacial da VR e permitir o acesso à incerteza do seu valor desconhecido em cada suporte, com base na percentagem de realizações que assumem ou ultrapassam um dado valor de referência. Segundo Journel (JOURNEL, 1974) a simulação condicional procede à simulação de um valor real, com base na simulação do respectivo erro de estimação. Basicamente a simulação e a estimação geoestatística diferem nos dois seguintes aspectos:

- i. utilizam diferentes conceitos sobre a precisão dos respectivos resultados, cabendo à estimação calcular a melhor solução local $z_i(x)$ sem atender à covariância entre os valores estimados, e à simulação reproduzir as características espaciais da função aleatória $Z(x)$ definida pela amostragem;
- ii. conduzem a modelos de natureza distinta, cabendo aos processos de estimação fornecer apenas o modelo que constitui a melhor estimativa para um determinado

suporte, e à simulação fornecer diferentes modelos onde cada um é uma representação da realidade, o que permite aceder à incerteza associada a cada suporte, a partir da respectiva função de distribuição de probabilidade.

Deste modo, as simulações estocásticas não produzem o cenário mais provável dos valores de uma VR, permitindo antes aceder a imagens equiprováveis com a mesma média, histograma e variabilidade espacial dos valores experimentais. Esta característica torna-os competentes para idealizar diferentes comportamentos (que podem ser extremos) das características de dispersão ou concentração de uma VR à custa de um conjunto lato de imagens.

2.4.2. SIMULAÇÃO SEQUENCIAL

O objectivo da simulação sequencial ($\{z^{(l)}(x_i); i = 1, \dots, N\}$ com $l = 1, \dots, L$) é gerar N valores nos N nós de uma malha, regular ou não, a partir da função de distribuição cumulativa marginal (função independente da localização dos valores) de um conjunto de n amostras experimentais ($\{z(x_i); i = 1, \dots, n\}$).

$$F(x_1, \dots, x_N; z_1, \dots, z_N | (n)) = \text{Prob} \{z(x_1) \leq z_1, \dots, z(x_N) \leq z_N | (n)\}$$

A simulação sequencial baseia-se na extensão do condicionamento a toda a informação existente na vizinhança do ponto a simular, fornecida tanto pelos valores experimentais como pelos valores anteriormente simulados. Este princípio é traduzido pela relação de Bayes, cuja aplicação sequencial pode ser descrita para dois nós condicionais, através dos dois seguintes passos sucessivos: simulação do primeiro valor z_1 , a partir da distribuição $F(x_1; z_1 | (n))$, a que se segue a simulação do segundo valor z_2 , a partir da distribuição condicional $F(x_2; z_2 | (n+1))$, onde $(n+1)$ representa o condicionamento aos dados experimentais acrescidos do valor simulado no primeiro passo. Este processo é traduzido pela seguinte expressão:

$$F(x_1, x_2; z_1, z_2 | (n)) = F(x_1; z_1 | (n)) \cdot F(x_2; z_2 | (n+1))$$

A generalização deste princípio a N nós pode ser efectuada através da aplicação recursiva da relação de Bayes, de onde resulta o acréscimo de um nível de condicionamento à função de probabilidade em cada nova iteração. Este modo de condicionamento conduz à seguinte expressão geral:

$$F(x_1, \dots, x_N; z_1, \dots, z_N | (n)) = F(x_1; z_1 | (n)) \cdot F(x_2; z_2 | (n+1)) \cdot \dots \\ \dots \cdot F(x_{N-1}; z_{N-1} | (n+N-2)) \cdot F(x_N; z_N | (n+N-1))$$

onde a última parcela do produto ($F(x_N; z_N | (n+N-1))$) se refere à condicionalização da função de distribuição cumulativa da simulação no ponto N , ao conjunto dos n dados experimentais e aos $(N-1)$ pontos anteriormente simulados. A primeira realização fica concluída após a simulação dos valores dos N nós, podendo o número de realizações ser incrementado de acordo com o considerado necessário, bastando para tal impor novos valores aleatórios e/ou diferentes percursos nas visitas a efectuar aos N nós, em cada uma das L vezes, que o processo é repetido. É esta possibilidade de repetição de realizações que permite obter os parâmetros estatísticos da função de distribuição cumulativa (média e variância) da VR simulada em cada nó, de modo semelhante ao que seria conseguido, caso fosse possível amostrar várias vezes um mesmo suporte.

A simulação sequencial apoia-se no método de Monte Carlo, desenvolvido por J. von Neumann e S. Ulam em 1949, com a finalidade de simular o comportamento de processos que dependem só de factores aleatórios. A sua aplicação assenta na possibilidade de construção da função de probabilidade de distribuição, que permite gerar de forma aleatória os valores dos parâmetros utilizados de acordo com o seu andamento.

2.4.3. SIMULAÇÃO SEQUENCIAL GAUSSIANA

A forma mais simples de viabilizar o tratamento conjunto de funções de distribuição cumulativas condicionais (*fdcc*) é através da assunção de modelos capazes de representar adequadamente todas as leis de distribuição, e traduzi-las analiticamente numa única

expressão passível de ser caracterizada por um pequeno número de parâmetros. Por outro lado, a utilização de modelos probabilísticos para prever as funções de distribuição de probabilidades (*fdp*) de uma VR, em pontos ou suportes não amostrados, só é viável desde que esta possa ser estimada a partir dos dados experimentais.

No entanto, o recurso a estes modelos, ainda que justificado pela escassa informação do comportamento real da VR, não determina o tipo de lei, mas apenas a sua média e variância, obrigando à introdução de hipóteses suplementares de modelos de permanência de lei, passando, se necessário, pela correcção da variância. Uma das formas de estimar *fdp* locais, pode ser exemplificada por um conjunto de variáveis aleatórias $\{Y(x), \in A\}$ que segue uma lei conjunta multigaussiana, e onde qualquer par de valores possui também uma lei de distribuição bigaussiana, caracterizada pela covariância $C_Y(x_1, x_2)$.

Assim, atendendo ao facto de qualquer combinação linear de VR normais seguir também uma distribuição normal, perfeitamente caracterizada em qualquer ponto x_0 pela esperança e variância condicionais

$$E\{Y(x_0) | Y(x_1), \dots, Y(x_N)\}$$

$$\text{var}\{Y(x_0) | Y(x_1), \dots, Y(x_N)\},$$

permite calcular a *fdp* gaussiana em x_0 :

$$G(x_0; y) = G \left[\frac{y - E\{Y(x_0) | Y(x_\alpha), \alpha = 1, \dots, N\}}{\sqrt{\text{var}\{Y(x_0) | Y(x_\alpha), \alpha = 1, \dots, N\}}} \right]$$

A adopção da hipótese de multigaussianidade conduz a uma situação particular, onde basta conhecer os dois primeiros momentos (média e variância) para caracterizar completamente a *fdcc* que, atendendo à estacionaridade da média em todo o domínio, permite recorrer aos estimadores lineares de krigagem simples com média global e à

respectiva variância de estimação para efectuar o seu cálculo (GOOVAERTS, 1997; SOARES, 2000):

$$E\{Y(x_0) | Y(x_i), i = 1, \dots, N\} = [Y_{KS}(x_0)]^* = m(x_0) + \sum_{\alpha=1}^n \lambda_{\alpha}^{KS} [Y(x_{\alpha}) - m(x_{\alpha})] = \sum_{\alpha=1}^n \lambda_{\alpha}^{KS} Y(x_{\alpha})$$

$$\text{var}\{Y(x_0) | Y(x_i), i = 1, \dots, N\} = \sigma_{KS}^2(x_0) = \sum_{\alpha=1}^n \lambda_{\alpha}^{KS} C_Y(x_{\alpha}, x_0)$$

ficando a *fdp* gaussiana em x_0 :

$$G(x_0; y) = G\left[\frac{y - [Y_{KS}(x_0)]^*}{\sigma_{KS}(x_0)}\right]$$

Existem, contudo, situações onde a média dos valores de um atributo não é estacionária e/ou depende de outros atributos. Nestes casos, pode recorrer-se aos processos de estimação não-estacionários, como a krigagem ordinária, a krigagem universal ou a krigagem com deriva externa, para calcular a média da função de distribuição gaussiana $G(x; y|n)$ num ponto x_0 . No entanto, a variância de krigagem terá sempre de ser calculada através da krigagem simples, uma vez que é a única calculada no pressuposto de existir uma média conhecida e constante em todo o domínio, tal como acontece com os valores gaussianos resultantes da transformação.

É a simplicidade de definir a *fdp*, em qualquer ponto do domínio em estudo, que torna atractiva a opção pela utilização dos modelos paramétricos gaussianos, em modelos de simulação sequencial. Inclui-se nesta categoria, como o próprio nome indica, a simulação sequencial gaussiana (SSG), que é realizada num “ambiente” onde a variável principal em estudo tem obrigatoriamente de ter uma distribuição gaussiana.

Este aspecto, levanta logo à partida a necessidade de transformar a função de distribuição dos dados experimentais numa função gaussiana, já que raramente as variáveis tratadas nas Ciências da Terra (principalmente quando representam variáveis de baixo teor

como U, Au, Ag, Cu, etc.) são possuidoras de uma distribuição deste tipo. Assim, a SSG requer, num primeiro passo, a transformação gaussiana dos valores experimentais $Z(x)$. Esta transformação (Φ)

$$Y(x_\alpha) = \Phi[Z(x_\alpha)], \quad \alpha = 1, \dots, N$$

é relativamente simples e pode ser efectuada por aproximação polinomial, através dos polinómios de Hermite ou de uma transformação gráfica. Este último, é um método mais simples que faz corresponder ao valor da *fdcc* de um atributo $Z(x)$ um valor da *fdcc* gaussiana $Y(x)$ que segue uma lei gaussiana de média nula e variância unitária (JOURNEL *et al*, 1978); (MUGE, 1982) e (SOARES, 2000).

Dadas as funções de distribuição destas variáveis

$$F(z) = \text{prob}\{Z(x) < z\} \quad \text{e} \quad G(y) = \text{prob}\{Y(x) < y\},$$

pode generalizar-se que o valor z , que corresponde ao valor gaussiano y , e satisfaz $F(z) = G(y)$, é dado pela expressão:

$$Y(x_\alpha) = \Phi[Z(x_\alpha)] = G^{-1}[F(Z(x_\alpha))], \quad \alpha = 1, \dots, N$$

A normalização promove uma nova variável cuja *fdp* univariada é gaussiana. No entanto, esta condição por si só não garante a multigaussianidade, persistindo ainda a necessidade de verificar se a variável transformada $Y(x_i)$ mantém a normalidade entre todos os bipontos, tripontos, ..., k -pontos separados pelos diferentes vectores h (bigaussianidade, trigaussianidade, tetragaussianidade, ..., k -gaussianidade). No entanto, a habitual escassez de informação, associada a malhas de amostragem com formas e passos irregulares, torna difícil a verificação da normalidade, para além dos bipontos existentes.

A simulação geoestatística, para além da reprodução das características estatísticas, tem por principal objectivo reproduzir a variabilidade estrutural das VR (SOUSA, 1983^a). Assim, e caso não exista qualquer indício em contrário, é comum assumir a

multigaussianidade durante o processo de SSG, e condicionar a admissibilidade dos respectivos resultados ao comportamento dos variogramas realizados sobre os dados simulados, que devem reproduzir o comportamento dos correspondentes variogramas experimentais.

A admissão de multigaussianidade de uma variável $Y(x)$ num determinado domínio D , e particularmente para o conjunto dos dados transformados (possuidores de média zero), é uma condição bastante conveniente, uma vez que nos casos em que a média possa ser considerada estacionária ($m_{Y(x)} = m \forall x \in D$), permite recorrer aos estimadores lineares da krigagem simples para estimar a média e a variância características da *fdcc* em qualquer ponto $x_0 \in D$ de uma forma bastante robusta.

Com esta assunção, o processo da SSG pode ser descrito da seguinte forma:

1. estabelecer um percurso aleatório de visita singular a cada um dos N nós da malha a simular;
2. estimar, por krigagem simples, a média e a variância da função de distribuição condicional cumulativa (gaussiana) no primeiro nó a visitar, à custa dos valores e variogramas ($\gamma(h)$) dos n $Y(x)$ valores resultantes da normalização dos valores experimentais:

$$[Y(x_1)]^* = E\{Y(x_1)|(n)\} = \sum_{i=1}^n \lambda_i Y(x_i)$$

$$\sigma_{KS}^2(x_1) = \text{var}\{Y(x_1)|(n)\}$$

3. gerar um valor aleatório p a partir de uma distribuição uniforme entre 0 e 1, e obter o valor gaussiano simulado de $Y_S(x_1)$ a partir da função inversa de distribuição condicional cumulativa gaussiana em x_1 , caracterizada no ponto anterior

$$Y_S(x_1) = G^{-1}([Y(x_1)]^*; \text{var}[Y(x_1)]^*; p)$$

$Y_S(x_1)$ passa a fazer parte dos $(n + 1)$ valores condicionantes da próxima iteração através da sua adição aos valores experimentais: $\{n + 1\} = \{n\} \cup \{Y_S(x_1)\}$;

4. repetir os pontos 2 e 3 para o nó seguinte x_2 , aleatoriamente escolhido no segundo ponto, com base no novo conjunto de $\{n + 1\}$ valores condicionantes da estimação. Este novo valor simulado $Y_S(x_2)$ passa também a integrar o novo conjunto condicionante de estimação

$$\{n + 2\} = \{n + 1\} \cup \{Y_S(x_2)\},$$

seguindo-se a repetição deste procedimento até ser simulado o N e último nó com base nos $(n + N - 1)$ valores condicionantes;

5. o processo é concluído através da transformação dos N valores gaussianos simulados $[Y_S(x_j)]$ em valores com uma função de distribuição semelhante à dos dados experimentais, através da aplicação da função inversa de transformação gaussiana dos dados experimentais

$$Z_S(x_j) = G^{-1}[Y_S(x_j)] \text{ com } j = 1, \dots, N$$

Como já foi referido, a construção de novos cenários pode ser efectuada através de novas simulações, em tudo semelhantes ao descrito nos pontos anteriores, desde que no ponto 3 se usem outros valores aleatórios p .

Para evitar que cada novo cenário simulado dependa sempre dos mesmos nós inicialmente simulados, é vulgar alterar aleatoriamente o percurso de visita dos N nós a simular.

2. Metodologia Proposta e Seus Fundamentos Teóricos

2.4. Simulação geoestatística
