Departamento de Engenharia Informática

Licenciatura em Engenharia Informática e de Computadores (LEIC)

Trabalho Final de Curso

(Ref. TFC/LEIC 01/02 nº39)

Relatório final

Sistema de Informação Geográfica para aplicações de Robótica Oceanográfica

Professor Orientador:
Carlos Jorge Ferreira Silvestre _____

Co-Orientador:
Paulo Jorge Coelho Ramalho Oliveira _____

Alunos:
44171 Rui Martins _____
44155 Luís de Sousa _____

RESUMO

Dado um conjunto finito de amostras de um dado fenómeno natural, distribuído no espaço (como a elevação ou concentração de ozono no ar), existem várias formas de interpolar uma superfície contínua, que o represente no seu todo. O aspecto espacial deste tipo de fenómenos, remete a sua representação num suporte informático, para os Sistemas de Informação Geográfica (SIG); área da informática desenvolvida ainda na década de 1960. Foi analisado o mercado presente dos SIG, considerando estes sistemas como ferramentas, não só de análise e visualização de informação espacial, como também de geração dessa mesma informação. Esta análise levou à escolha do ArcGis da ESRI, como suporte informático a usar. Passou-se depois à análise dos diferentes métodos a usar na interpolação, onde se destacaram os métodos Geoestatísticos, nomeadamente os métodos de Krigagem. Por último foram geradas alguns superfícies, num processo interactivo de refinamento dos modelos geoestatísticos, usando as técnicas de Validação e Validação Cruzada.

PALAVRAS CHAVE

Sistemas de Informação Geográfica (SIG)

Interpolação de Superfícies

Geoestatística

Krigagem

Análise Exploratória Espacial

Validação Cruzada

Validação

Erro de Medição

AGRADECIMENTOS

Gostaríamos de agradecer a colaboração do Doutor Carlos Silvestre e Doutor Paulo Oliveira pelo apoio e orientação fornecidos ao longo do desenvolvimento deste trabalho. Agradecimentos especiais ao Engenheiro Serralha pelo apoio e solicitude sempre demostrados ao longo destes últimos 13 meses.

Por último agradecemos a todos os colegas que alguma vez, de uma ou de outra maneira contribuíram para a progressão deste trabalho.

ÍNDICE

RESUMO……I
PALAVRAS CHAVE……I
AGRADECIMENTOS……II
ÍNDICE…...……...III
LISTA DE FIGURAS.....VI
LISTA DE TABELAS.....VIII
CAPÍTULO I - INTRODUÇÃO 1
INTRODUÇÃO.....1
1.1 Enquadramento.....1
1.2 Objectivos.....1
1.3 Estrutura 2
1.4 Resultados 3
CAPÍTULO II - SISTEMAS DE INFORMAÇÃO GEOGRÁFICA.....4
2.1 Introdução.....4
2.2 O que é um SIG? 4
2.2.1 Definição Genérica de SIG 5
2.2.2 Outras definições de SIG 5
2.3 Representação da Informação num SIG Vectores/*Raster*.....6
2.4 Tipos de Sistemas de Informação Geográfica.....9
2.4.1 Sistemas de Informação Geográfica 10
2.4.2 Funcionalidade do SIG.....10
2.4.3 *Desktop Mapper*.....10
2.4.4 Limitações do DM 12
2.4.5 CAD *Mapping Systems*.....12
2.4.6 Diferenças SIG-CAD 13
2.2.7 Resumo.....13
2.5 Para que serve um SIG? 14
CAPÍTULO III - ESTUDO DE MERCADO.....16
3.1 Introdução.....16
3.2 IDRISI 32 16
3.2.1 CARTA LINX.....19
3.2.2 Conclusões 19
3.3 ARCGIS 20
3.3.1 Aplicações do ArcGIS Desktop 21
3.3.2 Extensões do ArcGIS 21
3.4 GRASS 23
3.5 GEOMATICA.....24
3.5.1 FOCUS.....24
3.5.2 OrthoEngine.....26
3.5.3 PCI Modeler.....26
3.5.4 FLY!.....27
3.5.5 EASI.....27

3.4.6 Conclusões 28
3.6 Estrutura do Mercado dos SIG 29
3.6.1 Estrutura do Mercado em 1994 29
3.6.2 Estrutura do Mercado em 1999 31
3.6.3 Estrutura do Mercado em 2000 32
3.5.4 Outros dados referentes ao Mercado 33
3.6.5 Conclusões 34
3.7 Comparação e Escolha 34
CAPÍTULO IV - INTERPOLAÇÃO DE SUPERFÍCIES 36
4.1 Introdução 36
4.2 Métodos Determinísticos 36
4.2.1 Inverso da Potência das Distâncias (Inverse Distance Weighting - IDW) 37
4.2.2 Interpolação Polinomial Global 39
4.2.3 Interpolação Polinomial Local 40
4.2.4 Funções de Base Radial (*Radial Basis Functions* -RBF) 41
4.3.5 Conclusões 43
4.2 Geoestatística 43
4.2.1 Introdução 43
4.2.2 Modelo Geoestatístico 44
4.2.2.1 Função Aleatória Estacionária 45
4.2.2.2 Estacionariedade do Semivariograma e da Covariância espacial 45
4.2.3 Análise Estrutural 46
4.2.3.1 Parâmetros do Semivariograma 47
4.2.3.2 Modelo Esférico 49
4.2.3.3 Modelo Exponencial 49
4.2.3.4 Modelo Gaussiano 49
4.2.3.5 Modelos de Anisotropia 50
4.2.3.6 O que se vê num Semivariograma 51
4.2.4 Estimação – algoritmos de interpolação de krigagem 51
4.2.4.1 Krigagem Simples 53
4.2.4.2 Krigagem Ordinária 54
4.2.4.3 Krigagem Universal 54
4.2.5 Considerações sobre os ponderadores do estimador de krigagem 55
4.3 Conclusão 56
CAPÍTULO V - GERAÇÃO DE SUPERFÍCIES 58
5.1 Integração dos Dados 58
5.2 Análise Exploratória Espacial 62
5.2.1 Descrição Univariada das amostras 63
5.2.2 Descrição Espacial 65
5.2.3 Mapas de Voronoi 67
5.3 Geração de Superfícies 69
5.3.1 Modelos de Semivariograma 70
5.3.2 Validação Cruzada 73
5.3.3 Semivariograma Esférico 74
5.3.4 Semivariograma Exponencial 75
5.3.5 Semivariograma Gaussiano 76
5.3.6 Semivariograma Combinado 76
5.3.7 Melhores Modelos depois da Validação Cruzada 77

5.5.8 Validação 78
5.4 Zona Circundante do Monte da Guia 79
5.4.1 Amostra 79
5.4.2 Descrição univariada e espacial da zona do Monte da Guia 80
5.4.3 Resultados da Validação Cruzada 81
5.4.4 Validação 83
5.5 Zona de Amostragem Exaustiva a Leste 83
5.5.1 Amostra 83
5.5.2 Descrição univariada e espacial da zona a leste do Monte da Guia 84
5.5.3 Resultados da Validação Cruzada 85
5.5.4 Validação 87
5.6 Erro de Medição 88
5.6.1 Amostra Completa 90
5.6.2 Monte da Guia 91
5.6.3 Zona Leste 92
5.7 Conclusão 94
CAPÍTULO VI - CONCLUSÃO 96
6.1 Estudo de Mercado 96
6.2 Interpolação de Superfícies 96
6.3 Trabalho Futuro 103
ANEXO A – BREVE HISTÓRIA DOS SIG 105
ANEXO B - SEMIVARIOGRAMAS 107
B.1 Área de Trabalho Completa 107
B.1.1 Krigagem Ordinária 107
B.1.2 Krigagem Universal com deriva de 1ª ordem 108
B.1.3 Krigagem Universal com deriva de 2ª ordem 109
B.2 Zona do Monte da Guia 110
B.2.1 Krigagem Ordinária 110
B.2.2 Krigagem Universal de 1ª Ordem 111
B.2.3 Krigagem Universal com deriva de 2ª ordem 112
B.3 Zona a Leste do Monte da Guia 113
B.3.1 Krigagem Ordinária 113
B.3.1 Krigagem Universal com deriva de 1ª ordem 114
B.3.3 Krigagem Universal com deriva de 2ª ordem 115
ANEXO C - DERIVAS 116
C.1 Área de Trabalho Completa 116
C.2 Zona do Monte da Guia 116
C.3 Zona a Leste do Monte da Guia 117
REFERÊNCIAS 118

LISTA DE FIGURAS

Figura 2.1 – Representação Vectorial. 7
Figura 2.2 – Mapa de Densidade Populacional na Holanda, representado por vectores. 8
Figura 2.3 – Pormenor de uma camada *raster* representando o relevo da Serra de Gredos 8
Figura 2.4 – Diferenças entre as duas formas de representação do mundo num SIG 9
Figura 3.1 – Exemplo de Análise com Programação por Filtros 18
Figura 3.2 – Sistema ArcGis 20
Figura 3.3 – Escalabilidade do sistema ArcGis 20
Figura 3.4 – Horizontalidade das Extensões 22
Figura 3.5 – Janela do FOCUS 25
Figura 3.6 – PCI Modeler 26
Figura 3.7 – Gráfico da Estrutura de Mercado em 1994 30
Figura 3.8 – Gráfico da Estrutura de Mercado em 1999 31
Figura 3.9 – Gráfico da Estrutura de Mercado em 2000 32
Figura 4.1 – Vizinhança elíptica 38
Figura 4.2 – Superfície gerada por IDW 39
Figura 4.3 – Superfícies geradas por polinomiais globais 39
Figura 4.4 - Superfície gerada por polinomiais locais 40
Figura 4.5 – Superfície gerada por RBF 41
Figura 4.6 – RBF *versus* IDW 41
Figura 4.7 – Exemplo RBF com 3 amostras 42
Figura 4.8 – Parâmetros do Semivariograma 48
Figura 4.9 – Representação gráfica de modelos transitivos normalizados 49
Figura 4.10 - Representação gráfica da anisotropia geométrica 50
Figura 4.11 - Representação gráfica da anisotropia zonal 51
Figura 4.12 – Semivariogramas em amostras diferentes 51
Figura 4.13 – Krigagem Simples 53
Figura 4.14 – Krigagem Ordinária 54
Figura 4.15 – Krigagem Universal com deriva de 2ª Ordem 55
Figura 4.16 – Diferenças entre métodos Geostatísticos e Determinísticos 56
Figura 5.1 – Shapefiles linha da costa, batimetria, enrocamento e afloramento rochoso. 58
Figura 5.2 – Shapefiles e Mapa georeferenciado. 59
Figura 5.3 – Pontos de referência para projecção 60
Figura 5.4 – Shapefiles, Mapa georeferenciado e Pontos de leitura. 61
Figura 5.5 – Shapefiles, pontos de leitura, pontos de batimetria e altimetria 62
Figura 5.6 – Histograma inicial 63
Figura 5.7 – Histograma após a introdução de pontos 65
Figura 5.8 – Deriva nos dados iniciais 65
Figura 5.9 – Deriva após a inclusão de pontos 66
Figura 5.10 – Deriva Noroeste - Sudoeste 66
Figura 5.11 – Exemplo de um Mapa de Voronoi 67
Figura 5.12 – Mapas Voronoi Simples, inicial (esquerda) e com pontos adicionais (direita) . 68
Figura 5.13 – Mapas Voronoi IQR, inicial (esquerda) e com pontos adicionais (direita) 68
Figura 5.14 – Mapas Voronoi Aglomerado, inicial (esquerda) e pontos adicionais (direita) .. 69
Figura 5.15 – Mapas Voronoi Média, inicial (esq) e com pontos adicionais (dir) 69
Figura 5.16 – Krigagem ordinária com modelo exponencial e com modelo gaussiano 71

Figura 5.17 –Modelo gaussiano à esquerda e modelo exponencial à direita 71
Figura 5.18 – Krigagem Universal com deriva de primeira ordem 72
Figura 5.19 – Krigagem Universal com deriva de segunda ordem 73
Figura 5.20 – Gráfico de Validação Cruzada 73
Figura 5.21 – Sub-amostra do Monte da Guia 79
Figura 5.22 – Histograma da sub-amostra do Monte da Guia 80
Figura 5.23 – Deriva dos dados na zona do Monte da Guia 80
Figura 5.24 – Mapas Voronoi Aglomerado (esquerda) e IQR (direita) 81
Figura 5.25 – Sub-amostra a Leste do Monte da Guia 84
Figura 5.26 – Histograma do zona leste do Monte da Guia 85
Figura 5.27 – Deriva dos dados no zona leste do Monte da Guia 85
Figura 5.28 – RBF (esquerda), Polinomial Local (centro) e KU 1ª Ordem (direita) 87
Figura 5.29 – Conjunto de pontos com igual latitude e longitude 89
Figura 6.1 – Superfície final gerada com KO e semivariograma Combinado. 98
Figura 6.2 - Superfície final gerada com KU 1ª ordem e semivariograma Combinado. 99
Figura 6.3 – Monte da Guia gerado com KO e semivariograma Combinado. 100
Figura 6.4 – Monte da Guia gerado com KU 1ª ordem e semivariograma Esférico. 100
Figura 6.5 – Resultados na zona a Leste do Monte da Guia 101
Figura 6.7 – TIN do Monte da Guia gerado com KO e semivariograma Combinado 102
Figura 6.8 – TIN gerado com KU de 1ª ordem e semivariograma Combinado 103
Figura B.1 – Semivariogramas usados com KO na área de trabalho completa. 107
Figura B.2 – Semivariogramas usados com Krigagem Universal com deriva de 1ª ordem. . 108
Figura B.3 – Semivariogramas usados com Krigagem Universal com deriva de 2ª ordem. . 109
Figura B.4 – Semivariogramas usados com KO na área do Monte da Guia. 110
Figura B.5 – Semivariogramas da KU com deriva de 1ª ordem na zona do Monte da Guia. 111
Figura B.6 – Semivariogramas da KU com deriva de 2ª ordem na zona do Monte da Guia. 112
Figura B.7 – Semivariogramas da Krigagem Ordinária da zona a Leste do Monte da Guia. 113
Figura B.8 – Semivariogramas da Krigagem Universal com deriva de 1ª ordem. 114
Figura B.9 – Semivariogramas da Krigagem Universal com deriva de 2ª ordem. 115
Figura C.1 – Derivas usadas nos modelos da amostragem completa 116
Figura C.2 – Derivas usadas nos modelos da zona do Monte da Guia 116

LISTA DE TABELAS

Tabela 3.1 - Prefixos dos comandos existentes no Grass 24
Tabela 3.2 - Número de Licenças Activas em 1994 29
Tabela 3.2 - Factores de Escolha 35
Tabela 5.1 – Pontos auxiliares de projecção 60
Tabela 5.2 – Validação Cruzada para Semivariograma Esférico 75
Tabela 5.3 - Validação Cruzada para Semivariograma Exponencial 76
Tabela 5.4 - Validação Cruzada para Semivariograma Gaussiano 76
Tabela 5.5 - Validação Cruzada para Semivariograma Combinado 77
Tabela 5.6 - Resultados mais significativos 77
Tabela 5.7 – Validação para melhores modelos 78
Tabela 5.8 – Resultados mais significativos no Monte da Guia 82
Tabela 5.9 – Validação para a zona do monte da Guia 83
Tabela 5.10 – Resultados mais significativos na Zona a Leste 85
Tabela 5.11 – Validação Cruzada para Métodos Determinísticos 86
Tabela 5.12 – Validação para a zona Leste 87
Tabela 5.13 – Tratamento da variância fornecida para amostra completa 89
Tabela 5.14 – Validação Cruzada com e sem Erro de Medição na Amostra Completa 90
Tabela 5.15 – Validação com Erro de Medição na Amostra Total 91
Tabela 5.16 - Validação Cruzada com e sem Erro de Medição na zona do Monte da Guia 91
Tabela 5.17 – Validação com Erro de Medição para o Monte da Guia 92
Tabela 5.18 - Validação Cruzada com e sem Erro de Medição na Zona Leste 93
Tabela 5.19 –Validação com Erro Medição na zona Leste 93
Tabela B.1 – Parâmetros dos Semivariogramas de KO na área de trabalho completa 107
Tabela B.2 – Parâmetros dos Semivariogramas de KU com deriva de 1º ordem. 108
Tabela B.3 - Parâmetros dos Semivariogramas de KU com deriva de 2º ordem. 109
Tabela B.4 – Parâmetros dos Semivariogramas de Krigagem Ordinária. 110
Tabela B.5 – Parâmetros dos Semivariogramas da KU com deriva de 1ª ordem. 111
Tabela B.6 – Parâmetros Semivariogramas da KU com deriva de 2ª ordem. 112
Tabela B.7 – Parâmetros dos Semivariogramas da Krigagem Ordinária. 113
Tabela B.8 – Parâmetros Semivariogramas da KU com deriva de 1ª ordem. 114
Tabela B.9 – Parâmetros dos Semivariogramas da KU com deriva de 2ª ordem. 115

CAPÍTULO I

INTRODUÇÃO

1.1 Enquadramento

Desde sempre o Homem sentiu necessidade de representar o mundo que o rodeia. As primeiras representações do mundo conhecido começaram por ser desenhadas em placas de barro que depois eram cozidas. Os mapas passaram depois a ser desenhos em suportes mais familiares (como o papiro ou o couro), que evoluíram ao longo do tempo, quer nas técnicas de desenho usadas, quer nas dimensões e formas do mundo representado.

Lentamente os mapas tornaram-se em mais do que figuras, transformaram-se em ferramentas complexas de análise. Com o aparecimento dos computadores estas ferramentas passaram a ter suporte digital, e tornando-se, deste modo, acessíveis a mais pessoas. Esta evolução culmina no aparecimento dos primeiros Sistemas de Informação Geográfica (SIG), durante a década de 1960, depois de uma série de inovações tecnológicas, quer no *Hardware*, quer no *Software* dos computadores de então.

Os SIG fazem já parte do dia a dia moderno, embora tal não seja completamente perceptível. A maioria das empresas de grande ou média dimensão têm departamentos SIG, ou requerem serviços nesta área. Os SIG são usados em muitos países como poderosas ferramentas de gestão e governação, a nível das Câmaras Municipais. A sua flexibilidade permite aos SIG tratar de qualquer tipo de informação que tenha um componente geográfico, e nos dias de hoje quase todo o tipo de informação se pode ligar a informação geográfica.

O desenvolvimento dos SIG apresenta ainda grandes desafios às Ciências da Informática, não só na criação de novas ferramentas que beneficiem de tecnologias mais recentes, como na sua integração apropriada nos domínios de análise. Este é uma área em franco desenvolvimento e expansão, apresentando boa razões de investigação.

1.2 Objectivos

O Objectivo deste Trabalho Final de Curso é desenvolver um Sistema de Informação Geográfica, de modo a que este represente a informação tridimensional obtida, durante as

operações de diversos veículos robóticos autónomos. Esta informação será processada para mostrar a sua posterior utilização no sistema de controlo de missão dos veículos.

Com base num sistema SIG disponível comercialmente, assente numa base de dados pretende-se estudar as características dos dados disponíveis, e estudar a melhor forma de os representar.

1.3 Estrutura

Este documento encontra-se organizado em seis capítulos, o primeiro dos quais é constituído pela presente introdução. Aqui se apresenta o enquadramento e a motivação que deram origem a este trabalho, são apresentados os objectivos a atingir, é delineada a estrutura do relatório e enunciam-se os resultados do trabalho desenvolvido.

O segundo capítulo introduz conceitos básicos de SIG, características bases e funções deste tipo de sistemas. São aqui delimitados os diferentes tipos de SIG existentes no mercado, e apresentadas as diferentes orientações de cada um deles.

No terceiro capítulo são analisados alguns sistemas presentemente no mercado, tendo em vista os objectivos finais do trabalho. É igualmente analisada a estrutura passada e presente do mercado dos SIG. Por fim são comparados os diferentes SIG estudados e efectuada uma escolha perante certos critérios, de acordo com os objectivos a atingir.

O quarto capítulo apresenta as diferentes técnicas de interpolação de superfícies, a partir de um conjunto finito de amostras. São apresentados os métodos Determinísticos e Geoestatísticos, e o modelo matemático que permite efectuar a análise do caso de estudo.

No quinto capítulo, são postos em prática os modelos estudados no capítulo pretérito. São seleccionadas três zonas de análise na área de estudo, para as quais se procede à análise exploratória espacial. São então aplicados os métodos de interpolação, e filtrados os resultados usando, primeiramente a Validação Cruzada, e por última a Validação.

No último e sexto capítulo são discutidos os resultados obtidos, perante as opções tomadas. São igualmente apresentados alguns dos resultados finais.

1.4 Resultados

Este trabalho levou à obtenção dos seguintes resultados:

- ?? Ficheiros em formato TIN (*Triangulated Irregular Network*) das superfícies finas geradas
- ?? Camadas SIG geoestatísticas das superfícies geradas (formato Arc/Info - *layer*)
- ?? Conjuntos de teste e de treino usados na Validação, bem como as respectivas camadas vector geradas (formato Arc/Info - *shapefile*)

CAPÍTULO II

SISTEMAS DE INFORMAÇÃO GEOGRÁFICA

2.1 Introdução

Nos últimos anos diversas organizações têm vindo a investir grandes recursos no desenvolvimento de bases de dados referenciados geograficamente, e em Sistemas de Informação Geográfica. É previsível que durante os próximos anos se invistam vários milhões de Euros nesta área. Tudo isto está a acontecer num período de tempo relativamente curto, pois não há muito tempo os SIG eram ferramentas muito especializadas, só nas mãos de algumas, poucas organizações, sendo uma mera curiosidade para o público em geral. Podem ser dadas duas explicações a esta rápida expansão dos SIG:

- a primeira reside na queda dos custos dos equipamentos informáticos necessários, que cada dia os faz mais acessíveis para um maior número de utilizadores;
- a segunda, e de maior importância, é de que a geografia (e os dados que servem para a quantificar) já fazem parte do nosso mundo quotidiano; a maioria das decisões que tomamos diariamente estão em relação com, ou influenciadas por, um facto geográfico. Os auto-tanques de bombeiros, por exemplo, são enviados ao seu destino através da rota mais curta possível, as atribuições económicas dos governos aos entes locais baseiam-se frequentemente na distribuição geográfica da população, ou as doenças epidémicas que são estudadas graças à identificação das áreas onde se produzem e da velocidade a que se expandem.

2.2 O que é um SIG?

O uso dos Sistemas de Informação Geográfica aumentou bastante nas décadas de oitenta e noventa; passaram do total desconhecimento à prática quotidiana no mundo dos negócios, nas universidades e nos organismos governamentais, sendo usados para resolver problemas diversos. É lógico, portanto, que foram propostas várias definições.[20]

2.2.1 Definição Genérica de SIG

Sistema: um *framework* para manipular, pesquisar, analisar e disseminar informação.

Informação: dados processados de uma forma utilizável.

Geográfico: dados referenciados espacialmente.

Um SIG é uma base de dados digital no qual uma comum coordenada espacial é o principal meio de referência. Os SIG requerem[1]:

. Entradas de dados provenientes de mapas, fotos aéreas, satélites e outras fontes;

. Armazenamento e pesquisa de dados;

. Transformação de dados, análise e modelação, incluindo estatísticas espaciais;

. Saídas de dados, tal como mapas, relatórios e planos.

Três observações acerca desta definição:

Primeira: os SIG estão relacionados com outras aplicações de base de dados, mas com uma diferença importante. As bases de dados dos SIG usam referências geográficas como o principal meio de guardar e aceder a informação.

Segunda: os SIG integram tecnologia. Onde outras tecnologias podem ser usadas apenas para analisar fotos aéreas e imagens de satélite, de modo a criarem modelos estatísticos, ou para desenhar mapas, estas capacidades são conjuntamente oferecidas com um SIG.

Terceira: um SIG com o seu conjunto de funções, deve ser visto como um processo e não meramente como *software* ou *hardware*. Os SIG servem para tomar decisões. O modo como os dados são armazenados e analisados dentro de um SIG deve reflectir a maneira como a informação irá ser usada numa tarefa específica ou na tomada de uma decisão[1].

2.2.2 Outras definições de SIG

- Um sistema para capturar, armazenar, comprovar, integrar, manipular, analisar e visualizar dados que estão espacialmente georeferenciados.[2].

- Sistemas automatizados para a captura, armazenagem, composição, análise e visualização de dados espaciais[3].

- Um sistema de hardware, software e procedimentos desenhados para suportar a captura, gestão, manipulação, análise, modelado e visualização de dados espacialmente referenciados para resolver problemas complexos de planeamento e gestão[4].

Uma das razões porque é tão difícil existir acordo numa única definição de SIG é por existirem diversas vertentes em que podem ser aplicados, cada uma para diferentes propósitos e para diferentes tipos de tomada de decisões.

*AM/FM (Automated Mapping and Facilities Management)

AM/FM não são geralmente utilizados para análise espacial, e faltam-lhes a topologia das estruturas de dados dos SIG.

* CAD (Computer-Aided Design)

Sistemas projectados para desenho. Tratam dados espaciais como gráficos em vez de os tratarem como informação.

* LIS (Land Information System)

* *Multipurpose Cadastre*: Sistemas que guardam informação sobre parcelas de terra.

De um ponto de vista prático um Sistema de Informação Geográfica é um sistema informático capaz de realizar uma gestão completa de dados geográficos referenciados. Por referenciados entende-se que estes dados geográficos ou mapas têm coordenadas geográficas reais associadas, permitindo-nos estas manejar e fazer análises com dados reais como longitudes, perímetros ou áreas. Todos estes dados alfanuméricos associados aos mapas em conjunto com os que queiramos acrescentar são geridos por uma base de dados integrada no SIG.

O que distingue os SIG de outros tipos de sistemas de informação, tal como bases de dados e *spreadsheets*, é que os SIG lidam com informação espacial. Os SIG têm capacidade de relacionar diferentes camadas de informação (*layers*) para o mesmo ponto, combinando, analisando e, finalmente apresentar os resultados. A informação espacial usa a localização, num sistema de coordenadas, como a sua referência base.

2.3 Representação da Informação num SIG Vectores/*Raster*

Os dados nos SIG usualmente são guardados em mais do que uma camada de informação denominadas *layers*, de modo a que os problemas técnicos causados por se tratarem grandes volumes de informação em simultâneo sejam atenuados. É mais fácil trabalhar em problemas de análise espacial complexos com um *layer* de cada vez, do que ter que manipular todo um volume de informação redundante. O que os *layers* permitem é separar todo esse volume de

informação em camadas. Este é um aspecto importante dos SIG, que é conhecido como integração de dados.

Qualquer actividade económica ou social com uma dimensão espacial não pode ser correctamente entendida sem as suas qualidades espaciais serem referenciadas. A informação espacial tem duas partes essenciais: localização e propriedades (atributos).

Na localização além da latitude e longitude podem ser usados outros tipos de convenções.

Os atributos servem, como o próprio nome indica, para descreverem propriedades ou características de pontos geográficos. Esses atributos normalmente são guardados em tabelas.

Os SIG guardam e processam dados em dois formatos, vectorial e *raster* (este é um termo da língua inglesa, infelizmente não existe qualquer equivalente em português).

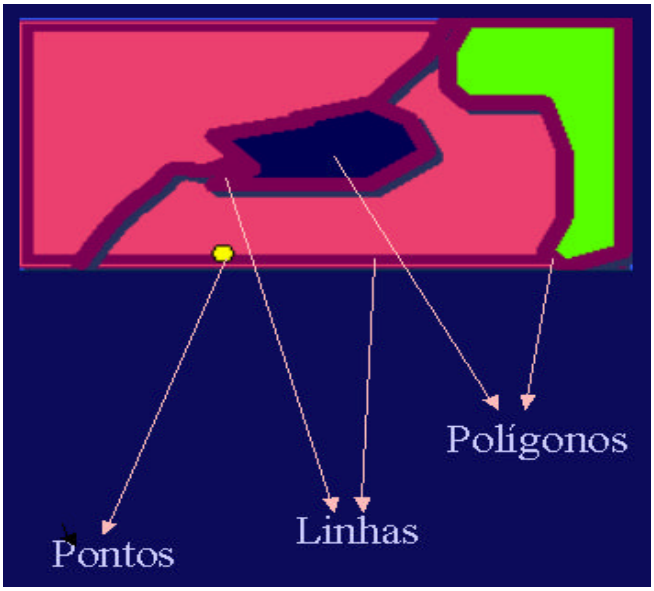

Figura 2.1 – Representação Vectorial.

No modelo vectorial, o mundo é simulado como um mosaico de linhas e pontos interligados representando a localização e fronteiras de entidades geográficas. Os dados são representados por arcos, polígonos, pontos e nós. A principal vantagem do formato de dados vectorial é que permite uma representação precisa, o que é muito útil em actividades tais como engenharia. Também é possível nos SIG baseados no modelo vectorial definir relações espaciais de conectividade e adjacência. Contudo, a maior desvantagem de dados vectoriais é que as fronteiras dos mapas de polígonos resultantes são discretas, onde na realidade os mapas de polígonos podem representar alterações graduais tal como mapas de solo.

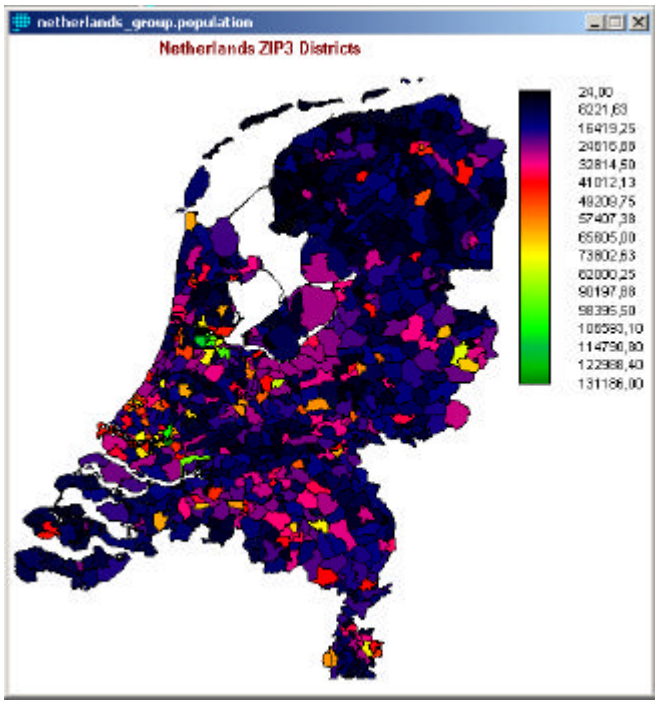

Figura 2.2 – Mapa de Densidade Populacional na Holanda, representado por vectores.

Ao invés, nos modelos raster (mapas de pixels), a resolução depende do tamanho do pixel, logo a resolução pode ir ate onde se desejar. Este modelo torna-se mais adequado quando o detalhe é essencial.

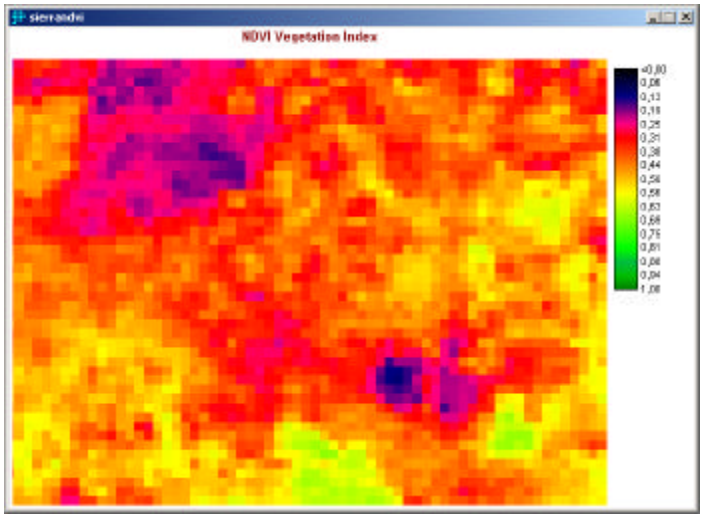

Figura 2.3 – Pormenor de uma camada *raster* representando o relevo da Serra de Gredos

Cada camada *raster* é na realidade uma matriz bidimensional, em que para cada célula existe um valor único que corresponde a uma única característica representada. Esta é a representação tradicional de imagens em formato digital, pois de facto uma imagem digital é sempre representada por um conjunto de pixels. Este formato é o mais usado para representar característica como a elevação do terreno.

A principal desvantagem das camadas *raster* em relação às camadas vectoriais é o facto de ocuparem muito mais espaço. Se por exemplo quisermos representar uma estrada teremos que ter um ficheiro que indique para uma dada zona quais os *pixels* que fazem parte da estrada e quais não fazem, isto equivale a uma tabela. Essa mesma estrada pode ser representado por uma simples linha num modelo vectorial, o que numa base de dados equivale unicamente a uma linha de uma tabela. Podemos ainda referir que para uma camada *raster* necessitamos de uma tabela para representar uma única característica geográfica, ao passo que numa camada vectorial, uma tabela pode conter dados a cerca de inúmeras características.

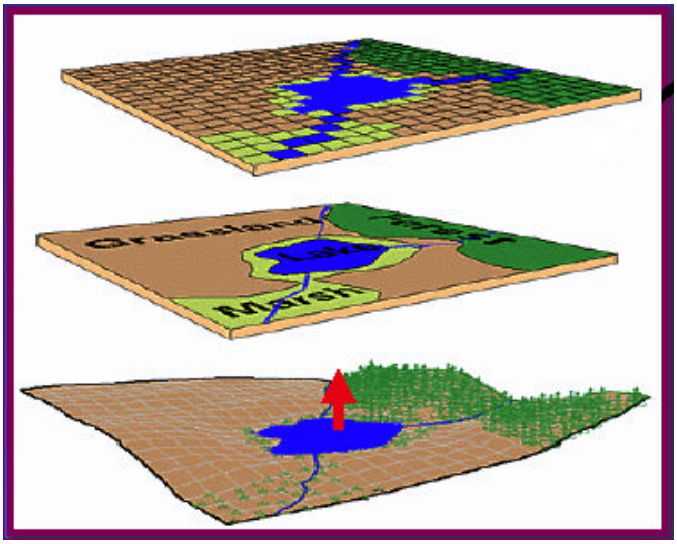

Figura 2.4 – Diferenças entre as duas formas de representação do mundo num SIG

2.4 Tipos de Sistemas de Informação Geográfica

O termo SIG abrange sistemas muito variados, aplicando-se muitas vezes a aplicações que não são propriamente um SIG. Podem-se distinguir três tipos de programas que, embora possam denominar-se conjuntamente SIG têm diferenças fundamentais no seu âmbito de aplicação. Em primeiro lugar distinguiremos um SIG propriamente dito, como um grande sistema informático que gere completamente uma base de dados geográficos. Por outro lado delimitaremos as aplicações, às quais foi dado o nome de *Desktop Mappers* (DM) - sistemas de análise e visualização integrados entre as aplicações Desktop de computador pessoal. Finalmente distinguiremos os sistemas de Desenho assistido por Computador (CAD) e sistemas semelhantes.

2.4.1 Sistemas de Informação Geográfica

Um SIG em sentido completo gere uma base de dados espacial. Permite a criação e estruturação dos dados partindo de fontes de informação como os mapas, a digitalização, bases de dados existentes, etc. Para além de possibilitar a análise, visualização e edição em mapas da base de dados, um SIG conta com ferramentas que permitem criar novos dados derivados dos existentes.

2.4.2 Funcionalidade do SIG

- Construir dados geográficos: Mediante dados geométricos existentes em CAD, ou capturando-os por digitalização, vectorização de imagens, GPS, etc. O sistema permite depurá-los e estruturá-los topologicamente, associando-os com bases de dados alfanuméricas. Desta forma obter-se-ão dados espaciais preparados para o seu uso na análise.

- Modelação cartográfica: Criação de novos mapas a partir de mapas existentes: Combinando atributos do terreno como pendentes, vegetação, tipo de solo, etc. Mediante um modelo matemático, podem-se criar novas variáveis, como um índice de erosão, de perigo de incêndios, etc.

- O SIG permite-nos analisar os mapas estruturados em combinação com bases de dados associadas. Podem ser interrogados para seleccionar os dados de interesse, ver os resultados interactivamente escolhendo a simbologia em função dos atributos associados e produzir cartografia de qualidade.

Também se podem preparar aplicações à medida, como um plano de controlo de incêndios, de avaliação de impactos ambientais, um modelo que preveja a evolução de um incêndio ou de um alagamento, aplicações verticais, como um sistema de gestão municipal ou uma aplicação para uma central eléctrica, etc.

2.4.3 *Desktop Mapper*

Recentemente têm aparecido aplicações simples de visualização e análise de dados com componente espacial para sistemas micro-computadores. Como é de esperar estes sistemas apresentam um custo inferior a um SIG completo. A denominação habitual destas aplicações em inglês é *Desktop Mapper* (DM), que significa algo como "Cartografia de escritório". Em primeiro lugar, estas ferramentas permitem o uso de dados espaciais (possivelmente criados e estruturados com um SIG) por parte de utilizadores que não são peritos em programação, cartografia, geologia, etc., da mesma forma como estes mesmos utilizadores empregam

processadores de textos, folhas de cálculo ou bases de dados, sem quaisquer conhecimentos de topografia, manutenção, teoria de computação, etc. Fundamentalmente os DM permitem ver e analisar a estrutura espacial dos dados. Por outro lado, estas aplicações servem também de veículo para a criação de aplicações concretas que trabalhem com dados espaciais com por exemplo uma aplicação que gira um inventário da fauna de uma certa zona protegida ou uma aplicação vertical para um banco.

Grande parte da capacidade de adequação destes produtos deriva das possibilidades de integração com outras aplicações através dos mecanismos fornecidos por ambientes de trabalho como Microsoft Windows, Mac OS, OS/2 PM, Windows NT e X Windows. Desta maneira, e conjuntamente com aplicações específicas, podem-se criar contornos de trabalho verticais que resolvam grande parte dos problemas de um determinado profissional, integrando-se a análise e visualização espacial com a preparação de documentos, modelos de cálculo, etc. Da mesma maneira, podem-se incorporar dados não geridos directamente pelo DM, como por exemplo, imagem de vídeo ou fotografias. Este tipo de aplicação tem um mercado potencial muito mais amplo que um SIG tradicional, pelas mesmas razões que o têm os processadores de texto, folhas de cálculo ou bases de dados. Permitem criar um modelo geográfico do funcionamento dum negócio ou organização. O facto de que grande parte das bases de dados existentes usem dados geográficos - estima-se que mais do 85 % contenham um componente geográfico - permite que o DM mostre padrões, relações e tendências que de outra maneira seriam dificilmente detectáveis.

Actualmente os DM usam-se em: departamentos de marketing, vendas, distribuição e retalho, telecomunicações, propriedade imobiliária, planeamento, seguros, serviços de urgência (bombeiros, polícia), saúde, e administração local. Um requisito básico para estas aplicações serem exploradas é que junto com a aplicação se administrem os dados espaciais básicos referidos à área de interesse do cliente, de forma já estruturada. (Mapas de municípios, infra-estruturas, dados topográficos, urbanos...). Assim vêm a ser ferramentas de exploração dos dados gerados por um SIG. Normalmente serão utilizadores de SIG os que disponibilizarão os dados e as aplicações para um sistema DM.

No que diz a respeito ao uso de um DM como complemento de um SIG, há que distinguir que algumas ferramentas DM trabalham directamente com as bases de dados dum SIG determinado, enquanto que outras requerem um formato próprio. No segundo caso, se se quer utilizar o DM como visualizador dum SIG determinado deve-se estabelecer primeiro um mecanismo que adapte os dados do SIG aos requerimentos e estruturação do DM.

2.4.4 Limitações do DM

- Não pode criar novos mapas por digitalização, vectorização, uso de GPS, ou incorporação de dados geométricos não estruturados, pois habitualmente não se conta com a capacidade de depuração dos dados, criação de topologia e transformação para a sua correcta localização espacial.

- Também não se poderão criar novos temas combinando dados existentes por análise de sobreposição. Não se conta com funções avançadas de manipulação da topologia nem de modelação cartográfica.

- Não se podem gerir dados tridimensionais. Isto exclui a criação de vistas perspectivas, análise de visibilidade, etc. (Embora se possam criar aplicações que implementem alguma funcionalidade deste tipo, como criar curvas de nível, aplicar modelos de iluminação, ...)

- Outra área na qual um DM é insuficiente é no tratamento de dados de detecção remota, que requer capacidade de processo de grandes quantidades de dados raster.

- Também não se podem esperar de um DM as capacidades mais avançadas dos SIG modernos, como "meta dados" de qualidade, controlo de concorrência dinâmico ou variabilidade temporal dos dados.

2.4.5 CAD *Mapping Systems*

Alguns vendedores de sistemas CAD (Computer Aided Design - programas de delineação e desenho) consideraram poder evoluir facilmente um destes sistemas, e converte-lo num SIG. Os sistemas que aparecem desta forma denominam-se CAD *Mapping Systems* (CMS), e habitualmente o vendedor chama-os SIG. Normalmente estes sistemas são o resultado da combinação de dois sistemas existentes no mercado, um CAD e um RDBMS (Relacional DataBase Management System - Sistema de gestão de bases de dados relacional). Desta forma os dados alfanuméricos contidos nas bases de dados associam-se com elementos gráficos dum arquivo - desenho de CAD. A verdade é que um CAD apresenta algumas características que dificultam o desenvolvimento dum SIG. Um SIG não é a soma de um CAD e um RDBMS; mais exactamente, o RDBMS encontra-se no coração do sistema, e as capacidades gráficas, subtilmente diferentes das do CAD, encontram-se intimamente ligadas ao RDBMS.

2.4.6 Diferenças SIG-CAD

Algumas das diferenças fundamentais entre a parte gráfica dum SIG e um CAD são:

- Propósitos diferentes: O SIG reflecte a realidade. O CAD desenha algo que ainda não existe.

- Ambos têm um estrato geométrico, mas a criação destes elementos é distinta: em CAD são criados por um delineador, com exactidão, em SIG são obtidos a partir de mapas ou do terreno, com certo erro e imprecisão inevitável, mas mensurável.

- O CAD segmenta os dados em arquivos independentes que não partilham um espaço de coordenadas global. Em SIG os dados formam um conjunto contínuo. (Isto implica diferentes formas de acesso e diferentes problemas de concorrência).

- A informação num SIG é pelo menos uma ordem de magnitude maior, os objectos são muito mais complexos. (Estima-se que a infra-estrutura municipal tipo requeira entre meio e um Gigabyte por cada cem mil habitantes).

- Os CAD habitualmente permitem a ligação com uma base de dados, mas não permitem uma integração suficiente, como por exemplo para responder a perguntas que combinem critérios alfanuméricos e espaciais. (Não contam com uma linguagem de consulta alfanumérico/espacial, nem a possibilidade de análise de sobreposição, e normalmente o conceito de topologia é muito pobre)

- Há alguns tipos de dados característicos do SIG que um CAD não gere: dados raster referenciados geograficamente e com atributos, como os de Detecção Remota, ou Modelos Digitais do Terreno, que não se podem implementar eficazmente com um programa de criação de sólidos.

- Um CAD separa as entidades geométricas em camadas ou níveis. Num SIG tal partição não deve existir, pois complica a manutenção dos dados.

2.2.7 Resumo

É importante conhecer as diferenças entre SIG e DM, porque é corrente cometer o erro - especialmente se se conta com recursos largos - de adquirir um sistema SIG quando o que se necessita realmente é um DM. O SIG requererá um enorme investimento em equipas, pessoal e treino, e não proporcionará a agilidade de visualização e análise que dá um DM nem a facilidade de uso duma aplicação vertical.

Deve ficar claro que SIG e DM não competem por um mesmo mercado, porque têm âmbitos distintos que se complementam e podem coexistir na mesma organização.

Por outro lado há que distinguir os sistemas CAD; de propósito muito diferente aos SIG e com limitações insolúveis para o seu uso como SIG. CAD e SIG podem coexistir numa organização, mas normalmente fá-lo-ão em departamentos diferentes. Por exemplo, um departamento de planeamento usará SIG para determinar a localização dum novo hospital, enquanto que o departamento de arquitectura usará CAD para desenhá-lo.

No que diz respeito aos chamados SIG, que se baseiam num sistema CAD e um DBMS, devem ser examinados com cuidado antes de ser adoptados. Possivelmente só são úteis para utilizadores com experiência e grandes quantidades de dados em CAD, e só para actividades muito concretas como a engenharia civil.

2.5 Para que serve um SIG?

Até agora descreveu-se um SIG por meio de definições formais e por meio da sua capacidade para satisfazer questões de cariz espacial, relacionando conjuntos de dados através da sua localização geográfica. Um SIG pode igualmente ser descrito enumerando o tipo de questões às quais pode responder. Um SIG suficientemente sofisticado pode responder a cinco perguntas genéricas:

Localização: O que há em...?

A primeira das perguntas, procura identificar o que se encontra numa localização determinada. A localização pode ser descrita de vários modos, por exemplo, pelo seu topónimo, pelo seu código postal, ou por referências geográficas como latitude e longitude.

Condição: Onde se encontra?

A segunda pergunta é a inversa da primeira e requer uma análise espacial. Em lugar de identificar o que se encontra num ponto o que se procura é um lugar que reuna certas condições (por exemplo, um terreno sem floresta, que tenha uma área maior que 2000 metros quadrados, que esteja a menos de 100 metros de uma estrada e que as suas condições geológicas lhe permitam suportar edifícios).

Tendência: O que mudou desde...?

Esta pergunta envolve as duas anteriores e a sua resposta estabelece quais são as diferenças que ocorrem numa área determinada através do tempo.

Distribuição: Que padrões de distribuição espacial existem?

Esta pergunta é mais complexa. Aparece ao querer determinar, por exemplo, se o cancro é uma causa importante de mortalidade entre as pessoas que residem nas proximidades de uma central nuclear. Ou também, ao querer conhecer quantas situações anormais se produzem numa determinada distribuição espacial e onde se localizam.

Modelização: Que sucede se...?

Questão que se traz à baila ao tentar conhecer o que se passa num sistema quando ocorre um facto determinado. Por exemplo, que ocorre a um sistema rodoviário se construirmos uma nova estrada, ou que sucederia se se produzisse uma determinada planta tóxica na rede de abastecimento de água potável. As respostas requerem, para além da informação geográfica, outras informações adicionais, como por exemplo, determinadas leis científicas.

CAPÍTULO III

ESTUDO DE MERCADO

3.1 Introdução

Ao se iniciar esta fase do trabalho algumas ideias estavam já esclarecidas. Ficou claro durante a fase anterior, que neste caso não seria suficiente um simples DM. No âmbito do TFC pretende-se criar mapas, armazenar e gerir dados disponíveis.

Vários foram os produtos com que se tomou contacto, visto o mercado ser relativamente alargado. Quatro foram os analisados mais profundamente através de versões de demonstração, o Idrisi 32, o ArcGIS(com uma versão de avaliação do ArcView), o Grass e o Geomatica. Ficaram no entanto alguns por analisar, dos quais destacamos o Manifold GIS, que apenas existe em versão comercial, visto não haver versão de demonstração; e o MGE da Intergraph, que apesar dos múltiplos pedidos de versões de demonstração, nunca foram recebidos. Dos restantes sistemas considerados estão, entre outros o Smallworlds da General Electric, e o Autodesk GIS.

3.2 IDRISI 32

O Idrisi é produzido pelos Clark Labs da Clark University desde 1989, sendo um dos mais antigos sistemas no mercado. Começa por ser um Desktop Mapper, mas apresenta igualmente muitas das funcionalidades de um SIG propriamente dito, permitindo fazer alguma análise topológica de considerável complexidade. O Idrisi apresenta igualmente uma faceta virada para o processamento de imagem, em especial, no tratamento e melhoramento de imagens raster. De resto este sistema privilegia claramente o trabalho com fotografia e camadas raster. Foi também pensado para ser um SIG cooperante com novas tecnologias em especial a Percepção Remota. Podemos chamar ao Idrisi um dois em um, pois tanto o podemos usar como um simples Desktop Mapper, ou como um SIG.

Como interface o Idrisi é bastante fraco, apresentando alguns problemas de usabilidade sérios, como o excesso de ícones e menus. A maior parte da interacção é feita através de menus e preenchimento de valores em tabelas, sendo este um modo de interacção muito lento e pouco intuitivo comparado com os métodos de *Apontar e Clicar*. De facto esta interface

estará alguns anos atrás das actuais encontradas nos ambientes Microsoft, fazendo lembrar algumas aplicações do Windows 3 e precedentes.

No capítulo da interacção é de realçar o sistema de programação por filtros disponibilizado pelo Idrisi, ao qual é dado o nome de *Macro Modeler*. Esta aplicação permite construir programas de análise topológica graficamente. As saídas e entradas são camadas, normalmente raster, os filtros são algoritmos de análise e classificação topológica. Em relação aos algoritmos propriamente ditos, estes são relativamente poucos e de reduzido poder analítico, sendo a maioria de classificação binária. Para operações mais complexas é necessário combinar vários algoritmos, ou então dar-lhes maior quantidade de informação. No final pode-se acabar com modelos muito complexos, com os mesmos ficheiros de camadas a serem entradas de muitos filtros, acrescentando ainda ficheiros auxiliares de cálculo.

O Idrisi é completamente aberto ao desenvolvimento. Usando a tecnologia COM de Invocação de Procedimentos Remotos é possível construir aplicações que interajam com o Idrisi. Segundo a Clark Labs o Idrisi é totalmente condescendente com esta tecnologia, sendo assim possível desenvolver todas as suas actuais faculdades.

Falando agora da portabilidade, o Idrisi encontra-se disponível para quase todas as plataformas Microsoft. Dado ser produzido há mais de dez anos, o Idrisi foi evoluindo ligado às diferentes plataformas da Microsoft que apareceram desde então. Para além das versões actuais do sistema operativo Windows, o Idrisi corre também nas versões pré sistema operativo desta aplicação, e mesmo em versões antigas do MS-DOS, indo até à versão 5.0 e anteriores.

Já aqui foi dito que o Idrisi não é apenas um *Desktop Mapper*, e que também apresenta características de um SIG real. Mas então porque não é este sistema, um dos grandes SIG do mercado? Porque está o seu preço relativamente longe dos sistemas de grande envergadura?

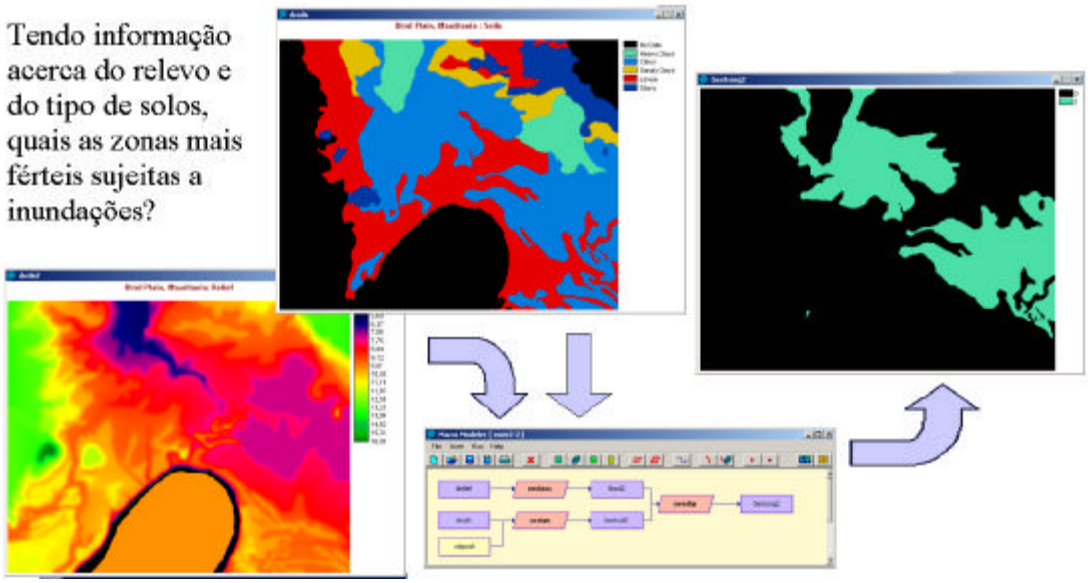

Figura 3.1 – Exemplo de Análise com Programação por Filtros

Como primeira grande desvantagem deste sistema, está a ausência de paralelismo, o Idrisi não tem qualquer funcionalidade multi-utilizador. Todo o acesso às Bases de Dados e Mapas é feito por apenas uma aplicação não havendo qualquer hipótese de edição simultânea. Esta falha é preponderante na incapacidade do Idrisi ser o sistema para uma grande equipa SIG.

Na mesma linha aparece a ausência de uma linguagem de desenvolvimento própria. Apesar de ser 100% condescendente com a tecnologia COM, uma linguagem de desenvolvimento é uma das ferramentas indispensáveis a um SIG de grande envergadura, para uma grande equipa. O desenvolvimento por Invocação de Procedimentos Remotos tem duas desvantagens. Primeiro é necessário haver algum conhecimento da arquitectura do sistema em causa. Segundo é necessário usar um ambiente diverso para o desenvolvimento, o que sem dúvida o complica, em relação à programação no ambiente próprio. Há ainda que referir que a tecnologia COM, está ela própria a cair em desuso, havendo outras tecnologias a serem usadas maioritariamente.

Por fim há que referir a total ausência de análise a Três Dimensões (3D). Esta pode-se tornar na falha principal. No mercado esta é uma área onde o desenvolvimento dos SIG está mais aceso, havendo já extensões 3D para os principais sistemas. O Idrisi permite visualizar camadas raster de relevo a três dimensões, mas mais nada para além disto, o que se revela algo relativamente inútil. Esta barreira parece ser bastante difícil de transpor, pois a interface do Idrisi não está de modo algum virada para este tipo de funcionalidades. A dificuldade de apresentar representações 3D, aliada a uma interface com o utilizador ineficiente é provavelmente a maior deficiência do Idrisi.

3.2.1 CARTA LINX

Trata-se da única extensão disponibilizada pela Clark Labs para o Idrisi. Tem como principal função a elaboração de mapas electrónicos, no entanto é apresentado como sendo igualmente capaz de desenvolver Bases de Dados e de editar a topologia de um mapa. Concentra algumas das funções núcleo de mais baixo nível de um SIG propriamente dito. Infelizmente esta extensão não está disponível em versão de demonstração, pelo que não foi experimentada.

No capítulo da criação de mapas esta ferramenta parece ser extremamente eficaz. É capaz de capturar a informação a partir de várias fontes, como mesas digitalizadoras, GPS, e introdução de Coordenadas Geográficas. Através de GPS é possível registar a topologia terrestre viajando através do terreno em estudo e marcando as coordenadas GPS e a respectiva altura em pontos escolhidos. Tendo um conjunto de pontos, é aplicado um algoritmo estatístico e criado um mapa *raster* da elevação do terreno.

3.2.2 Conclusões

O Idrisi pretende ser um SIG de baixo custo, oferecendo bastante funcionalidade a um preço reduzido. O seu preço base é de 600 dólares, fora manuais. O Carta Linx custa 400 dólares, mas é oferecido um desconto de 50% a utilizadores do Idrisi, não incluindo manuais.

Se à primeira vista este parece ser um bom negócio, olhando para este produto e para os seus concorrentes do mercado verifica-se que tal não é assim. De facto, produtos como o Manifold GIS apresentam características muito semelhantes por menos de metade do preço. Se for tido em conta o facto de que em termos de interface e interacção com utilizador o Idrisi é quase de certeza o pior de todos a escolha dificilmente pode recair sobre o Idrisi.

A esta altura da análise de mercado, para além de ter ficado patente que o Idrisi não seria uma opção, começou a ficar igualmente claro que os sistemas de menor envergadura estão demasiado inclinados para a análise terrestre. Fica também um apontamento para o facto de para esse tipo de sistemas ser bastante difícil encontrar versões de demonstração, a excepção foi mesmo o Idrisi.

3.3 ARCGIS

O ArcGis é um sistema de informação geográfico integrado, composto por três partes: ArcGis Desktop Software, ArcSDE Gateway e ArcIMS Server. O ArcGis Desktop é um termo que remete para um conjunto integrado de poderosas aplicações SIG. As aplicações oferecidas pela ESRI são: O ArcView, o ArcEditor e o ArcInfo e as Extensões ArcGis. O ArcSDE Gateway é uma interface para gerir bases de dados geográficas multi-utilizador enquanto que o ArcIMS Server é um SIG baseado na *internet* para permitir que esses dados e serviços sejam distribuídos.

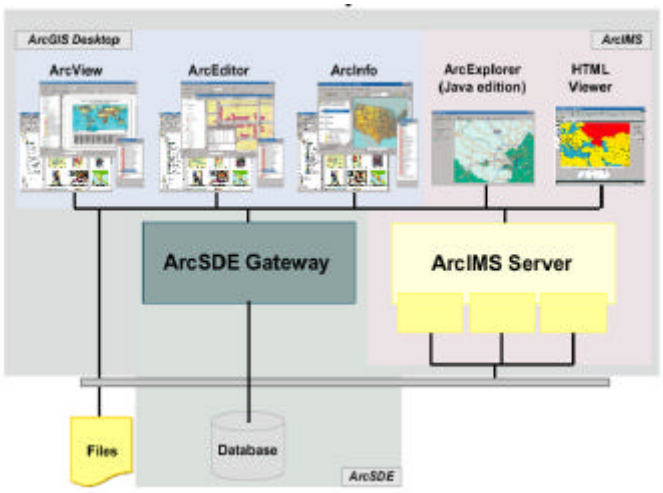

Figura 3.2 – Sistema ArcGis

O ArcGis fornece todo um esqueleto que permite implementar um SIG tanto para um simples utilizador como para muitos utilizadores. É um sistema escalável que se pode tornar num sistema SIG de qualquer tamanho, e que pode assentar sobre um único computador ou sobre uma rede.

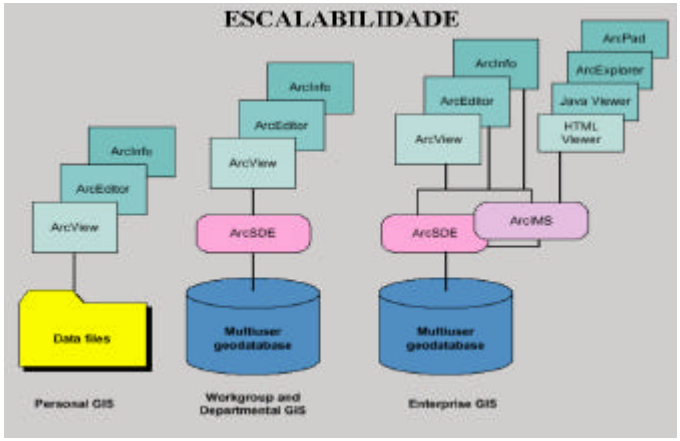

Figura 3.3 – Escalabilidade do sistema ArcGis

A opção para um Desktop SIG pode passar pelo ArcView, ArcEditor ou pelo ArcInfo visto que ambos partilham o mesmo "core" de aplicações. As aplicações comuns são: ArcMap, ArcCatalog, ArcToolbox.

A grande vantagem de os pacotes partilharem o mesmo core de aplicações é que as diferenças entre os pacotes podem ser vistas como incrementos de funcionalidades . Se for escolhido o ArcView e se houver necessidade no futuro de mudar para o ArcInfo será muito mais fácil pois os pacotes partilham a mesma interface e os mesmos conceitos operacionais. Consequentemente haverá um aumento de usabilidade.

3.3.1 Aplicações do ArcGIS Desktop

O ArcMap é a aplicação central de qualquer um dos pacotes do ArcGis Desktop e é usado para efectuar todas as principais tarefas como cartografia, análise e edição de mapas. O ArcCatalog permite organizar e gerir toda a informação num sistema SIG. Inclui ferramentas para pesquisar e encontrar informação geográfica, gravar e ver meta data, visualizar rapidamente qualquer tipo de dados, etc.

O ArcToolbox como o próprio nome indica é uma caixa de ferramentas essencialmente para conversão e gestão de dados. Existem duas versões desta aplicação. Uma vem com o ArcView/ArcEditor e outra, mais completa, com o ArcInfo.

3.3.2 Extensões do ArcGIS

Antes do aparecimento das extensões um utilizador que pretendesse efectuar uma análise raster teria que licenciar diferentes tipos de software consoante o pacote SIG que possuísse. Se possuísse o ArcView licenciava o ArcView Spatial Analyst mas se possuísse o ArcInfo licenciava o ArcGrid. O objectivo das extensões foi incorporar as melhores características de versões anteriores das extensões do ArcView e ArcInfo e torna-las comuns a todos os pacotes. Com as extensões, são reduzidos os custos de aprendizagem, operacionalidade e de aquisição.

As extensões existentes são: **ArcGis Spatial Analyst**, **ArcGis 3D Analyst**, **ArcGis Geostatistical Analyst**, **ArcGis StreetMap**, **ArcPress for ArcGis**, **MrSID Encoder for ArcGis**, **TIFF/LZW Compression**.

O **ArcGis Spatial Analyst** permite obter informação acerca dos dados, identificar relações espaciais, encontrar localizações apropriadas, calcular o custo acumulado de viajar de um ponto até outro.

Figura 3.4 – Horizontalidade das Extensões

O **ArcGis 3D Analyst** permite visualizar uma superfície de vários pontos de vista, determinar o que é visível de um determinado ponto, etc.

O poder do **ArcGis Geostatistical Analyst** reside na capacidade para criar superfícies contínuas através da medição dispersa de pontos. Possui ferramentas para análise de erro, comparação de modelos, etc.

O **ArcGis StreetMap** é um parente do software para a visualização de ruas.

O **ArcPress** permite obter uma melhoria na cor e impressão de documentos e é utilizado no sector da Imprensa.

O **MrSID Encoder** produz imagens MrSID de outras imagens até 500 MB.

Por último a extensão **TIFF/LZW Compression** permite usar compressão TIFF/LZW tecnologia esta que é patenteada pela UNISYS.

Chegados a este ponto uma das questões relevantes que se levanta é: **quais são as vantagens dos produtos ArcGis em relação a outros SIG existentes no mercado?**

Uma das vantagens é que a sua escalabilidade é comparável à da Microsoft, apesar de os seus produtos serem licenciados separadamente, quer isto dizer que para passar, por exemplo, do ArcView para o ArcInfo só se tem que proceder a um *upgrade*. A sua usabilidade é também comparável aos produtos da Microsoft pois permite que utilizadores pouco treinados adquiram uma rápida aprendizagem.

A família ArcGis foi construída usando um esqueleto COM (Component Object Model) isto torna possível o desenvolvimento de aplicações com os ArcObjects utilizando linguagens como o VB e C++. No entanto a principal razão é o baixo custo para Investigação e Ensino.

No momento de fazer escolhas dentro da família de produtos da ESRI outra questão que se coloca é: Porquê o ArcInfo e não o ArcView?

Esta questão fica automaticamente respondida se tivermos em conta que a aplicação ArcToolbox traz mais 170 ferramentas, o ArcInfo permite vectorizar imagens raster

(vectorização) e a diferença de preços entre o ArcInfo e o ArcView para instituições dedicadas a Investigação/Ensino é muito pouco significativa.

3.4 GRASS

O Grass também referenciado por Geographic Resources Analysis Support System é igualmente um SIG utilizado para gestão de informação, processamento de imagem, produção de gráficos, análise espacial e visualização de muitos tipos de dados. É um software grátis produzido sob licença da GNU.

Foi originalmente escrito pelos laboratórios de pesquisa e engenharia do exército dos Estados Unidos como uma ferramenta para gestão de território e planeamento ambiental usada pelos militares. O Grass transformou-se num utilitário poderoso com um vasto alcance em diferentes áreas de pesquisa científica.

O Grass pode ser corrido usando quer uma linha de comandos quer uma interface gráfica. Devido a ser um sistema *open source* a sua interface por comandos é mais completa do que a interface gráfica. Os novos módulos que vão surgindo são disponibilizados primeiro na interface por comandos.

Actualmente existe uma versão do Grass para correr no Windows NT/2000 mas ainda se encontra em fase experimental.

A linguagem de desenvolvimento é C.

Os sistemas operativos que suportam o Grass são: Linux, Sun Solaris, Solaris x86 etc.

O Grass permite análise espacial e visualização 2D e 3D tal como qualquer outro SIG e tem bastantes modelos de geração de dados. Esses modelos assentam sobretudo nas áreas de erosão e hidrologia. A principal desvantagem de se usar um sistema de informação geográfica como o Grass reside na muito fraca documentação visto ser um sistema em constante evolução. Outra grande desvantagem é o facto de o modo de interface por comandos ser muito pesado para utilizadores ainda não ambientados com o Grass pois requer um grande treino e memorização. Apesar de os comandos terem como prefixo uma letra que elucida o utilizador sobre qual a operação que pretende efectuar, não esconde de forma nenhuma a fraca usabilidade.

Tabela 3.1 - Prefixos dos comandos existentes no Grass

d.*	- display commands for graphical screen output
g.*	- general file management commands
i.*	- image processing commands
r.*	- raster processing commands
v.*	- vector processing commands
s.*	- site processing commands (point data)
m.*	- miscellaneous commands
p.* / ps.*	- map creation commands
...	- unix scripts

3.5 GEOMATICA

O Geomatica não pretende ser de modo algum um Desktop Mapper, este sistema é a todos os níveis um SIG completo. Está projectado para grandes instituições onde o SIG será uma ferramenta de trabalho de grande importância, consumindo muitos recursos e uma quantidade significativa de técnicos. Não é tão pouco um único programa, mas sim um agregado de diferentes aplicações, cada qual com a sua funcionalidade específica, dirigida a perfis de utilizadores diferentes.

É também um sistema inclinado para o trabalho com imagens de satélite, normalmente camadas *raster*, a partir das quais vão sendo criadas outras camadas, vectoriais e raster, e as bases de dados correspondentes. Ao iniciar o trabalho a equipa SIG, não tem só como objectivo criar as bases de dados e camadas específicas para posterior análise, como também deve desenvolver o próprio sistema no sentido de facilitar esse trabalho.

Vai ser dada agora uma visão geral dos principais módulos do Geomatica e as suas funções características.

3.5.1 FOCUS

É a janela principal, o nosso primeiro olhar para dentro do Geomatica.

Aqui vemos os nossos mapas e camadas, e fazemos a maioria das operações mais simples através de facilidades de *apontar e clicar*.

Como interface, e principal janela para o sistema, está bem concebida, a esta altura o utilizador ainda não se apercebe do "monstro" que reside detrás desta janela. A janela divide-se em três partes, a zona de botões e menus (barras superiores), árvore de ficheiros e mapas (lateral esquerda), zona de visualização de mapas (todo o resto, ocupa a maior parte da

janela). Os menus são poucos mas os suficientes, os botões estão bem agrupados, as suas representações são aceitáveis e quando não estão disponíveis aparecem desfocados.

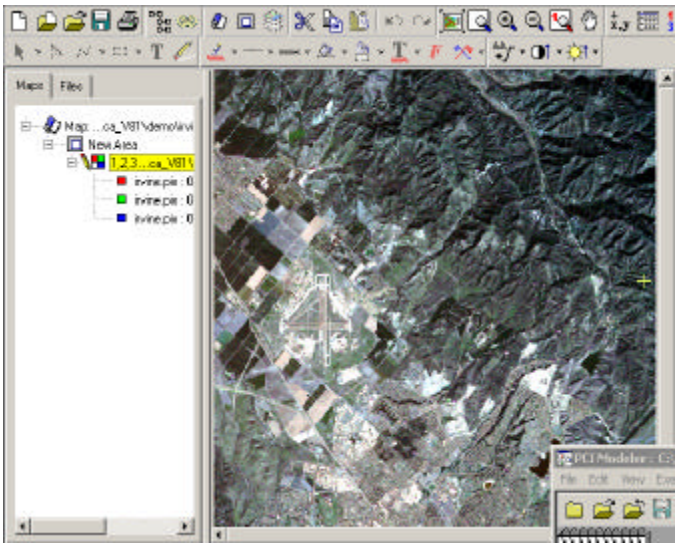

Figura 3.5 – Janela do FOCUS

A sua principal função é visualizar os mapas e as suas respectivas camadas. Mas este visualizar é bastante profundo, podem ser vistos um número indefinido de camadas sobrepostas, mapas de 8, 16 e 32 bits, de qualquer tamanho ou resolução; estando todas as camadas adequadamente geo-referenciadas. Mas uma vez é de realçar que apesar de estarem muitos mapas carregados no Focus, não se chega a sentir confusão em lidar com todos eles, dado a ser relativamente navegar entre eles, na árvore de mapas, com a ajuda do rato.

Outra função importante é o processamento e melhoramento gráfico das imagens disponíveis, normalmente fotografias de satélite. Está ao dispor do utilizador uma biblioteca de algoritmos de processamento de imagem, incluindo :

?? Algoritmos de filtragem de imagem;

?? Utilitários de Interpolação;

?? Algoritmos de Classificação;

?? Análise Espacial;

?? Análise de Modelos Digitais de Elevação (*DEM*).

Mas o Focus não se fica por aqui, permite fazer inspecção de dados através de histogramas, ou de geração de perfis. Pode-se, por exemplo, calcular distâncias ou perímetros apenas com alguns *cliques* do rato; ou editar as tabelas de atributos directamente. Existem ainda algoritmos específicos para a depuração da classificação de imagens; algoritmos que

produzem mapas cartográficos bastante avançados, com referencias de latitude e longitude muito apuradas, *et coetra*, *et coetra*.

3.5.2 OrthoEngine

É a ferramenta que vai permitir criar mapas e extrair informação a duas e três dimensões das fontes mais comuns de informação de um SIG, as imagens de Fotogrametria. As imagens fontes para um SIG são as fotografias tiradas em voo, dos satélites temos fotografias, assim como as imagens de radar e imagens digitais do terreno. O Orthoengine vai permitir, a partir destas fontes criar modelos digitais do terreno, modelos de elevação (DEM), assim como modelos a três dimensões. Tudo isto de forma automática, recorrendo a algoritmos de extracção de informação avançada. Este pacote vem ainda munido de um conjunto de funcionalidades, que, para além de permitirem fazer este trabalho, também o facilitam poupando tempo e esforço.

3.5.3 PCI Modeler

Mais uma vez estamos perante um ambiente de programação gráfica por filtros. A sua principal função é a mesma que o Macro Modeler do Idrisi, mas à medida que é explorado, torna-se claro que o seu âmbito é muito maior.

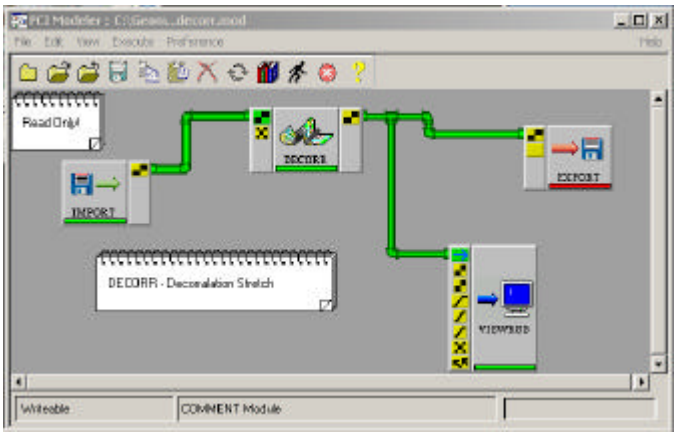

Figura 3.6 – PCI Modeler

O PCI Modeler começa pois por ser um ambiente que permite fazer análise topológica a partir de uma ou mais camadas, a partir das quais se pode gerar outras camadas, com o sistema de filtro já explicado, os mapas ou camadas são saídas ou entradas de filtros que são algoritmos de análise. A primeira diferença que surge em relação ao Macro Modeler é o facto de este aparecer como uma aplicação completamente independente, com uma janela própria e menus próprios, este e outros factores vão dar uma maior usabilidade ao PCI Modeler.

Graficamente o PCI Modeler é bastante avançado, não só se pode montar um esquema de análise com o rato, como as caixas que representam os filtros são de algum modo transparentes (os filtros do Macro Modeler são graficamente compactos), dando a possibilidade de definir alguns parâmetros com apenas alguns cliques.

Mas, conforme se vai aprendendo a trabalhar com o PCI Modeler, vai-se igualmente percebendo a dimensão das suas possibilidades. Em primeiro lugar está à disposição uma biblioteca com mais de trezentos algoritmos, e só este factor marca a diferença, para além de que nem todos eles se dedicam à análise topológica. Existem algoritmos que importam ficheiros em formatos externos ao Geomatica, que os voltam a importar para outros formatos, algoritmos que ajudam e permitem fazer alguma classificação. É claro que o factor mais importante é a existência de um número gigantesco de algoritmos disponíveis. A única forma de aproveitar todo este poder de análise e processamento é ter várias pessoas a usar este módulo do Geomatica.

3.5.4 FLY!

Este é o módulo que permite levar a cabo a análise a três dimensões. Tendo por base ficheiros DEM, ou mesmo vectores, o FLY! é capaz de criar um modelo a três dimensões do terreno em causa, do qual se podem visualizar diferentes perspectivas, quase em tempo real.

A principal característica que se deve salientar em relação a esta ferramenta de análise 3D é a simulação de voo. O FLY! dá a possibilidade de explorar o modelo a três dimensões do terreno, simulando que o utilizador voa por cima do mesmo. Com o auxílio do rato o utilizador controla a direcção e velocidade de voo, escolhendo trajectórias e pontos de observação. Mas é ainda possível simular condições atmosféricas, ou por exemplo simular o efeito do nevoeiro na visibilidade do voo. Resta ainda referir que tudo isto é feito recorrendo às mais recentes técnicas de *rendering*.

3.5.5 EASI

Este é o módulo de desenvolvimento do Geomatica, que tal como todos os outros apresenta características que o levam para além do desenvolvimento. Ao programador é apresentada uma linguagem de programação baseada na sintaxe do C, o que de facto facilita o desenvolvimento. Para além de esta linguagem poder ser usada para expandir o Geomatica e as suas funcionalidades, é também uma excelente linguagem de *scripting*, dando a possibilidade de criar automatismos de trabalho de grande utilidade.

O EASI começa por ser um ambiente baseado numa linha de comandos que permite efectuar toda a interacção utilizador/sistema independentemente do sistema operativo envolvente. Serve de ponte entre os diferentes sistemas base (ex: Sun / Microsoft), eliminado diferenças de editores dos sistemas operativos e formatos de próprios entre outros. Daqui, como já referido, parte-se para a possibilidade de criar *scripts*, e para o desenvolvimento do ambiente propriamente dito.

Assim descritos os principais módulos do Geomatica, onde muito ficou por dizer, devem ser referidas algumas das principais extensões que poderão fazer parte do pacote a adquirir, consoante as áreas de interesse.

EASI/PACE - Oferece funcionalidades avançadas no tratamento de imagens de satélite, quer de origem óptica, ou de radar. Incluindo:

SPANS ou PAMAP - Permitem ir ainda mais além na análise geográfica terrestre, em particular para bacias hidrográficas e na área da geologia.

ACE - Extensão das funcionalidades de cartografia. O Geomatica já é na sua base bastante poderoso neste capítulo, com o ACE pode-se chegar ainda mas longe.

Análise ***Hyperspectral*** - É possível, usando uma extensão PACE, efectuar análise terrestre baseada em espectros e ângulos espectrais. Esta extensão vem adicionar possibilidades de classificação diferentes das convencionais, podendo capturar informação que nos modos mais comuns (Análise de radar e óptica), não será possível.

3.4.6 Conclusões

Esta foi sem dúvida a solução de maiores dimensões que foi possível estudar. Os trinta dias de avaliação foram evidentemente escassos para conhecer por completo todo o ambiente do Geomatica.

Em primeiro lugar fica claro que a dimensão do Geomatica não se enquadra com aquilo que procuramos. De facto, para ter um sistema deste tipo a funcionar convenientemente são necessários muitos técnicos e operadores, sendo não só útil ter vários familiarizados com todo o sistema, como ter vários especialistas para cada um dos módulos de trabalho. Este é dos factores que vai levar a que a escolha não recaia sobre o Geomatica.

Finalmente, e mais relevante, serão os preços destes produtos. Só o Geomatica Prime custa 7200 €, custando cada uma das sua extensões 3200 €. O preço das extensões para

OrthoEngine varia entre os 3200 € e os 4300 €. O Sistema na sua totalidade com todas as extensões orça em mais de 40000 €.

3.6 Estrutura do Mercado dos SIG

Segue-se agora uma curta análise da Estrutura do Mercado SIG. Tentando-se perceber a evolução deste mercado nos últimos anos, de modo a tirar algumas conclusões sobre o seu futuro, e o modo como este pode influenciar a decisão da compra de um produto.

3.6.1 Estrutura do Mercado em 1994

Mostrou-se bastante difícil encontrar dados sólidos acerca da estrutura do Mercado SIG, mais ainda, reportando-se a anos mais afastados. Estes foram os dados mais antigos encontrados, retirados do *International SIG Sourcebook* de 1994. Baseiam-se no número de licenças emitidas e ainda activas, por cada fabricante. É de notar que há altura existiam já algumas centenas de empresas em estado embrionário, ou com expressão marginal no mercado, que aqui não aparecem quantificadas[6].

Tabela 3.2 - Número de Licenças Activas em 1994

Fabricante	Número	Quota
ESRI	800000	0,88
Intergraph	100000	0,11
Spans	2500	0,003
ERDAS	2000	0,002
Graphics Design System	1200	0,001
GenaMap	350	> 0,001
Topologic	30	> 0,001
Total	906080	

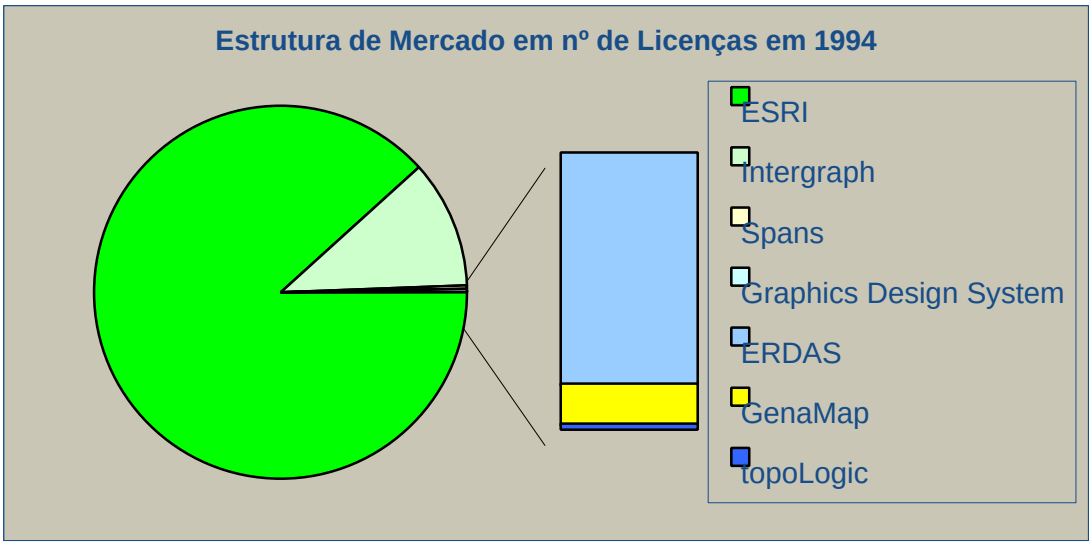

Figura 3.7 – Gráfico da Estrutura de Mercado em 1994

O número de licenças activas pode não ser o melhor indicador da estrutura de Mercado, pois uma licença de ArcView não têm o mesmo preço que uma de Spans. Os valores de facturação monetária são indicadores mais correctos mas não foram encontrados para anos anteriores a 1994. Há ainda a acrescentar que neste estava em actividade a empresa que fornecia o System 9, um SIG bastante difundido na época, mas para qual não existia registo do número de licenças activas.

Outro aspecto importante é facto de a SmallWorld, apesar de estar já em actividade, não ter sido ainda sido adquirida pela General Electrics. Pode-se referir ainda que estava por esta altura eminente a compra da ERDAS pela ESRI, passando a haver um cooperação tecnológica entre as duas empresas, embora continuando no mercado com os dois nomes.

O Mercado assim caracterizado mostra claramente uma situação de Monopólio[17,18] por parte da ESRI, ressalvando novamente o facto de o número de licenças não ser o melhor indicador da sua estrutura. A ESRI declara ter, num dia normal, por volta de um milhão de utilizadores dos seus produtos SIG. É claro que o tipo de produtos influência estes números, a esta altura empresa como a ESRI ou a Intergraph, operavam sobretudo no mercado dos sistema *Desktop*, enquanto que outras como a Spans (empresa que veio a evoluir o Geomatica), operavam mais no mercado dos SIG de grande dimensão.

Estes dados do início da década de 1990 (época ainda de alguma embrionagem no mercado SIG), são pouco conclusivos, mas para anos mais recentes foram encontrados dados mais sólidos, baseados nas receitas anuais das empresas no Mercado.

3.6.2 Estrutura do Mercado em 1999

Os dados referentes a 1999 e 2000, foram publicados pela DaraTech, um organismo independente de análise de alguns mercados de software[7].

As quotas de mercado aqui apresentadas referem-se às receitas de cada empresa no ano de 1999, que na globalidade do mercado perfizeram um valor de 845 milhões de dólares.

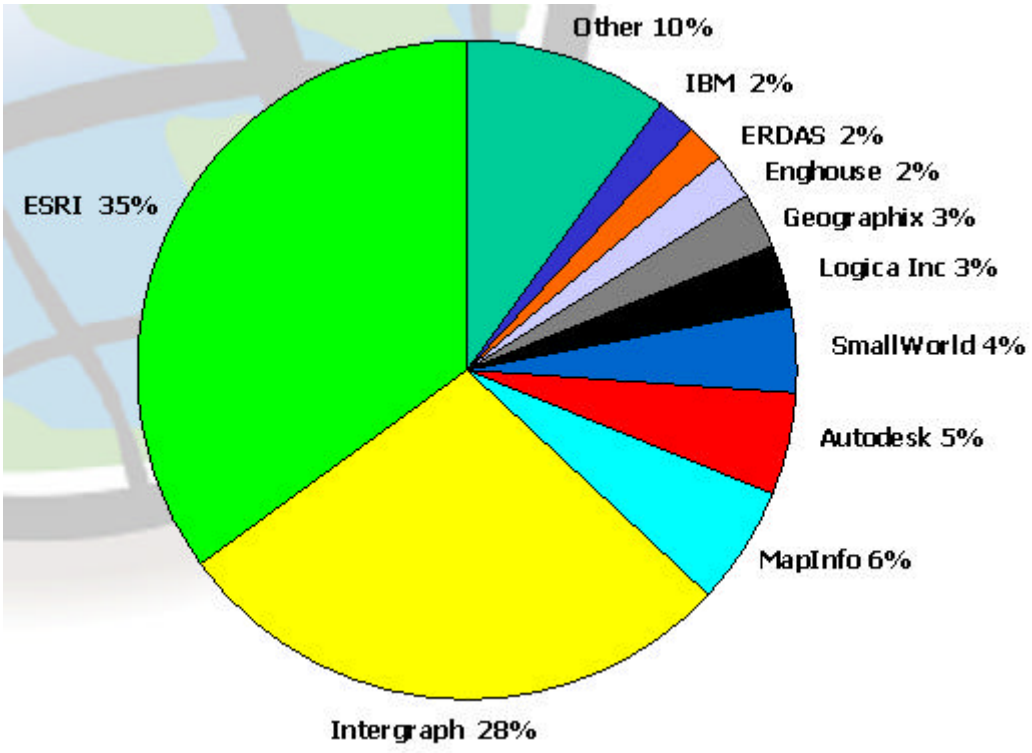

Figura 3.8 – Gráfico da Estrutura de Mercado em 1999

Passa-se agora a uma análise mais cuidada, recorrendo ao cálculo dos índices de Mercado.

Índice de Herfindahl [17,18] :

$H = 0.35^2 + 0.28^2 + 0.06^2 + 0.05^2 + 0.04^2 + 2x\ (0.03^2)\ 3x(0.02^2) =$

$= 0{,}1225 + 0{,}0784 + 0{,}0036 + 0{,}0025 + 0{,}0016 + 0{,}0018 + 0{,}0012 =$

= 0,2116

Índices de Concentração

C1 = 0.35 = **0.35**

C2 = 0.35 + 0.28 = **0.63**

C3 = 0.35 + 0.28 + 0.06 = **0.69**

C4 = 0.35 + 0.28 + 0.06 + 0.05 = **0.74**

Por estes índices se constata, que o mercado era em 1999 um Oligopólio [17,18]. O índice de Herfindahl encontra-se ligeiramente acima de 0,2 mas mesmo assim acima. Pelos índices de concentração observa-se que as duas maiores empresas detêm 63% do mercado, e as quatro maiores, quase três quartos do mesmo.

A liderança do mercado parece bastante sólida, pela parte da ESRI e da Intergraph, mas não se aproxima em nada do Monopólio verificado pelo dados relativos a 1994. Também é de sublinhar o aparecimento de bastantes empresas com quotas de mercado entre 2 e 10 por cento. Podem-se então desenhar dois cenários, as empresas de topo lideram o mercado generalista, apresentando soluções flexíveis e escaláveis; por outro lado existe um grande conjunto de empresas que se movimentam num mercado de Concorrência Monopolista, onde encontram pequenos nichos de mercado, que abarcam apresentando soluções específicas. Estima-se mais uma vez que as empresas com quotas de mercado inferior a 2% sejam várias centenas.

3.6.3 Estrutura do Mercado em 2000

As receitas totais do Mercado subiram neste ano para 939 milhões de dólares. No entanto dá-se uma viragem na sua estrutura, uma vez que a Intergraph perde quase metade da sua quota.

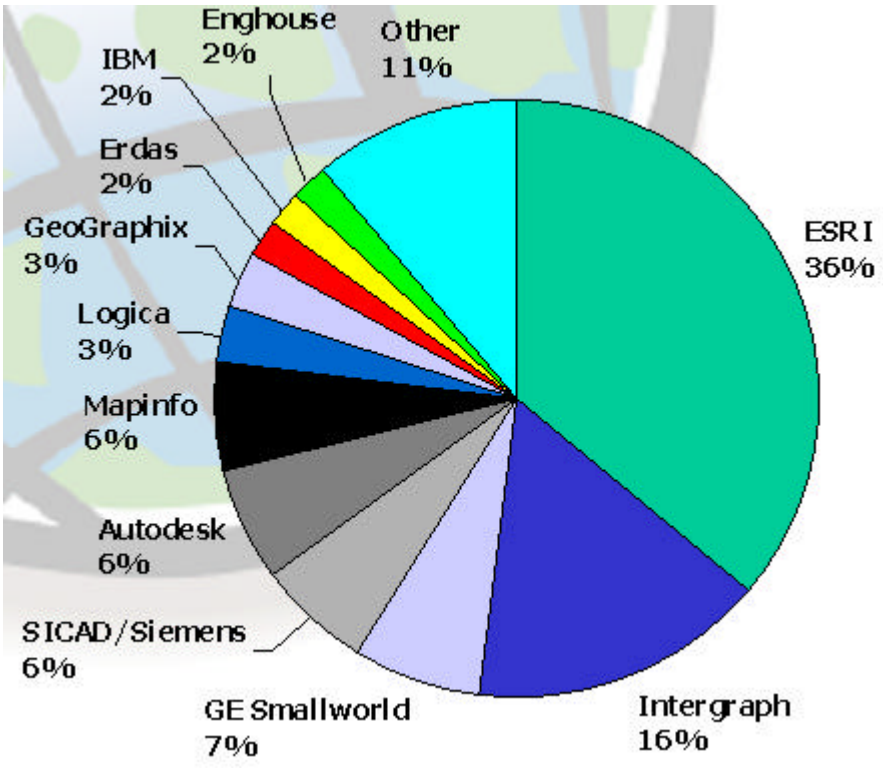

Figura 3.9 – Gráfico da Estrutura de Mercado em 2000

Mais uma vez calculam-se os índices de Mercado [17,18].

Índice de Herfindahl

$H = 0.36^2 + 0.16^2 + 0.07^2 + 3x(0.06^2) + 2x(0.03^2) + 3x(0.02^2) =$

$= 0,1296 + 0,0256 + 0,0049 + 0,0108 + 0,0018 + 0,0012 =$

= 0,1739

Índices de Concentração

C1 = 0.36 = **0.36**

C2 = 0.36 + 0.16 = **0.52**

C3 = 0.36 + 0.16 + 0.07 = **0.59**

C4 = 0.36 + 0.16 + 0.07 + 0.06 = **0.65**

Verifica-se então uma mudança curiosa de 1999 para 2000, o mercado passa repentinamente de um Oligopólio, para um Mercado de Concorrência Monopolista [17,18]. Aparentemente não se põe o caso de um Mercado de Concorrência Perfeita, pois as receitas cresceram 94 milhões de dólares, mais de 10%, e não houve um aumento visível do número de empresas a operar no mercado.

Apesar de toda esta movimentação a ESRI conseguiu solidificar ainda mais a sua posição no mercado. Passou para os 36% de quota, mas este aumento traduziu-se em 34 milhões de dólares de ganho em relação ao ano anterior. De 304 milhões de dólares de receitas em 1999, passou para 338 milhões em 2000.

A Intergraph parece estar a sofrer os efeitos da entrada de dois grandes gigantes do sector Electrotécnico no mercado, a General Electric, que anos antes tinha adquirido a SmallWorld e a Simens, com uma entrada de gigante nos sistemas SIG ligados à distribuição eléctrica. Esta ascendência pode indicar que a situação de Mercado de Concorrência seja apenas uma situação de transição, enquanto cai um dos gigantes e se espera que alguém ocupe o seu lugar.

3.5.4 Outros dados referentes ao Mercado

Já em 2001 foi publicado pela Geospatial Information and Technology Association (GITA) um estudo intitulado “The Geospatial Technology Report 2001.”, que trouxe informações interessantes acerca do mercado dos SIG. Este estudo teve em conta as respostas dadas a questionários por parte de 147 empresas[8-9].

O estudo dá conta que a liderança de mercado varia de sector para sector, o que reforça de algum modo a estrutura de Concorrência Monopolista. Concretamente, no mercado do Gás, lidera a General Electrics, no mercado das Telecomunicações a Intergraph, nos mercados da Electricidade, das Condutas e Hidrográfico lidera a ESRI. No entanto a fatia de mercado fora destes segmentos, representa a fatia de leão, para o qual não há informação.

O estudo apresenta também informação acerca dos sistemas operativos usados como suporte aos SIG. O Microsoft Windows lidera largamente com 77% do mercado, a seguir estão os sistemas híbridos Windows/UNIX com 16%, e os puramente UNIX com apenas 4%. O estudo conjectura ainda que com a introdução dos novos sistemas Windows XP e Windows.NET, desapareçam do mercado os sistemas UNIX.

3.6.5 Conclusões

O Mercado dos SIG é constituído por um pequeno grupo de empresas com bastante implantação no mercado e um sem número de pequenas empresas que lutam entre si pelos nichos de mercado onde não chegam os gigantes. Neste grupo maior é frequente aparecerem e desaparecerem pequenas empresas que nunca chegam a adquirir quotas de mercado significantes. Surge então aqui um primeira indicação, ao seleccionar um SIG que se pretenda usar durante um período de tempo relativamente grande, corre-se o risco de o seu fabricante sair do mercado entretanto. Ao optar-se por um SIG completo, que deverá estar em constante evolução, pode ser uma má decisão a parceria com uma empresa mais pequena.

A situação do ano de 2000 é muito certamente transitória, sendo de esperar a adaptação da estrutura do Mercado novamente a uma situação de Oligopólio. A ESRI lidera claramente o Mercado, mas deixa livre essa liderança em alguns sectores importantes, permitindo assim uma concorrência algo saudável.

À data da realização deste texto não existiam ainda dados concretos acerca do ano de 2001, no entanto é de referir que os primeiros números apontavam já para um novo crescimento global do Mercado SIG na ordem dos 10%, ultrapassando claramente 1 bilião de dólares de receitas.[9]

3.7 Comparação e Escolha

Depois de tomado contacto com estes quatro sistemas, houve que optar por um como sistema de suporte à fase seguinte do trabalho. Segue-se uma tabela onde estão equacionados os principais factores de escolha.

Tabela 3.2 - Factores de Escolha

Factor de Escolha	**IDRISI32**	**ARCINFO**	**GEOMATICA PRIME**	**GRASS**
Importação de dados do exterior	Bom	Bom	Muito Bom	Inexistente
Vectorização	Fraca	Bom	Muito Boa	Muito Fraca
Usabilidade	Fraca	Muito Boa	Boa	Muito Fraca
Linguagem Própria	Não tem	Tem, sintaxe própria - AML	Tem, sintaxe comum - C	Não tem, é *open source- C/Java*
Número de operadores	Reduzido	Escalável	Muito Alto	Muito Reduzido
Assistência	Apenas nos EUA	Representado em Portugal	Representado no Reino Unido	Não Representado
Documentação	Boa	Muito Boa	Boa, dispersa	Fraca
Preço	Razoável	Bom	Mau, > 25000 E	Grátis

Os sistemas que sobressaem à primeira vista são o Geomatica e o ArcInfo, aparecendo como os únicos com dimensão suficiente para trabalho em vista. Os factores que determinaram a escolham entre estes dois foram o preço, sobremaneira, o número de operadores e técnicos necessários ao seu correcto funcionamento, a documentação e a assistência. O Geomatica apesar de ser um sistema mais completo em algumas áreas, como o processamento de imagem e a vectorização, revelou-se um sistema muito pesado para as necessidades. O ArcInfo para actividades de Investigação e Ensino revela-se um sistema de custos muito inferiores. Em termos de número de pessoas envolvidas o Geomatica é um sistema muito mais exigente, requerendo, como referido anteriormente grande número de técnicos. Pelo contrário o ArcInfo revela-se neste aspecto bastante flexível, podendo ser explorado quer por um número muito reduzido de técnicos (talvez mesmo só um), quer por um número grande de técnicos ligados a uma Base de Dados centralizada, onde se ligam várias estações de trabalho.

A juntar a todos estes elementos está a existência da extensão ao ArcInfo, para interpolação de superfícies, o *Geostatistical Analyst*, que como se verá mais será crucial para es trabalho. Apesar de existirem outras aplicações de interpolação no mercado, o *Geostatistical Analyst* é neste momento o 'estado da arte' no que diz respeito a aplicações informáticas de Geoestatística [10].

Com estes elementos a escolha recaiu sobre o ArcInfo, que foi posteriormente adquirido com as extensões *ArcGis Spatial Analyst*, *ArcGis 3D Analyst* e ArcGis *Geostatistical Analyst*.

CAPÍTULO IV

INTERPOLAÇÃO DE SUPERFÍCIES

Escolhido o sistema a usar passou-se a uma nova fase de trabalho, a integração, análise e tratamento dos dados em questão. Este conjunto de dados consiste numa colecção de leituras batimétricas de profundidade do mar, com o qual se pretende produzir uma representação da área em causa. A melhor forma de conseguir tal representação é criar uma superfície com base nestas amostras. De seguida são discutidos alguns dos métodos de interpolação de superfícies implementadas pela extensão ao ArcGIS, *Geostatistical Analyst*.

4.1 Introdução

Tendo um conjunto de amostras de elevação (ou de profundidade neste caso), e querendo gerar uma superfície contínua, ha que definir um método de preencher os espaços vazios onde não foi recolhida informação. Partindo dos pontos conhecidos é necessário atribuir valores aos pontos para os quais é desconhecido o seu valor. Tal é conseguido através de técnicas de interpolação. No entanto esta interpolação deve ser o mais precisa possível, de modo a ser gerada uma superfície o mais próxima possível da realidade, e de modo a ser usada correctamente noutros formas de análise e modelação GIS, tal como na sua modelação a três dimensões (3D).

Existem dois tipos de métodos de interpolação usados na geração de superfícies, os métodos de interpolação Determinística e os métodos de interpolação Geoestatística. As técnicas Determinísticas calculam superfícies baseadas na extensão de semelhança entre os pontos, ou no grau de suavidade pretendido, através de funções matemáticas. As técnicas Geoestatísticas geram superfícies baseando-se em estatísticas e modelos de previsão de superfícies que incluem erros e incerteza na previsão[12].

4.2 Métodos Determinísticos

A solução dada por estes métodos é relativamente simples, a superfície é gerada através de uma função matemático que modele as variações da mesma. Ao calcular o valor de um ponto consideram-se as várias amostras em seu redor, e é válido assumir que os pontos mais próximos têm valores semelhantes ao do ponto desconhecido. Mas existem várias questões a

considerar, incluem-se todos os pontos da amostra no cálculo, ou apenas alguns? E nesse caso quantos pontos incluir? E devem ser todos os pontos considerados com igual influência no cálculo do valor do ponto em causa? É possível que um ponto demasiado afastado esteja numa zona completamente diferente da zona em causa.

Dadas estas questões existem dois grandes grupos de métodos Determinísticos, os Globais, que calculam a predição num dado ponto usando informação de todos os pontos conhecidos, e os Locais, que calculam a predição baseada numa dada vizinhança.

Apresenta-se agora uma breve descrição de alguns dos métodos Determinísticos.

4.2.1 Inverso da Potência das Distâncias (Inverse Distance Weighting - IDW)

Esta solução selecciona um certo número de pontos conhecidos em redor do ponto a estimar. Como indica o próprio nome, neste método é atribuído um peso a cada amostra considerada no cálculo, em função da sua distância ao ponto a calcular. Quanto mais afastada a amostra estiver menor será o seu peso, e vice-versa. O IDW assume que cada ponto exerce um influência local que diminui com a distância, desta forma este método segue a assunção de que os valores medidos mais próximos uns dos outros tendem a ser mais semelhantes do que aqueles medidos a distâncias maiores[11].

A fórmula usada é a que se segue:

$$Z(X_0)^* = \sum_{i=1}^{N} \lambda_i \, Z(X_i) \qquad (4.1)$$

Onde $Z(x_0)^*$ é o valor que se está a calcular para a posição x_0 , N é o número de pontos conhecidos e considerados na vizinhança de , $?_i$ são os pesos associados a cada ponto conhecido, e $Z(x_i)$ é o valor observado no ponto conhecido s_i.

A fórmula que determina cada um dos pesos é a seguinte:

$$\lambda_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^{N} d_{i0}^{-p}} \qquad \sum_{i=1}^{N} \lambda_i = 1 \qquad (4.2)$$

O valor d_{i0} representa a distância entre o ponto a calcular, X_0, e os pontos de valor conhecido, s_i, com o aumento da distância o peso é reduzido por um factor p. Este factor indica a forma como varia o peso dos pontos com a distância, quanto maior for o valor da potência mais rapidamente decresce o peso dos pontos mais afastados. Neste método, a soma dos pesos de todos os pontos considerados é ajustada de modo a ser sempre 1.

O número de pontos considerados na estimação deve ser suficiente para se obter um valor preciso, mas ao mesmo tempo prático para a interpolação. Este número de pontos pode variar conforme a distribuição e número de amostras, e com a própria configuração da superfície. Se as amostras estão distribuídas equitativamente e não houver mudanças drásticas das características da superfície pode-se obter uma superfície de precisão razoável.

Antes de usar as fórmulas e necessário escolher qual o tipo de vizinhança a usar nos cálculos. No caso de não serem conhecidas influências direccionais nos dados recolhidos, o melhor será considerar pontos de igual modo em todas as direcções, sendo assim a vizinhança a usar será um círculo. Mas se forem conhecidas influências direccionais, como por exemplo o vento, já será mais correcto usar outra forma de vizinhança. Usando o exemplo da influência do vento, sendo a direcção deste conhecida, a vizinhança ideal passaria a ser uma elipse com o eixo maior paralelo a esta direcção. É de esperar que amostras alinhadas com a direcção do vento sejam mais semelhantes do que as alinhadas prependicularmente a este[11].

Figura 4.1 – Vizinhança elíptica

Um superfície gerada usando IDW é influenciada pela escolha do factor p e pela forma da vizinhança escolhida. O IDW é um interpolador exacto, onde os valores máximos e mínimos da superfície são sempre valores da amostra.

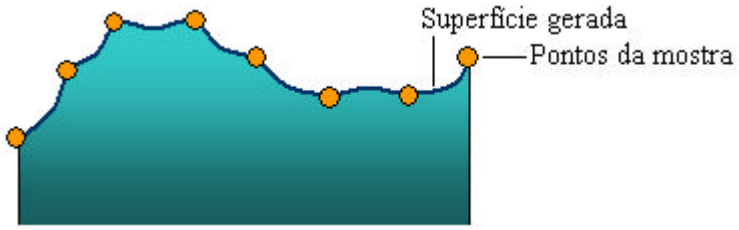

Figura 4.2 – Superfície gerada por IDW

4.2.2 Interpolação Polinomial Global

Este método é especialmente útil quando existe uma variação tendencial nos dados recolhidos. Se por exemplo as amostras forem recolhidas numa encosta, terão valores mais baixos no sopé da encosta e maiores no seu cume. Estas tendências globais podem não ser detectadas em modelos Locais[11].

Numa Interpolação Polinomial Global, como o próprio nome indica, é procurada um função polinomial que melhor se enquadre no conjunto global de amostras. Para uma encosta com uma inclinação relativamente constante, será usada um polinomial de primeira ordem, caso se esteja a modelar um vale, será usada uma polinomial de segunda ordem positiva. Ou seja para modelar um superfície com uma inflexão é necessária uma polinomial de segunda ordem, duas inflexões, terceira ordem e assim por diante.

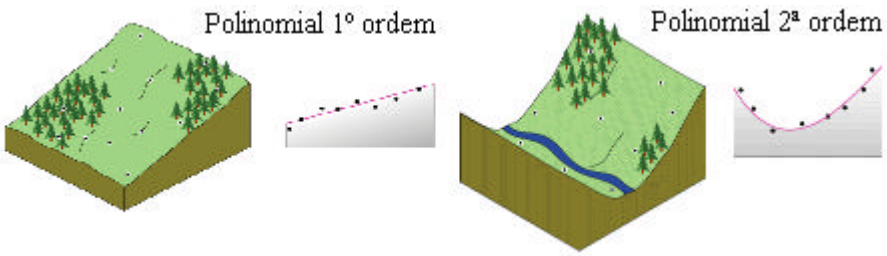

Figura 4.3 – Superfícies geradas por polinomiais globais

Este tipo de interpoladores não é exacto, a superfície gerada não coincide com todos os pontos recolhidos (pode até não coincidir com qualquer um). Embora estes métodos obtenham resultados globais satisfatórios, ignoram pequenas alterações como pequenas colinas ou pequenas depressões, características a que se dá o nome de Variação Local[11]. De modo a evitar estas variações recorre-se por vezes a polinomiais mais complexas, mas na maioria dos casos a complexidade crescente da polinomial tende a produzir resultados que não têm correspondência nos fenómenos físicos reais.

4.2.3 Interpolação Polinomial Local

Imagine-se que foram colhidas amostras numa área constituída por uma encosta que termina planalto, do qual parte um nova encosta. Uma predição Global seria aqui desastrosa. Mas se em vez de uma polinomial forem usadas várias mais pequenas, obtém-se uma superfície constituída por vários planos adjacentes que modelarão mais aceitavelmente a zona de estudo.

A interpolação Polinomial Global usa então, não uma, mas várias polinomiais, cada uma calcula numa dada vizinhança. Tal como visto na interpolação IDW, a configuração desta vizinhança pode ser ajustada, de acordo com a informação disponível sobre a zona de estudo. Mais ainda, pode-se também, definindo um parâmetro de potência *p*, influenciar o peso de cada ponto da vizinhança no cálculo da polinomial local, em função da sua distância ao centro da vizinhança[11]. Para cada ponto da superfície a gerar são escolhidos quais os pontos vizinhos a considerar, mediante a configuração da vizinhança definida. De seguida são atribuídos os pesos a estes pontos e calculada a polinomial de ordem definida. O valor da polinomial no centro da vizinhança é o valor atribuído ao ponto da predição.

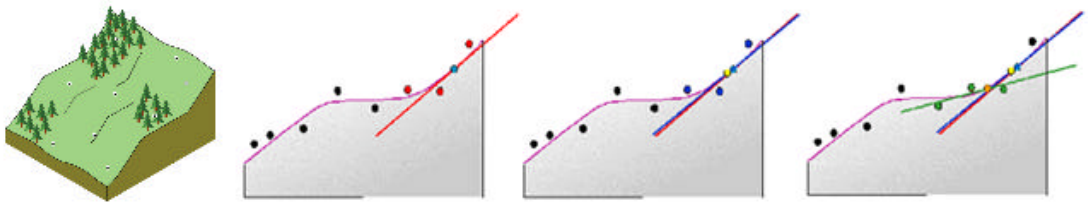

Figura 4.4 - Superfície gerada por polinomiais locais

Neste exemplo, na primeira secção do terreno, é calculado o valor do ponto azul usando uma vizinhança que compreende os pontos vermelhos, que geraram polinomial de 1ª ordem a vermelho. Numa localização seguinte, o ponto amarelo, é calcula uma nova polinomial, deste feita, embora a vizinhança compreenda os mesmo pontos, a polinomial produzida é ligeiramente diferente, dado haver uma pequena mudança na distribuição dos pesos. A interpolação segue assim por diante, até chegar o um novo ponto, de cor laranja, onde a vizinhança passa a compreender outro conjunto de pontos, de cor verde, e onde é gerada mais uma polinomial, também a verde.

As superfícies geradas por interpolação Polinomial Local conseguem modelar variações dos dados a curtas distâncias, algo que não é possível com uma Polinomial Global. Apesar de

tudo, este interpolador continua a não ser exacto, existindo amostras pelas quais as superfícies geradas não passam.

4.2.4 Funções de Base Radial (*Radial Basis Functions* -RBF)

Este método permite criar superfícies que incluem as tendências globais e as variações locais. A ideia base é torcer e esticar a superfície de modo a que esta passe por todos os pontos da amostra. Existem várias maneiras de encontrar a forma desta superfície. Por exemplo, forçar a superfície a formar curvas suaves (*thin-plate spline*), ou controlar a forma como são puxadas as pontas da superfície. Este método pode ser exemplificado como ajustar uma membrana de borracha de forma a passar por todos os pontos da amostra[11].

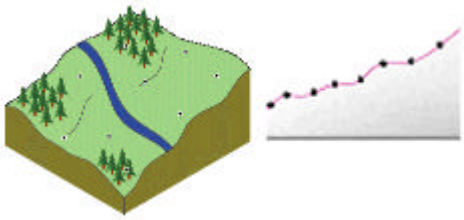

Figura 4.5 – Superfície gerada por RBF

Tal como o método IDW, o método RBF é exacto, ao contrário dos métodos Polinomiais. Comparativamente ao IDW, a grande novidade que traz o RBF é o facto de as amostras deixarem de ser os pontos máximos e mínimos da superfície gerada. A superfície gerada apresenta maior suavidade, e modela melhor certos fenómenos físicos como a elevação do terreno.

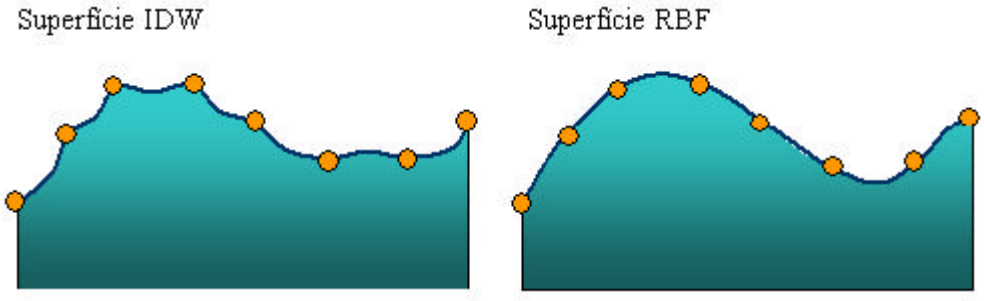

Figura 4.6 – RBF *versus* IDW

Uma função RBF é uma função que muda consoante a distância, quanto maior a distância, maior o seu valor. Na figura que se segue existem três amostras (nos pontos de

coordenadas (4,6), (6,5) e (6,6)), para os quais estão representadas as respectivas RBF, cones neste caso, que representam simplesmente a distância ao ponto da amostra.

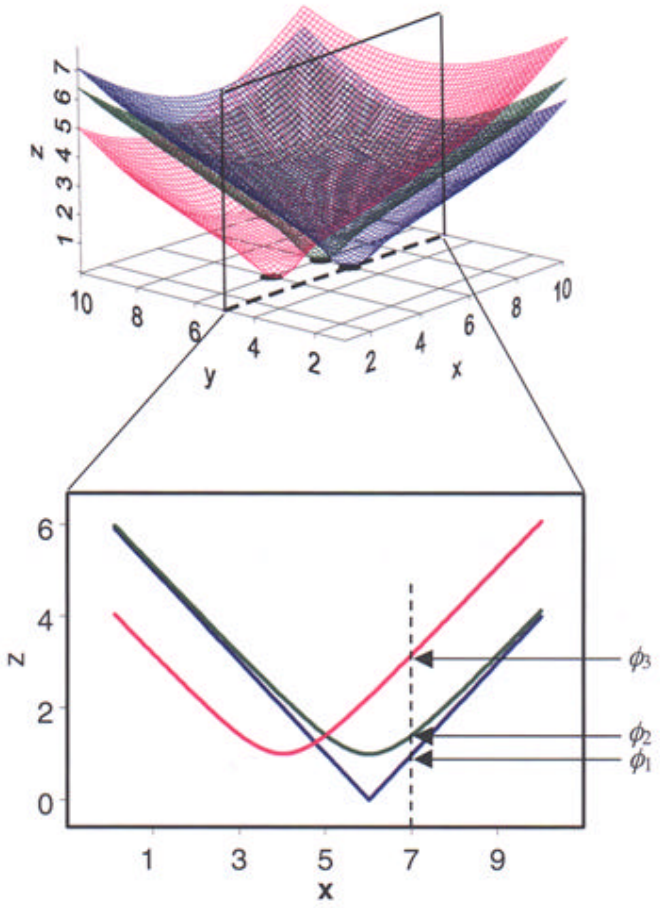

Figura 4.7 – Exemplo RBF com 3 amostras

Na parte de baixo da figura temos uma secção do espaço no plano y=5, onde estão representadas as projecções das três RBF. Imagine-se agora que se está a interpolar o valor do ponto (7,5), os valores de cada RBF são neste ponto representados por $?_1$, $?_2$, e$?_3$. O valor do ponto (7,5) será a média ponderada dos valores de cada uma das RBF nesta localização, $w_1 ?_1 + w_2 ?_2 + w_3 ?_3$. Para completar os cálculos falta apenas conhecer os valore dos diferentes pesos w_1, w_2, w_3, para tal usa-se o conhecimento de que nos pontos de amostragem a predição é exacta. Tendo *N* pontos conhecidos e *N* pesos desconhecidos, é construído um sistema de *N* equações a *N* incógnitas, que pode ser resolvido calculando os valores correctos de cada peso *w*.

Existem várias realizações possíveis do método RBF, variando a forma das funções usadas nas amostras. O exemplo aqui apresentado usa RBF multiquadráticas, calculadas

simplesmente em função da distância, outras funções podem ser usada como as RBF multiquadráticas inversas, ou funções de *spline*[11].

4.3.5 Conclusões

Estes métodos, chamados determinísticos, baseados em funções puramente matemáticas, ignoram todo um conjunto de factores que estão na origem dos fenómenos físicos. O seu objectivo é simplesmente obter uma interpolação espacial dos valores observados. Este é um dos problemas que os métodos geoestatísticos se propõem solucionar, recorrendo a uma metodologia baseada em probabilidades[13].

Apesar desta característica métodos como o RBF e o Polinomial Local podem obter resultados interessantes em zonas de grande densidade de amostragem[11], não ficando muito atrás dos métodos geoestatísticos, e apresentando um solução mais simples, e menos pesada computacionalmente.

4.2 Geoestatística

4.2.1 Introdução

"*Em geral, coisas que estão mais próximas tendem a ser mais semelhantes que coisas que estão mais afastadas. Este é um princípio fundamental da Geografia*."

Tobler, 1970

Na natureza existem vários fenómenos distribuídos no espaço, que devido à sua complexidade apresentam um elevado grau de incerteza. O conhecimento existente em relação a estes fenómenos é, em geral, bastante reduzido e resume-se a um conjunto de observações[13]. Como exemplos, refere-se o número de árvores por hectare numa floresta, o nível de poluente num aquífero, a medição da temperatura numa dada região, ou como no caso em estudo, a profundidade do mar.

A Geoestatística surgiu nos anos 60, como uma metodologia para quantificar a incerteza espacial de variáveis que caracterizam os fenómenos naturais no espaço[13]. Foi inicialmente aplicada na industria mineira para a avaliação de jazigos naturais[14]. A Geoestatística permite não só considerar a distância entre as amostras e o seu peso na estimação, como também a disposição global das mesmas[11]. Os métodos geoestatísticos incluem auto-correlação (relações estatísticas entre as amostras), e podem dar medidas de certeza ou exactidão da estimações geradas[11].

Quando se pretende adquirir conhecimento acerca de um fenómeno físico contínuo (como a profundidade, a elevação, a composição dos solos, ou do ar), não é possível, na grande maioria dos casos, medir valores para todos os pontos da área da estudo (o número de pontos numa área de estudo relativamente grande pode ser considerado infinito numerável). As medições são feitas num conjunto finito de N pontos, para os quais é conhecida a sua localização espacial.

Estas variáveis, que incorporam localização espacial não são completamente aleatórias, pois apresentam um comportamento estruturado que reflecte a semelhança de valores medidos em posições próximas no espaço [15]. A este tipo de variáveis aleatórias, Matheron deu a designação de Variáveis Regionalizadas (VR) que, para além do aspecto aleatório da variação imprevisível dos valores entre os pontos e o espaço, compreendem um aspecto estruturado que reflecte as ligações (correlações) entre os pontos do espaço onde o fenómeno ocorre [16]. A partir de um conjunto restrito de valores que representam a realização de uma ou mais VR e do conhecimento qualitativo adquirido sobre um fenómeno espacial em estudo (experiência pessoal, conselho de especialistas, etc), pode ser desenvolvido um modelo que descreve a aleatoriedade espacial do fenómeno e permite quantificar a sua incerteza espacial. Neste sentido, a Geoestatística associa a aleatoriedade com a VR através de um modelo estocástico, considerando que a VR é uma de várias realizações possíveis de uma função aleatória [15].

4.2.2 Modelo Geoestatístico

Como foi visto anteriormente, a abordagem determinística associa a qualquer ponto do espaço não amostrado uma única estimativa sem fornecer uma medida de estimação; a abordagem probabilística, por seu turno, associa à estimação uma probabilidade de ocorrência [14]. Enquanto os modelos deterministas geralmente assentam no conhecimento físico do fenómeno, a informação utilizada nos modelos probabilistas é proveniente, na maior parte dos casos, de dados experimentais [14].

Nos modelos geoestatísticos, os dados amostrais são interpretados como provenientes de um processo aleatório. O facto de estes modelos incorporarem a incerteza na sua concepção não significa que o fenómeno em si tenha resultado de um processo aleatório, mas serve somente de base metodológica à inferência espacial ou estimação de grandezas em áreas não amostradas e à quantificação da incerteza associada ao estimador[13].

A geoestatística é baseada em processos aleatórios com dependência. Num contexto espacial ou temporal, tal dependência é designada auto-correlação. Com apenas um conjunto

de valores medidos, não existe esperança de se conhecerem as regras de dependência.

Contudo, através de observações contínuas de várias amostras, as dependências tornam-se aparentes. Um valor localizado espacialmente em x_1 (denominação genérica de um conjunto de coordenadas geográficas) é interpretado como uma realização $z(x_1)$ da variável aleatória $Z(x_1)$. No espaço A, no qual se dispersa o conjunto de amostras, temos as realizações das N variáveis aleatórias $Z(x_1)$, $Z(x_2)$, ..., $Z(x_n)$, correlacionadas entre si.

4.2.2.1 Função Aleatória Estacionária

O conjunto de variáveis aleatórias $Z(x_i)$, i=1, ..., N, correlacionadas entre si, constituem uma função aleatória, da qual só se conhece uma realização $Z(x_i)$ – o conjunto dos dados experimentais.

A solução proposta no formalismo geoestatístico consiste em assumirem-se diversos graus de estacionariedade da função aleatória, de modo a que a inferência de alguns parâmetros estatísticos seja possível. Assim, se for admitido que todas as variáveis aleatórias têm a mesma média,

$$E\{Z(x_1)\} = E\{Z(x_2)\} = ... = E\{Z(x_i)\} = E\{Z(x)\} = m, \quad (4.3)$$

este parâmetro, que passa a ser independente da localização de xi, pode ser estimado à custa de uma média aritmética dos valores das realizações das variáveis aleatórias:

$$m = \frac{1}{N} \sum_{\alpha=1}^{N} Z(X_\alpha) \quad (4.4)$$

Esta hipótese de estacionariedade da média é parte integrante e fundamental do modelo probabilista geoestatístico e não pode nunca ser validada ou refutada na prática, uma vez que, na realidade, só existe uma realização da função aleatória.

Julgar esta hipótese de estacionariedade como apropriada é julgar a média das amostras como representativa da área A, ou que os valores das amostras são suficiente homogéneos para validar aquela representatividade[13].

4.2.2.2 Estacionariedade do Semivariograma e da Covariância espacial

A ideia da estacionariedade é usada para se obter a replicação necessária[11]. Existem dois tipos de estacionariedade. O primeiro, como já foi visto, é designado por estacionariedade da média, isto é, a média é constante entre amostras e é independente da

localização. O segundo tipo de estacionariedade é designado por estacionariedade do 2º momento para a covariância e estacionariedade intrínseca para semivariogramas. A estacionariedade do 2º momento é a suposição de que a covariância é a mesma entre quaisquer dois pontos que estejam à mesma distância e direcção. A covariância é dependente da distância entre quaisquer dois valores e não da sua localização. Para os semivariogramas, a estacionariedade intrínseca é a suposição que a variância da diferença é a mesma entre quaisquer dois pontos que estejam à mesma distância e direcção independentemente dos pontos que se escolham [11].

A hipótese de estacionariedade em relação a estas estatísticas é definida do seguinte modo: a correlação (4.5) e a variância (4.6) entre duas variáveis aleatórias depende somente da distância espacial que as separa e é independente da sua localização. Dado um vector h que distancia duas variáveis aleatórias, temos que:

$$C(Z(x_1), Z(x_2)) = C(Z(x_i), Z(x_i + h)) = C(h), \quad (4.5)$$

$$?(Z(x_1), Z(x_2)) = ?(Z(x_i), Z(x_i + h)) = ?(h) \quad (4.6)$$

Tal como para a média, a hipótese de estacionariedade do segundo momento permite a inferência daquelas medidas – o semivariograma (4.7) e a covariância (4.8) com base nas *N* amostras $Z(x_a)$, $a = 1, ..., N$

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [Z(X_\alpha) - Z(X_\alpha + h)]^2 \quad (4.7)$$

$$C(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} [Z(X_\alpha) - Z(X_\alpha + h)] - m(X_\alpha)\, m(X_\alpha + h) \quad (4.8)$$

sendo $m(X_\alpha) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(X_\alpha)$ e $m(X_\alpha + h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(X_\alpha + h)$.

4.2.3 Análise Estrutural

O elemento estruturante Biponto (par de pontos separados por uma distância h) é um instrumento geoestatístico de medida da continuidade espacial das características de um recurso natural ou fenómeno espacial. Esta medida desempenha um papel crucial na metodologia geoestatística, cumprindo, fundamentalmente, dois objectivos: a caracterização e a quantificação do modo como se dispersam espacialmente as características de um recurso –

o grau de continuidade espacial, o modo diferencial como a grandeza se dispersa no espaço (anisotropia) – prática que se denomina, na linguagem geoestatística, análise estrutural; por outro lado serve de base aos processos de inferência espacial/estimação e simulação.

O biponto apesar de não abranger a noção de conjunto conexo (dois pontos podem pertencer simultaneamente a X, e este não ser conexo), é, no entanto, o elemento estruturante privilegiado na prática geoestatística de análise de continuidade espacial de um recurso natural, quando este provem de um conjunto de amostras ou observações, normalmente discreto, limitado e, na maioria das vezes, escasso.

Para diferentes valores de h (designado em terminologia anglo–saxónica por *lag size*) obtêm-se as diferentes nuvens, que são representativas da correlação linear entre os valores Z(x) das amostras localizadas em x e os valores das amostras distanciadas de h, Z(x + h).

O conjunto destas nuvens, para os diferentes passos h, contém a quase totalidade da informação disponível, a partir do biponto, relativa ao grau de dispersão/continuidade da variável Z(x). No entanto, para uma melhor interpretação e posterior utilização, é necessário sintetizar as nuvens em instrumentos que, resumindo a sua dispersão, permitem visualizar a sua evolução com h [13].

4.2.3.1 Parâmetros do Semivariograma

A Figura 4.8 ilustra um semivariograma experimental com características muito próximas do ideal. O seu padrão representa o que, intuitivamente, se espera de dados de campo, isto é, que as diferenças $\{Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})\}$ decresçam à medida que $\mathbf{h}$, a distância que os separa decresce. É esperado que observações mais próximas geograficamente tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Desta maneira, é esperado que $?(\mathbf{h})$ aumente com a distância $\mathbf{h}$. Os parâmetros do semivariograma podem ser observados directamente da Figura 4.8 :

1. ***Alcance (a)***: distância dentro da qual as amostras se apresentam correlacionadas espacialmente.
2. ***Patamar (C)***: é o valor do semivariograma correspondente ao seu *alcance (a).* Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras (Var[Z(**x**) - Z(**x**+**h**)]) torna-se invariante com a distância.

3. ***Pepita (c0***): idealmente, *?* **(**0)=0. Na prática, à medida que **h** tende para 0 (zero), *?* **(**h)

aproxima-se de um valor positivo chamado *Efeito Pepita*, que revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. Parte desta descontinuidade pode ser também devida a erros de medição [11], mas é impossível quantificar se a maior contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem.

4. ***Contribuição (C_1)***: é a diferença entre o patamar (C) e o Efeito Pepita (C_o).

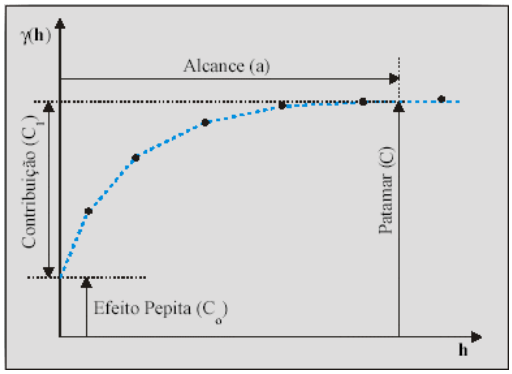

Figura 4.8 – Parâmetros do Semivariograma

O gráfico do semivariograma experimental, $\gamma(h)$, calculado através da equação (4.7), é formado por uma série de valores, conforme ilustra a Figura 4.8, sobre os quais se objectiva ajustar uma função. É importante que o modelo ajustado represente a $\gamma(h)$ em relação a h. Deste modo, as estimativas obtidas serão mais exactas e, portanto mais confiáveis. Os modelos aqui apresentados são considerados modelos básicos, denominados de modelos isotrópicos (quando uma função depende apenas da distância entre dois locais e não da posição relativa desse locais). Estão divididos em dois tipos: modelos com patamar e modelos sem patamar. Modelos do primeiro tipo são referenciados na geoestatística como modelos transitivos. Alguns dos modelos transitivos atingem o patamar *(C)* assimptoticamente. Para tais modelos, o alcance *(a)* é arbitrariamente definido como a distância correspondente a 95% do patamar. Modelos do segundo tipo não atingem o patamar, e continuam a aumentar enquanto a distância aumenta. Tais modelos são utilizados para modelar fenómenos que possuem capacidade infinita de dispersão, é o caso do modelo Potência. Os modelos transitivos mais utilizados são: modelo esférico, modelo exponencial e modelo gaussiano. Estes modelos estão apresentados na figura 4.9 com o mesmo alcance (a).

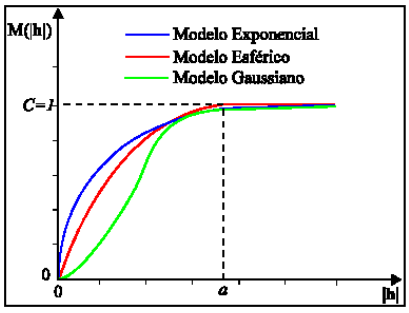

Figura 4.9 – Representação gráfica de modelos transitivos normalizados

4.2.3.2 Modelo Esférico

Trata-se de um dos modelos mais usados em geoestatística, função de dois parâmetros: um patamar C, limite superior para o qual tendem os valores do semivariograma com o aumento dos valores de h (que neste exemplo toma o valor de 1), e o alcance (a), distância a partir da qual os valores de *?(h)* param de crescer e são iguais a um patamar que é normalmente coincidente com a variância de *Z(x)* [13].

O modelo esférico tem a seguinte expressão:

$$\mathrm{Sph}(|\mathbf{h}|)=\begin{cases}0 & , \quad |\mathbf{h}|=0\\ 1{,}5\left(\frac{|\mathbf{h}|}{a}\right)-0{,}5\left(\frac{|\mathbf{h}|}{a}\right)^3 & , \quad 0<|\mathbf{h}|\le a\\ 1 & , \quad |\mathbf{h}|>a\end{cases} \tag{4.9}$$

4.2.3.3 Modelo Exponencial

É uma função dos mesmos parâmetros do modelo esférico. O patamar corresponde ao valor para o qual o semivariograma tende assimptoticamente. O valor do alcance (a) é a distância em que o modelo atinge 95% do patamar: ?(a) = 0.95C.

O modelo exponencial tem a seguinte expressão:

$$\mathrm{Exp}(|\mathbf{h}|)=\begin{cases}0 & , \quad |\mathbf{h}|=0\\ 1-\exp\left(-\frac{|\mathbf{h}|}{a}\right), & |\mathbf{h}|\neq 0\end{cases} \tag{4.10}$$

4.2.3.4 Modelo Gaussiano

Os modelos apresentados anteriormente têm um crescimento relativamente rápido junto da origem, como se pode constatar pela figura 4.9, denunciando um comportamento típico de fenómenos relativamente irregulares. No entanto existem outros fenómenos, bastante mais

regulares e contínuos, que denunciam um crescimento lento de ?(h) originando um comportamento parabólico na origem.

O modelo Gaussiano tem a seguinte expressão:

$$\text{Gau}(|\mathbf{h}|) = \begin{cases} 0 & , \quad |\mathbf{h}| = 0 \\ 1 - \exp\left(-\frac{|\mathbf{h}|}{a}\right)^2 & , \quad |\mathbf{h}| \neq 0 \end{cases} \tag{4.11}$$

O alcance neste modelo é igualmente a distância para o qual o modelo atinge 95% do patamar.

4.2.3.5 Modelos de Anisotropia

A continuidade espacial de um recurso natural, por exemplo, pode variar com as diferentes direcções do espaço. Uma dada característica de um recurso natural diz-se que tem uma estrutura de continuidade espacial isótropa quando o semivariograma (ou covariância) tem o mesmo comportamento em todas as direcções (isto é, ?(h) depende somente do módulo do vector h).

Na situação em que a variável em estudo se estende preferencialmente, de um modo mais contínuo, numa direcção particular, dá origem a uma estrutura anisótropa. A anisotropia estrutural pode ser entendida como a variabilidade ou continuidade diferencial dependente das direcções do espaço[13].

Os modelos mais comuns de anisotropia são: a anisotropia geométrica e a anisotropia zonal.

1. Anisotropia Geométrica – Semivariogramas estimados para diferentes direcções apresentam o mesmo modelo, o mesmo patamar, mas amplitudes diferentes – a amplitude máxima verifica-se na direcção perpendicular à amplitude mínima.

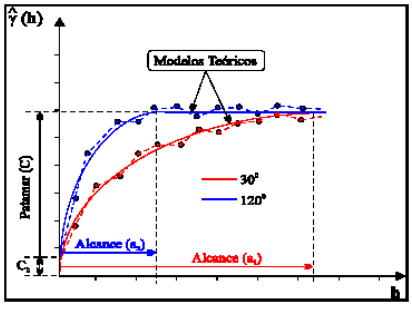

Figura 4.10 - Representação gráfica da anisotropia geométrica

2. Anisotropia Zonal – Quando o patamar do semivariograma é diferente para a direcção perpendicular à considerada.

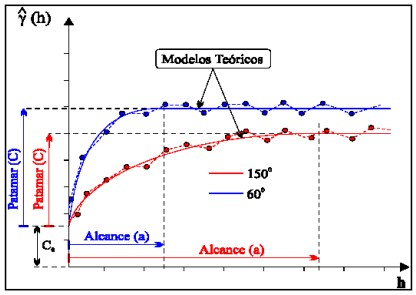

Figura 4.11 - Representação gráfica da anisotropia zonal

4.2.3.6 O que se vê num Semivariograma

Com um simples exemplo gráfico, pode-se perceber aquilo que a forma de um Semivariograma representa.

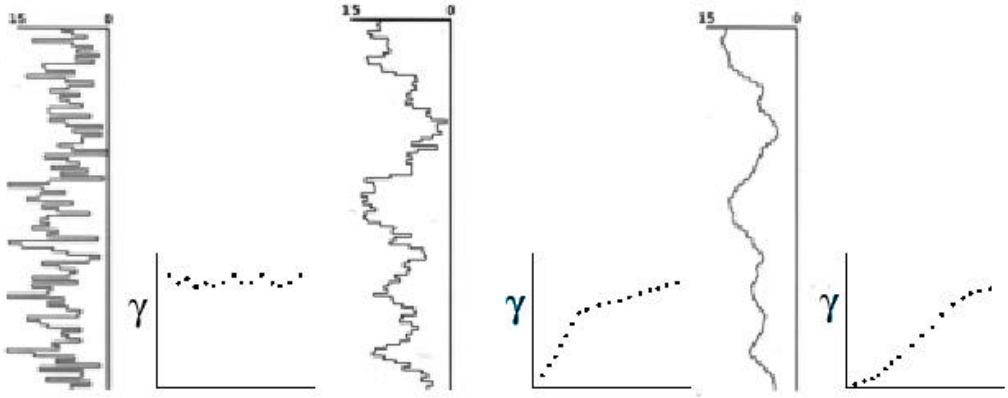

Figura 4.12 – Semivariogramas em amostras diferentes

No conjunto de amostras da esquerda existe uma grande diferença entre medições feitas em locais adjacentes, o que proporciona um Semivariograma quase horizontal e constante. No centro temos um conjunto de amostras com alguma correlação espacial, o efeito de pepita é já quase zero, existe já um patamar quase definido. No conjunto da direita, a correlação espacial entre as amostras é muito maior, pelo que o Semivariograma gerado é quase "perfeito".

4.2.4 Estimação – algoritmos de interpolação de krigagem

A partir da definição do modelo global de semivariograma é, então, possível estimar valores do fenómeno espacial em pontos desconhecidos da área de estudo. Neste tipo de situações é habitual utilizarem-se os métodos de interpolação de krigagem. O nome de krigagem designa um conjunto de algoritmos de regressão linear, com base num critério de

mínimos quadrados, em homenagem ao investigador Daniel Kridge que se apercebeu, durante o seu trabalho de prospecção nas minas sul africanas, que a variância de dispersão dos teores reais de ouro num painel de pesquisa de uma mina é sempre mais fraca que a variância de dispersão dos resultados de uma amostragem interior, estando subjacente um erro sistemático da sobrestimação dos painéis ricos e subestimação dos pobres. Apenas se poderá evitar este tipo de erros, se a estimação do teor médio de determinada zona incluir toda a informação disponível e ponderada por coeficientes que tomem em consideração a estrutura espacial do fenómeno, e portanto a sua irregularidade[16].

Os algoritmos de krigagem têm por base o estimador linear de regressão [$Z(x_0)^*$], definido por:

$$Z(X_0)^* - m = \sum_{i=1}^{N} \lambda_i [Z(X_i) - m_i] \qquad (4.12)$$

Onde $?_i$ são ponderadores a afectar às amostras e $Z(x_i)$ representa os valores da VR em vários pontos das amostras do espaço i=1, 2, ..., N na vizinhança de x_0. m e m_i são os valores esperados de $z(x_0)$ e $z(x_i)$, respectivamente. Este estimador é construído com base nas seguintes condições:

1. A média dos desvios entre os valores reais e os estimadores deve ser nula – condição de não enviezamento universal: $E[Z(x_0) - Z(x_0)^*] = 0$, de modo a desagregar os agrupamentos preferenciais de amostras. Amostras agregadas têm menos peso que amostras isoladas.
2. A variância de estimação, $s_E^2 = Var[Z(x_0) - Z(x_0)^*]$, deve ser mínima.

A krigagem é actualmente utilizada devido principalmente ao facto de requerer apenas a modelação do semivariograma e possibilitar o cálculo da variância de estimação através de ponderadores, não dependendo dos valores da VR. Apresenta como principais propriedades :

1) Exactidão. As estimativas das amostras $Z(x_a)$ em locais conhecidos são idênticas aos seus valores experimentais:
2) Regularidade. Os interpoladores de krigagem fornecem sempre uma solução única e
3) Variância de estimação não depende dos valores das amostras $Z(x_a)$. Isto significa que a mesma configuração amostral (com as mesmas relações geométricas entre amostras e o ponto a estimar) reproduz sempre a mesma variância de estimação em

qualquer parte da área A. Assim, a variância de estimação não mede a incerteza da estimação resultante da variabilidade local [13].

As diferenças existentes entre os interpoladores de krigagem assentam na definição do modelo de função aleatória Z que é habitualmente decomposta numa componente residual R e numa componente de deriva m: Z = R + m [14]. A componente residual é modelada como uma função aleatória estacionária de média zero e covariância $C_R(h) = E[R(x)*R(x+h)]$, o que leva a que o valor esperado da FA Z na posição x_0 seja dado pelo valor da componente de deriva naquele local, ou esperança de Z. Consoante o modelo de deriva m escolhido, três variantes da krigagem podem ser consideradas: Krigagem Simples (KS); Krigagem Ordinária (KO) e Krigagem Universal (KU).

4.2.4.1 Krigagem Simples

Neste tipo de krigagem a componente de deriva é modelada como a média estacionária dos valores da amostra. Esta média m é conhecida e constante na área de estudo. Deste modo, com base na expressão (4.12) podemos definir o estimador de krigagem simples como:

$$Z(X_0)^* = \sum_{i=1}^{N} \lambda_i [Z(X_i) - m] + m \tag{4.13}$$

Os N ponderadores $?_i^{KS}$ são determinados de modo a minimizar a variância do erro s_E^2 de acordo com a segunda condição do estimador geral de krigagem. Este algoritmo é normalmente aplicado em situações em que o formalismo teórico do modelo impõe o conhecimento da média da função aleatória, ou então quando existe um bom conhecimento da tendência (deriva) do fenómeno. Nestes casos podem assumir-se os valores dessa deriva (quando conhecida em todo o campo) como a média das variáveis aleatórias referentes aos valores amostrados e aos pontos no espaço não amostrado [13].

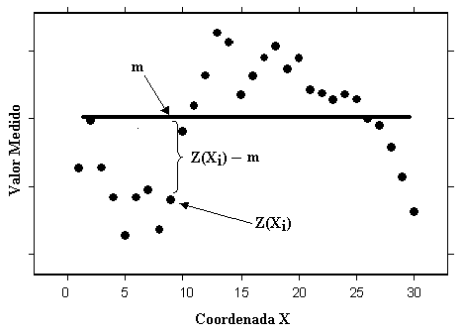

Figura 4.13 – Krigagem Simples

4.2.4.2 Krigagem Ordinária

Na KO a média m é desconhecida e varia localmente na área de estudo. Isto significa que a estacionariedade da média é limitada a uma dada vizinhança V, centrada na posição a estimar x_0 [14]. Apesar de desconhecida, a média m pode ser filtrada da estimação forçando a que a soma dos ponderadores de krigagem seja igual a 1 [14]. Assim, o estimador de krigagem é expresso por

$$Z_{KO}(X_0) = \sum_{i=1}^{N} \lambda_i^{KO} Z(X_i) \qquad \text{com} \qquad \sum_{i=1}^{N} \lambda_i^{KO} = 1 \tag{4.11}$$

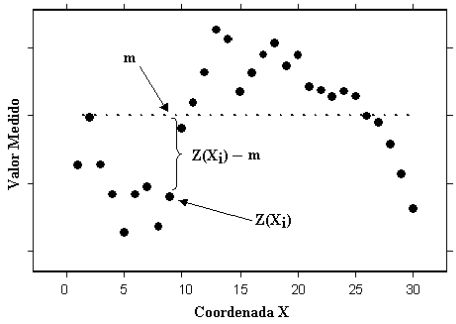

Figura 4.14 – Krigagem Ordinária

4.2.4.3 Krigagem Universal

Neste caso, o modelo de deriva considera que a média varia localmente dentro de uma determinada vizinhança V em função das coordenadas, segundo a expressão geral

$$m = \sum_{k=1}^{K} a_k f_k \tag{4.15}$$

onde as funções f_k são conhecidas e os coeficientes a_k desconhecidos e constantes.

Deste modo a deriva fica modelada por uma polinomial. Assim, utilizando a expressão (4.12) podemos descrever o estimador de KU, como:

$$Z(X_0)^* = \sum_{i=1}^{N} \lambda_i Z(X_i) + \sum_{k=1}^{K} a_k \left[f_k(X_0) - \sum_{i=1}^{N} \lambda_i f_k(X_i) \right] \tag{4.16}$$

Contudo, para o cálculo dos coeficientes de deriva a_k é necessário impor os seguintes constrangimentos:

$$\sum_{i=1}^{N} \lambda_i \, f_k(X_i) = f_k(X_0) \tag{4.17}$$

com k=0, 1 ..., K;

o que leva à seguinte expressão para o estimador de KU

$$Z_{KU}(X_0)^* = \sum_{i=1}^{N} \lambda_i^{KU} Z(X_i) \tag{4.18}$$

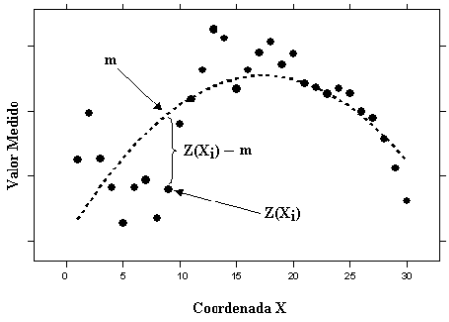

Figura 4.15 – Krigagem Universal com deriva de 2ª Ordem

4.2.5 Considerações sobre os ponderadores do estimador de krigagem

A determinação dos ponderadores que irão ser usados na estimação do ponto desconhecido resumem dois factores:

- A distância estrutural (segundo membro): quanto mais próximas estiverem as amostras do ponto desconhecido maior peso terão no estimador;

- A desagregação (primeiro membro): quanto mais correlacionadas estiverem as amostras, maior o efeito de agrupamento ou redundância e menor será o seu peso individual na construção do estimador. [13]

$$[?] = [K]^{-1} * [M] \tag{4.19}$$

O primeiro membro desta equação é a inversa da matriz que resulta de se aplicar o modelo escolhido através do semivariograma. Essa matriz não é mais do que a variância existente entre todas as amostras vizinhas ao ponto que se pretende estimar.

O segundo membro surge aplicando a cada ponto vizinho conhecido do ponto desconhecido o modelo determinado no semivariograma, em função da distância a cada amostra vizinha.

4.3 Conclusão

Existem dois grupos de algoritmos de interpolação de superfícies, os Determinísticos e os Geoestatísticos. Enquanto que os primeiros tentam ajustar ao conjunto de pontos conhecidos uma função matemática que modele as características do fenómeno em causa, os segundos usam estatísticas e modelos de previsão que podem modelar incerteza e erros de medição. A Geoestatística compõe-se de duas fases principais, a Variografia, onde se quantifica a forma como os dados variam consoante a distância entre si, e a Predição, onde se estimam valores para os pontos desconhecidos. No primeiro processo tenta-se encontrar o modelo que melhor descreva a auto-correlação existente entre os dados, modelo esse que se traduz num Semivariograma. É o Semivariograma que vai permitir estimar valores desconhecidos; existem diversos tipos de Semivariogramas, entre os quais se destacam três, Esférico, Exponencial e Gaussiano.

Dentro dos métodos Geostatísticos, consideraram-se a Krigagem Simples, a Krigagem Ordinária e a Krigagem Universal. A Krigagem Universal comporta derivas polinomiais, ao passo que os outros dois métodos não, usando apenas médias dos valores medidos. No entanto a Krigagem Ordinária considera médias locais em redor do ponto de estimação, enquanto que a Krigagem Simples considera apenas a média global, para toda a área de estimação. Por esta razão, a Krigagem Simples foi excluída dos modelos a usar, pois não se revelou realista em relação à amostra deste estudo (este facto será aprofundado no Capítulo V).

Os métodos Geostatísticos comportam melhor as características de variáveis que modelam fenómenos físicos naturais, como é o caso. Tome-se por exemplo a amostra da figura 4.16;

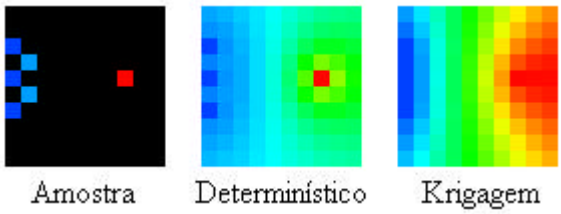

Figura 4.16 – Diferenças entre métodos Geostatísticos e Determinísticos

Existem cinco medições à esquerda com valores baixos, e uma com um valor alto à direita, é de esperar que a amostra da direita tenha uma influência maior na superfície gerada. Usando um método determinístico (IDW), a influência dos pontos mais baixos estende-se bastante pela superfície, e o ponto mais alto aparece como um pico no meio de uma zona de valores médios. Com um método Geoestatístico, não só é dada maior influência ao ponto medido isolado, como é identificada uma deriva (uma inclinação) da que sobe da esquerda para a direita da superfície.

CAPÍTULO V

GERAÇÃO DE SUPERFÍCIES

5.1 Integração dos Dados

Como exemplo significativo de um conjunto de dados a processar com as ferramentas de SIG anteriormente discutidas, foram postos à nossa disposição vários conjuntos de dados, obtidos no âmbito do projecto MAROV. Estes dados, dividem-se em três grupos, primeiro um conjunto de *shapefiles* (*layers* vectoriais do ArcGIS), com a linha da costa das ilhas do Pico e do Faial, linhas de batimetria em torno das mesmas ilhas, e enrocamentos dos molhes e afloramentos rochosos na zona circundante ao Monte da Guia. Depois um conjunto de *rasters* com mapas hidrográficos tanto das duas ilhas como pormenores da zona do Monte da Guia nas escalas 1:500 e 1:3500. Por último um conjunto de mais de 400 mil leituras de profundidade recolhidas por sonar.

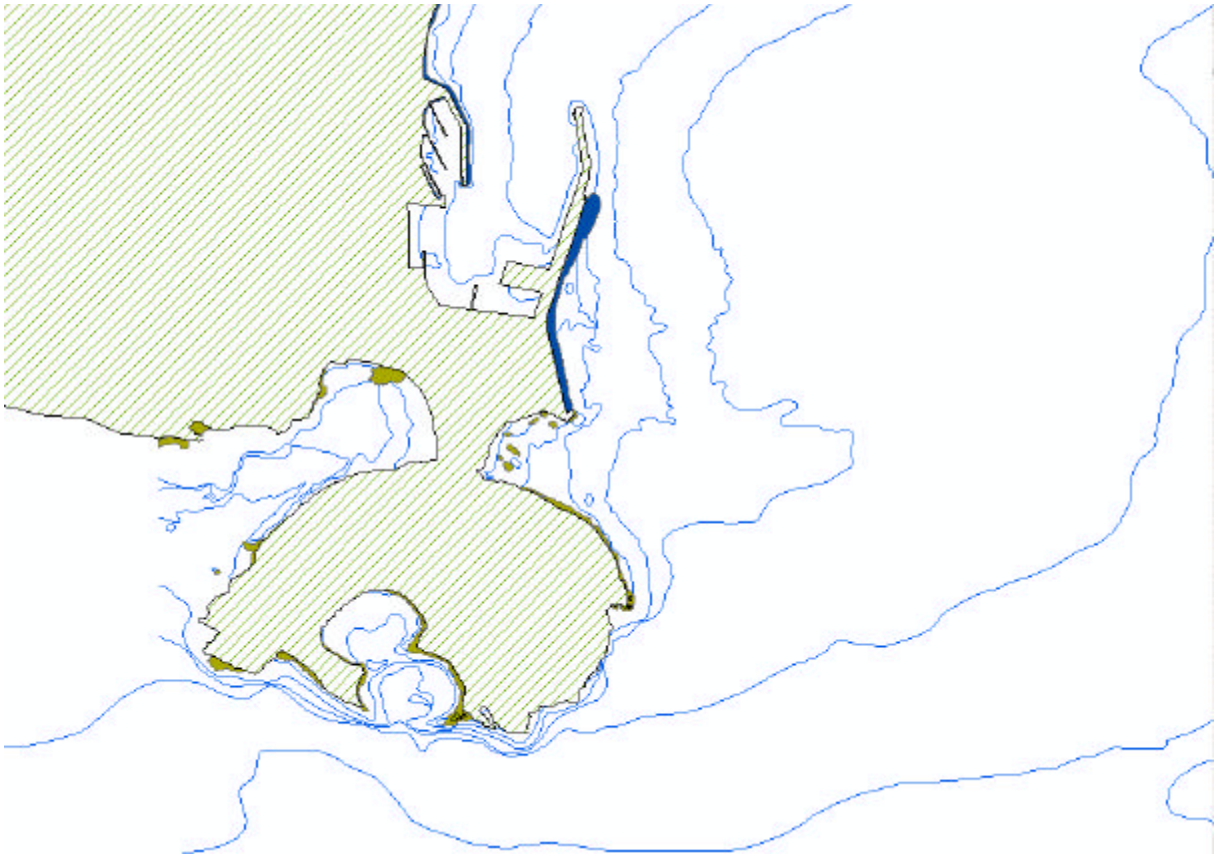

Figura 5.1 – Shapefiles linha da costa, batimetria, enrocamento e afloramento rochoso.

As *shapefiles* encontram-se numa projecção de Mercator no paralelo 38º33'30", num *Datum* WGS84. O passo seguinte foi a georeferenciação do *raster* de um mapa do Monte da Guia e respectiva área circundante numa escala de 1 para 3750. Esta georeferenciação foi levado a cabo em pontos de morfologia bem distinta como a ponta norte do molhe leste, a Ponta da Greta, a Ponta das Furnas, a ponta dos Peixes Reis e a Ponta das Garças.

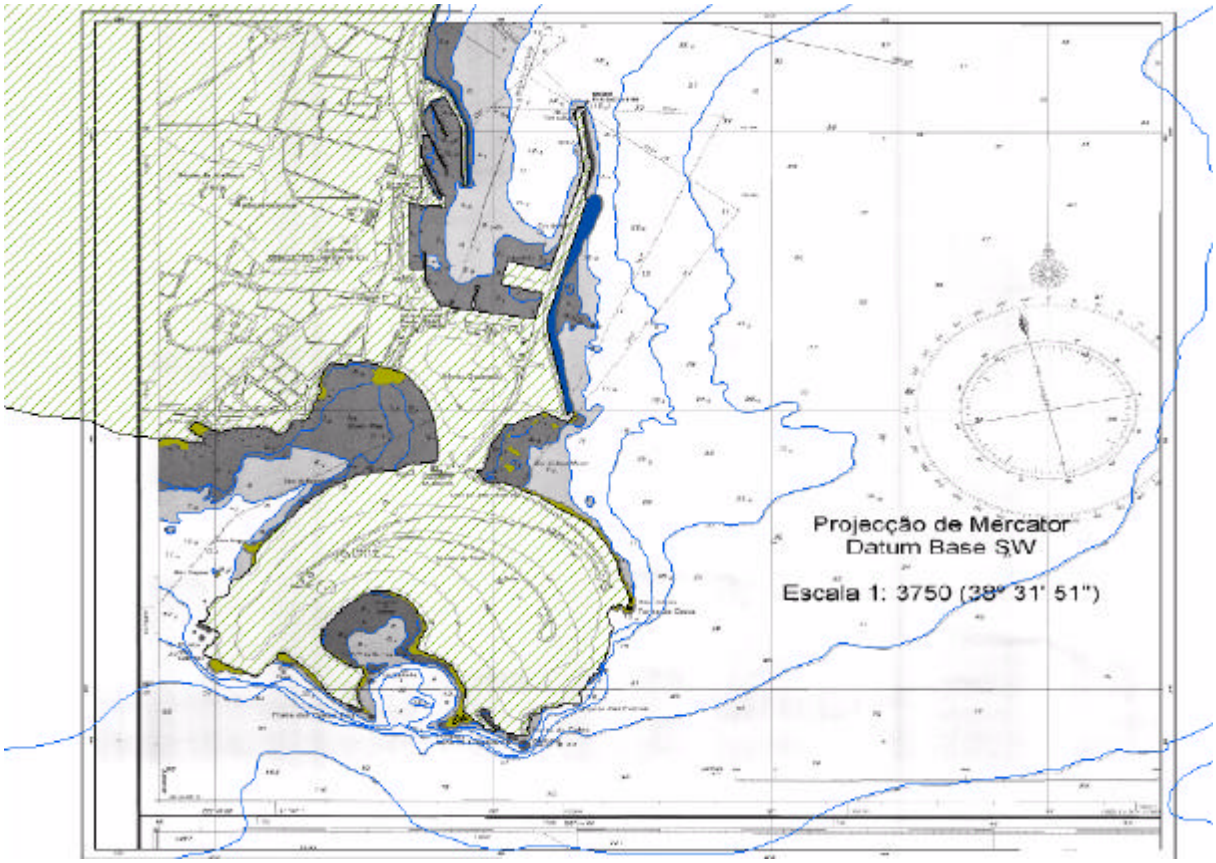

Figura 5.2 – Shapefiles e Mapa georeferenciado.

Com estes dois elementos já integrados, ficou a faltar a inclusão das amostras de profundidade no conjunto. As amostras encontravam-se inicialmente num ficheiro de texto, com quatro campos diferentes por amostra, um valor em longitude (eixo xx), um valor em latitude (eixo yy), a leitura de profundidade no ponto (eixo zz) e um valor de variância.

A primeira operação foi transportar estes dados para o formato ArcGIS, criando com eles uma Base de Dados própria para a continuação do trabalho. O ArcGIS interpreta automaticamente os valores de x e y como medidas de posição e pode já a esta altura mostrar, cada registo da Base de Dados criada, como um ponto num *layer* vectorial. Apesar destes automatismos, a integração deste conjunto de dados seria bem mais laboriosa, uma vez que não se poderia fazer através de uma georeferenciação como no *raster*, teria que ser efectuada uma projecção no espaço dos dados já integrados.

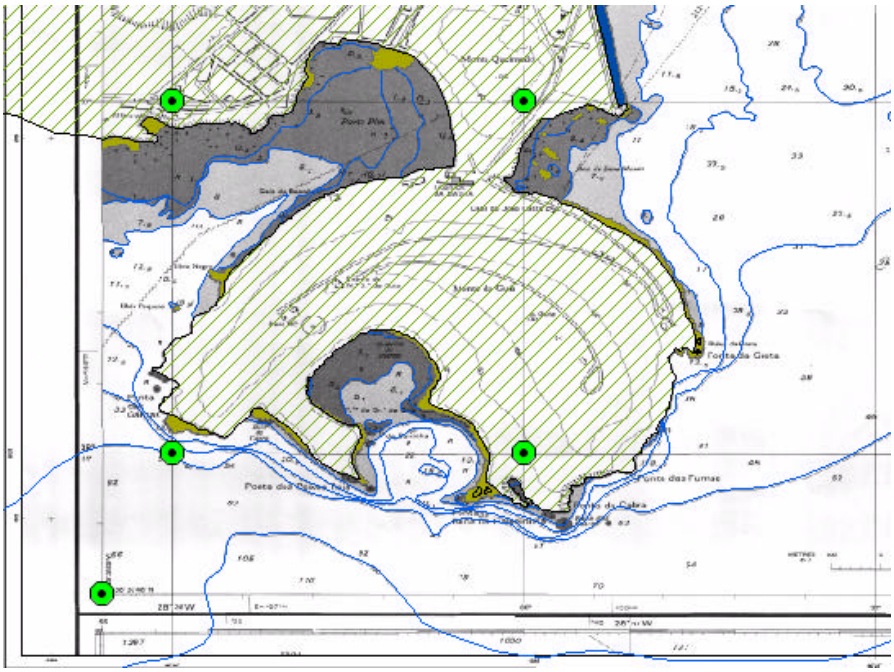

Figura 5.3 – Pontos de referência para projecção

Os valores de posição indicam a distância, em metros, de cada amostra relativamente ao ponto de coordenadas 38º30’48” Norte e 28º38’06” Oeste. Este ponto coincide com o canto inferior esquerdo do mapa já integrado, e representa intersecção dos eixos xx e yy, o ponto de coordenadas (0,0). Com esta informação adicionamos à Base de Dados pontos de referência nas intersecções dos meridianos e paralelos representados no mapa *raster*.

Tabela 5.1 – Pontos auxiliares de projecção

Longitude	Latitude	X	Y
28º 38’ 06" W	38º 30’ 48" N	0	0
28º 38' 00" W	38º 31' 00" N	145,081519	371,0648
28º 38' 00" W	38º 31' 30" N	145,081519	298,72691
28º 37' 30" W	38º 31' 00" N	870,489113	371,06483
28º 37' 30" W	38º 31' 30" N	870,489113	1298,72691

Tendo estes pontos como guias, passou-se à projecção deste *layer*. Seguindo a projecção do *layer* do vector da linha da costa, usou-se uma Projecção de Mercator no paralelo 38º 33' 30" Norte, no mesmo *Datum*, WGS84.

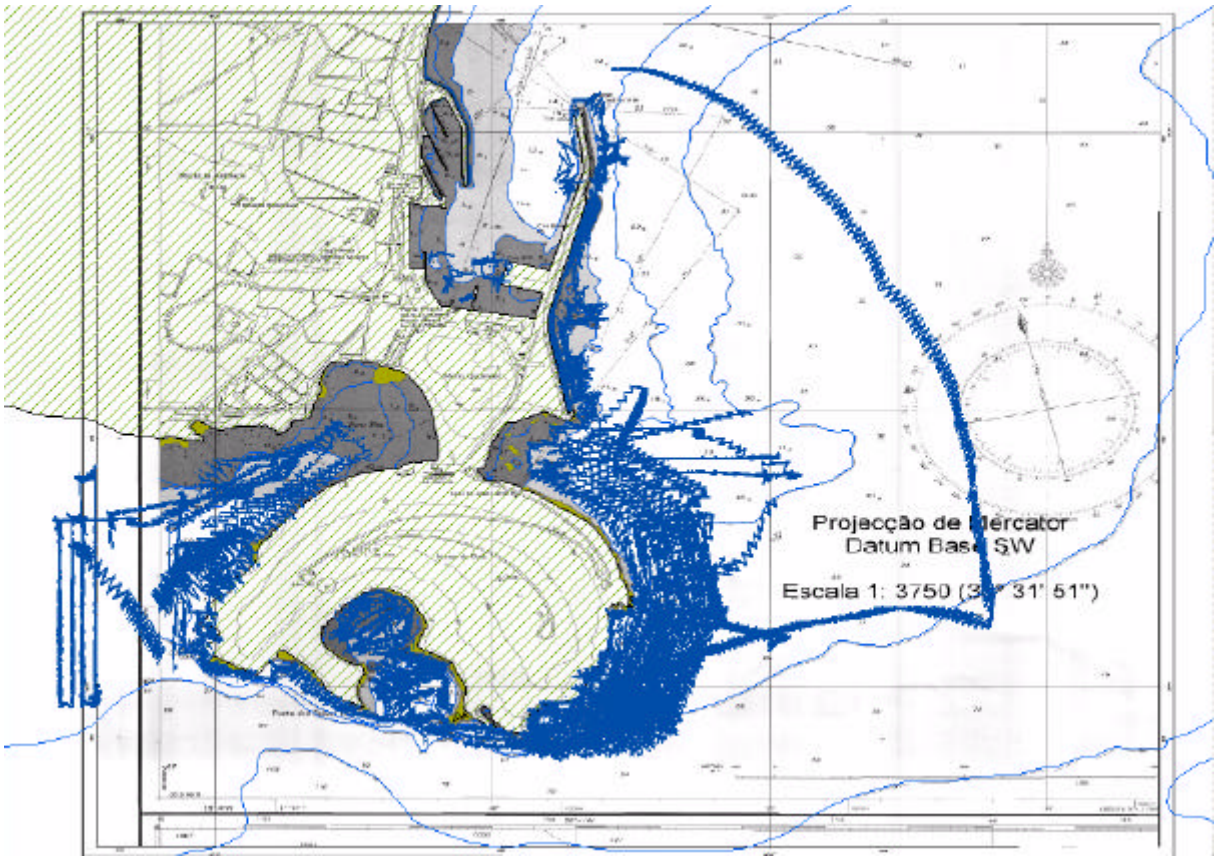

Figura 5.4 – Shapefiles, Mapa georeferenciado e Pontos de leitura.

Como se pode observar pela figura, apesar de haver um grande quantidade de amostras, existiam ainda zonas consideráveis sem informação de profundidade na Base de Dados, tanto no mar como em terra, sendo completamente inexistente na última. Recorrendo à informação de altimetria e batimetria, presente no mapa *raster* sob a forma de curvas de nível, foi enriquecida a Base de Dados com esta informação adicional. Foram introduzidos, entre outros, pontos sobre a linha da costa, nos 10, 50 e 100 metros de altura, e nos 20, 30, 50 e 100 metros de profundidade. Este conjunto final de amostras abrange uma zona com aproximadamente 2271 metros de ordenadas por 2612 metros de coordenadas.

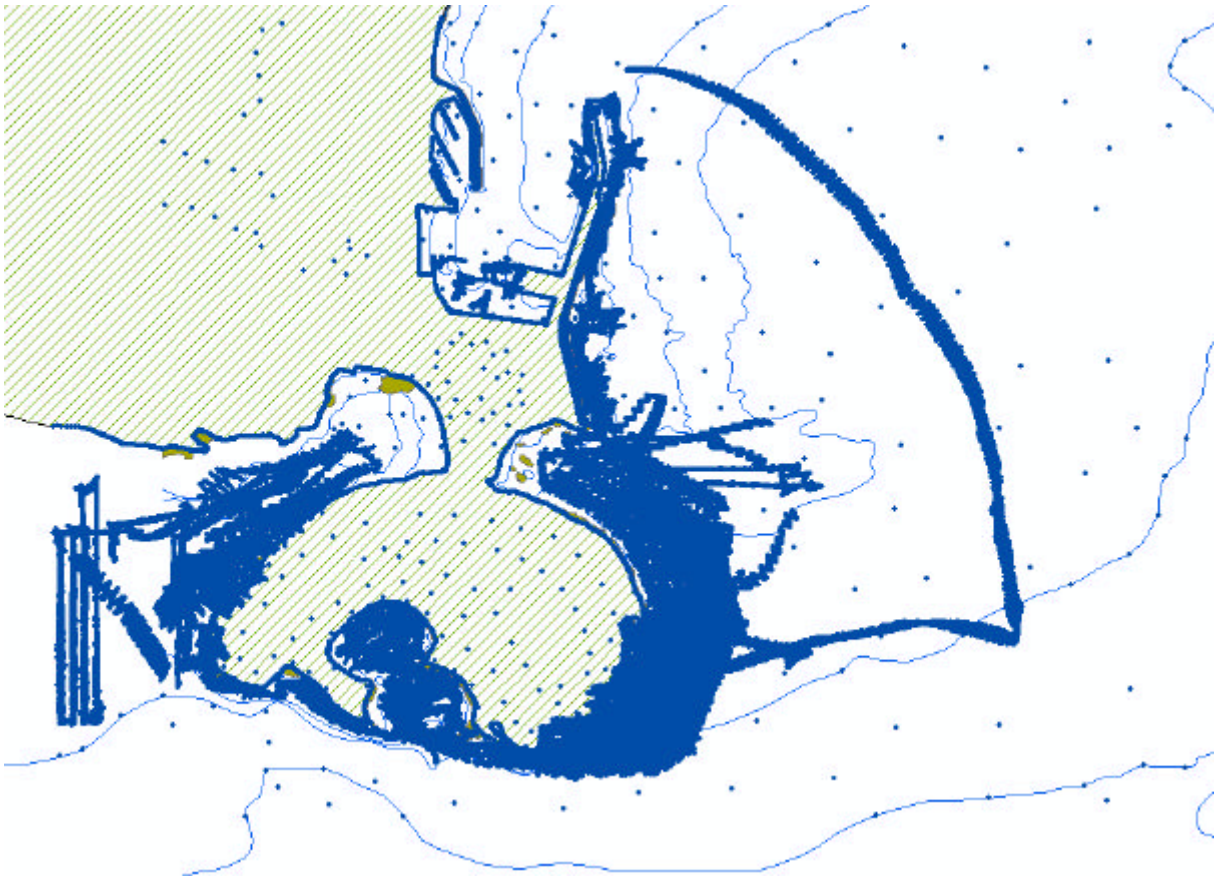

Figura 5.5 – Shapefiles, pontos de leitura, pontos de batimetria e altimetria

Com esta nova informação adicionada ao conjunto de informação puderam-se finalmente aplicar os modelos geoestatísticos explorados no capítulo IV.

5.2 Análise Exploratória Espacial

Antes de se dar início à interpolação de superfícies, é parte integrante do trabalho de Geoestatística a fase de análise exploratória espacial. É nesta primeira fase que é tomado conhecimento com a distribuídos os dados, sendo possível chegar a algumas conclusões acerca dos métodos que se enquadram melhor no conjunto de amostras.

De forma a facilitar este trabalho o ArcGIS Geostatistical Analyst põe ao dispor a ferramenta Exploratory Spatial Data Analysis (ESDA), que incluí um conjunto de utilitários que mostram uma série de facetas dos dados em estudo. O ESDA cria Histogramas, mapas de Voronoi, modelos a três dimensões da distribuição de amostras, entre outros, que fornecem um conjunto vasto de características preciosas acerca dos dados. Infelizmente um destes utilitários, o mapa de nuvens de Semivariância/Covariância, não pode ser usado dado que apenas está preparado para trabalhar com conjuntos de no máximo trezentos pontos, muito aquém dos quatrocentos mil em estudo.

De seguida é apresentada e interpretada a informação dada pelo ESDA acerca da amostragem em estudo.

5.2.1 Descrição Univariada das amostras

A Mediana, na figura 5.6, indica que 50% dos valores estão acima de -15,16 metros e que os outros 50% estão abaixo de -15,16 metros, é o "meio" da distribuição. A vantagem da mediana em relação à média é que a mediana não é sensível a valores extremos também denominados *outliers*. Como a mediana e a média, neste caso onde existe uma grande amostragem, estão relativamente próximas, indicam que não existem muitos desvios de valores nas amostras. Comprovado igualmente pelos valores máximo e mínimo presentes na amostragem.

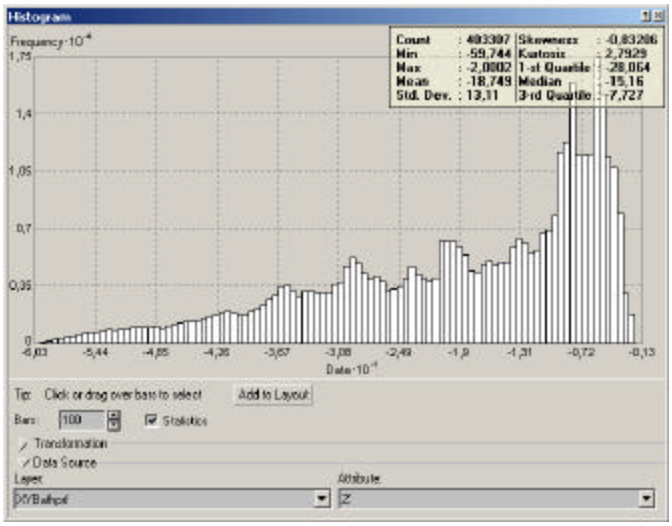

Figura 5.6 – Histograma inicial

Analisando o desvio padrão pode-se constatar que apesar da média ser de -18,749 metros, como existe um grande desvio padrão pode-se garantir que existe alguma inclinação. Esta inclinação corresponde á tendência que as amostras têm de apresentar valores mais profundos à medida que se afastam da costa.

Pela análise do coeficiente de assimetria (*skewness*) verifica-se que a distribuição não é simétrica, caso em que a média é menor que a mediana. Mais propriamente está-se perante uma distribuição assimétrica negativa com o valor de assimetria de -0,83206. Quando assim acontece as médias locais baixam e as variâncias aumentam pois existe maior discrepância entre as amostras. Este fenómeno é justificado pela influência que tem a distância a que estão o máximo e mínimo de uma distribuição assimétrica. São estas situações caracterizadas por

um aumento das variâncias locais que normalmente implicam a inferência espacial com maior incerteza e maiores erros [13].

Analisando agora os quartis pode-se concluir que as amplitudes entre o primeiro quartil e a mediana, e a mediana e o terceiro quartil são aproximadamente equidistantes. No entanto, o número de amostras não se encontra bem distribuído face ao intervalo considerado, isto é, o intervalo entre o valor máximo e mínimo presentes nas amostras. Analisando ainda a amplitude inter-quartil (Q¾ - Q¼) concluí-se que 50% dos elementos do meio da amostra, estão contidos num intervalo com uma amplitude de 20,337 metros, o que vai de encontro à área de estudo do fenómeno físico.

O valor *kurtosis* é uma medida que revela se a distribuição dos dados é achatada ou pontiaguda quando comparada com a distribuição normal, cujo valor é 3, e indica com que probabilidade a distribuição irá criar *outliers*. Em termos grosseiros indica a distância a que estão os valores máximo e mínimo da média da distribuição. O valor *kurtosis* de uma distribuição normal é de 3, e é dado pela seguinte expressão:

$$kurtosis = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^4}{(N-1)s^4} \qquad (5.1)$$

onde x_i são os valores das amostras, x é a média, N o número total de amostras e s o desvio padrão.

Com os dados em questão, verifica-se que o valor de *kurtosis* é de 2,7929, o que representa uma distribuição relativamente achatada quando comparada com uma distribuição normal. Após terem sido introduzidos os 6203 pontos adicionais, mostrados na figura 5.6, verificou-se que as únicas alterações significativas foram, por um lado o aumento do valor de *kurtosis* e por outro um pequeno aumento quer do primeiro quartil, quer da mediana. Os pontos introduzidos foram alguns pontos de batimetria, altimetria e pontos da linha da costa onde a cota é zero.

A alteração que se verificou no valor do *kurtosis* foi pouco significativa e apenas tornou a distribuição ligeiramente pontiaguda quando comparada com uma distribuição normal. As alterações ocorridas com o primeiro quartil e com a mediana podem ser justificadas com a diferente proporção de valores que foram introduzidos. Foram introduzidos mais valores de altimetria e da linha da costa, valores esses que são superiores à mediana, e que provocaram um ligeiro desvio desta, de -15,16 para -14,813.

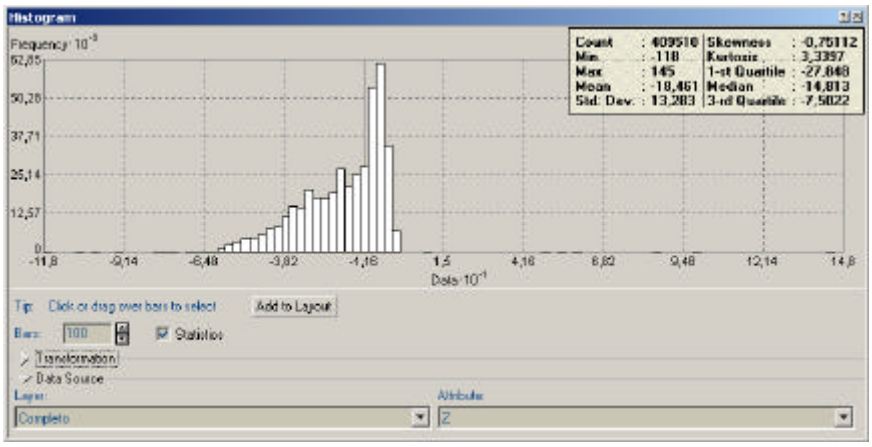

Figura 5.7 – Histograma após a introdução de pontos

5.2.2 Descrição Espacial

Uma das ferramentas que é incluída no ESDA para descrição espacial é o analisador de derivas. Esta ferramenta após por no plano xy as amostras faz uma projecção dos pontos nos planos xz e yz. De seguida ajusta funções polinomiais que permitam descrever as derivas a que os dados estão sujeitos.

Com os dados iniciais, figura 5.8, constata-se a existência de derivas, quer no sentido Norte – Sul (linha azul) quer no sentido Este - Oeste (linha verde) pois as linhas não são horizontais.

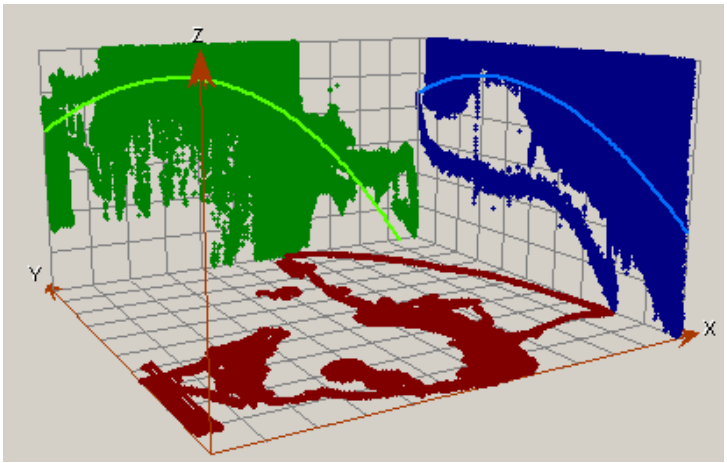

Figura 5.8 – Deriva nos dados iniciais

Com a inclusão dos pontos de batimetria, altimetria e linha da costa, figura 5.9, as derivas continuaram a ser claramente visíveis. De modo a apurar em que sentido se faz sentir exactamente a deriva, rodaram-se os dados, no sentido contrário ao dos ponteiros do relógio, e verificou-se que a deriva existe e se faz sentir no sentido Noroeste – Sudoeste, claramente

visível na figura 5.10. Pela sua forma em "U" o melhor ajustamento será através de um polinómio de segunda ordem.

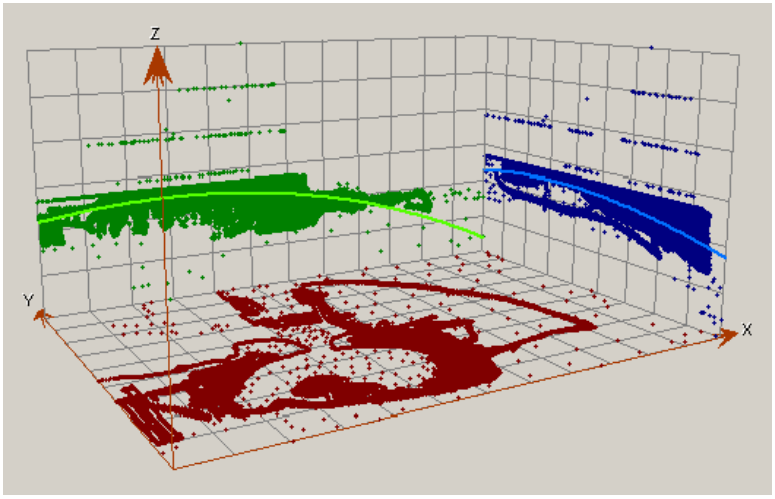

Figura 5.9 – Deriva após a inclusão de pontos

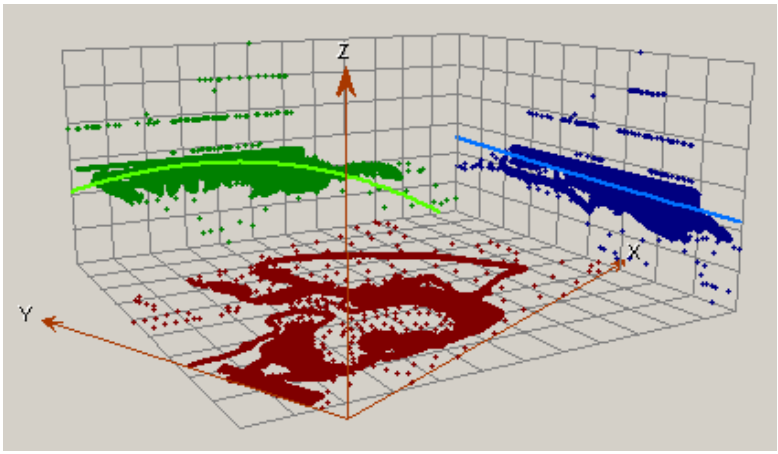

Figura 5.10 – Deriva Noroeste - Sudoeste

A existência de uma deriva é igualmente confirmada se for tido em linha de conta o campo onde estas amostras foram recolhidas. No campo, podemos observar, à medida que nos deslocamos de Noroeste para Sudeste, que começamos com valores de altimetria bastante elevados e chegamos a valores de profundidade que vão aumentando gradualmente nesse sentido.

Uma das implicações imediatas desta conclusão é a ineficácia do método de Krigagem Simples para este conjunto de dados, pois existe uma deriva. Este método trabalha com um média constante ao longo de toda a área de estudo, o que é irreal quando existe uma tendência clara nos dados.

5.2.3 Mapas de Voronoi

Os mapas de Voronoi são compostos por uma série de polígonos formados em torno de cada ponto de amostragem. Os polígonos de Voronoi são criados de tal maneira que, cada ponto dentro do polígono esteja mais perto do ponto de amostragem contido no polígono, do que de qualquer outro. Depois de criados os polígonos são definidos os vizinhos, sendo estes todos os que partilham arestas com o polígono da a amostra em causa.

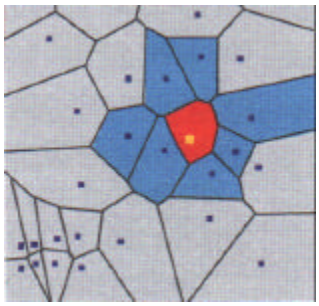

Figura 5.11 – Exemplo de um Mapa de Voronoi

Neste exemplo podemos ver uma amostra a amarelo cercada por um polígono de Voronoi a vermelho. Os polígonos azuis são os vizinhos da amostra amarela. Todos os pontos do polígono vermelho estão mais perto da amostra amarela do que qualquer outra.

Usando estas definições podem-se calcular uma série de diferentes estatísticas úteis. Por exemplo a média local encontra-se calculando a média dos pontos nos polígonos azuis e vermelho, ou seja a média local num polígono é média do seu valor e dos valores dos seus vizinhos. Seguidamente apresentam-se os métodos de cálculo dos mapas de Voronoi utilizados neste estudo.

Simples – o valor atribuído a um polígono é o mesmo do ponto de amostragem nele contido. Dá indicações acerca da influência local de cada ponto.

Alcance Inter-Quartil (IQR) – são calculados o primeiro e terceiro quartil a partir da distribuição dos valores das amostras do polígono e dos seus vizinhos. O valor atribuído ao polígono de Voronoi é a diferença entre o valor do primeiro e do terceiro quartis. Dá indicações acerca da variabilidade local.

Aglomerado – os polígonos são todos agrupados em cinco classes, se a classe de um polígono for diferente de todas as dos seus vizinhos este é considerado neutro (representado a cinzento). Este método permite identificar *outliers*.

Média – o polígono toma o valor da média do seu valor com os dos seus vizinhos. Permite asseverar da "suavidade" local.

Foram usados alguns dos tipos de mapas de Voronoi gerados pelo ESDA, quer para tomar conhecimento das características dos dados, quer para perceber o modo como a inclusão de valores de altitude e batimetria afectaram estas características.

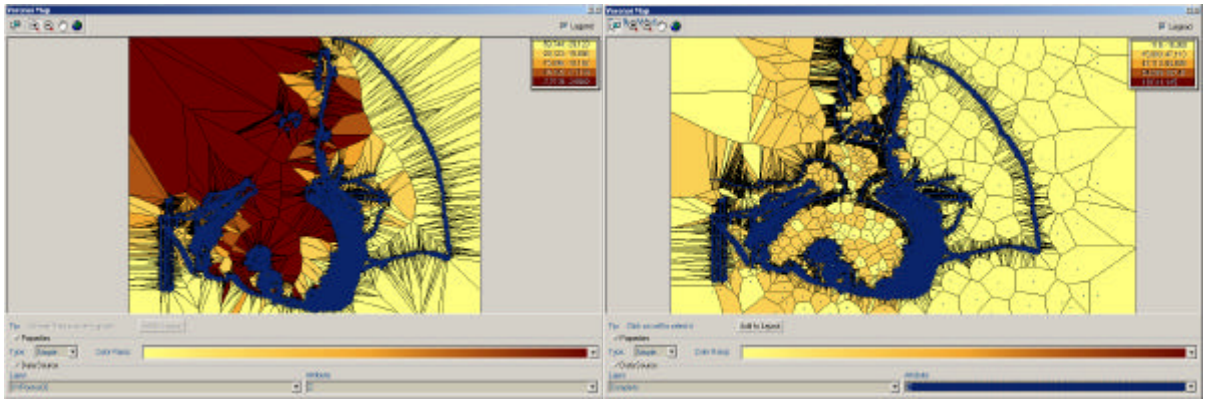

Figura 5.12 – Mapas Voronoi Simples, inicial (esquerda) e com pontos adicionais (direita)

A introdução dos novos dados vem, como esperado, tirar influência a algumas das amostras do conjunto inicial. Principalmente em terra, onde a informação era totalmente ausente, as novas amostras vêm retirar peso a algumas das amostras obtidas perto da linha da costa, que como tal dariam um indicação bastante errada acerca da evolução da superfície em terra. Daqui se concluí que neste capítulo a inclusão de amostras adicionais foi bem sucedida.

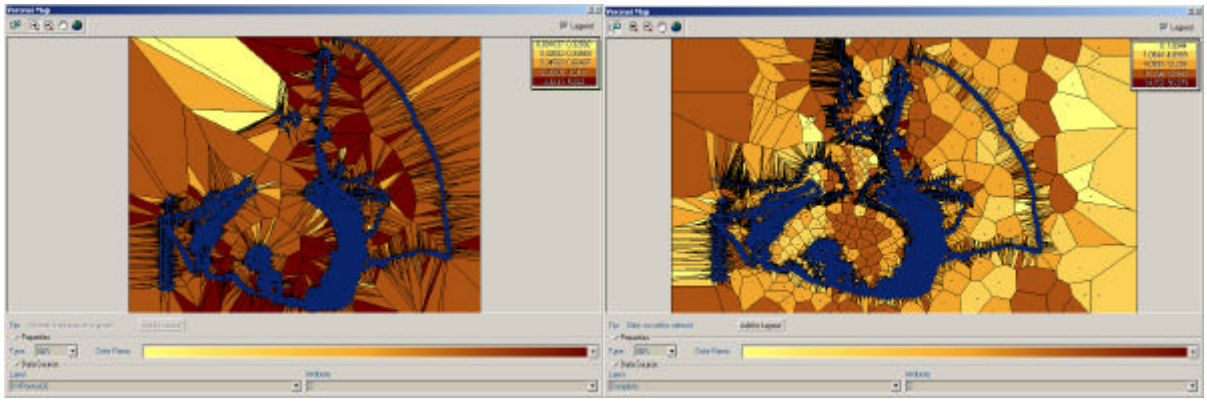

Figura 5.13 – Mapas Voronoi IQR, inicial (esquerda) e com pontos adicionais (direita)

Outro dos objectivos da inclusão de amostras adicionais fica igualmente patente como bem sucedido, olhando para os mapas de Voronoi IQR. Em praticamente todo o espaço da amostragem há uma diminuição da variabilidade local.

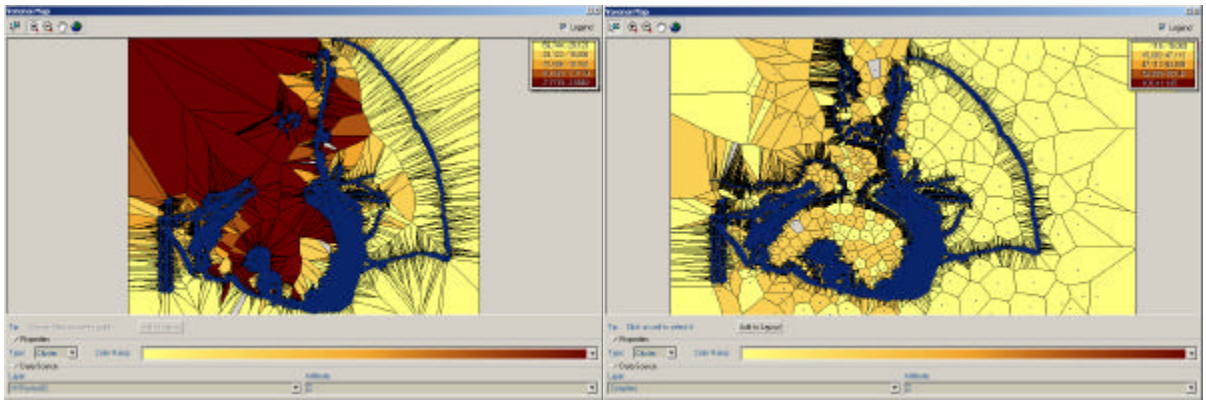

Figura 5.14 – Mapas Voronoi Aglomerado, inicial (esquerda) e pontos adicionais (direita)

Tal como visto na análise dos histogramas, a introdução de novos dados não mudou significativamente a probabilidade de existência de *outliers*. Pode-se no entanto constatar, nos mapas de Voronoi Aglomerado, que a existência destes deixa de ser tendencialmente no mar para passarem a estar em terra.

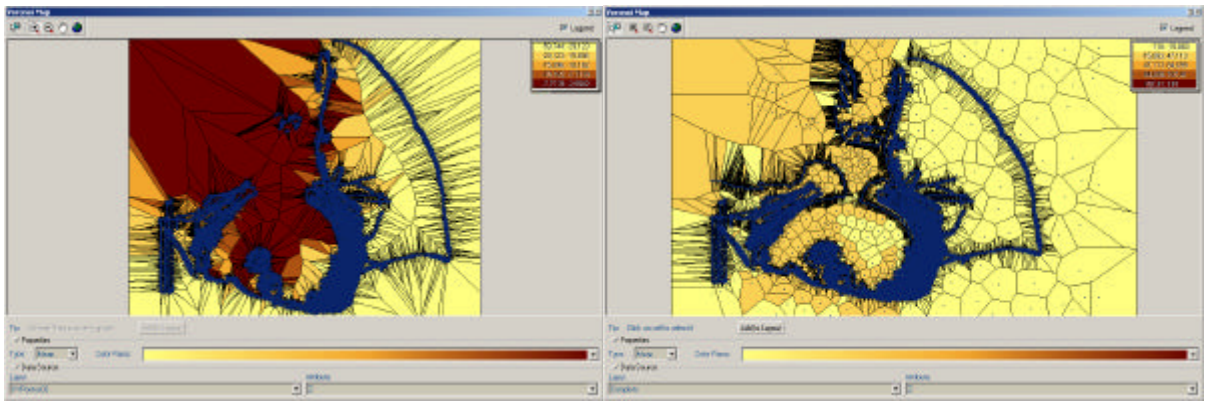

Figura 5.15 – Mapas Voronoi Média, inicial (esq) e com pontos adicionais (dir)

Nestes mapas fica patente que as novas amostras incluídas propiciam uma superfície mais suave, dado que vão adicionar informação em zonas para as quais não existia. Este era igualmente um dos objectivos da inclusão de amostras adicionais, e pode ser dado como cumprido.

5.3 Geração de Superfícies

Depois de concluída a análise exploratória espacial dos dados passou-se à geração de superfícies de acordo com os conhecimentos obtidos nesta análise. Os tipos de Krigagem usados foram a Krigagem Ordinária e a Krigagem Universal com derivas de Primeira e de Segunda Ordem. Cada um destes métodos foi usado com três modelos de Semivariograma diferentes, Esférico, Exponencial e Gaussiano, e ainda um modelo de combinação dos três.

Todos estes modelos foram usados com Anisotropia. As polinomiais de deriva usadas nos métodos de Krigagem Universal, foram as geradas pela ferramenta ESDA, que como se pode ver nas figuras 5.9 e 5.10, seguem a deriva identificada anteriormente na análise de estrutura espacial.

Para a exucução dos algoritmos de interpolação foi usada uma máquina com um processador INTEL Pentium III, com uma frequência de relógio de 933 MHz, e com 512 Mbytes de memória RAM.

Este sub-capítulo mostra o caminho levado desde a escolha dos modelos de Semivariograma a usar até às superfícies finais, passando pelos processos de Validação e Validação Cruzada, quer permitem seleccionar os modelos mais correctos para a amostra em causa.

5.3.1 Modelos de Semivariograma

Depois de se ter chegado à conclusão que não seria realístico aplicar a krigagem simples, restaram três hipóteses: krigagem ordinária, krigagem universal de segunda ordem e krigagem universal de primeira ordem. Apesar de se ter verificado que para se tratar a forma em "U" da deriva seria mais correcto utilizar um polinómio de segundo grau, decidiu-se fazer também algumas experiências com a primeira ordem, visto que a estimação é um processo interactivo e que varia muito de caso de estudo para caso de estudo.

Na krigagem ordinária verificou-se que o maior alcance é o mesmo para todos os modelos e que só o menor alcance varia. O modelo que apresenta um valor maior para este último é o modelo exponencial, seguido do modelo esférico e sendo o menor deles no modelo gaussiano. Este fenómeno, devido a existir uma anisotropia mais acentuada no modelo guassiano, provoca uma forma elíptica mais acentuada porque a amplitude entre o maior e menor alcance é maior, como se observa na figura 5.16.

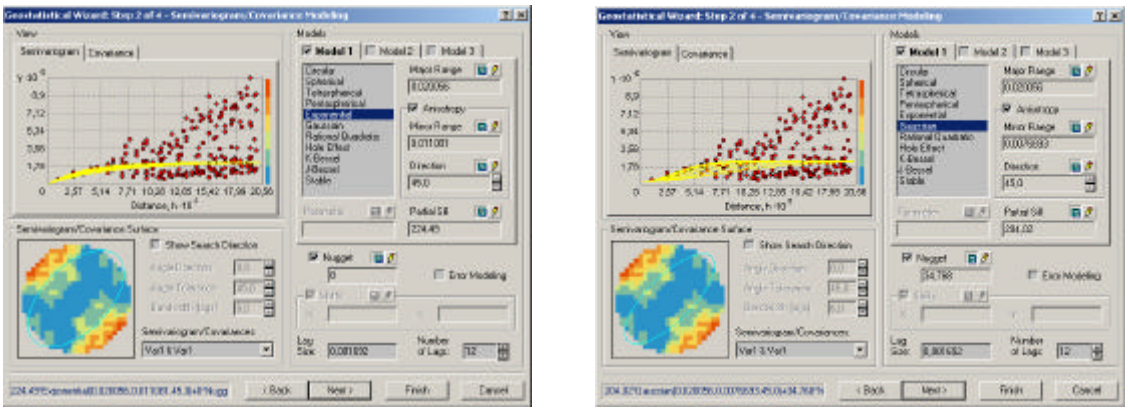

Figura 5.16 – Krigagem ordinária com modelo exponencial e com modelo gaussiano

A esta redução incremental do menor alcance quando percorremos os modelos do exponencial para o gaussiano passando pelo modelo esférico, corresponde um aumento incremental do efeito de pepita quando se percorrem os diferentes modelos no mesmo sentido. Quanto maior for o alcance de menor raio menor será o efeito de pepita fazendo com que as amostras vizinhas exerçam uma maior influência nos pontos a estimar.

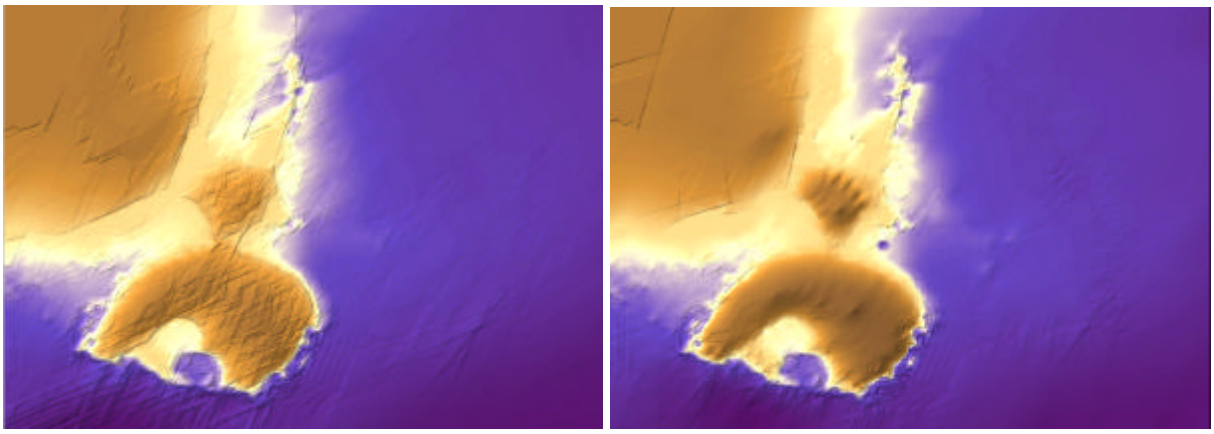

Figura 5.17 –Modelo gaussiano à esquerda e modelo exponencial à direita

Quanto menor for o efeito de pepita menor será a influência do conjunto total de amostras. Zonas onde exista uma grande quantidade de pontos o modelo exponencial (com menor efeito de pepita logo sofre uma maior influência de pontos vizinhos) será o melhor a aplicar, tal como se poderá verificar em cima pela figura da direita. Na zona do monte da Guia onde existe uma grande amostragem de pontos foi notório o desaparecimento de algum do padrão de amostragem que o mapa tinha junto da linha da costa. Por outro lado zonas onde se deverá fazer sentir uma maior influência do conjunto total de amostras, o modelo gaussiano deverá ser o melhor modelo. Neste caso, onde se deveria ter em linha de conta um conjunto de pontos mais dispersos, o modelo exponencial realçou o padrão de amostragem na zona onde foi feita uma viagem mais longe da costa e já em alto mar. Nesta zona o modelo

gaussiano foi claramente superior pois conseguiu esconder esse padrão de amostragem aproveitando melhor a influência dispersa que o conjunto de amostras naquela zona consegue oferecer.

Na krigagem Universal de primeira ordem observou-se o mesmo efeito, isto é, os modelos que têm um efeito de pepita mais elevado apresentam um valor maior no alcance menor. Nesta experiência o resultado que saltou mais à vista foi o do modelo exponencial ser praticamente isotrópico. No modelo gaussiano a direcção da elipse variou cerca de 342º onde se verificou também uma diminuição do patamar alcançado, que provavelmente indicará que num patamar entre a dos modelos exponencial e esférico e entre a do modelo gaussiano a auto-correlação sofre uma alteração na direcção em que se faz sentir, figura 5.18.

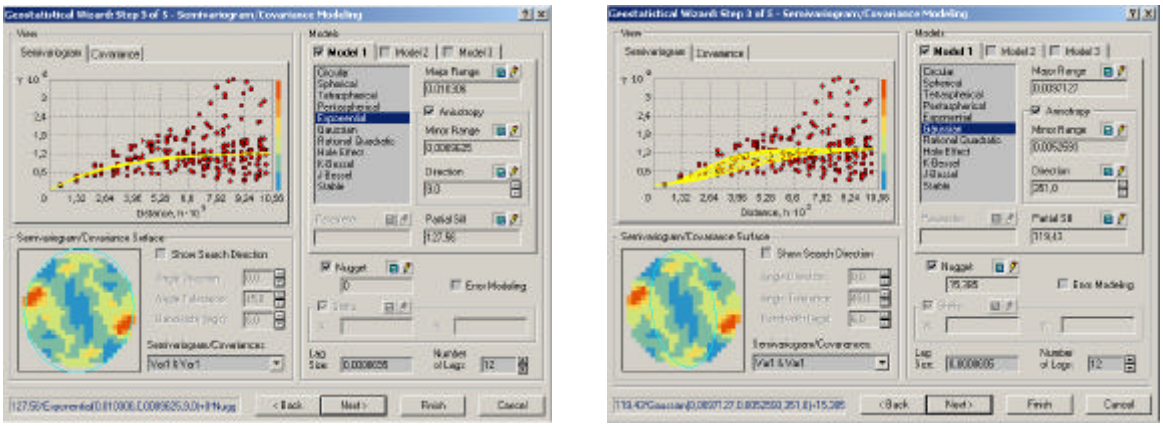

Figura 5.18 – Krigagem Universal com deriva de primeira ordem

Na krigagem Universal mas agora com uma deriva de segunda ordem, verificou-se o mesmo efeito que anteriormente, ou seja, o modelo que apresenta um menor raio de menor alcance tem um efeito de pepita superior. Nesta experiência foi no modelo exponencial que se verificou uma alteração de 315º na direcção da elipse. Tal ocorrência pode-se dever ao facto de o seu alcance ser maior, mas como apresenta uma isotropia quase perfeita, a direcção da elipse não irá influenciar a escolha da melhor linha para modelar o semivariograma, como se pode ver na figura 5.19.

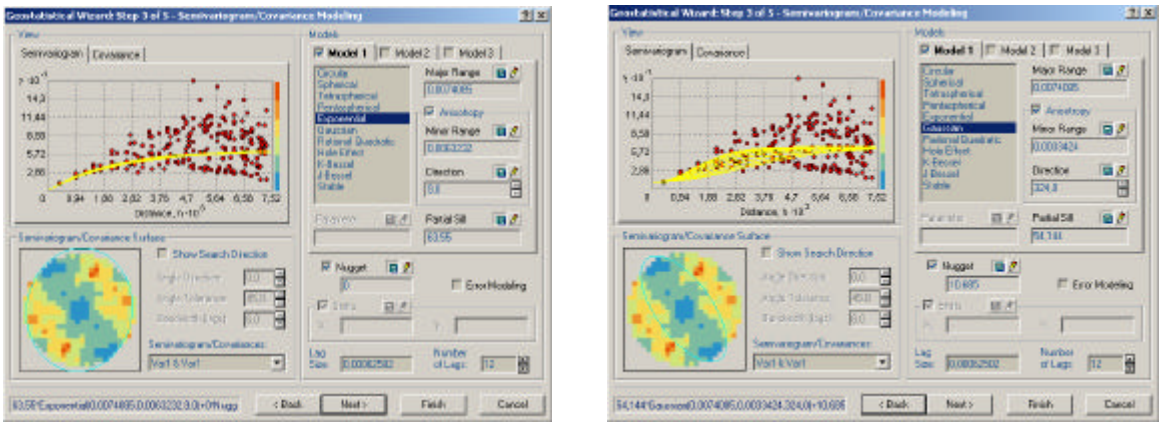

Figura 5.19 – Krigagem Universal com deriva de segunda ordem

De modo a encontrar o modelo mais adequado aos dados em questão recorreu-se aos valores de Validação Cruzada, proporcionados pela ferramenta ESDA.

5.3.2 Validação Cruzada

A Validação Cruzada é construída usando todos os dados conhecidos do modelo. Para cada ponto de amostragem conhecido retira-se o seu valor do conjunto de pontos, aplica-se então o modelo nesta localização espacial e estima-se um valor hipotético, ignorando que a amostra existe. Finalmente compara-se este valor com o valor medido. Este procedimento é repetido para todos os pontos da amostragem, um de cada vez, obtendo-se no final uma regressão linear com os valores medidos contrapostos aos valores estimados para as mesmas localizações.

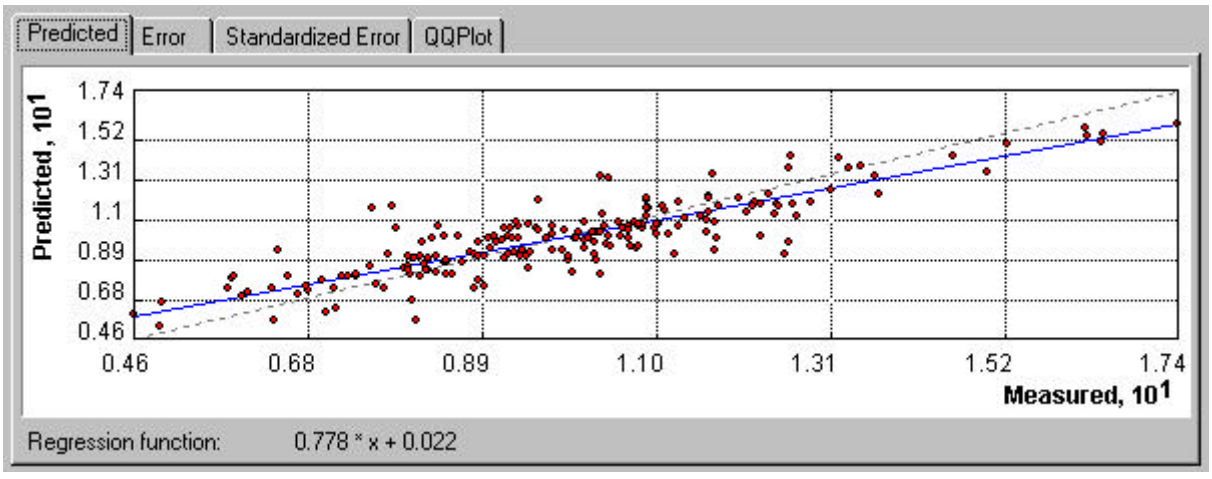

Figura 5.20 – Gráfico de Validação Cruzada

Nos testes de validação cruzada o ESDA apresenta um conjunto de estatísticas que permitem comparar formalmente diferentes modelos. Segue-se uma breve descrição das indicações dadas por estes valores.

Erro Médio (EM) – deve estar perto de zero e indica se a estimação é ou não imparcial. No entanto este erro depende da escala dos dados.

Raiz da Média Quadrada (RMQ) – raiz quadrada da média das distâncias quadradas da estimação à recta da regressão linear da validação cruzada. Este valor dá uma indicação da proximidade a que estão os valores estimados dos medidos. Quanto menor for melhor é o modelo empregue.

Média dos Erros Padrão (MEP) – Deve estar igualmente o mais perto de zero possível. No entanto dá também uma preciosa indicação de como está a ser tratada a variabilidade na estimação, se esta média estiver perto do Raiz da Média Quadrada. Se a MEP estiver acima do RMQ está a ser sobrestimada a variabilidade das estimações, caso contrário (MEP < RMQ) a variabilidade está a ser subestimada.

Erro Médio Aferido (EMA) – é a média dos erros de estimação aferidos. Deve estar perto de zero, dá igualmente indicação da imparcialidade da estimação, mas agora não dependente da escala dos dados (erro aferido é o resultado da divisão do erro de estimação pelo seu erro de estimação padrão).

Raiz da Média Quadrada Aferida (RMQA) – é outra forma de perceber a forma como está a ser tratada a variabilidade. Divide-se cada erro de estimação pelo seu erro padrão estimado, o resultado destas divisões deve ser, em média, semelhante, e a Raiz da Média Quadrada Aferida deve estar perto de um. Se este valor estiver acima de um está a ser subestimada a variabilidade, caso contrário (RMQA < 1) está a ser sobrestimada.

Como primeiro contacto com os valores de validação cruzada, passa-se a analisar os resultados obtidos por cada tipo de Semivariograma usado.

5.3.3 Semivariograma Esférico

Para este tipo de Semivariograma a Krigagem Universal de 2ª Ordem é a que apresenta melhores resultados, este revelar-se-á mesmo como o melhor modelo nesta fase. É o modelo que obtém estimações mais próximas dos valores medidos e o que trata melhor a variabilidade dos dados. Neste capítulo a proximidade entre a RMQ e a MEP e a RMQP dão os melhores valores de todas as simulações levadas a cabo.

Tabela 5.2 – Validação Cruzada para Semivariograma Esférico

	KO	KU 1ª Ord	KU 2ª Ord
Erro Médio	-0,02277	0,01581	-0,01863
Raiz da Média Quadrada	2,075	1,779	1,732
Média dos Erros Padrão	3,058	0,3561	2,124
Erro Médio Aferido	-0,00753	0,02003	-0,01045
RMQ Aferida	0,6701	6,224	0,8294
Diferença RMQ MEP	0,983	1,4229	0,392

Melhor Médio Pior

Quanto aos outros métodos de Krigagem, o método Ordinário obteve aqui diferenças grandes entre valores estimados e medidos, mas revelou uma melhor consideração de variabilidade e incerteza que a Krigagem Universal de 1ª Ordem. Em relação à imparcialidade da estimação embora a KU 1ª ordem tenha um EM mais baixo, este valor é empolado depois de aferido, passando a KO a ter o melhor valor, no entanto é de assinalar a igual descida do EM para o EMA verificado na KU de 2ª ordem.

A inclusão do factor de deriva de 2ª Ordem beneficia aqui bastante a KU, tal como aferido na análise espacial. No entanto, como vamos ver esta situação apenas se verifica para o Semivariograma Esférico.

5.3.4 Semivariograma Exponencial

Para este Semivariograma a KO revela-se novamente menos eficaz que a KU, entre os dois métodos de KU o de 1º ordem revela-se agora ligeiramente melhor. Apesar de o EM e o EMA serem menores para a KU de 2ª ordem, também existe um decrescimento para o valor aferido na KU de 1ª ordem, se bem que não tão acentuado. A RMQ é menor para a KU de 1ª ordem, sendo este último valor o mais acertado acerca da proximidade entre valores medidos e estimados. Quanto à variabilidade o modelo de KU de 1ª ordem é claramente o melhor, mas todos obtém um valor de RMQA bastante alto, dando a indicação de que este tipo de Semivariograma subestima em demasia a variabilidade dos dados.

Tabela 5.3 - Validação Cruzada para Semivariograma Exponencial

	KO	KU 1ª Ord	KU 2ª Ord
Erro Médio	0,01697	0,0165	0,009902
Raiz da Média Quadrada	1,847	1,758	1,813
Média dos Erros Padrão	0,4988	0,4663	0,5121
Erro Médio Aferido	0,01523	0,01549	0,002979
RMQ Aferida	4,591	4,712	5,633
Diferença RMQ MEP	1,3482	1,2917	1,3009

Melhor Médio Pior

5.3.5 Semivariograma Gaussiano

Mais uma vez a KO fica claramente atrás da KU, e novamente a KU de 1ª ordem obtém melhores resultados que a KU de 2ª ordem. A KU de 2ª ordem tem agora um valor mais baixo de RMQ, mas tem uma diferença muito grande deste valor para a MEP. Em relação à imparcialidade das estimações se bem que a KU tem o melhor EM, melhor EMA pertence, por pouca margem, à KO; no entanto é de realçar a descida drástica do EM para o EMA, que se verifica nos três modelos.

Tabela 5.4 - Validação Cruzada para Semivariograma Gaussiano

	KO	KU 1ª Ord	KU 2ª Ord
Erro Médio	-0,01608	-0,01171	-0,01733
Raiz da Média Quadrada	2,433	2,152	1,801
Média dos Erros Padrão	6,043	4,035	4,82
Erro Médio Aferido	-0,00268	-0,00283	-0,00537
RMQ Aferida	0,4022	0,5329	0,4956
Diferença RMQ MEP	3,61	1,883	3,019

Melhor Médio Pior

Com o Semivariograma Gaussiano existe uma tendência para a subestimação da variabilidade dos dados, por oposição ao que acontece com o Semivariograma Exponencial. No entanto é de verificar como a KU de 1ª ordem lida bastante melhor com esta característica deste Semivariograma, obtendo valores bem razoáveis de variabilidade.

5.3.6 Semivariograma Combinado

Com a combinação dos três Semivariogramas são esbatidos alguns dos problemas encontrados na aplicação dos Semivariogramas Exponencial e Gaussiano, no entanto a KU de 2ª Ordem falha redondamente. Tal como tinha lidado melhor com os Semivariogramas

anteriores a KU de 1ª ordem mostra-se mais uma vez eficaz, e finalmente são obtidos resultados para a KO que não ficam muito atrás dos valores da KU. Nomeadamente o valor da RMQ é bastante abonatório para a KO, o que não acontece com a KU de 2ª ordem. A KU de 2ª ordem falha igualmente na imparcialidade da estimação, registando um EMA bastante acima do EM, ao contrário do que se passa nos outros dois modelos.

Tabela 5.5 - Validação Cruzada para Semivariograma Combinado

	KO	KU 1ª Ord	KU 2ª Ord
Erro Médio	0,00943	0,009045	0,00072
Raiz da Média Quadrada	1,73	1,705	1,782
Média dos Erros Padrão	0,7035	0,5785	0,5908
Erro Médio Aferido	0,005621	0,005567	-0,01245
RMQ Aferida	2,373	2,87	3,675
Diferença RMQ MEP	1,0265	1,1265	1,1912

Melhor Médio Pior

5.3.7 Melhores Modelos depois da Validação Cruzada

Seguem-se os modelos com melhores resultados de Validação Cruzada, escolhidos entre os doze aqui apresentados anteriormente. A escolha dos melhores segue uma lógica de eliminação de modelos com valores anormalmente altos, maioritariamente encontrados para a MEP e a RMQA. São também eliminados modelos com valores de RMQ acima de 2,1 (neste caso só um esteve acima deste valor, nos testes subsequentes, no Monte da Guia, este valor será mais alto).

Tabela 5.6 - Resultados mais significativos

	KO	KU 1ª Ord	KU 2ª Ord
Semivariograma	Combin.	Combin.	Esférico
Erro Médio	0,00943	0,009045	-0,01863
Raiz da Média Quadrada	1,73	1,705	1,732
Média dos Erros Padrão	0,7035	0,5785	2,124
Erro Médio Aferido	0,005621	0,005567	-0,01045
RMQ Aferida	2,373	2,87	0,8294
Diferença RMQ MEP	1,0265	1,1265	0,392

-Melhor -Segundo -Terceiro

São estes os modelos que melhores resultados de Validação Cruzada obtiveram, sendo igualmente os três com valores mais baixos de RMQ. De destacar ainda um valor

relativamente alto da MEP no modelo de KU de 2ª Ordem com Semivariograma Esférico, e valores também não muito bons de RMQA para os outros dois modelos aqui presentes.

5.5.8 Validação

A Validação, tal como a validação cruzada, fornece uma ideia de como o modelo prevê valores para as localizações desconhecidas. A validação cruzada não "demonstra" que o modelo de semivariograma é correto, meramente que ele não está grosseiramente incorreto[17].

A validação por sua vez fornece uma visão mais apropriada da exactidão do modelo. Do conjunto inicial de pontos criam-se dois conjuntos de pontos os quais somados são exactamente o conjunto de pontos inicial. Esses conjuntos de pontos criados designam-se por conjunto de teste e conjunto de treino. De seguida criam-se os modelos desejados utilizando o conjunto de treino. Ao contrário da validação cruzada, não é apenas retirado um ponto mas um conjunto de pontos (conjunto de teste). Para calcular a precisão da estimação comparam-se os conjuntos de pontos (teste e treino) e determinam-se os erros provenientes da estimação.

Para amostra completa foi feita a validação para os modelos que tinham apresentado melhores resultados na validação cruzada, ou seja, KO com semivariograma combinado, KU com deriva de primeira ordem com semivariograma combinado e KU com deriva de segunda ordem com semivariograma esférico.

Tabela 5.7 – Validação para melhores modelos

	KO Comb.	KU 1ª Comb	KU 2ª Esf
Erro Mínimo	0	0	0
Erro Máximo	45,565	46,739	49,917
Média do Erro	0,601	0,599	0,667
Soma do Erro	37910,137	37779,538	42068,418
Desvio Padrão	1,589	1,616	1,633

Melhor Médio Pior

Na validação com a amostra completa foram utilizados para o conjunto de teste 63072 pontos, cerca de 15% do conjunto total de pontos. Os valores apresentados tentam dar alguma estatística sobre o erro decorrente da validação.

Assim verifica-se que o modelo que apresenta um erro médio menor é a KU com deriva de primeira ordem com um valor de 0,598991 metros. O modelo de KU com deriva de segunda ordem esférico revelou-se o pior modelo apesar de ter tido uma validação cruzada

muito boa. Ao revelar-se o pior modelo nesta fase veio confirmar o facto de na geração da superfície ter introduzido zonas onde o erro de estimação foi de cerca de 1000 metros. Onde é notório o pior desempenho deste modelo é na soma dos erros de estimação. Com este modelo a soma foi de 42608 metros contra os 37780 metros da Krigagem Universal com deriva de 1ª ordem.

A validação na Krigagem Ordinária deu um resultado muito semelhante ao melhor modelo, o que vem confirmar mais uma vez que existe uma deriva acentuada nos dados e que deve ser tida em conta na geração de superfícies.

5.4 Zona Circundante do Monte da Guia

5.4.1 Amostra

O conjunto de dados inicial compreende uma área bastante extensa, composta por sub-áreas com características bastante diversas. Algumas destas sub-áreas estão claramente pouco caracterizadas, e não são, de todo, o objectivo central do amostragem. Parte-se agora para um exame mais pormenorizado da área objectivo desta amostragem, a caldeira e zonas circundantes do vulcão adormecido do Monte da Guia.

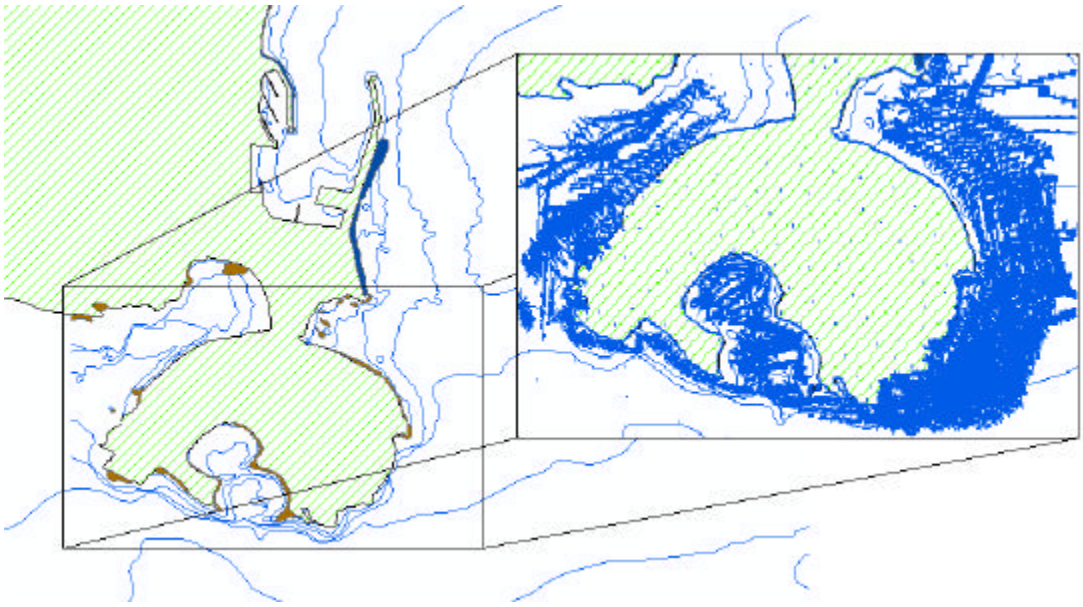

Figura 5.21 – Sub-amostra do Monte da Guia

O subconjunto obtido com esta selecção de amostras compreende 303040 pontos, cobrindo uma área com 1200 por 1214 metros de lado, apresentando uma densidade bastante maior de informação conhecida.

5.4.2 Descrição univariada e espacial da zona do Monte da Guia

Na figura 5.22 podemos observar a descrição univariada dos dados na zona do Monte da Guia. Observamos que nos 303040 pontos incluídos o ponto máximo de altimetria tem 145 metros e o máximo de batimetria tem -145 metros. Os valores de *kurtosis* e do coeficiente de assimetria (*skewness*), bem como as outras medidas de variabilidade não sofrem alterações significativas relativamente ao histograma gerado depois de se introduzirem os pontos de batimetria, altimetria e linha da costa existentes no *raster*.

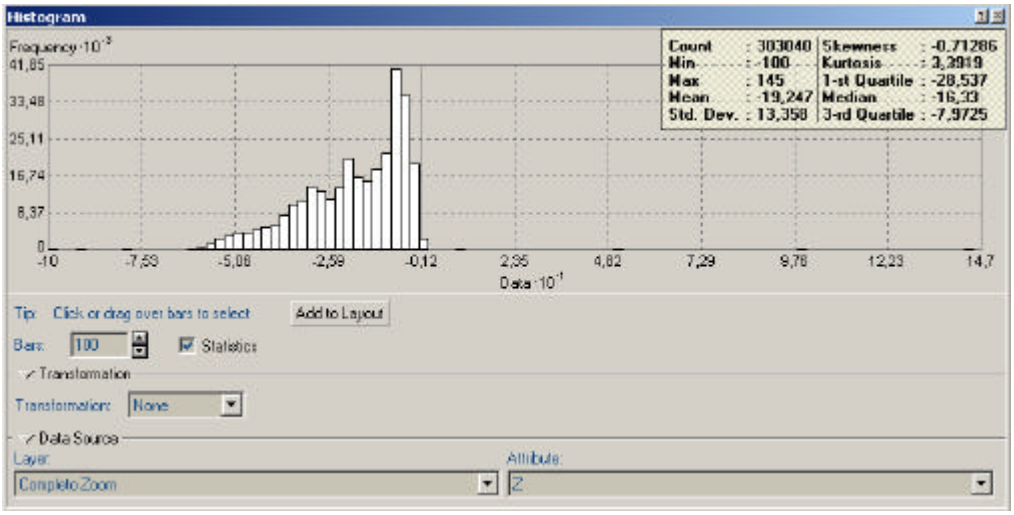

Figura 5.22 – Histograma da sub-amostra do Monte da Guia

Mais uma vez a deriva calcula pela ferramenta de análise de derivas existente no ESDA, comprova que o fenómeno em estudo apresenta um tendência na amostragem. As amostras à medida que nos afastamos da linha da costa começam a ser mais profundas. Em particular, forma em "U" é particularmente acentuada pois atravessamos o Monte da Guia.

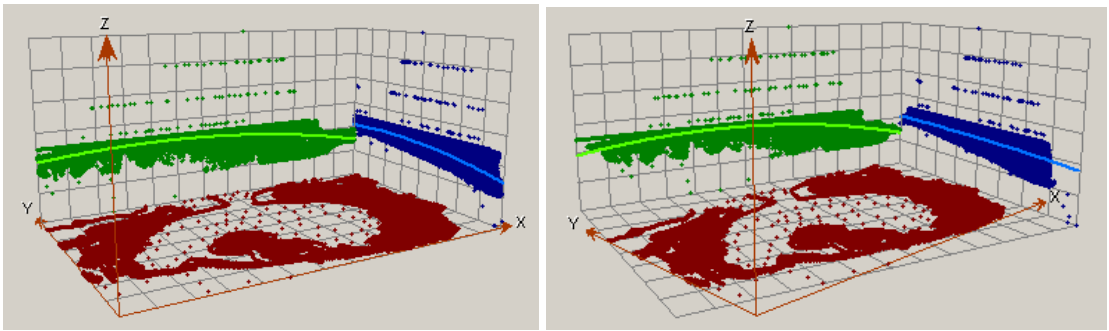

Figura 5.23 – Deriva dos dados na zona do Monte da Guia

Os mapas de Voronoi apresentaram algumas mudanças em relação ao conjunto global de amostras. O mapa de Aglomerado volta a identificar a possibilidade de ocorrência de *outliers* em terra, desta vez mais demarcadamente. Mas a existirem são pontos com valores muito

longe da média, e positivos sendo natural que apresentem este comportamento nos mapas Voronoi. O mapa IQR apresenta um variabilidade bastante diferente da original, tal deve-se à existência, mais uma vez, de pontos com valores muito longe da média, quer em terra quer no mar.

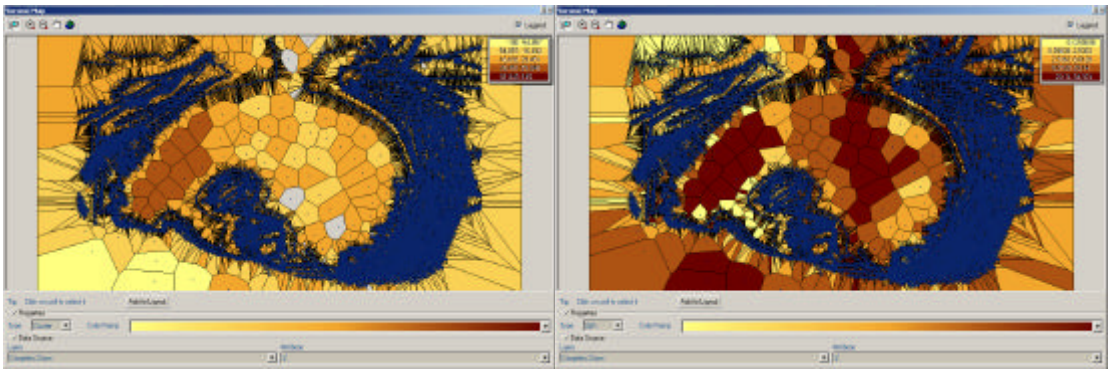

Figura 5.24 – Mapas Voronoi Aglomerado (esquerda) e IQR (direita)

Os outros mapas de Voronoi não apresentaram informação significantemente diferente da encontrada nos mapas do conjunto completo de amostras.

5.4.3 Resultados da Validação Cruzada

Apesar de existirem algumas semelhanças com as características do conjunto de dados completo, os resultados obtidos para este subconjunto são algo diversos. Em primeiro lugar há a assinalar um aumento do valor RMQ em todos os modelos aplicados. Tal deve-se ao facto de haver amostras com bastante influência local, como se pôde ver nos mapas de Voronoi, e de haver amostras em terra com valores muito distantes dos valores encontrados no grosso das amostras.

Depois é igualmente de realçar o facto de o Semivariograma Esférico obter melhores resultados para a KU de 1ª Ordem do que com a KU de 2ª, tal como acontecera com o conjunto global. Esta mudança pode se explicar pela deriva que se registou para esta zona, embora semelhante na orientação e forma à encontrada para o conjunto completo de dados, é menos acentuada. A maioria dos pontos de maior profundidade não se encontram nesta área, o que "aplana" a forma da deriva.

Tabela 5.8 – Resultados mais significativos no Monte da Guia

	KO		KU 1ª Ord.	KU 2ª Ord	
Semivariograma	Esférico	Combin.	Esférico	Esférico	Combin.
Erro Médio	-0,02561	0,01033	0,003497	-0,01303	-0,02476
Raiz da Média Quadrada	2,26	1,921	1,893	1,93	1,946
Média dos Erros Padrão	2,796	0,7135	0,8208	0,9409	1,384
Erro Médio Aferido	-0,00938	0,005419	-0,000898	-0,02022	-0,02008
RMQ Aferida	0,7955	2,557	2,207	2,104	1,391
Diferença RMQ MEP	0,536	1,2075	1,0722	0,9891	0,562

-Melhor -Segundo -Terceiro

Mantêm-se as características do conjunto completo neste subconjunto, zonas com grande profusão de informação e zonas com esparsa densidade de amostras. As dificuldades mantém-se para os Semivariogramas Exponencial e Gaussiano, tal como visto na área de estudo anterior.

Os modelos com a KO continuaram a ter pouco sucesso, salvo mais uma vez no Semivariograma Combinado, único com o qual a RMQ chega abaixo de 2. O Semivariograma Combinado funciona também, desta feita, com a KU de 2ª Ordem, e não com a KU de 1ª Ordem, ao contrário do que se passou no conjunto de dados completo. Mais uma vez aparece um modelo que se destaca de certo modo de todos os outros, desta vez a KU de 1ª Ordem, que apenas apresenta um valor de RMQA um pouco alto (2,207), mas com excelentes resultados em todos os outros parâmetros. Aparentemente dá-se, para este subconjunto das amostras, uma troca de posições entre os modelos de KU de 1ª Ordem e o de 2ª Ordem, deixando todo o resto muito parecido.

Antes de se avançar na análise de resultados é necessário responder a uma questão. Porque razão na amostra completa o modelo com melhores resultados foi com uma deriva de 1ª ordem e no Monte da Guia o modelo com melhores resultados foi com uma deriva de 2ª ordem? A razão desta mudança é exactamente o Monte da Guia. A deriva considerada nos mapas é global, isto é, não foi considerada a existência de derivas locais. Quando se faz esta restrição, o que antes era modelado com uma deriva de 1ª ordem (devido ao mapa estender-se de uma zona montanhosa a uma zona com alguma profundidade) é melhor modelado com uma deriva de 2ª ordem, precisamente pela transição ao passar-se a linha da costa (efeito "U").

5.4.4 Validação

Nesta experiência foram utilizados 151760 pontos, cerca de 50% dos 303520 pontos incluídos nesta sub-amostra.

Tabela 5.9 – Validação para a zona do monte da Guia

	KO Esf.	KO Comb.	KU 1ª Esf	KU 2ª Esf
Erro Mínimo	0	0	0	0
Erro Máximo	79,742	70,695	83,967	281,433
Média do Erro	0,929	0,753	0,719	0,803
Soma do Erro	141106,438	114341,637	109183,053	121824,561
Desvio Padrão	2,143	1,983	2,021	2,221

Melhor 2º melhor Mau

Numa primeira análise à tabela de Validação da zona do Monte da Guia, o que se verifica é que o melhor modelo foi o da KU de 1ª ordem seguido pela KO com semivariograma combinado. A KO com semivariograma esférico e a KU com deriva de 2ª ordem mostraram resultados anormais, pelo que foram nesta fase descartados.

Apesar de se ter verificado nos teste de Validação Cruzada uma melhor adequação da deriva de 2ª ordem para esta sub-amostra, este modelo falha novamente os teste de Validação. Os valores de erro anormais (são estimados valores positivos em zonas onde a profundidade deveria rondar os 100 metros!), são provavelmente gerados por um aumento do número de ponderadores negativos. Assim sendo podem ser estimados valores muito para além dos máximos e mínimos admissíveis do ponto de vista do fenómeno físico[13].

De modo a perceber melhor o que se passa em zonas mais densamente amostradas, partiu-se para a análise de uma área contendo apenas amostras originais.

5.5 Zona de Amostragem Exaustiva a Leste

5.5.1 Amostra

A primeira selecção de amostras não revelou grandes novidades, apesar de se reportar a uma área mais estrita. As características essenciais do conjunto de amostras manteve-se aproximadamente o mesmo, registando-se apenas um ligeiro esbatimento da deriva, o que veio a permitir aos modelos de KU de 1ª Ordem obter melhores resultados.

Parte-se agora para a análise de uma nova sub-amostra, ainda mais reduzida, compreendendo apenas pontos conhecidos originais, numa zona onde a amostragem foi exaustiva. Este conjunto compreende 18853 pontos de leitura, numa área de 128 por 323 metros de lado, batendo as profundidades a Leste do Monte da Guia.

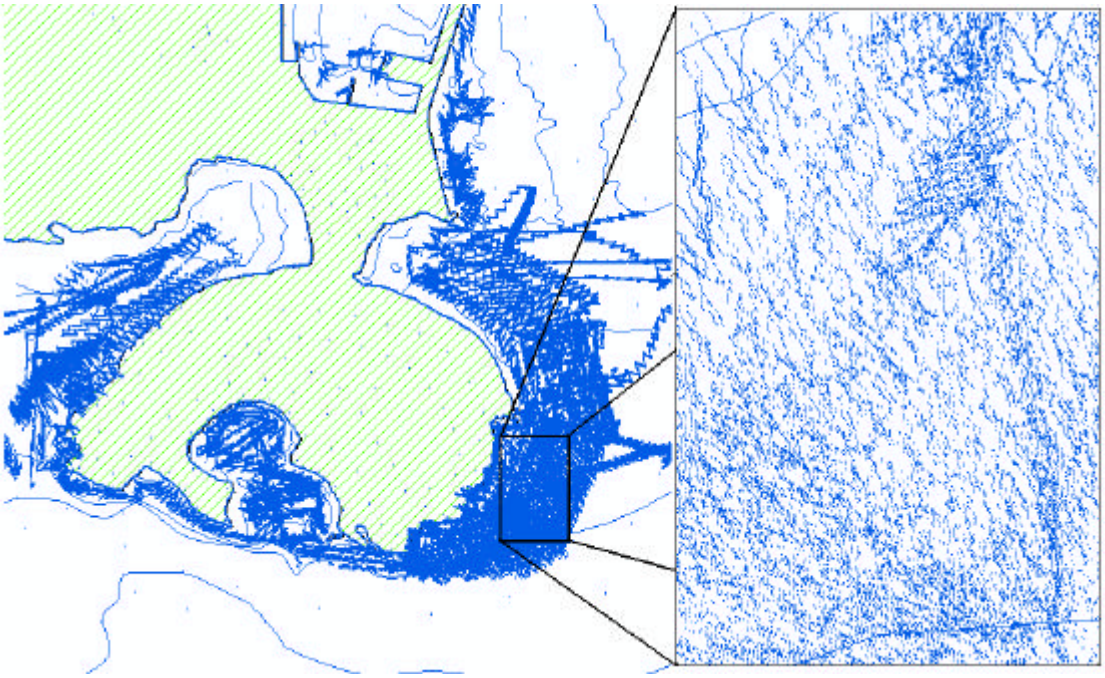

Figura 5.25 – Sub-amostra a Leste do Monte da Guia

5.5.2 Descrição univariada e espacial da zona a leste do Monte da Guia

Nesta experiência foi utilizada uma pequena amostra do conjunto de pontos totais, 18853 pontos. Os seus valores máximo e mínimo são de -2,1939 m e -57,615 m respectivamente, já que esta sub-amostra foi restringida a amostras de profundidade. No entanto apesar de este intervalo entre o valor máximo e mínimo ser relativamente elevado, metade das amostras estão entre os -46,463 m e os -38,417m. As amostras apresentam pois uma mediana de -42,139m e uma média muito próxima de -41,973m. O valor de *kurtosis* é de 9,7799 e indica precisamente que esta distribuição é pontiaguda relativamente a uma distribuição normal, isto é, apresenta um pico de valores e depois decresce muito rapidamente tendo valores muito afastados da média. O valor do coeficiente de assimetria (*skewness*) é de 1,6835, isto é, justifica o facto de a distribuição apresentar uma assimetria positiva.

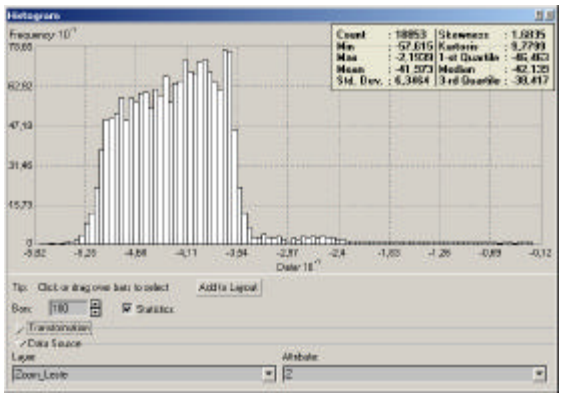

Figura 5.26 – Histograma do zona leste do Monte da Guia

Quanto à existência de derivas é claro pelas figuras que se está presente uma deriva em que a modelação mais correcta passa por polinómio de primeiro grau.

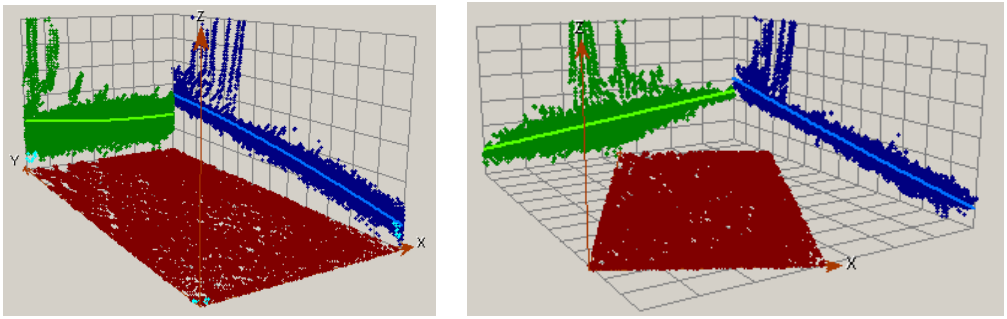

Figura 5.27 – Deriva dos dados no zona leste do Monte da Guia

5.5.3 Resultados da Validação Cruzada

As primeiras impressões retiradas dos valores de Validação Cruzada mostram mais ligeira subida dos valores da RMQ, mas por contraponto, uma melhor adaptação dos modelos à variabilidade das amostras da área de estudo. Mas o facto mais significativo é o sucesso do modelo com Semivariograma Exponencial, face ao modelo com Semivariograma Esférico.

Tabela 5.10 – Resultados mais significativos na Zona a Leste

	KO	KU 1ª Ord.		KU 2ªOrd.
Semivariograma	Combin.	Esférico	Exponen.	Combin.
Erro Médio	0,06138	0,01783	0,008738	0,0349
Raiz da Média Quadrada	2,274	2,085	1,999	1,919
Média dos Erros Padrão	2,53	2,388	2,12	2,822
Erro Médio Aferido	0,02432	0,007697	0,004981	0,0134
RMQ Aferida	0,8989	0,875	0,9487	0,6852
Diferença RMQ MEP	0,256	0,303	0,121	0,903

-Melhor -Segundo -Terceiro

O Semivariograma Exponencial têm um melhor desempenho em áreas onde a amostragem é mais exaustiva, tal como apresentado no sub-capítulo 5.3.1, o seu sucesso aqui é total, apresentando os melhores valores para quase todos os parâmetros de Validação Cruzada. Embora o Semivariograma Esférico não seja tão bom, continua a mostrar bons resultados, demonstrando mais uma vez a sua flexibilidade. O mesmo se pode constatar para o Semivariograma Combinado, que consegue até o valore mais baixo de RMQ.

Por último, é de referir a superioridade dos modelo de KU de 1ª Ordem face aos outros. Tal como seria de esperar nem o modelo de KO, nem o modelo de KU de 2ª Ordem se adaptam tão bem ao tipo de deriva encontrada, que define relativamente bem um plano inclinado de Noroeste para Sudoeste.

Como referido no capítulo 4, neste tipo de amostragem é de esperar que alguns dos métodos Determinísticos consigam bons resultados, nomeadamente o método de Polinomiais Locais e o RBF. Foram então usados estes métodos, e também o IDW, para calcular superfícies que se pudessem comparar com as geradas pelos métodos de Krigagem. Em seguida apresentam-se os resultados de Validação Cruzada para estes modelos.

Tabela 5.11 – Validação Cruzada para Métodos Determinísticos

	IDW	RBF	Pol. Local
Erro Médio	-0,08143	-0,03817	-0,00623
Raiz da Média Quadrada	2,083	2,025	1,879

Estes valores são bastante interessantes, em especial para a Polinomial Local, onde a RMQ é inferior a qualquer das calculadas para os modelos de Krigagem. O modelo IDW já tem valores menores, como esperado, mas mesmo assim que não estão ao nível dos valores obtidos para a Krigagem. No entanto as comparações que se possam fazer entre estes métodos e os de Krigagem são sempre deficientes dado que a Validação Cruzada dá apenas valores de EM e RMQ para os primeiros.

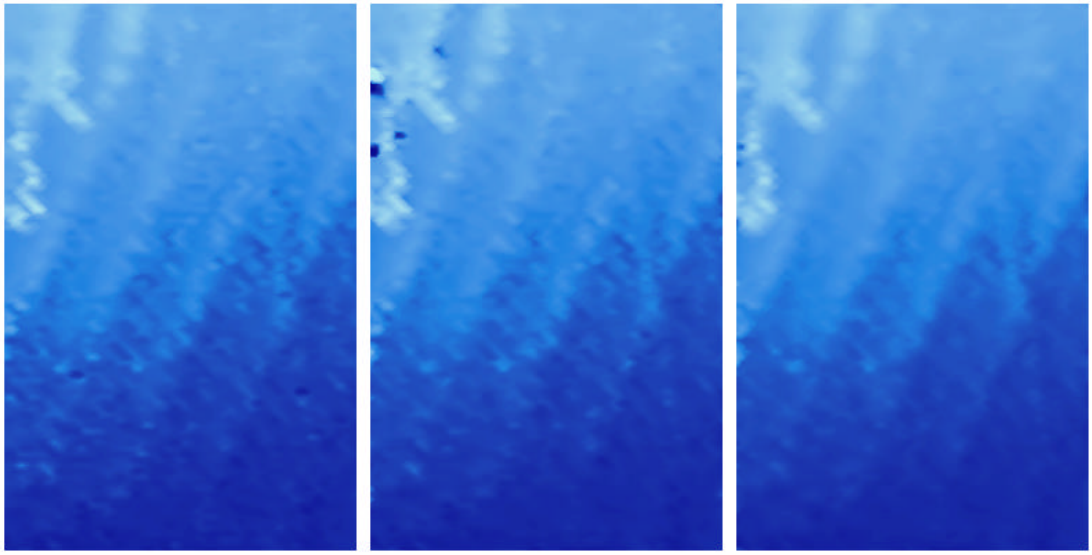

Figura 5.28 – RBF (esquerda), Polinomial Local (centro) e KU 1ª Ordem (direita)

Na figura percebe-se claramente uma evolução, no sentido da esquerda para a direita, da suavidade da superfície gerada. Não existem, na superfície gerada pelo método de Krigagem, mudanças bruscas de profundidade, a evolução dos valores interpolados é sempre suave e progressiva, mais próxima do fenómeno físico real. Fica aqui bem patente a grande diferença entre métodos Determinísticos e Geoestatísticos.

5.5.4 Validação

Na zona a Leste do Monte da Guia, para fazer os testes de validação foram utilizados 9427 pontos, cerca de 50% dos pontos existentes nesta sub-amostra.

Tabela 5.12 – Validação para a zona Leste

Semivariograma	KO Comb.	KU 1ª Esf	KU 1ª Exp	KU 2ª Comb
Erro Mínimo	0,001	0,001	0,001	0,001
Erro Máximo	28,562	29,635	29,157	36,627
Média do Erro	1,359	1,284	1,263	1,131
Soma do Erro	12815,548	12107,818	11906,844	10660,658
Desvio Padrão	2,118	1,962	1,896	1,652

-Melhor -Segundo -Terceiro

Aqui os resultados a que se chegaram não foram de modo algum surpreendentes. Como a zona tem dimensões reduzidas, o grau da deriva tem menos influência daí que atrás da KU de 2ª ordem com semivariograma combinado os melhores resultados tenham sido alcançados com KU de 1ª ordem. Entre o segundo e terceiro melhores modelos nesta experiência (ambos

com deriva de 1ª ordem) o que apresentou melhores resultados foi o que utilizou um semivariograma exponencial, indicado para zonas de intensa amostragem pois apresenta menor efeito de pepita. O modelo que utilizou semivariograma esférico não ficou muito atrás, o que indica a boa adequabilidade do semivariograma esférico para a maioria das ocasiões. Em último lugar ficou o modelo, KO com semivariograma combinado, que não utiliza qualquer espécie de deriva para a estimação dos pontos.

Os diferentes desempenhos alcançados pelos quatro métodos também são claramente visíveis nos gráficos. À medida que o método vai baixando de desempenho as cores nos gráficos não se sobrepõem tanto.

5.6 Erro de Medição

Três dos métodos de Krigagem, ordinário, simples e universal, permitem a modelação do erro de medição. O erro de medição ocorre quando se têm várias observações na mesma localização e elas diferem. No entanto, quando a variância da precisão do instrumento é documentada pode-se introduzir directamente a variação da medição no ArcGis Geostatistical Analyst.

Nos modelos com erro de medição, o efeito de pepita é então modelado como a soma de dois componentes; a micro-estrutura e o erro de medição. A micro-estrutura corresponde ao efeito de pepita a que se chega quando se aplica um modelo de semivariograma. O segundo componente pode ser calculado directamente pelo ArcGis, quando se tem mais do que uma medição para a mesma localização, ou então introduzido directamente como uma variância proveniente do aparelho de medição. Essa variância é menor quanto mais pontiaguda for a distribuição e quanto mais perto estiverem os valores máximo e mínimo da média dessa distribuição; o que indica que o desvio de valores em relação à média é pequeno e que o aparelho de medição é bastante preciso.

Quando não é modelado o erro, a krigagem é exacta. Quer isto dizer que quando é feita uma predição para um local onde tenha sido medido o valor, o resultado será o mesmo. Falta referir que para cada uma das componentes é possível escolher a percentagem que vão ter para o calculo final do efeito de pepita.

No caso de estudo foram detectadas doze localizações, com a mesma latitude e longitude. Cada uma dessas localizações apresenta uma segunda leitura para a mesma localização perfazendo um total de vinte e quatro pontos, como se vê na figura 5.29.

	X	Y	Z	VARIANCIA
	715,177992	1810,159535	5,90445	0,616353
	715,177992	1810,159535	5,908569	0,616353
	715,186864	1810,20805	5,920931	0,610825
	715,186864	1810,20805	5,925051	0,610825
	715,328922	1811,268551	5,267903	0,540831
	715,328922	1811,268551	5,268674	0,540831
	715,336517	1811,31022	5,282342	0,537937
	715,336517	1811,31022	5,283734	0,537937
	715,355732	1811,120634	6,272522	0,535164
	715,355732	1811,120634	6,273829	0,535164
	715,416882	1811,484375	6,296353	0,518932
	715,416882	1811,484375	6,289286	0,518932
	715,603804	1812,834782	5,572094	0,502273
	715,603804	1812,834782	5,573255	0,502273
	715,606241	1812,228591	5,41937	0,509169
	715,606241	1812,228591	5,418201	0,509169
	715,662272	1812,394937	5,419714	0,504653
	715,662272	1812,394937	5,420841	0,504653
	715,756816	1813,751503	5,594402	0,535164
	715,756816	1813,751503	5,596575	0,535164
	715,801392	1814,029009	5,566707	0,553643
	715,801392	1814,029009	5,56598	0,553643
	716,463389	1814,750587	5,593638	0,600297
	716,463389	1814,750587	5,594202	0,600297

Figura 5.29 – Conjunto de pontos com igual latitude e longitude

Como são muito poucas as leituras que temos para a mesma localização, a estimação efectuada pelo ArcGis não é muito realística dando uma valor irrisório na ordem dos nano-metros.

A moda da Variância situa-se no intervalo [0,5 – 0,55], onde se situam os valores para mais de trezentas mil leituras; valores desta ordem são igualmente irrisórios para serem usados como erros de medição. Ao contrário a média da Variância é da ordem de 10^{23}, número completamente inadequado para ser usado como erro de medição (este valor esta muito acima de qualquer patamar dos modelos de Semivariograma vistos até aqui). Tal valor de média, deve-se a registos muito altos de Variância, para um conjunto restrito de pontos. Para determinar um valor realístico para o erro de medição fez-se uma análise da variância de cada ponto e verificou-se o seguinte:

Tabela 5.13 – Tratamento da variância fornecida para amostra completa

VARIÂNCIA MÁXIMA	MÉDIA DA VARIÂNCIA	NÚMERO DE PONTOS
1000	5,014262	518
10000	8,127179	77
20000	9,415062	22
100000	11,647405	9
200000	13,795281	5
500000	15,931938	3

Esta tabela foi obtida dando igual valor de Variância para todos os pontos com valores superiores ao mostrado na coluna da esquerda. Desta forma foi possível baixar drasticamente

a média da Variância, para os valores que se podem ver na coluna do centro. Na coluna da direita está registado o números de pontos encontrados com variância superior ao tecto assumido na coluna da esquerda.

Com variâncias acima de mil ainda se encontra um número razoável de pontos, mas acima de cem mil este número é quase residual. Podem-se considerar estes valores acima de cem mil como meros casos pontuais, para os quais a variância é infinita. Seguindo este raciocínio, pode-se considerar a Variância Média deste conjunto de pontos como estando no intervalo [9,4 – 11,6], antes do qual se ignoram alguns pontos com variâncias não infinitas, e para lá do qual a Variância Média começa a comportar pontos de variância infinita.

Neste intervalo optou-se, empiricamente, pelo valor de Variância Média de dez. No entanto é de salientar que qualquer outro valor escolhido dentro do intervalo [9,4 – 11,6] chegaria a resultados muito semelhantes aos apresentados seguidamente.

5.6.1 Amostra Completa

Depois de chegar a um valor de Variância, com o qual se poderiam obter resultados visíveis, foi aplicado este valor nos modelos de Krigagem mais bem sucedidos. A tabela 5.14 mostra as diferenças de Validação Cruzada obtidas para estes modelos, relativamente aos valores sem erro.

Tabela 5.14 – Validação Cruzada com e sem Erro de Medição na Amostra Completa

Semivariograma	KO Combinado		KU 1ª Ord. Comb.	
	Com Erro	Sem Erro	Com Erro	Sem Erro
Erro Médio	-0,02693	0,00943	-0,01664	0,009045
Raiz da Média Quadrada	2,008	1,73	2,025	1,705
Média dos Erros Padrão	3,296	0,7035	3,275	0,5785
Erro Médio Aferido	-0,0082	0,005621	-0,00498	0,005567
RMQ Aferida	0,6017	2,373	0,614	2,87
Diferença RMQ MEP	1,288	1,0265	1,25	1,1265

Claramente se percebe uma evolução negativa dos valores de Validação Cruzada. Como seria de esperar, o aumento do efeito de pepita faz igualmente aumentar os valores da RMQ, pois vai produzir uma superfície mais lisa, que não passa exactamente por todos os pontos medidos. Há ainda a registar uma melhoria dos valores de RMQA, dando a entender uma melhor modelação da variabilidade, mas esta melhoria resulta em grande parte do grande aumento dos valores da MEP.

Tabela 5.15 – Validação com Erro de Medição na Amostra Total

	KO Comb.		KU 1ª Comb	
	Sem Erro	Com Erro	Sem Erro	Com Erro
Erro Mínimo	0	0	0	0
Erro Máximo	45,565	36,507	46,739	36,835
Média do Erro	0,601	0,834	0,599	0,821
Soma do Erro	37910,137	52596,393	37779,538	51764,867
Desvio Padrão	1,589	1,842	1,616	1,818

Na tabela 5.15 estão os resultados de Validação dos modelos com erro de medição, e o método de Krigagem Universal de 1ª ordem com semivariograma combinado continuou a apresentar melhores resultados face à Krigagem Ordinária. No entanto com a introdução do erro verificou-se que não existiu melhoria, muito pelo contrário a soma dos erros de estimação passou de 37780 metros para 51765 metros. Tal deve-se ao facto de o efeito de pepita ter aumentado, o que gera superfícies mais lisas e logo com maior imprecisão.

5.6.2 Monte da Guia

Para este conjunto de dados, foram utilizados os mesmos métodos de análise da Variância. Novamente se deparou com alguns pontos de variância infinita, em condições muito semelhantes às encontradas na Amostra Completa; moda da Variância muito perto de 0,5; média na ordem de 10^{23}, e diferenças irrisórias em pontos com mais de uma medição. Esta semelhança quase total levou à escolha do mesmo valor de Variância Média a aplicar como Erro de Medição.

Tabela 5.16 - Validação Cruzada com e sem Erro de Medição na zona do Monte da Guia

Semivariograma	KO Esférico		KO Combinado		KU 1ª Ord. Comb.	
	Com Erro	Sem Erro	Com Erro	Sem Erro	Com Erro	Sem Erro
Erro Médio	-0,03048	-0,02561	-0,03413	0,01033	-0,01876	0,003497
Raiz da Média Quadrada	2,305	2,26	2,227	1,921	2,0201	1,893
Média dos Erros Padrão	3,299	2,796	3,296	0,7135	3,293	0,8208
Erro Médio Aferido	-0,00934	-0,00938	-0,01039	0,005419	-0,00564	-0,0009
RMQ Aferida	0,6895	0,7955	0,6668	2,557	0,6616	2,207
Diferença RMQ MEP	0,994	0,536	1,069	1,2075	1,2729	1,0722

As semelhanças continuam nos resultados de Validação Cruzada, com uma igual subida dos valores de RMQ e MEP, e no caso da KO com Semivariograma Combinado, uma ligeira melhoria de modelação da variabilidade.

Na tabela 5.17 são apresentados os resultados de Validação para os modelos com erro de medição do zona do Monte da Guia. Ao se introduzir um erro de dez, o efeito de pepita aumentou, e a consequência imediata é a maior influência do conjunto total de amostras na estimação de um ponto. O facto de ser modelado com uma deriva de 1ª ordem faz com que supere os resultados alcançados com os modelos de Krigagem Ordinária.

Tabela 5.17 – Validação com Erro de Medição para o Monte da Guia

	KO Comb.		KU 1ª Esf	
	Sem Erro	Com Erro	Sem Erro	Com Erro
Erro Mínimo	0	0	0	0
Erro Máximo	70,695	77,426	83,967	84,701
Média do Erro	0,753	1,026	0,719	0,998
Soma do Erro	114341,637	155730,236	109183,053	151458,293
Desvio Padrão	1,983	2,258	2,021	2,226

A introdução do erro de medição continua a gerar resultados piores em toda a linha. Tal como nos modelos usados para a amostra completa, os efeitos de pepita iniciais eram todos muito baixos, pelo que este erro diminui consideravelmente a influência de medições mais próximas do ponto de estimação.

5.6.3 Zona Leste

Por último foi feita a mesma experiência com o erro para a zona a leste do Monte da Guia e os resultados a que se chegou estão apresentados nas tabelas que se seguem. O erro utilizado foi de 1,125, pois desta feita não foram encontrados pontos de variância infinita. Este valor é a média aritmética das variâncias de todos os pontos da sub-amostra da zona Leste.

Tabela 5.18 - Validação Cruzada com e sem Erro de Medição na Zona Leste

Semivariograma	KO Combinado		KU 1ª Ord. Esférico		KU 1ª Ord. Exponen.	
	Com Erro	Sem Erro	Com Erro	Sem Erro	Com Erro	Sem Erro
Erro Médio	-0,06118	0,06138	-0,01737	0,01783	-0,0062	0,008738
Raiz da Média Quadrada	2,274	2,274	2,079	2,085	1,971	1,999
Média dos Erros Padrão	2,464	2,53	2,337	2,388	1,915	2,12
Erro Médio Aferido	-0,02488	0,02432	-0,0077	0,007697	-0,0045	0,004981
RMQ Aferida	0,9225	0,8989	0,8919	0,875	1,038	0,9487
Diferença RMQ MEP	0,19	0,256	0,258	0,303	0,056	0,121

Ao contrário do que de passava nos casos anteriores, o ArcGis fez simplesmente um ajuste ao efeito de pepita quando se introduziu directamente o erro. Nos casos anteriores, onde o erro de dez era superior ao efeito de pepita calculado através do modelo do semivariograma, era atribuído dez ao efeito de pepita. Sendo o valor do Erro de Medição bastante baixo, poucas alterações se verificaram, no entanto em alguns casos registaram-se pequenas melhorias em algubs dos valores de Validação Cruzada.

Tabela 5.19 –Validação com Erro Medição na zona Leste

	KO Comb.		KU 1ª Esf	
	Sem Erro	Com Erro	Sem Erro	Com Erro
Erro Mínimo	0,001	0,001	0,001	0,001
Erro Máximo	28,562	28,539	29,635	27,114
Média do Erro	1,359	1,356	1,284	1,254
Soma do Erro	12815,548	12810,675	12107,818	11822,451
Desvio Padrão	2,118	2,116	1,962	1,892

	KU 1ª Exp.		KU 2ª Comb.	
	Sem Erro	Com Erro	Sem Erro	Com Erro
Erro Mínimo	0,001	0,001	0,001	0,001
Erro Máximo	29,157	25,112	36,627	23,532
Média do Erro	1,263	1,172	1,131	1,184
Soma do Erro	11906,844	11052,218	10660,658	11163,306
Desvio Padrão	1,896	1,717	1,652	1,668

Os resultados de Validação também não apresentam grandes mudanças causadas pela introdução do Erro de Medição. Na KO com semivariograma Combinado, não existem quase

diferenças; para os modelos de KU de 1ª existem melhorias ligeiras ao ser introduzido o erro; enquanto que para a KU de 2ª ordem o erro provoca pequenos aumentos dos valores.

Em todos os modelos, para os quais foi introduzido erro de medição, não se obtiveram superfícies finais com diferenças visíveis. Na zona a Leste do Monte da Guia, as superfícies geradas, são mesmo idênticas, não se observando quaisquer mudanças com a introdução do erro.

5.7 Conclusão

Depois de estudados os métodos a usar na interpolação de superfícies, foram seleccionados a KO e a KU cm derivas de primeira e segunda ordem. A estes métodos foram aplicados quatro semivariogramas diferentes, Esférico, Exponencial, Gaussiano e Combinado (dos anteriores três).

Estes diferentes métodos foram aplicados a três áreas diferentes (todos no mesmo espaço de estudo), tendo sido observados resultados distintos em cada uma delas. De forma a apreciar a adequação dos métodos empregues em cada área de estudo, foram usadas a Validação Cruzada e a Validação.

Através da geração de superfícies conseguiu-se perceber e chegar a alguns resultados, sobre a influência do efeito de pepita na produção de superfícies. Verificou-se que zonas onde o padrão de amostragem é mais intenso consegue-se camuflar melhor esse padrão utilizando um modelo de semivariograma com um efeito de pepita maior (a superfície resultante é mais polida). Quando o efeito de pepita é menor o conjunto total das amostras tem uma influência menor na geração da superfície.

No caso de estudo está-se perante uma área onde existem zonas com um intenso padrão de amostragem e zonas onde a amostragem é mais dispersa, daí que não seja de estranhar que os modelos do semivariograma que apresentaram melhores resultados na validação cruzada, tenham sido aqueles onde se combinaram os modelos com menor e maior efeito de pepita, exponencial e gaussiano respectivamente, juntamente com o modelo esférico. O modelo esférico revela-se uma escolha equilibrada, pois como tem um crescimento intermédio consegue gerar melhores superfícies. Este semivariograma revela-se a melhor escolha, quando existe pouca informação sobre a distribuição espacial das amostras, confirmando-se assim a razão de ser o mais usado. Por último resta referir que o semivariograma Combinado gerado

pelo *Geostatistical Analyst* obteve resultados satisfatórios em quase todas as situações, apresentando-se como o mais versátil dos semivariogramas usados.

Após a validação teve-se a confirmação que existe uma deriva e que deve ser considerada, pois o KU de 1ª ordem apresentou melhores resultados que a KO. Finalmente é de realçar que a validação, na área de trabalho, descartou a KU de 2ª ordem comprovando assim que existiam estimações incorrectas nas superfícies geradas.

Na aproximação ao Monte da Guia, os melhores resultados alcançados foram a KO com o modelo de semivariograma combinado e a KU com deriva de 1ª ordem esférica, visto aqui também existirem zonas que têm muitas amostras e zonas onde as amostras estão mais dispersas. Assim sendo, na validação, os melhores modelos foram os que conseguem que a influência das amostras seja intermédia ao longo da área em estudo. É de salientar que novamente a KU de 2ª ordem foi o pior interpolador pois apresentou erros muito elevados na estimação.

Na zona de amostra intensiva a leste do Monte da Guia, o melhor resultado na validação cruzada foi alcançado com o semivariograma exponencial, que se adequa melhor a zonas onde se deve fazer sentir mais a influência de amostras na sua vizinhança, precisamente porque tem a tendência para ter um menor efeito de pepita. O algoritmo de Krigagem que melhor resultados apresentou com este semivariograma foi o KU de 1º ordem, o que prova que existe uma deriva e que deve ser tida em conta. Na validação o melhor resultado foi alcançado com a KU de 2ª ordem com semivariograma combinado. Pode-se justificar este resultado devido essencialmente à zona de amostragem estar numa vertente, ou seja, uma deriva modelada com um polinómio de 2º grau constitui uma melhor aproximação à realidade, sendo por isso o modelo que apresenta um erro médio menor.

Por fim, resta referir o efeito que a introdução do erro teve nestes três casos de estudo. Nos dois primeiros casos de estudo, área de trabalho e aproximação à zona do Monte da Guia, a introdução do erro piorou os modelos pois o efeito de pepita aumentou diminuindo a influência de medições mais próximas do ponto de estimação. Deste facto resultam superfícies mais lisas e logo com maior imprecisão. Na zona a leste do Monte da Guia só em alguns casos da validação se notaram ligeiras melhorias, muito pouco significativas, e que se devem aos ajustes dos efeitos de pepita nos modelos com a introdução dos erros, feitos pelo ArcGis.

CAPÍTULO VI

CONCLUSÃO

6.1 Estudo de Mercado

A primeira parte deste relatório, onde foi tomada uma decisão de qual o melhor SIG a adoptar, acabou por ser uma parte primordial no desenvolvimento deste TFC. A escolha, que recaiu sobre o ArcGis, revelou-se uma escolha acertada por inúmeras razões; o ArcGis possuí um leque de ferramentas que possibilitam realizar bastantes operações em voga nos SIG. Além disso traz consigo uma enorme quantidade de documentação, que descreve as funcionalidades e bastantes exercícios que se revelaram indispensáveis na compreensão de todas essas funcionalidades. Aliada a isto, vem igualmente, toda uma interface de alto nível e de fácil manuseamento e que se enquadra nas chamadas *WYSIWYG*. Outro dos factores que contribuiu para o ArcGis deixar de ser uma opção e passar a ser uma escolha foi a sua enorme escalabilidade. É possível percorrer toda a gama de produtos existentes na família ArcGis sem nunca ser perceptível essa mudança, tal como é possível adoptar um sistema de informação persistente, que pode ir de um simples sistemas de ficheiros a uma base de dados com múltiplos utilizadores, acessível inclusive pela internet.

Para chegar a esta escolha, foi feito um estudo do mercado. Chegou-se à conclusão que o mercado está em expansão e que actualmente é liderado pela ESRI . Não deixa contudo, de existir uma grande quantidade de empresas de pequenas dimensões que exploram soluções específicas para alguns sectores e que se conseguem manter activas nesses nichos de mercado. No entanto a adopção de um SIG, que está em constante evolução, com uma empresa de pequenas dimensões poderá revelar-se uma escolha fatal.

O ArcGis adquirido inclui com três extensões: o *3D Analyst*, o *Geostatistical Analyst* e o *Spatial Analyst*. Cedo se tornou claro que a ferramenta que iria ser preponderante no trabalho seria o *Geostatistical Analyst*.

6.2 Interpolação de Superfícies

Para se compreender a melhor maneira de modelar os dados que foram cedidos, no capítulo IV, foi apresentado o modelo das funções aleatórias. É este modelo que vai guiar o

processo de estimação ou simulação de valores da variável, identificada com o atributo em estudo, em zonas não amostradas. O conjunto discreto e limitado de dados é interpretado como uma realização de um conjunto de variáveis aleatórias, mais precisamente de uma função aleatória estacionária. Julgar a hipótese de a função aleatória ser estacionária em relação ao primeiro momento é julgar a média das amostras como representativa de toda a área. A KS impõe a todos os pontos estimados uma média conhecida e constante, ao contrário da KO, não reproduzindo tão bem as variações locais de Z(x), isto é, tende a atenuar os valores estimados em torno da média conhecida. A KO reproduz bastante melhor as médias locais e, por este motivo, é o algoritmo mais usado na prática geoestatística. A versão estacionária da KS é usada somente, quando se pretende impor uma média global para os valores estimados. Por outro lado, julgar a hipótese de a função aleatória ser estacionária em relação aos segundos momentos, a covariância e o semivariograma, é decidir que o instrumento usado para medir a continuidade ou variabilidade espacial das amostras, o semivariograma experimental, é válido e representativo de toda a área coberta pelas amostras[13]. As amostras deixam assim de estar relacionadas por uma média constante em toda a área de trabalho e passam a estar correlacionadas.

Ainda neste capítulo foram apresentados os interpoladores determinísticos e geoestatísticos. Logo aí ficou demonstrado a superioridade destes últimos pois além de terem em conta a aproximação aos pontos vizinhos, têm em conta também o modo como os pontos vizinhos estão correlacionados.

No capítulo seguinte procedeu-se à geração de superfícies e a todo o processo de iteração na determinação das melhores superfícies. Foram usados quatro semivariogramas (Esférico, Exponencial, Gaussiano e Combinado), para três métodos de Krigagem (Ordinária, Universal com deriva de 1ª ordem e Universal com deriva de 2ª ordem).

Mais uma vez as ferramentas do ArcGis Geostatiscal Analyst revelaram-se de grande importância, pois permitiram efectuar a triagem de modo a que se chegasse às melhores superfícies. Como primeiro filtro usou-se a Validação Cruzada, apesar deste método não asseverar exactamente a correcção de um modelo, permite identificar modelos errados, para a amostragem em causa. Desta feita a Validação Cruzada permitiu identificar a inadquação dos modelos com semivariograma Gaussiano a todos as amostragens estudadas. Permitiu também perceber que o semivariograma Exponencial se enquadra melhor em áreas mais densamente amostradas. Ficou nesta fase clara a versatilidade dos semivariogramas Esférico e Combinado que passaram os testes de Validação Cruzada em todas as diferentes áreas de estudo.

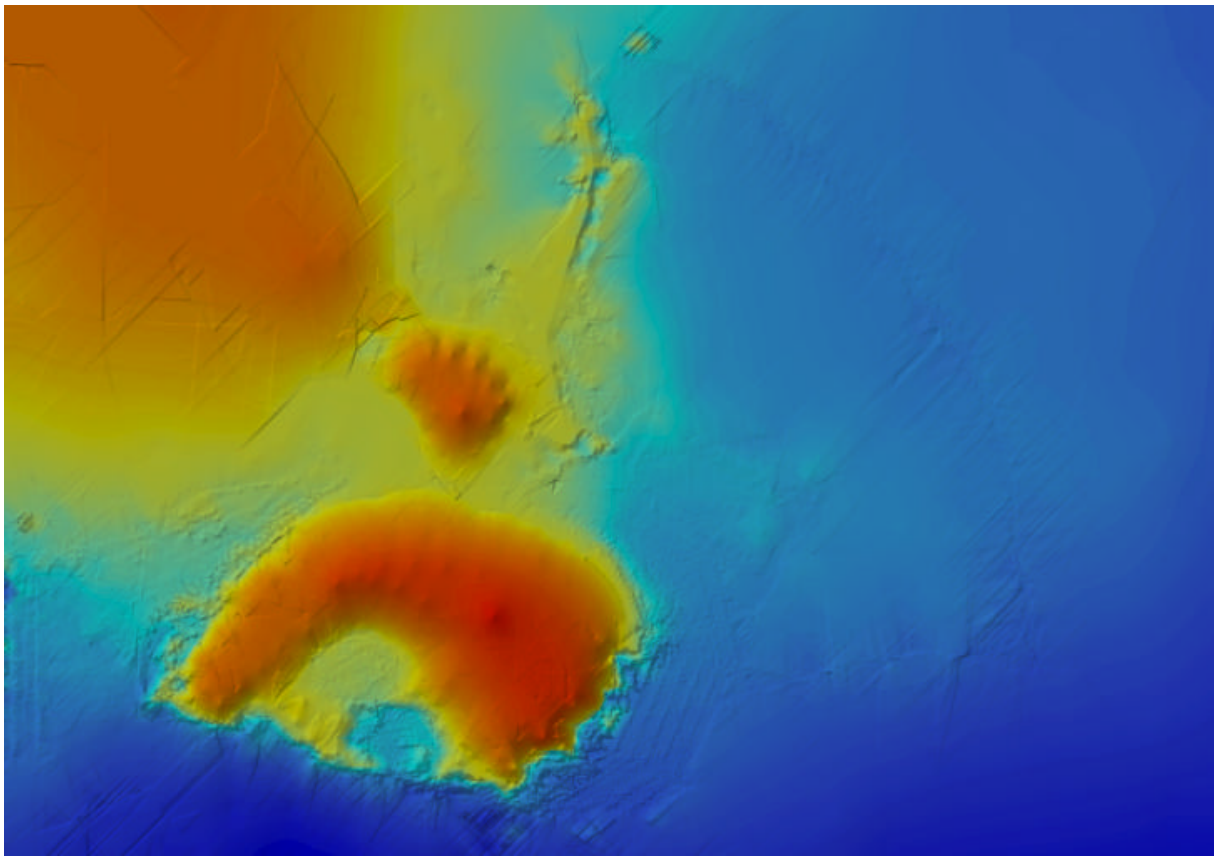

Figura 6.1 – Superfície final gerada com KO e semivariograma Combinado.

Uma segunda filtragem foi efectuada usando a Validação, sendo desta feita identificadas incorrecções nos métodos de Krigagem Universal. Estes métodos tendem a gerar valores extraordinariamente fora do intervalo de amostragem, em zonas de fraca amostragem. Todos os modelos de Krigagem Universal com deriva de 2ª ordem foram aqui descartados, e mesmo com deriva de 1ª ordem se encontraram anomalias deste género. A identificação destas anomalias permite chegar a duas conclusões, primeira, a inadquação da Krigagem Universal a amostragens esparsas, e segunda, a limitação, já enunciada da Validação Cruzada. Ficou aqui patente que a Validação Cruzada não avalia por si só, a qualidade das superfícies geradas. É por isso imperioso usar, em qualquer processo geoestatístico semelhante, os testes de Validação, mesmo sendo estes muito mais morosos.

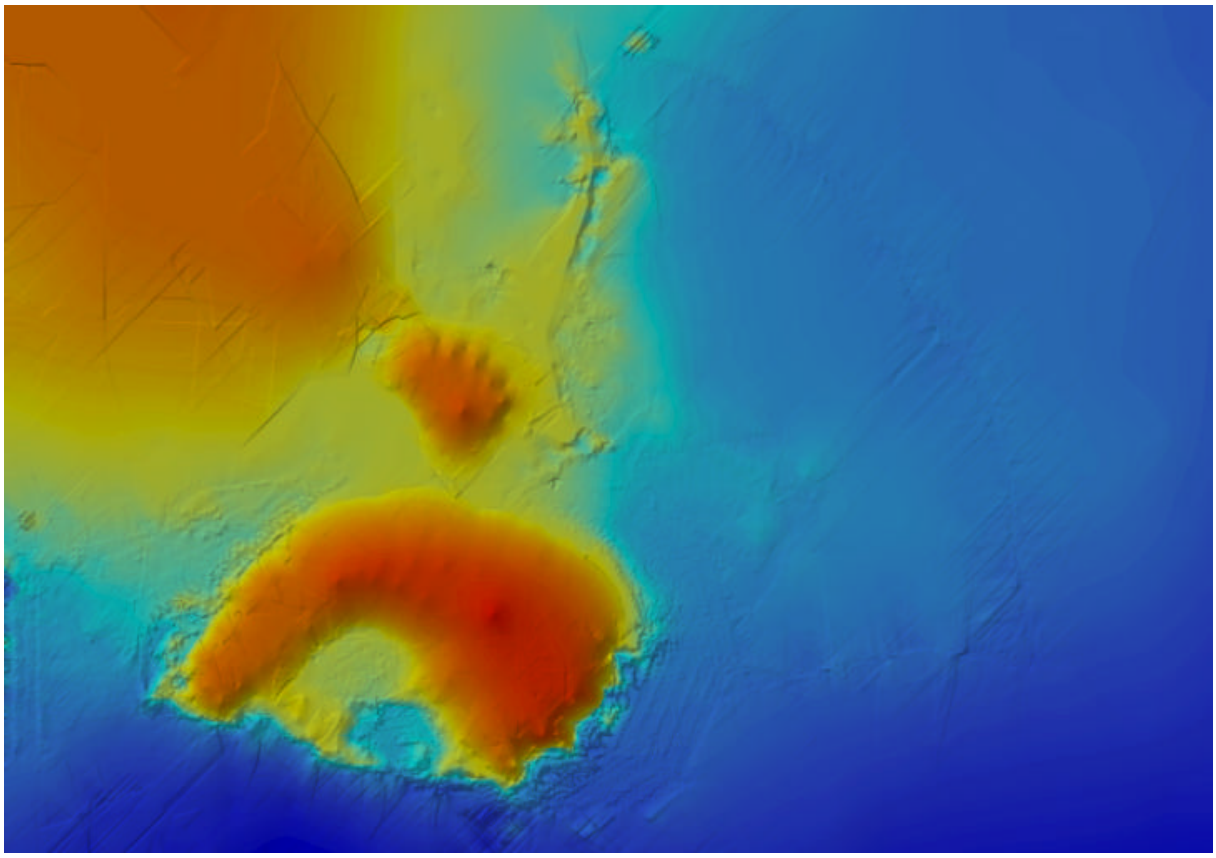

Figura 6.2 - Superfície final gerada com KU 1ª ordem e semivariograma Combinado.

Os melhores resultados na área de trabalho completa foram alcançados com a KO e com a KU com deriva de 1ª ordem. Ambos os modelos utilizaram semivariogramas combinados, que tendem a simular melhor o facto de existirem zonas com grande intensidade de amostragem e zonas onde a amostragem é mais dispersa. Na KU, a deriva, isto é, a tendência que os dados mostram, revelou que este caso de estudo, será modelada mais correctamente com um polinómio de 1º grau.

Na zona do Monte da Guia, os melhores modelos produzidos foram alcançados com a Krigagem Ordinária e semivariograma Combinado e com a Krigagem Universal com deriva de 1ª ordem e semivariograma Esférico. As características desta amostragem revelaram-se bastante semelhantes à amostragem completa, com a Krigagem Universal de 2ª ordem a ser mais uma vez descartada pela existência de valores interpolados anómalos.

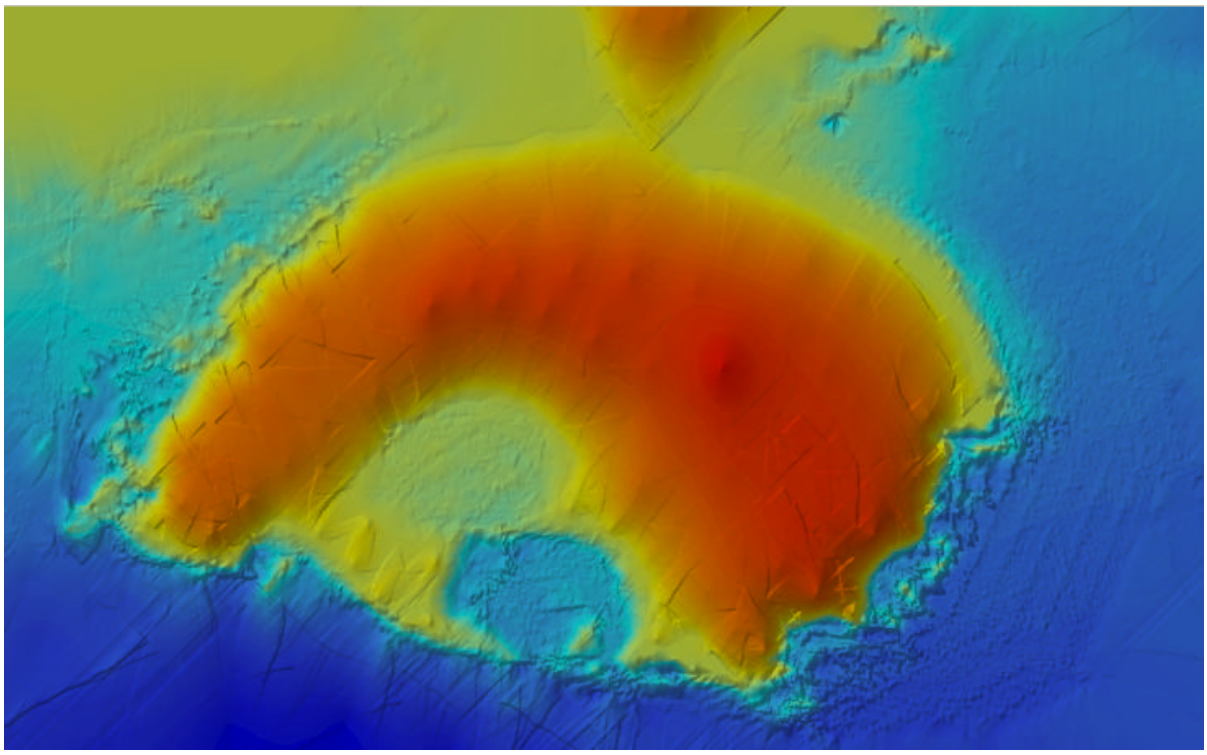

Figura 6.3 – Monte da Guia gerado com KO e semivariograma Combinado.

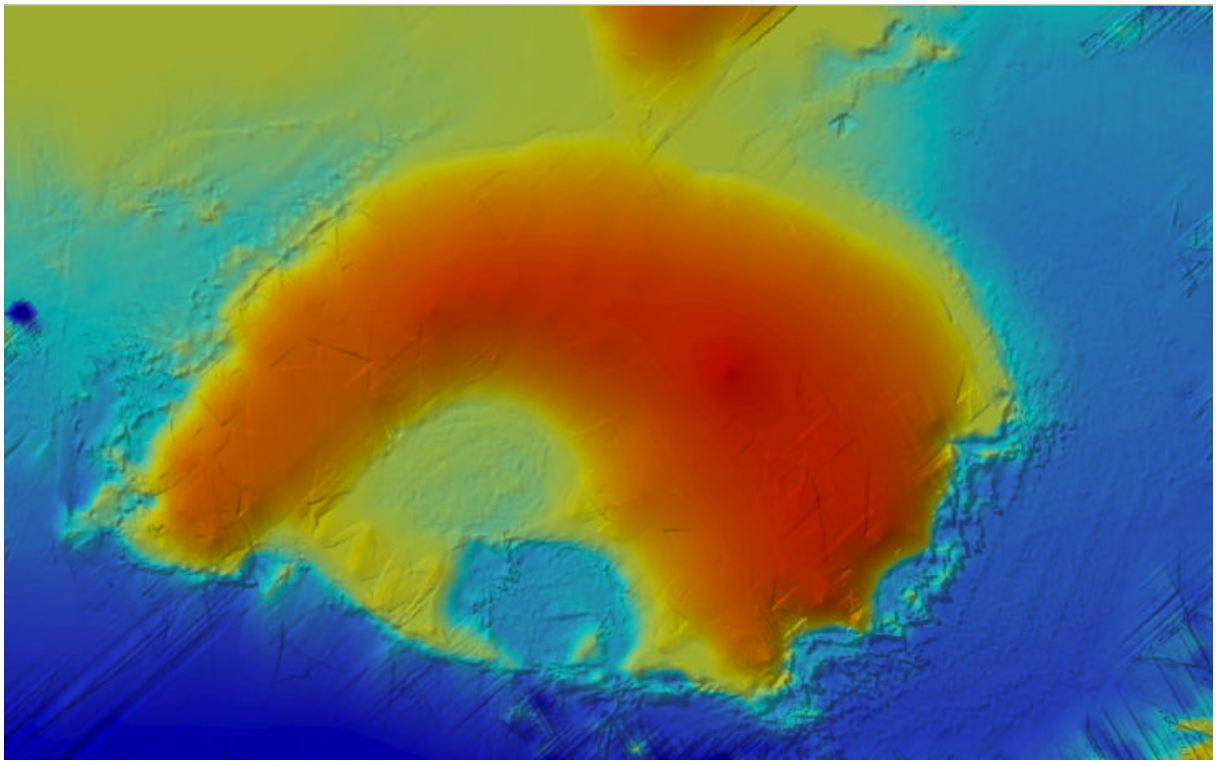

Figura 6.4 – Monte da Guia gerado com KU 1ª ordem e semivariograma Esférico.

Na zona de amostragem intensiva a Leste do Monte da Guia encontraram-se condições algo diversas dos conjuntos de amostras anteriores. As anomalias provocadas pelo método de Krigagem Universal foram bastante menores, obtendo este método bons resultados com

deriva de 1ª ordem e semivariograma Exponencial, e com deriva de 2ª ordem e semivariograma Combinado. Novamente a Krigagem Ordinária com semivariograma Combinado apresentou bons resultados. O aumento da intensidade de amostragem permitiu aos métodos que anteriormente falharam obter aqui os melhores resultados, sendo de destacar os sucessos do semivariograma Exponencial e da deriva de 2ª ordem, que apenas se registaram nestas condições.

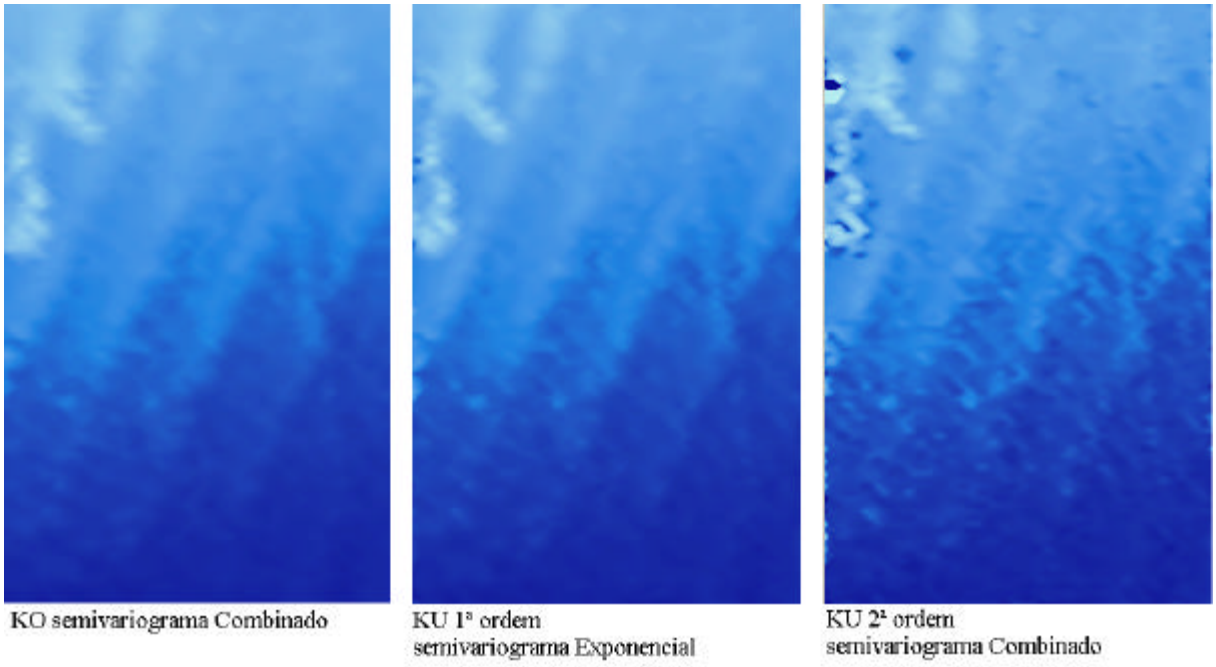

Figura 6.5 – Resultados na zona a Leste do Monte da Guia

Depois da triagem terminada é de destacar o sucesso, em todos os diferentes conjuntos de amostras, da Krigagem Ordinária com semivariograma Combinado. Este é o mais versátil de todos modelos, tendo bons desempenhos em condições, quer de grande, quer de fraca intensidade de amostragem. O semivariograma Combinado tinha sido já identificado como tal, dado combinar da melhor maneira as características dos semivariogramas Exponencial e Gaussiano. A versatilidade do método de Krigagem Ordinária vem do facto de ser considerada a deriva apenas numa vizinhança limitada, que se vai adaptando, ao longo da área de interpolação, às características dessa mesma vizinhança.

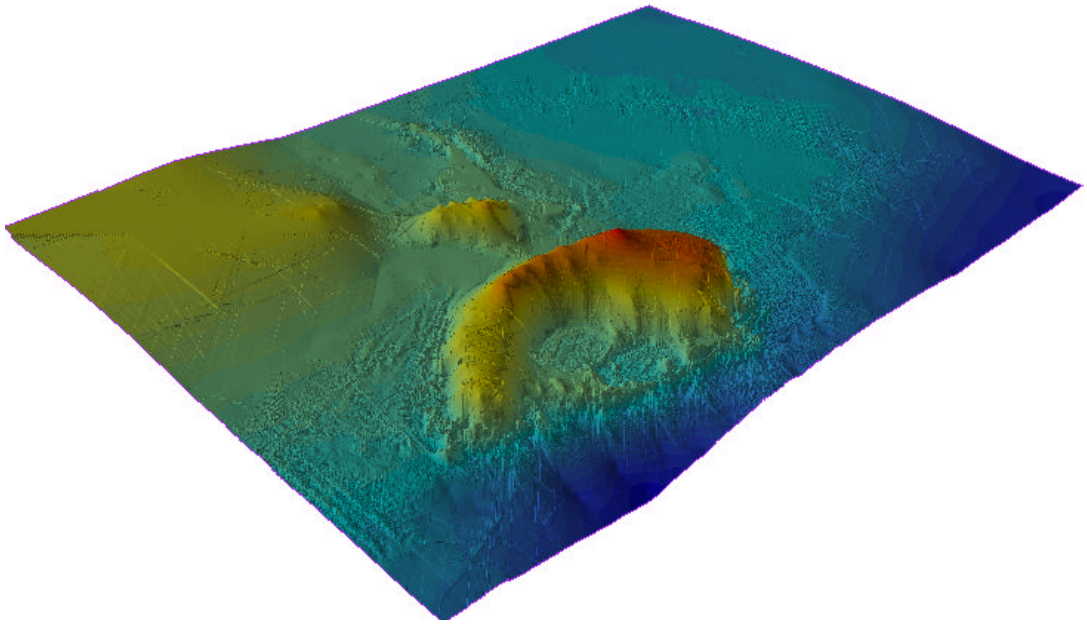

Figura 6.6 – TIN gerado com KO e semivariograma Combinado

Por último, foram gerados os melhores modelos para cada conjunto de amostras, considerando um erro de medição com base na variância registada para cada amostra. Como seria de esperar, a introdução do erro permitiu aplainar as superfícies geradas, mas com o custo de resultados piores quer de Validação quer de Validação Cruzada. Na zona a Leste do Monte da Guia a introdução do erro teve resultados diferentes, não se registando um impacto negativo na qualidade dos modelos, mas as melhorias observadas foram diminutas.

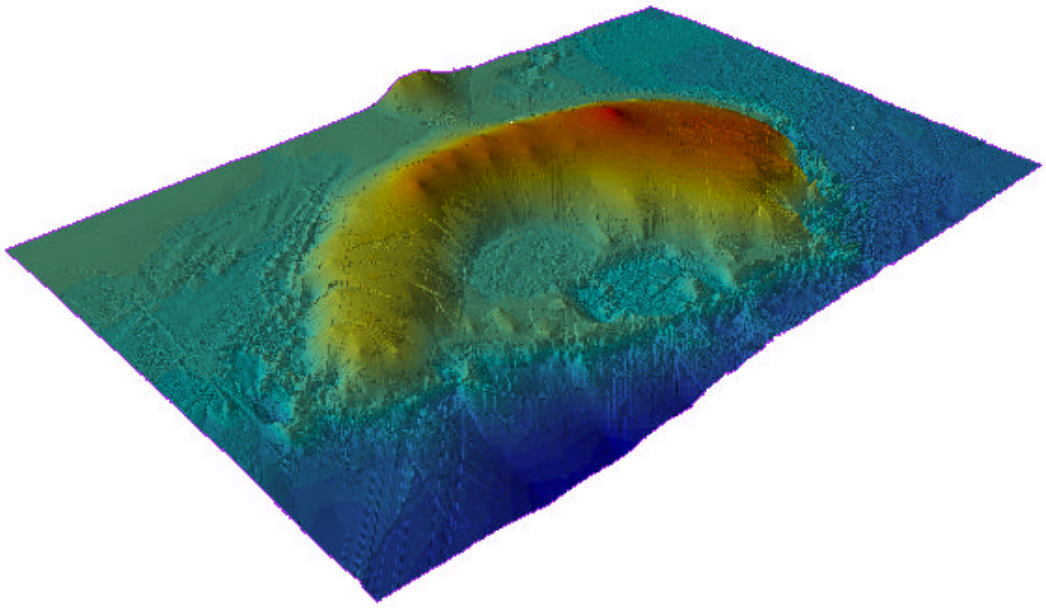

Figura 6.7 – TIN do Monte da Guia gerado com KO e semivariograma Combinado

6.3 Trabalho Futuro

Pode-se concluir que os objetivos traçados no início deste estudo foram alcançados, no entanto muitas questões foram levantadas, que podem ser desenvolvidas.

Primeiramente podem-se referir os tectos de definição alcançados nas áreas de estudo maiores, na zona do Monte da Guia, as superfícies finais apresentam uma célula de aproximadamente 1,2 metros de lado; na área total as superfícies finais têm uma célula de 1,7 metros. Estes tectos existem não só pelo tempo que é necessário ao calculo da superfície (cerca de 12 horas nos piores casos), como na quantidade de memória virtual necessária. Para resoluções maiores pode-se facilmente esgotar a memória virtual da máquina usada, bem como chegar a estados onde não é possível ter toda a memória virtual mapeada na memória RAM da máquina. Neste caso a interpolação pode gerar acessos ininterruptos ao disco rígido. Será interessante estudar a geraração das superfícies finais em máquinas diferentes da usada.

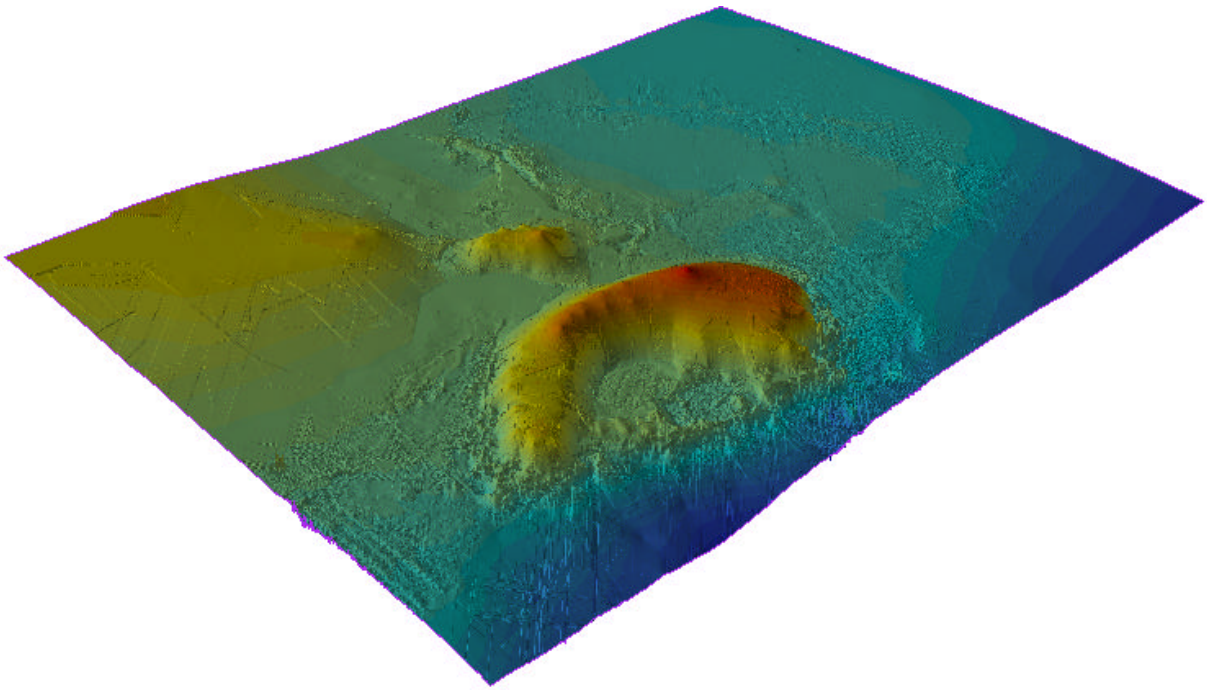

Figura 6.8 – TIN gerado com KU de 1ª ordem e semivariograma Combinado

A questão mais importante levantada foi o falhanço da Krigagem Universal com deriva de 2ª ordem nas áreas de amostragem maiores. Especialmente na zona envolvente do Monte da Guia seria de esperar que este modelo obtivesse os melhores modelos, dado existir um ponto de altitude dominante. Mas este método gerou aqui valores anómalos, muito para além do admitido. Estes anomalias estão documentadas nos manuais de Geoestatística, e praticamente incapacita a Krigagem Universal para conjuntos de amostras mais esparsos. Mais grave ainda

é o facto de no estudo do fundo dos oceanos a informação ser quase sempre pouco abrangente devido à sua extensão. È pois pertinente o estudo aprofundado de formas de contornar esta incapacidade da Krigagem Universal.

A representação digital da elevação em terra, é estudada ha bastante tempo, tendo sido uma das primeiras áreas de interesse no mundo dos SIG. Nos dias de hoje, com o auxílio de técnicas como a Fotografia Aérea, poucas serão as áreas terrestres não documentadas digitalmente. Mas para o fundo do Mar, passa-se quase o oposto. Primeiro, só recentemente se tornou possível registar informação no Mar, como se faz em terra com fotografia aérea, e só recorrendo a satélites. Depois há o facto de existirem ainda áreas consideráveis de fundo dos oceanos por explorar, e acerca dos quais a informação é muito escassa, ou mesmo inexistente. Por estas razões o estudo de formas de representar e guardar informação do fundo dos oceanos poderá estar ainda nos primeiros passos.

ANEXO A

BREVE HISTÓRIA DOS SIG

A história dos SIG começa com as primeiras tentativas de representar o mundo num mapa, inicialmente em placas de barro que seriam posteriormente cozidas. Daí se evoluiu para os mapas em papiro e mais tarde em papel. Entretanto foram ficando na História nomes como Erathotenes (raio da Terra sec. II A.C.), Ptolomeu (mapa Mediterrâneo sec. II D.C.), até à chegada da era das descobertas. Nos séculos XV e XVI, sucedem uma série de inovações que culminam no Atlas de Mercator, considerado como o pai da Cartografia Moderna.

No século XVIII novas informações são recolhidas um pouco por todo Mundo, as viagens marítimas de investigação levadas a cabo por Cook, a exploração do continente Americano nas colónias que vão tornando independentes, e as primeiras explorações do coração do continente Africano. No século seguinte, dão-se uma nova série de progressos, impulsionados pela disseminação dos Caminhos de Ferro, tanto na Europa como na América. No final deste século é fundada a *National Geografic Society*, e são produzidos os primeiros mapas com informação demográfica em níveis de cor, em 1890 é fundada a *International Business Machines Cooporation* (IBM).

No início do século XX começam a ser usados mapas no estudos das ciências sociais, apoiando áreas como a Educação e Economia. Ainda antes da Primeira Guerra Mundial começa a ser usada a técnica de Fotografia Aérea, originalmente de balões ou dirigíveis, viria a permitir a produção dos primeiros Fotomosaicos. Depois de 1918 dá-se a primeira grande revolução automóvel, a sua rápida implantação leva ao aparecimento dos primeiros mapas de estradas.

Durante a Segunda Guerra Mundial é criado o Colossus, considerado o primeiro computador da História, totalmente mecânico, em 1946 fica pronto o ENIAC o primeiro computador eléctrico. Em 1949 é inventado o primeiro transístor, na década de cinquenta dá-se a primeira fase de miniaturização electrónica e aparecem os primeiros dispositivos de digitalização.

No início da década de sessenta, ao se deparar com grandes dificuldades em lidar com o crescente volume informação geográfica disponível, e necessitando de tomar decisões acerca da gestão de recursos naturais, o Governo do Canadá inicia um processo que vai levar ao desenvolvimento de um conjunto de programas informáticos para compilar toda esta informação. Este é considerado o primeiro Sistema de Informação Geográfica (SIG), e levou à criação do *Canada Geographic Information Systems*. Começam a ser desenvolvidos uma série de aplicações de gestão de recursos e de modelação de fenómenos físicos, ainda nesta década aparecem os primeiros programas de desenho assistido por computador.

Nos anos setenta começa a implantar-se a área de investigação SIG, aparecem as primeiras grandes companhias como a ESRI e ERDAS, são lançados vários programas de investigação em conjunto com Universidades nos Estados Unidos e são levadas a cabo as primeiras conferências de SIG.

Em 1982 é posto à venda o primeiro SIG comercial o ARC/INFO 1.0, em 1986 este programa passa a estar disponível para computadores pessoais, em 1987 saem para o mercado o SPANS e o IDRISI. Em 1992 está à venda o ArcView 1.0, os SIG começam a estar abertos a pessoas sem formação SIG. Quatro anos depois aparece o ArvView 2.0, entra-se em definitivo no época dos *Desktop Mappers*, proliferam inúmeras empresas especializadas em SIG, em especial nos Estados Unidos. Por esta altura os SIG de maior dimensão já possuem sistemas de gestão de base de dados próprios. Em 1999 os SIG começam a tornar-se sistemas distribuídos, funcionando com bases de dados centralizados a que se ligam inúmeras estações de trabalho.

No início do século XXI, os SIG tomam a *internet* como suporte de trabalho. Aparecem inúmeros sites informativos, e são disponibilizadas bases de dados com mapas detalhados de todo o planeta. Estão a ser desenvolvidos projectos no sentido de aproveitar a tecnologia UMTS, para a criação uma nova geração de SIG prontos a serem usados para qualquer tarefa onde quer que se esteja.

ANEXO B

SEMIVARIOGRAMAS

B.1 Área de Trabalho Completa

Nas figuras que se seguem os Semivariogramas Combinado, Exponencial, Gaussiano e Esférico, são apresentados pela mesma ordem, de cima para baixo e da direita para esquerda.

B.1.1 Krigagem Ordinária

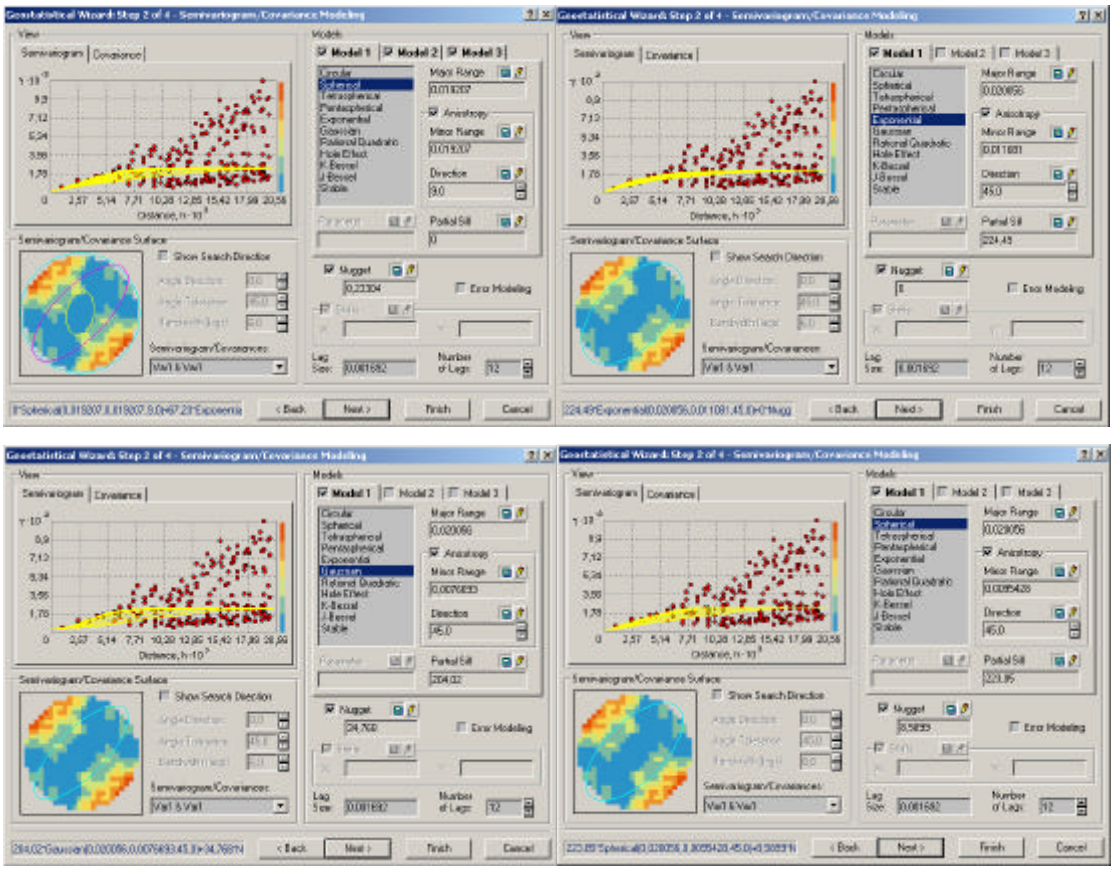

Figura B.1 – Semivariogramas usados com KO na área de trabalho completa.

Tabela B.1 – Parâmetros dos Semivariogramas de KO na área de trabalho completa

ÁREA DE TRABALHO				
	Semivariograma Esférico	Semivariograma Exponencial	Semivariograma Gaussiano	Semivariograma Combinado
Maior Alcance	0.020056	0.020056	0.020056	0.019207
Menor Alcance	0.0095428	0.011081	0.0076693	0.019207
Patamar	223.85	224.49	204.02	0
Efeito Pepita	8.5899	0	34.768	0.23304

B.1.2 Krigagem Universal com deriva de 1ª ordem

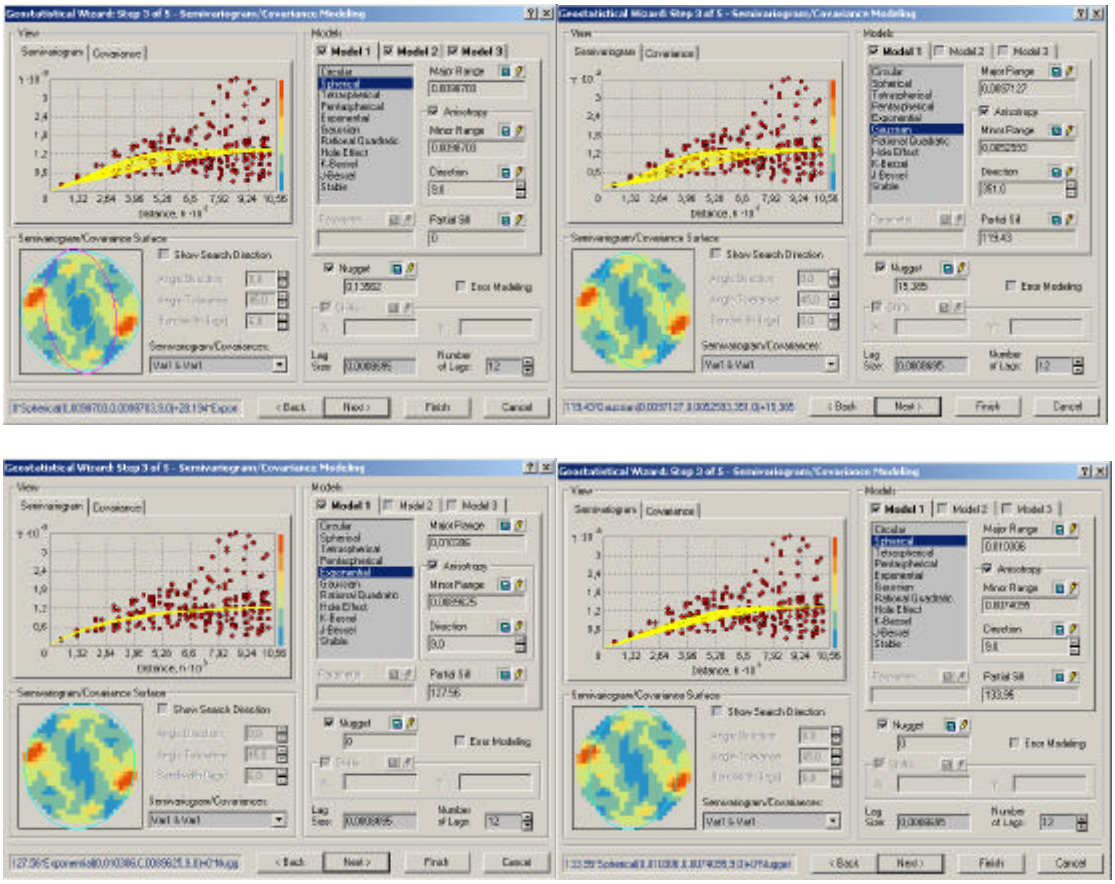

Figura B.2 – Semivariogramas usados com Krigagem Universal com deriva de 1ª ordem.

Tabela B.2 – Parâmetros dos Semivariogramas de KU com deriva de 1º ordem.

ÁREA DE TRABALHO				
	Semivariograma Esférico	Semivariograma Exponencial	Semivariograma Gaussiano	Semivariograma Combinado
Maior Alcance	0.010306	0.010306	0.0097127	0.0098703
Menor Alcance	0.0074095	0.0089625	0.0052593	0.0098703
Patamar	133.95	127.56	119.43	0
Efeito Pepita	0	0	15.385	0.13562

B.1.3 Krigagem Universal com deriva de 2ª ordem

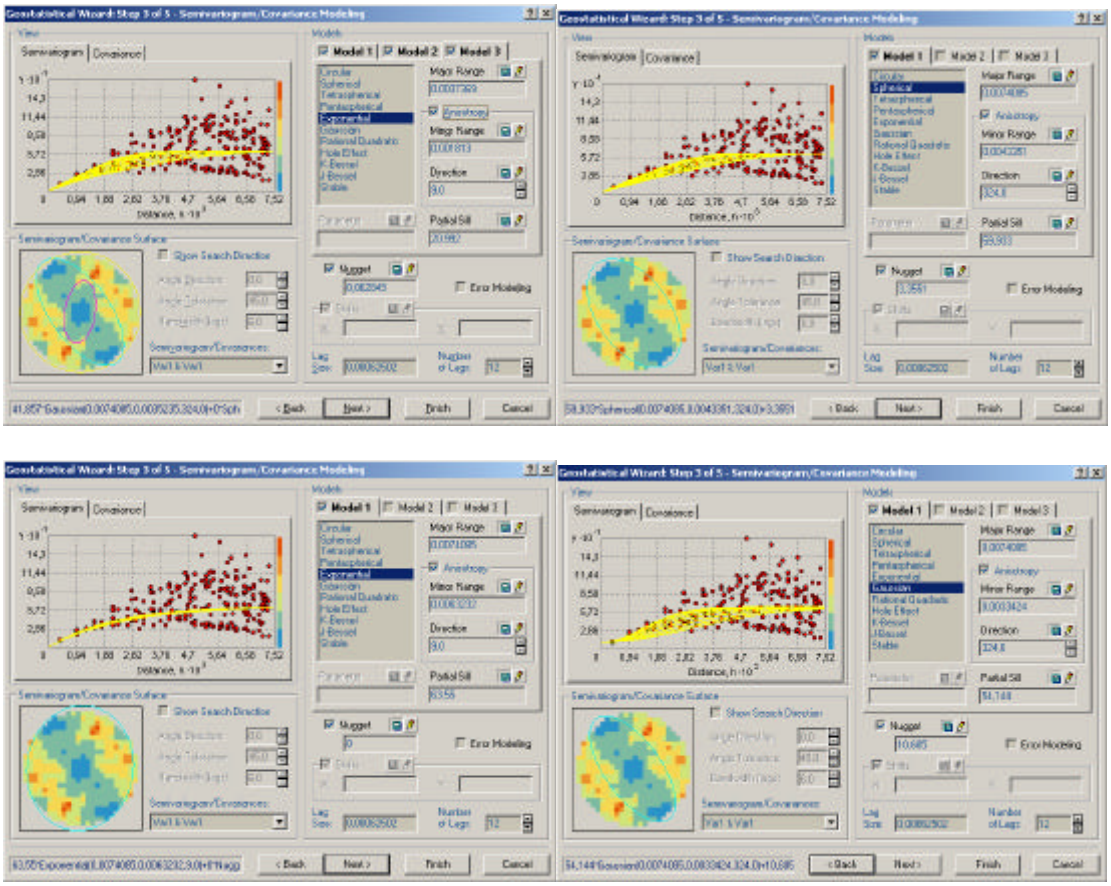

Figura B.3 – Semivariogramas usados com Krigagem Universal com deriva de 2ª ordem.

Tabela B.3 - Parâmetros dos Semivariogramas de KU com deriva de 2º ordem.

ÁREA DE TRABALHO				
	Semivariograma Esférico	Semivariograma Exponencial	Semivariograma Gaussiano	Semivariograma Combinado
Maior Alcance	0.0074085	0.0074085	0.0074085	0.0037369
Menor Alcance	0.0043351	0.0063232	0.0033424	0.001813
Patamar	59.933	63.55	54.144	20.992
Efeito Pepita	3.3551	0	10.685	0.062849

B.2 Zona do Monte da Guia

B.2.1 Krigagem Ordinária

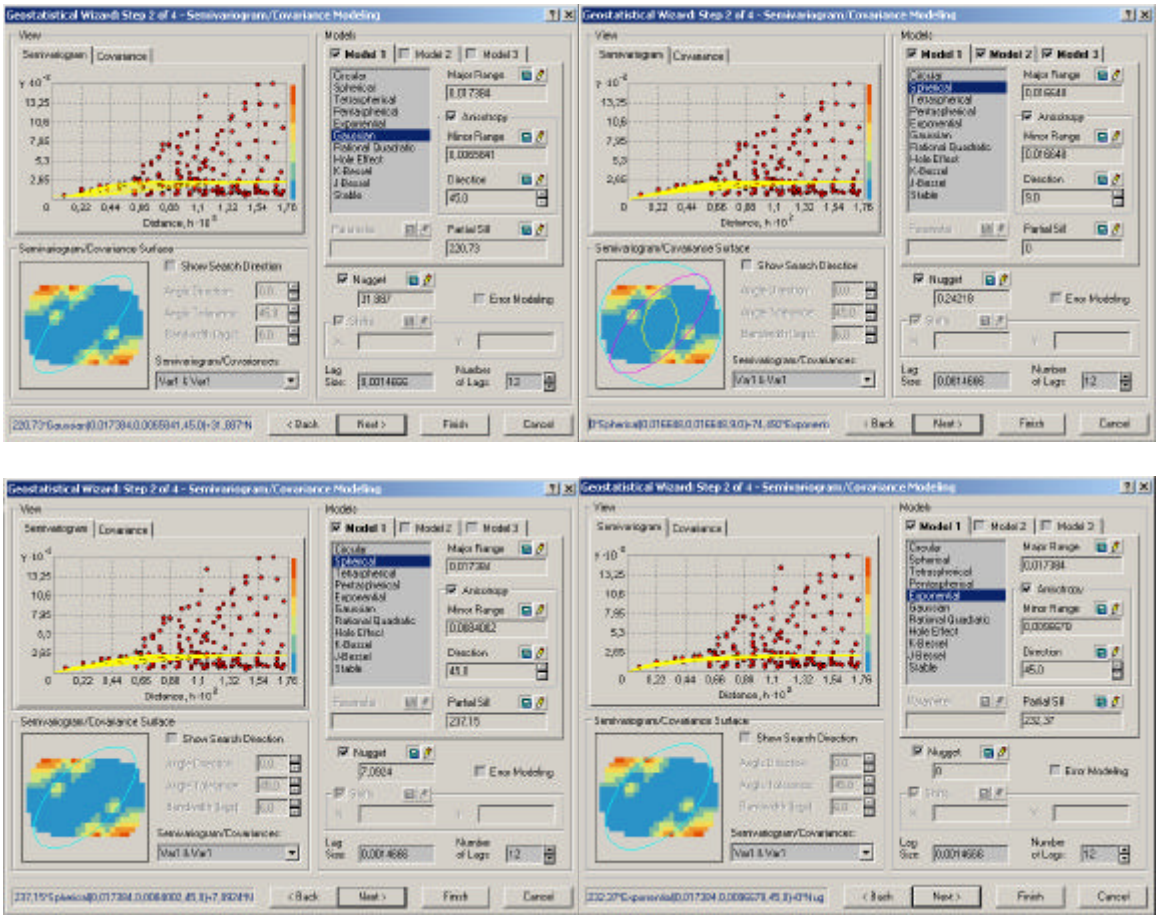

Figura B.4 – Semivariogramas usados com KO na área do Monte da Guia.

Tabela B.4 – Parâmetros dos Semivariogramas de Krigagem Ordinária.

ZONA DO MONTE DA GUIA				
	Semivariograma Esférico	Semivariograma Exponencial	Semivariograma Gaussiano	Semivariograma Combinado
Maior Alcance	0.017384	0.017384	0.017384	0.016648
Menor Alcance	0.0084002	0.0096678	0.0065841	0.016648
Patamar	237.15	232.37	220.73	0
Efeito Pepita	7.0924	0	31.887	0.24218

B.2.2 Krigagem Universal de 1ª Ordem

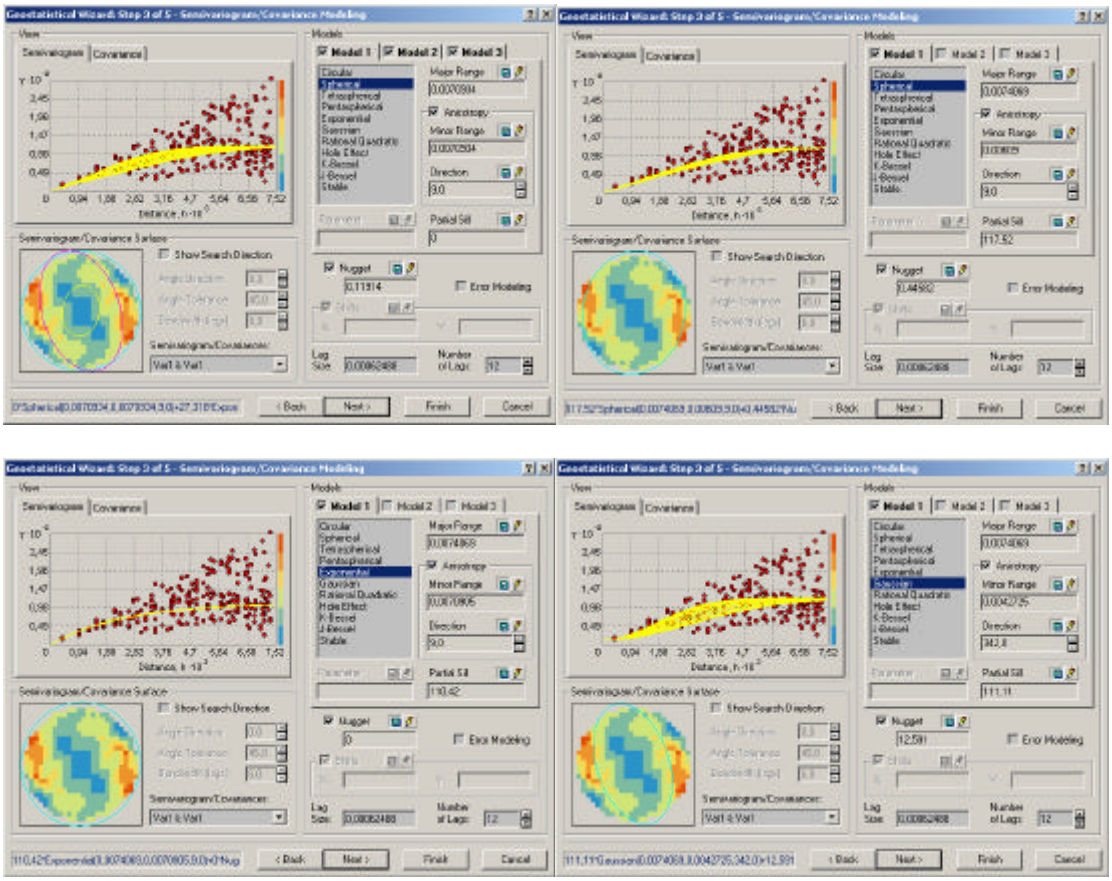

Figura B.5 – Semivariogramas da KU com deriva de 1ª ordem na zona do Monte da Guia.

Tabela B.5 – Parâmetros dos Semivariogramas da KU com deriva de 1ª ordem.

ZONA DO MONTE DA GUIA				
	Semivariograma Esférico	Semivariograma Exponencial	Semivariograma Gaussiano	Semivariograma Combinado
Maior Alcance	0.0074069	0.0074069	0.0074069	0.0070934
Menor Alcance	0.00609	0.0070905	0.0042725	0.0070934
Patamar	117.52	110.42	111.11	0
Efeito Pepita	0.44582	0	12.591	0.11914

B.2.3 Krigagem Universal com deriva de 2ª ordem

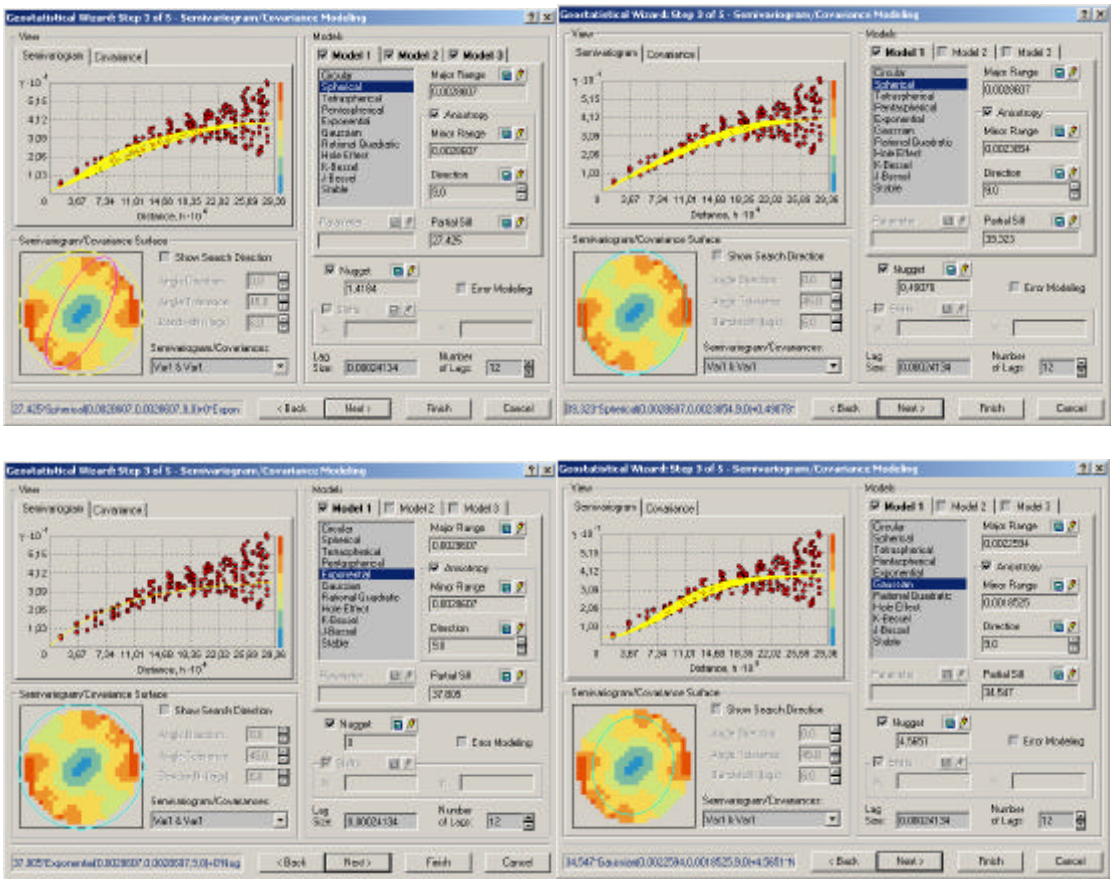

Figura B.6 – Semivariogramas da KU com deriva de 2ª ordem na zona do Monte da Guia.

Tabela B.6 – Parâmetros Semivariogramas da KU com deriva de 2ª ordem.

ZONA DO MONTE DA GUIA				
	Semivariograma Esférico	Semivariograma Exponencial	Semivariograma Gaussiano	Semivariograma Combinado
Maior Alcance	0.0028607	0.0028607	0.0022594	0.0028607
Menor Alcance	0.0023854	0.0028607	0.0018525	0.0028607
Patamar	39.0323	37.805	34.547	27.425
Efeito Pepita	0.49078	0	4.5651	1.4184

B.3 Zona a Leste do Monte da Guia

B.3.1 Krigagem Ordinária

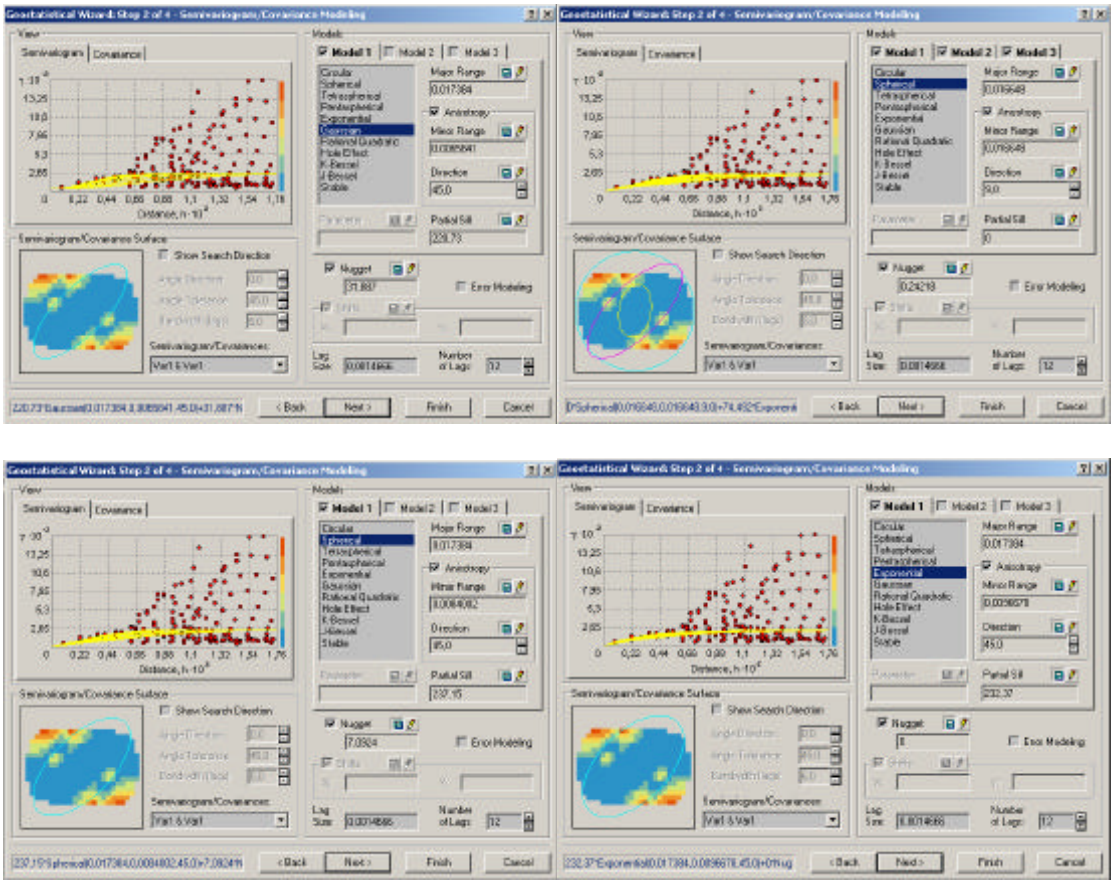

Figura B.7 – Semivariogramas da Krigagem Ordinária da zona a Leste do Monte da Guia.

Tabela B.7 – Parâmetros dos Semivariogramas da Krigagem Ordinária.

ZONA A LESTE DO MONTE DA GUIA				
	Semivariograma Esférico	Semivariograma Exponencial	Semivariograma Gaussiano	Semivariograma Combinado
Maior Alcance	319.24	319.24	319.24	305.74
Menor Alcance	245.34	319.24	219.98	305.74
Patamar	60.239	56.723	60.65	0
Efeito Pepita	0	0	6.0902	6.0902

B.3.1 Krigagem Universal com deriva de 1ª ordem

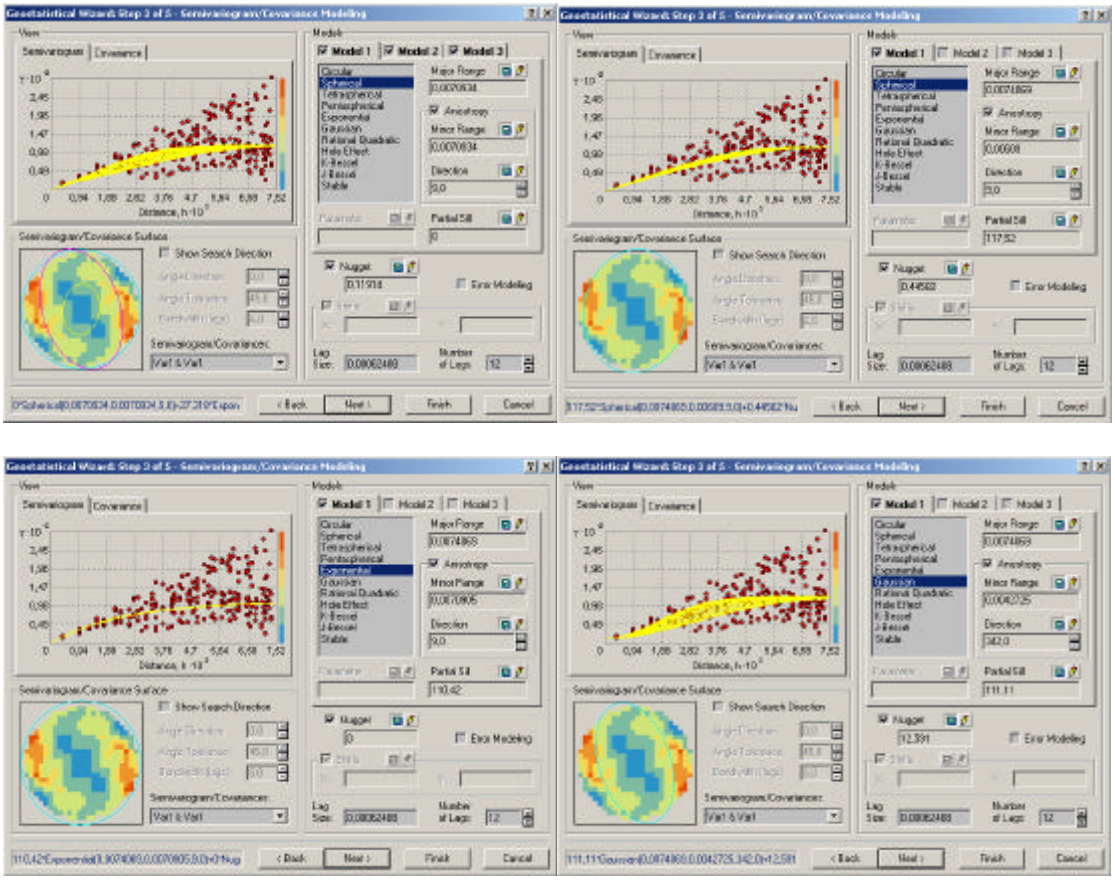

Figura B.8 – Semivariogramas da Krigagem Universal com deriva de 1ª ordem.

Tabela B.8 – Parâmetros Semivariogramas da KU com deriva de 1ª ordem.

ZONA A LESTE DO MONTE DA GUIA				
	Semivariograma Esférico	Semivariograma Exponencial	Semivariograma Gaussiano	Semivariograma Combinado
Maior Alcance	133.06	133.06	133.06	133.06
Menor Alcance	111.14	97.473	98.681	98.681
Patamar	7.8781	9.144	8.2389	8.2389
Efeito Pepita	5.236	3.8914	5.773	5.773

B.3.3 Krigagem Universal com deriva de 2ª ordem

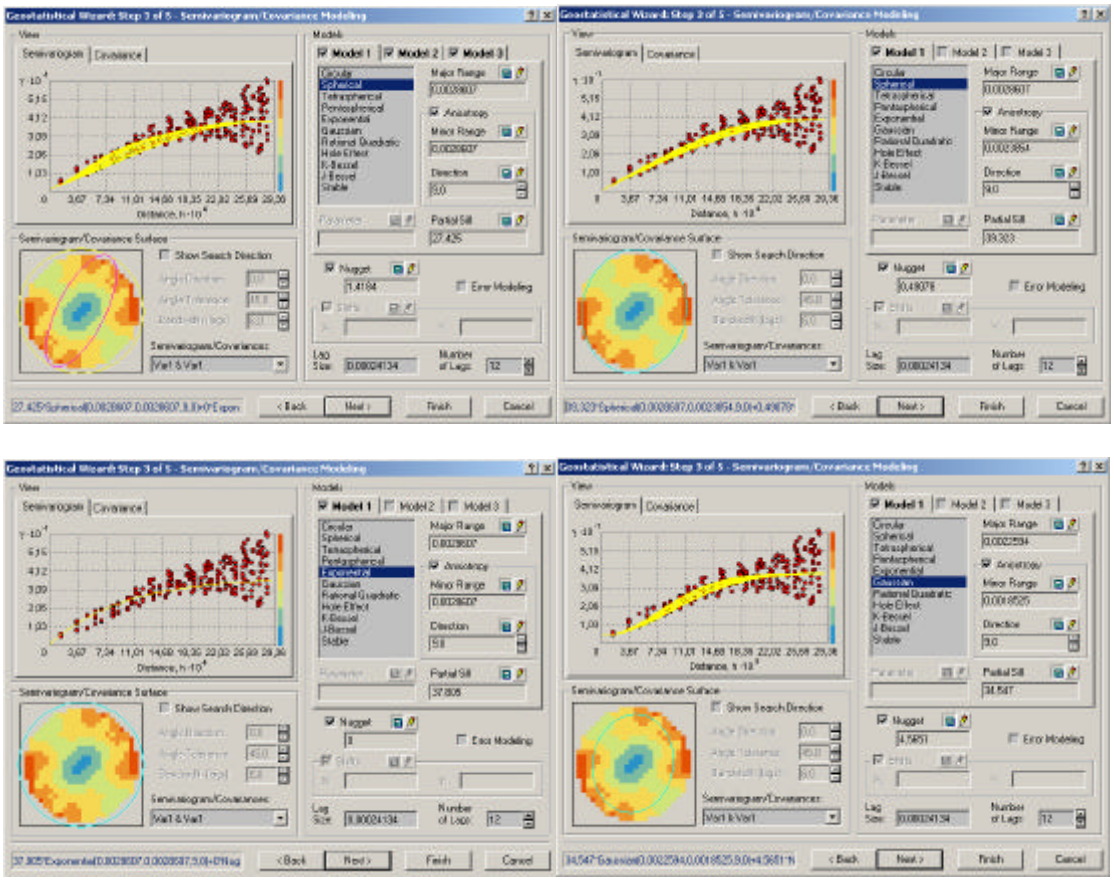

Figura B.9 – Semivariogramas da Krigagem Universal com deriva de 2ª ordem.

Tabela B.9 – Parâmetros dos Semivariogramas da KU com deriva de 2ª ordem.

ZONA A LESTE DO MONTE DA GUIA				
	Semivariograma Esférico	Semivariograma Exponencial	Semivariograma Gaussiano	Semivariograma Combinado
Maior Alcance	27.522	27.522	27.522	26.357
Menor Alcance	4.2299	4.2299	4.2299	26.357
Patamar	0.38251	0.65663	0.41985	0
Efeito Pepita		6.1571	6.395	6.395

ANEXO C

DERIVAS

C.1 Área de Trabalho Completa

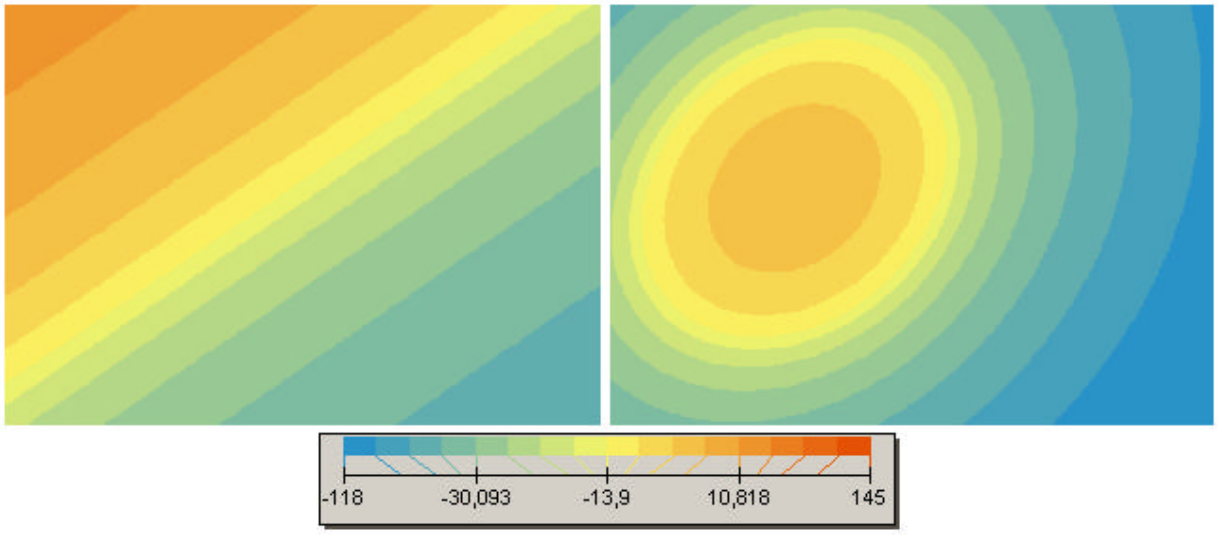

Figura C.1 – Derivas usadas nos modelos da amostragem completa

C.2 Zona do Monte da Guia

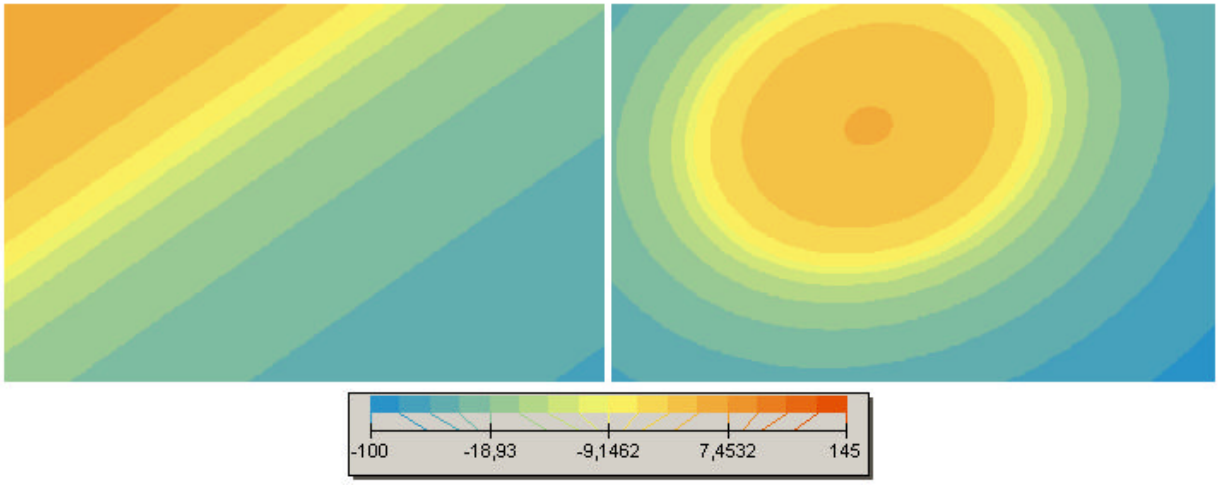

Figura C.2 – Derivas usadas nos modelos da zona do Monte da Guia

C.3 Zona a Leste do Monte da Guia

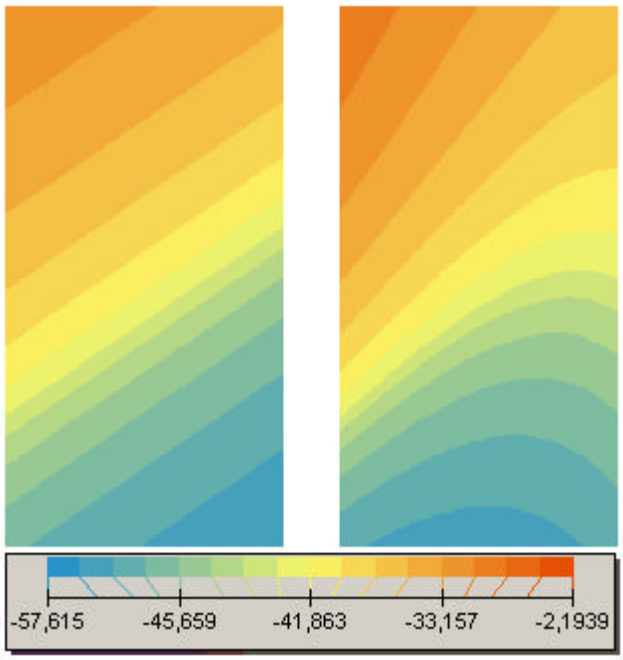

Figura C.3 – Derivas usadas nos modelos da zona a Leste do Monte da Guia

REFERÊNCIAS

[1] Kenneth E. Foote e Margaret Lynch. The Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder. *Geographic Information Systems as an Integrating Technology: Context, Concepts, and Definitions*, 1995.

[2] R. Chorley Handling Geographic Information. *Report to the Committee of Enquiry*, chaired by Lord Chorley HM Stationary Office publication, 1987.

[3] Clarke, K. C. *Analytical and Computer Cartography*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[4] http://www.cesga.es/

[5] David Cowen. NCGIA Notional Conference for Geographical Information Analysis, *International Journal of Geographic Information Systems*, pp117-136, 1989.

[6] *1994 International GIS Sourcebook*, GIS World, Inc., Fort Collins, CO., 1994.

[7] http://www.daratech.com

[8] *The Geospatial Technology Report 2001 -GITA*

[9] Joe Francica, *A review of the Geospatial Technology Report 2001 by GITA*, Directions Magazine, 16 de Janeiro 2002.

[10] C. A. Felgueiras, A. V. Monteiro, E. G. Camargo, G. C. Neto e S. D. Fuks. Integrating Geostatistical Tools in Geographical Information Systems. Em *I Brazilian Workshop on GeoInformatics GEO-INFO*, Outubro 1999.

[11] Kevin Johnston, Jay M. Ver Hoef, Konstantin Krivoruchko, Neil Lucas. *Using ArcGis Geostatistical Analyst*, ESRI, 2001.

[12] ArcGIS Geostatistical Analyst: Statistical Tools for Data Exploration, Modeling, and Advanced Surface Generation. *An ESRI White Paper*, ESRI, Redlands, CA, Agosto 2001.

[13] Amílcar Soares, *Geoestatística para as ciências da terra e do ambiente*. Colecção ensino da ciência e da tecnologia, IST Press, Lisboa, 2000.

[14] P. Goovaerts, *Geostatistic for natural resources evaluation*. Oxford University Press, New York, 1997.

[15] J. P. Chilés e P. Delfiner, *Geostatistics. Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics. Wiley – Interscience Publication, 1999.

[16] A .J. Sousa. e F. Muge, *Elementos de geoestatística. Laboratório de Mineralugia e Planeamento Mineiro*, Instituto Superior Técnico, Universidade Técnica de Lisboa, 1990.

[17] N. Cressie, *The Origins of Kriging*. Journal Mathematical Geology – 22, pp. 239 – 252, 1990.

[18] José Mata, *Economia da Empresa*, Fundação Calouste Gulbenkian, Lisboa, 2000.

[19] Paul A. Samuelson e William D. Nordhaus, *Economia*, McGraw Hill, Amadora, 1999.

[20] Pat Halpin e outros, http://www.env.duke.edu/landscape, Landscape Ecology Lab Janeiro 2002.

Alguns dos sítios consultados na internet que contribuíram para o desenvolvimento deste trabalho:

http://www.gis.com

http://www.marinegis.com

http://www.gisportal.com

http://www.opengis.org

http://www.cesga.es

http://www.usgs.gov

http://www.esri.com

http://www.pcigeomatics.com

http://www.clarklabs.org

http://www3.baylor.edu/grass

http://www.intergraph.com

http://www.manifold.net

http://www.gepower.com