

**A modelação geoestatística e os  
Sistemas de Informação Geográfica - uma  
abordagem exploratória aplicada em Saúde Pública**

por

Manuel Luís Castro Ribeiro

Dissertação apresentada como requisito  
parcial para a obtenção do grau de

Mestre em Estatística e Gestão de Informação

pelo

Instituto Superior de Estatística e Gestão de Informação  
da  
Universidade Nova de Lisboa

**A modelação geoestatística e os  
Sistemas de Informação Geográfica - uma  
abordagem exploratória aplicada em Saúde Pública**

por

Manuel Luís Castro Ribeiro

Orientadora: Doutora Maria João Correia Colunas Pereira

Dissertação apresentada como requisito  
parcial para a obtenção do grau de

Mestre em Estatística e Gestão de Informação

pelo

Instituto Superior de Estatística e Gestão de Informação  
da  
Universidade Nova de Lisboa

Setembro de 2006

# A modelação geoestatística e os Sistemas de Informação Geográfica - uma abordagem exploratória aplicada em Saúde Pública

## Errata

**Página 81** na definição do coeficiente de variação,  $CV(s)$ , deve ler-se:

$$CV(s) = \frac{\sigma(s)}{m(s)}$$

**Página 104** na fórmula do mapa de probabilidades médio,  $m(s)$ , deve ler-se:

$$m(s) = \frac{\sum_{l=1}^L P[Z'(s,t) > 40]}{L}$$

**Página 104** na fórmula do mapa de variâncias,  $\sigma^2(s)$ , deve ler-se:

$$\sigma^2(s) = \frac{\sum_{l=1}^L \{P[Z'(s,t) > 40] - m(s)\}^2}{L}$$

## **Abstract**

In this study a data driven approach based on exploratory spatial data analysis is used with geographic information technologies in order to formulate hypothesis about effects of particulate matter concentration ( $PM_{10}$ ) on child respiratory diseases and allergies. In order to integrate the spatial data in a common environment and to correlate the studied variables, a geographical referenced database was created for every variable with a Geographic Information Systems (GIS) tool with capabilities to store, manage and perform spatial modelling considering one or more variables. An exploratory spatial data analysis based on graphical methods and user interaction for analysis of spatial structures was conducted for both environmental and health variables and geostatistical modelling was performed to build child respiratory diseases and allergies and air pollution concentration, uncertainty and local correlation maps.

Given the scarcity of data and the spatial location of both variables, a co-regionalization model was not possible to be driven from the previous analysis. Hence a stochastic spatial simulation was conducted independently for each variable.  $PM_{10}$  concentrations were related to child respiratory diseases and allergies incidence by local correlation coefficients obtained with simulated realizations of air pollution and health data. In the last step, local correlation maps associated to critical values were obtained.

**Keywords:** Environment, Geostatistics, Health, GIS

## Resumo

Neste estudo aplica-se uma abordagem exploratória de dados espaciais com utilização de tecnologias de informação geográfica com o objectivo de formular hipóteses sobre os efeitos das partículas inaláveis ( $PM_{10}$ ) na incidência de alergias e doenças respiratórias em crianças. De modo a integrar os dados espaciais num só ambiente com o objectivo de correlacioná-los, foi criada uma base de dados geográficos para todas as variáveis com uma ferramenta de Sistemas de Informação Geográfica (SIG) com capacidades de armazenamento, gestão e modelação de dados espaciais. Procedeu-se a uma análise exploratória de dados espaciais baseada em métodos gráficos e de interacção com o utilizador para análise de estruturas espaciais das variáveis do ambiente e da saúde, e aplicaram-se modelos geoestatísticos para criar mapas de alergias e doenças respiratórias e de concentração de  $PM_{10}$ , mapas de avaliação de Incerteza espacial e mapas de correlação locais.

Dada a escassez de dados no espaço em ambas as variáveis, não foi possível aplicar um modelo de co-regionalização, pelo que se optou por um modelo de simulação estocástica calculado de modo independente para cada uma das variáveis. As concentrações de  $PM_{10}$  obtidas foram relacionadas com a incidência de alergias e doenças respiratórias, através do cálculo de coeficientes correlação locais das simulações obtidas com os dados de poluição e da saúde. Por fim, criaram-se mapas de correlação locais associados a valores de corte definidos.

Palavras-chave: Ambiente, Geoestatística, Saúde, SIG

## **Agradecimentos**

À minha orientadora científica, Professora Doutora Maria João Pereira, pelas críticas e sugestões, pelo incentivo, entusiasmo e disponibilidade com que orientou este trabalho.

Ao Professor Doutor Amílcar Soares, do Departamento de Engenharia de Minas do Instituto Superior Técnico pelas sugestões enriquecedoras e pela disponibilidade que sempre demonstrou ao longo das várias fases deste trabalho.

Ao Doutor José Rosado Pinto, Director do Serviço de Imunoalergologia do Hospital D. Estefânia, que colaborou neste trabalho por amavelmente ter disponibilizado os dados sobre o International Study of Asthma and Allergies in Childhood para a região de Lisboa.

Ao Dr. José Rufino, dos Correios de Portugal S.A, e ao Dr. Rui Mestre Cardoso, do Ministério da Educação, pela total disponibilidade e auxílio prestado na aquisição dos dados geográficos e tabulares usados na elaboração deste trabalho.

Ao António Pinto de Carvalho, Manuel Margarido Ribeiro e Maria da Luz Castro pelo apoio moral e disponibilidade incedível que sempre manifestaram.

À Sara pela tolerância dos momentos em que não estive disponível.

## Índice

Lista de Anexos .....	vi
Lista de Tabelas .....	vii
Lista de Figuras .....	viii
1 Introdução .....	1
1.1 Enquadramento .....	1
1.2 Objectivos do estudo .....	2
1.3 Metodologia .....	3
2 Poluição atmosférica e saúde pública .....	5
2.1 Epidemiologia .....	5
2.2 O problema da poluição atmosférica .....	8
2.3 Os efeitos da poluição atmosférica na saúde pública .....	10
2.4 Os mapas de incidência ou prevalência de doenças .....	13
3 Sistemas de informação geográfica (SIG) .....	16
3.1 Introdução .....	16
3.2 A integração de dados geográficos .....	16
3.3 Análise espacial em saúde pública .....	19
4 Dados e análise exploratória .....	22
4.1 Variáveis da saúde .....	22
4.1.1 O International Study of Asthma and Allergies In Childhood (ISAAC) .....	22
4.1.2 A razão padronizada de incidência (SIR) .....	23
4.1.3 A dimensão da amostra .....	24
4.1.4 Área de estudo .....	24
4.1.5 Análise descritiva de dados .....	26
4.2 Variável de qualidade do ar .....	35
4.2.1 Base de dados Qualar .....	35
4.2.2 Área de estudo .....	35
4.2.3 Análise descritiva de dados .....	37

4.3	Comentários .....	44
5	Modelação geoestatística .....	45
5.1	Introdução ao formalismo geoestatístico .....	45
5.2	Análise da continuidade espacial e temporal .....	47
5.2.1	Introdução .....	47
5.2.2	Variáveis da saúde .....	50
5.2.3	Variável da qualidade do ar .....	55
5.3	Inferência espacial .....	67
5.3.1	Introdução aos métodos de estimação geoestatísticos .....	67
5.3.2	Variáveis da saúde .....	70
5.3.3	Variável de qualidade do ar .....	73
5.4	Avaliação da incerteza espacial .....	75
5.4.1	Introdução à simulação estocástica geoestatística .....	75
5.4.2	Variáveis da saúde .....	77
5.4.3	Variável da qualidade do ar .....	85
5.4.4	Síntese dos resultados obtidos .....	105
6	Análise das relações entre a qualidade do ar e a saúde .....	105
6.1	Introdução .....	105
6.2	Análise e apresentação de resultados .....	107
6.2.1	Mapas de probabilidades para as variáveis da saúde e qualidade do ar .....	107
6.2.2	Mapas de correlações locais .....	111
7	Conclusões .....	119
8	Referências bibliográficas .....	121
	Anexos .....	124

## Lista de Anexos

Anexo A - Diagramas de dispersão das PM10 - Período de 01/01/2001 a 31/12/2003..	124
Anexo B - Análise do efeito proporcional nas variáveis da saúde.....	129
Anexo C - Variogramas temporais das PM10- Período de 01/01/2001 a 31/12/2003.....	133
Anexo D - Variogramas das simulações das variáveis da saúde.....	135
Anexo E - Histogramas das simulações das variáveis da saúde.....	155
Anexo F - Mapas de probabilidade, variância e coeficientes de variação das PM10 por período (valor de corte=40 $\mu\text{g}/\text{m}^3$ ).....	175
Anexo G - Mapas de probabilidades e de variâncias das variáveis da saúde (valores de corte= P.50 e P.75.....	181
Anexo H - Mapas de correlações locais com buffers de 2500, 5000 e 10000 metros.....	188

## Lista de tabelas

Tabela 4.1- Questões do ISAAC consideradas no âmbito do estudo.....	22
Tabela 4.2- Variáveis de saúde consideradas no âmbito do estudo .....	22
Tabela 4.3- Proporção (P) das populações inquiridas que declararam já terem tido pieira, asma, rinite ou eczema (retirado e adaptado de Pinto e Almeida, 2003) e respectivas variâncias ( $S^2$ ) corrigidas.....	24
Tabela 4.4- Parâmetros amostrais das variáveis da saúde.....	27
Tabela 4.5- Estações de medição de $PM_{10}$ consideradas no âmbito do estudo.....	35
Tabela 4.6- Medidas de localização e dispersão das concentrações médias diárias de $PM_{10}$ .....	39
Tabela 4.7- Coeficientes de correlação de Pearson das concentrações médias diárias de $PM_{10}$ .....	41
Tabela 5.1- Modelos de variograma e respectivos parâmetros, ajustados às variáveis de saúde .....	55
Tabela 5.2- Definição das dimensões X, Y e Z do domínio a simular .....	85
Tabela 5.3- Definição da dimensão Z para cada periodo de tempo .....	85
Tabela 5.4- Valores conhecidos das médias em cada estação, para os vários periodos. .....	93
Tabela 5.5- Valores-limite estabelecidos pela 1ª Directiva-Filha 1999/30/CE para as $PM_{10}$ .....	103
Tabela 5.6- Estatísticas descritivas dos mapas de probabilidade para um valor de corte de $40 \mu\text{g}/\text{m}^3$ .....	105
Tabela 6.1- Dados considerados na análise das relações entre as variáveis da saúde e qualidade do ar .....	107
Tabela 6.2- Cenários escolhidos para análise das relações entre as variáveis da saúde e qualidade do ar.....	108
Tabela 6.3- Valores de corte definidos para as variáveis da saúde .....	108

## Lista de figuras

Figura 2.1- Deposição de partículas em função do seu diâmetro, em várias regiões dos pulmões (retirado de Boubel, Fox et al., 1994).....	12
Figura 3.1- Operações num SIG segundo o seu grau de complexidade de implementação.....	18
Figura 3.2- O enquadramento dos SIG na análise exploratória em saúde pública (adaptado de Douven e Scholten, 1995).....	21
Figura 4.1-Localização das escolas consideradas no estudo.....	25
Figura 4.2- Ambiente envolvente em duas das escolas inseridas no estudo a) escola em ambiente rural e b) escola em ambiente urbano.....	25
Figura 4.3- Histogramas dos valores das variáveis da saúde.....	27
Figura 4.4- Distribuição espacial dos SIR (.a) e das respectivas variâncias (.b) das variáveis ash_13 (1), ash_6 (2), ecz_13 (3), ecz_6 (4), hay_13 (5), hay_6 (6), pnose_13 (7), pnose_6 (8), rash_13 (9), rash_6 (10), whez_13 (11), whez_6 (12). .....	34
Figura 4.5- Localização das estações de medição consideradas no estudo.....	36
Figura 4.6- Localização de duas das estações de medição a) de fundo (E6) e b) de tráfego (E1).....	36
Figura 4.7- Histograma e alguns parâmetros amostrais de concentrações médias diárias de PM <sub>10</sub> observadas no conjunto das estações de medição. ....	37
Figura 4.8- Histogramas das concentrações médias diárias de PM <sub>10</sub> .....	38
Figura 4.9- Séries temporais das concentrações médias diárias de PM <sub>10</sub> .....	40
Figura 4.10- Distribuição geográfica e concentrações médias de PM <sub>10</sub> .....	42
Figura 4.11- Correlograma (correlação linear de Pearson) das concentrações médias diárias de PM <sub>10</sub> .....	43
Figura 4.12- Correlograma (autocorrelação de Moran I) das concentrações médias diárias de PM <sub>10</sub> .....	44
Figura 5.1- Variograma experimental e ajustamento de modelo esférico (ash_13).....	52
Figura 5.2- Variograma experimental e combinação linear de modelos esféricos (ash_6).....	53
Figura 5.3- Variograma experimental e ajustamento de modelo esférico (hay_13).....	53
Figura 5.4- Variograma experimental e ajustamento de modelo exponencial (rash_13) .....	54
Figura 5.5- Variograma experimental e ajustamento de modelo esférico (pnose_13). 55	55

Figura 5.6- Modelos temporais ajustados, das concentrações médias diárias de $PM_{10}$	57
Figura 5.7- Variogramas temporais das concentrações médias diárias de $PM_{10}$ nas estações de tráfego (1.) e de fundo (2.)	58
Figura 5.8- Variograma espacial das concentrações médias diárias de $PM_{10}$	59
Figura 5.9- Variograma espacial das concentrações médias diárias de $PM_{10}$ estandardizadas	61
Figura 5.10- Coeficientes de correlação linear de Pearson entre estações de medição ao longo do tempo por períodos: A – de 15/08/2003 a 31/12/2003; B – 30/04/2003 a 15/08/2003; C – 11/03/2003 a 30/04/2003; D - 09/02/2003 a 11/03/2003; E – de 17/01/2003 a 09/02/2003; F – de 17/12/2002 a 17/01/2003; G – de 02/09/2002 a 17/12/2002; H – de 14/07/2002 a 02/09/2002; I – de 25/04/2002 a 14/07/2002; J – de 24/02/2002 a 25/04/2002	62
Figura 5.11- Variogramas espaciais (a) e temporais (b) das concentrações médias diárias de $PM_{10}$ estandardizadas, ao longo do tempo por períodos: A – de 15/08/2003 a 31/12/2003; B – 30/04/2003 a 15/08/2003; C – 11/03/2003 a 30/04/2003; D - 09/02/2003 a 11/03/2003; E – de 17/01/2003 a 09/02/2003; F – de 17/12/2002 a 17/01/2003; G – de 02/09/2002 a 17/12/2002; H – de 14/07/2002 a 02/09/2002; I – de 25/04/2002 a 14/07/2002; J – de 24/02/2002 a 25/04/2002	65
Figura 5.12- Modelos esféricos obtidos para caracterizar a variografia espacial	66
Figura 5.13- Mapas das estimativas do risco relativo de incidência de ash_13 (a.) ash_6 (b.) hay_13 (c.) rash_13 (d.) e pnose_13 (e.) obtidos com o estimador linear de Krigagem Normal	73
Figura 5.14- Mapas de concentrações medias diárias de $PM_{10}$ obtidos com o estimador linear de Krigagem Normal, para os períodos de tempo A (a.), C (b.), D (c.), E (d.), F (e.), G (f.), H (g.) e J (h.)	74
Figura 5.15- Exemplos de realizações equiprováveis simuladas para as variáveis da saúde	80
Figura 5.16- Mapas de médias (a), variâncias (b) e coeficientes de variação (c) obtidos pelo método de simulação sequencial directa: ash_13 (1.); ash_6 (2.); hay_13 (3.); rash_13 (4.) e pnose_13 (5.)	84
Figura 5.17- Três realizações dos dias 27, 28 e 29 de Setembro de 2003	86
Figura 5.18- Histogramas e estatísticas descritivas dos dados experimentais e das simulações #7, #11 e #20 do período A	87
Figura 5.19- Variogramas espaciais (a.) e temporais (b.) das simulações #7, #11 e #20 do período A	88

Figura 5.20- Três realizações dos dias 25, 26 e 27 de Março de 2003.....	89
Figura 5.21- Histogramas e estatísticas descritivas dos dados experimentais e das simulações #3, #6 e #23 do período C. ....	91
Figura 5.22- Variogramas espaciais (a.) e temporais (b.) das simulações #3, #6 e #23 do período C.....	91
Figura 5.23- Variografia espacial das médias e das variâncias para o período C. ....	94
Figura 5.24- Mapas das médias e desvios-padrão locais estimados para o período C95	
Figura 5.25- Variogramas espaciais e temporais de 5 das 30 simulações da variável transformada obtidas para o período C. ....	96
Figura 5.26- Histogramas e estatísticas descritivas dos dados experimentais (transformados) e de 5 simulações, para o período C. ....	98
Figura 5.27- Variogramas espaciais e temporais de 5 das 30 simulações da variável original obtida por transformação inversa, para o período C.....	99
Figura 5.28- Histogramas e estatísticas descritivas dos dados experimentais e de 5 simulações (após transformação inversa), para o período C.....	101
Figura 5.29- Um dos mapas de probabilidades obtido para um valor de corte=40 $\mu\text{g}/\text{m}^3/\text{dia}$ (Período A).....	103
Figura 6.1-Domínio espacial considerado na análise.....	106
Figura 6.2- Mapas de probabilidades e de variâncias da variável Ash_13 para os valores de corte definidos.....	109
Figura 6.3- Mapas de probabilidades e de variâncias de PM <sub>10</sub> para os valores de corte definidos (Período G) .....	111
Figura 6.4- Mapa de correlações locais entre a excedência da razão padronizada de incidência de ash_13 =1.07 e da concentração de PM <sub>10</sub> = 20 $\mu\text{g}/\text{m}^3$ .....	113
Figura 6.5- Mapas de correlações locais obtidos a. Ash_13 (P75) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), b. Ash_13 (P75) e PM10 (40 $\mu\text{g}/\text{m}^3$ ), c. Ash_13 (P50) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), d. Ash_13 (P50) e PM10 (40 $\mu\text{g}/\text{m}^3$ ), e. Ash_6 (P75) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), f. Ash_6 (P75) e PM10 (40 $\mu\text{g}/\text{m}^3$ ), g. Ash_6 (P50) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), h. Ash_6 (P50) e PM10 (40 $\mu\text{g}/\text{m}^3$ ), i. Hay_13 (P75) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), j. Hay_13 (P75) e PM10 (40 $\mu\text{g}/\text{m}^3$ ), l. Hay_13 (P50) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), m. Hay_13 (P50) e PM10 (40 $\mu\text{g}/\text{m}^3$ ), n. Pnose_13 (P75) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), o. Pnose_13 (P75) e PM10 (40 $\mu\text{g}/\text{m}^3$ ), p. Pnose_13 (P50) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), q. Pnose_13 (P50) e PM10 (40 $\mu\text{g}/\text{m}^3$ ), r. Rash_13 (P75) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), s. Rash_13 (P75) e PM10 (40 $\mu\text{g}/\text{m}^3$ ), t. Rash_13 (P50) e PM10 (20 $\mu\text{g}/\text{m}^3$ ), u. Rash_13 (P50) e PM10 (40 $\mu\text{g}/\text{m}^3$ ). ....	117

# 1 Introdução

## 1.1 Enquadramento

O estudo das relações entre a poluição atmosférica e a incidência de doenças respiratórias e alérgicas em crianças reveste-se actualmente de elevado interesse, dado verificar-se ao longo das últimas décadas, um aumento da prevalência destas doenças em grupos da população mais vulneráveis (as crianças, os idosos e as pessoas com problemas cardio-respiratórios) associada a factores ambientais. Esta evidência, resultado de estudos conduzidos nas últimas décadas por todo o mundo (uma boa síntese de vários estudos publicados pode ser consultada em Tamburini, Ehrenstein et al., 2002), desencadeou na Europa um processo de planeamento a longo prazo, que passa por um esforço concertado envolvendo os principais organismos e instituições europeias, no sentido de promover políticas de protecção e prevenção da saúde dos efeitos nocivos provocados por factores ambientais. As declarações expressas nas últimas Conferências Ministeriais do Processo Ambiente e Saúde, organizadas pela OMS (Organização Mundial de Saúde) e realizadas em Londres (1999) e em Budapeste (2004), evidenciaram o reconhecimento da importância destas políticas para a saúde das crianças e gerações futuras, e culminaram com a apresentação, em Budapeste, do Plano de Acção Europeu sobre Ambiente e Saúde para as Crianças (CEHAPE) que suporta o Programa "Ar limpo para a Europa" (CAFE), lançado pela Comissão Europeia em 2001. Neste plano de acção incluem-se a criação de um sistema de integração das informações sobre o estado do ambiente e da saúde, para desenvolver um conjunto de iniciativas destinadas a aprofundar o conhecimento e a compreensão sobre a relação entre ambos, e a determinar o modo como a exposição ambiental provoca efeitos epidemiológicos.

Na sequência dos compromissos assumidos por Portugal na conferência de Budapeste, encontra-se em fase de desenvolvimento um Programa Nacional em Saúde Ambiental inserido no Plano Nacional de Saúde 2004-2010 (<http://www.dgsaude.min-saude.pt/pns>), sob a coordenação da Direcção Geral da Saúde (DGS) e com a colaboração dos Centros Regionais de Saúde Pública onde se incluem acções específicas para protecção e prevenção da saúde das crianças, cuja implementação deverá ter início em 2007.

## 1.2 Objectivos do estudo

Apesar de existirem inúmeros estudos sobre os efeitos da poluição atmosférica na incidência e prevalência de asma e alergias em crianças, nem sempre os resultados são consistentes quanto à incidência ou grau de exposição observados, pelo que ainda não existe uma ideia clara sobre o modo como ambos se relacionam. Os métodos e abordagens usados geralmente nos estudos com dados agregados (as observações são contagens indexadas, por exemplo, a áreas administrativas, centros de saúde ou escolas), recorrem geralmente a ferramentas estatísticas para quantificar os graus de associação entre a exposição à poluição atmosférica e a incidência de asma e alergias sem considerar a distribuição espacial dos fenómenos, cuja variação tem actualmente uma importância considerável no estudo de incidência das doenças (Westlake, 1995; Lawson, Biggeri et al., 2000).

Com este trabalho apresenta-se uma metodologia baseada na análise exploratória de dados, isto é, onde não são impostas condições sobre modelos subjacentes aos dados (também frequentemente designada por abordagem *data-driven*), para descrever a distribuição espacial da incidência e prevalência da asma e alergias em crianças e formular hipóteses sobre a sua etiologia em função da exposição a partículas inaláveis de diâmetro inferior a 10 µm. Por um lado, recorrem-se a Sistemas de Informação Geográfica para integrar bases de dados relacionais de dados geográficos sobre variáveis do ambiente e da saúde, e por outro, a modelos geoestatísticos para a construção de mapas de incidência de doenças respiratórias e alergias em crianças e mapas de exposição à poluição atmosférica. Num estudo de Carrat e Valleron (1992) aplicaram-se modelos geoestatísticos a ocorrências de gripe (*influenza*) observadas em França, com o objectivo de representar o risco de ocorrência num mapa de isolinhas. Collins (1998), publicou um estudo no âmbito do projecto *Small Areas Variations in Air Quality and Health* (SAVIAH), onde aplicou o estimador geoestatístico de Krigagem Normal a séries temporais obtidas numa rede de 120 estações de medição de NO<sub>2</sub> dispostas ao longo de uma rede viária em Huddersfield (Grã Bretanha) para estimar a concentração média do poluente nas áreas residenciais adjacentes à rede de medição montada, em que os resultados sugerem que o estimador de Krigagem pode ser usado na modelação da poluição atmosférica à escala local. O trabalho de Berke (2004), desenvolvido num âmbito exploratório, utilizou os modelos geoestatísticos para interpolar o risco de ocorrência do síndrome da morte súbita infantil (SIDS) na região da Carolina do Norte (E.U.A). Goovaerts e Jacquez (2004) apresentam uma metodologia

baseada no modelo geoestatístico de Krigagem Factorial para estimar a incidência de vários tipos de cancro na região de Nova Iorque (E.U.A).

A estrutura deste trabalho divide-se num capítulo inicial onde são tecidas considerações sobre o contexto, definição, limitações deste estudo, e contributo para o desenvolvimento das metodologias geoestatísticas e aplicação de Sistemas de Informação Geográfica na área da saúde pública.

No segundo capítulo, dedicado à revisão do estado actual sobre o conhecimento nas áreas da epidemiologia geográfica e da poluição atmosférica, definem-se os seus conceitos e discutem-se os efeitos nocivos, nomeadamente as doenças respiratórias e alergias, provocados na saúde pública pela exposição à poluição atmosférica.

No terceiro capítulo, são abordados os conceitos teóricos sobre os SIG que sustentam o trabalho exploratório a realizar, e os princípios que os regem.

No quarto capítulo apresenta-se a descrição dos dados recolhidos numa perspectiva exploratória e das técnicas estatísticas usadas para sua interpretação.

No quinto capítulo introduzem-se as técnicas geoestatísticas usadas actualmente na análise de fenómenos naturais e ambientais, e apresentam-se os resultados da aplicação dessas técnicas aos dados experimentais da qualidade do ar e da saúde.

No sexto capítulo são analisados e sintetizados os resultados obtidos, enquanto que no sétimo capítulo apresentam-se as conclusões retiradas dos resultados e sugeridas recomendações para trabalhos futuros.

### **1.3 Metodologia**

Sendo o tópico central deste trabalho a análise exploratória de dados espaciais para modelação geoestatística de partículas inaláveis ( $PM_{10}$ ) e de incidências de asma e alergias em jovens e crianças, as principais fontes de dados são alicerçadas num conjunto de dados referenciados geograficamente sobre ambas as variáveis. Os dados relativos à caracterização espacial das concentrações de  $PM_{10}$  ( $\mu m/m^3$ ) foram fornecidos sob a forma de séries de observações ordenadas no tempo, registadas entre 2001 e 2003, em oito estações de medição localizadas na região de Lisboa. Os dados

relativos à caracterização espacial das variáveis da saúde foram fornecidos por um inquérito realizado em 2002 a jovens de 13/14 anos e a pais de crianças com 6/7 anos, alunos em 143 escolas da região da Grande Lisboa, no âmbito de um estudo epidemiológico realizado a nível internacional sobre asma e outras doenças alérgicas na infância designado por ISAAC - International Study of Asthma and Allergies In Childhood (<http://isaac.auckland.ac.nz/>).

A análise inicial dos dados consistiu na definição da escala de observação adequada ao estudo (definição das unidades de observação usadas ao longo do trabalho) e no tratamento e descrição dos dados. Procedeu-se à inspecção e exploração visual dos dados da saúde e do ambiente, e elaborou-se uma análise descritiva das variáveis (medidas de localização e de dispersão, e histogramas). A análise exploratória prosseguiu com a análise das estruturas de continuidade espacial, onde se identificaram e caracterizaram as estruturas de covariância espaciais evidenciadas pelos dados. Foram experimentados vários parâmetros estatísticos para análise da estrutura de continuidade espacial, aos quais se ajustaram modelos teóricos adequados a cada variável. Os resultados obtidos foram usados nos processos de inferência espacial baseados em modelos lineares de estimação de Krigagem e nos processos de simulação sequencial estocástica para avaliação da incerteza espacial. Os mapas obtidos permitiram descrever os padrões de incidência evidenciados e forneceram indicações sobre associações com os factores de risco considerados. A metodologia aplicada pode sintetizar-se nas seguintes etapas:

- i. Recolha e tratamento dos dados da saúde e ambiente;
- ii. Elaboração de mapas da saúde e do ambiente para evidenciar padrões de incidência e de exposição à poluição;
- iii. Quantificação e modelação das relações espaciais evidenciadas pelos padrões de incidência e de exposição à poluição;
- iv. Avaliação da incerteza espacial associada ao conjunto de dados observados;
- v. Interpretação dos resultados obtidos nas alíneas anteriores com o intuito de fornecer indicações sobre as regiões susceptíveis de estudos futuros e de formulação de hipóteses.

A metodologia aplicada foi desenvolvida numa estreita ligação entre as ferramentas de SIG - ArcView 3.1 e 8.1 com o módulo Spatial Analyst (desenvolvido e comercializado pela ESRI) - e pacotes estatísticos adequados à análise estatística e modelação

geoestatística. Para apresentação das estatísticas descritivas, análise de continuidade espacial e temporal assim como a visualização dos resultados obtidos, quer por estimação quer por simulação foram realizados no software geoestatístico geoMS- Geostatistical Modelling Software, desenvolvido no Centro de Modelização de Reservatórios Petrolíferos do Instituto Superior Técnico. A análise estatística dos resultados obtidos foi realizada com o software SAS Enterprise Guide, um componente do SAS System.

## **2 Poluição atmosférica e saúde pública**

### **2.1 Epidemiologia**

Nas definições encontradas na literatura, é consensual a ideia de que a epidemiologia é a área da ciência que estuda a incidência, distribuição e factores determinantes de infecções e doenças observadas na população humana. Last (2001), acrescenta que a definição de epidemiologia inclui ainda o estudo das medidas mais eficazes e efectivas necessárias à minimização ou controlo dos problemas de saúde observados na população. Retém-se daqui que o âmbito da epidemiologia é bastante vasto. De facto, é actualmente aceite que o estudo epidemiológico pode ser aplicado com propriedade a todas as doenças, situações e acontecimentos ligados à doença (Mausner e Kramer, 1984). Assim, é costume considerar-se que os estudos epidemiológicos se dividem em duas grandes categorias: o estudo da quantidade e distribuição das doenças numa comunidade, segundo as pessoas, lugar e tempo, designado por epidemiologia descritiva, e o estudo mais concreto dos determinantes da doença ou razões que explicam a sua frequência, relativamente alta ou baixa, em grupos específicos da população, designado por epidemiologia analítica.

Um estudo epidemiológico inicia-se assim com a identificação de variações não aleatórias na distribuição de uma doença (através dos estudos descritivos), prossegue com a formulação de hipóteses sobre a associação entre um agente ou factor de risco suspeito e a doença, e finaliza com um estudo para verificação das hipóteses consideradas (através de estudos analíticos).

No âmbito desta dissertação abordam-se as etapas do estudo epidemiológico referentes aos estudos descritivos e à formulação de hipóteses, e que se enquadram num tipo de estudos de epidemiologia designado por estudos ecológicos. Estes

definem-se como uma via para explorar as correlações espaciais ou temporais, entre as variações na prevalência de uma doença e as variações no grau de exposição a um agente ou factor de risco suspeito (geralmente um agente ambiental ou estilo de vida). Os estudos ecológicos caracterizam-se assim por ser mais do tipo "gerador de hipóteses" do que do tipo "verificador de hipóteses", na medida em que a unidade de comparação é o grupo e não o indivíduo, isto é, a exposição a um determinado factor de risco suspeito e a ocorrência da doença não são medidas nos mesmos indivíduos. Morgenstern (citado por Mausner e Kramer, 1984) chama a atenção para o perigo de se retirarem conclusões falaciosas sobre os riscos individuais, a partir de riscos observados em grupos (também designado por "falácia ecológica"), dado que não se analisam os comportamentos individuais que podem influenciar o risco.

Conforme foi referido, os estudos ecológicos abrangem os estudos epidemiológicos de âmbito descritivo onde se identificam as variações não aleatórias na distribuição de doenças, permitindo assim ao investigador criar hipóteses verificáveis sobre a sua etiologia. Para ser possível identificar variações não aleatórias na distribuição de doenças, é necessário descrever a ocorrência de uma doença especificando pessoas, tempo e lugares, ou seja, é necessário conhecer as características demográficas, distribuição no tempo e no espaço dos casos de doença observados.

As pessoas podem ser caracterizadas em função de um número quase infinito de variáveis, como por exemplo a idade, sexo, grupo étnico, classe social, ocupação, estado civil, estilos de vida, entre outros. Para os estudos epidemiológicos, é rotina especificar, pelo menos, três características de qualquer pessoa - a idade, sexo e grupo étnico - devido ao facto de existir geralmente uma relação qualquer entre estas variáveis e as doenças (Mausner e Kramer, 1984).

O tempo é uma variável fundamental na análise epidemiológica, na medida em que permite por um lado caracterizar a prevalência das doenças num determinado momento, e por outro permite conhecer a sua evolução ou variação ao longo do tempo.

O lugar como variável determinante na epidemiologia descritiva, reside no facto da frequência das doenças variar muitas vezes segundo a localização geográfica (e período de tempo). Os factores do ambiente físico (onde se inclui entre outros a meteorologia, a qualidade do ar e da água) e biológico (que inclui agentes infecciosos, as plantas e animais, entre outros) que caracterizam as regiões podem tomar certas

doenças prevalentes. Isto significa que as variações de frequência das doenças descritas em função do local de ocorrência, podem fornecer explicações ou pistas sobre sua etiologia. Os exemplos clássicos sobre o papel fundamental do conhecimento da localização geográfica na epidemiologia descritiva são os de John Snow (1854), que desenvolveu a hipótese sobre o modo de transmissão da cólera na cidade de Londres, com base na distribuição espacial dos casos de cólera e sua relação com a distribuição espacial dos poços de água; e de Palm (1892) quando observou que a incidência geográfica de raquitismo estava relacionada de forma inversa com a quantidade de exposição solar.

O desenvolvimento de metodologias estatísticas para análise de fenómenos do domínio espaço-temporal foi impulsionado na década de 30 e 40, altura em que começaram a existir dados publicados de cobertura nacional, sobre as ocorrências de doenças (por exemplo, com o aparecimento em 1933, do registo obrigatório dos óbitos nos EUA, ou a criação do *National Health Service* no Reino Unido, em 1948).

Por outro lado, a partir dos finais da década de 50 e década de 60, o desenvolvimento das tecnologias de informação e em particular dos sistemas de informação geográfica (SIG), veio enriquecer o lote de ferramentas de cálculo para análise de ocorrências no espaço e no tempo.

O contributo dos SIG para o desenvolvimento dos estudos ecológicos assentou desde o início na facilidade de cartografar e identificar padrões espaciais de morbilidade ou mortalidade e formular hipóteses sobre factores de risco suspeitos. Mais recentemente, o aparecimento de ficheiros de dados referenciados sobre recenseamentos da população permitiu expor as taxas de prevalência ou incidência de doenças e identificar áreas onde os riscos relativos de incidência são mais elevados (Haining, 1998).

Na década de 90, as melhorias das velocidades de cálculo computacional permitiram desenvolver ferramentas SIG capazes de incorporar métodos estatísticos de análise espacial para a saúde pública (entre outras áreas) com utilização de métodos de Monte Carlo para modelos bayesianos ou para modelos geoestatísticos. Numerosos autores (Gatrell e Loytonen, 1998; Beyea e Hatch, 1999; Jacques, 2000; Hertz-Picciotto e Brunekreef, 2001; Nuckols, Ward et al., 2003; Elliott e Wartenberg, 2004) são unânimes em considerar que o próximo passo das ferramentas SIG para contribuir para o desenvolvimento dos estudos ecológicos tem de assentar na implementação de sistemas de informação que permitam indexar os dados no tempo e no espaço para

modelar processos espácio-temporais. Jaquez (2000) aborda a necessidade de se desenvolverem sistemas STIS (*Space-Time Information Systems*) que permitam estimar parâmetros de modelos espácio-temporais, como meio indispensável para o avanço da ciência da epidemiologia espacial.

## **2.2 O problema da poluição atmosférica**

A poluição atmosférica pode ser definida como uma condição atmosférica onde o ambiente apresenta concentrações de um qualquer elemento ou composto químico, de origem natural ou criada pelo Homem, acima de valores considerados normais, ao ponto de produzir efeitos nocivos mensuráveis no Homem, fauna, flora ou materiais (Seinfeld, 1986). Estas substâncias podem existir na atmosfera no estado gasoso (gases), líquido (gotículas) ou sólido (partículas). Seguindo a proposta de Seinfeld (1986), considera-se neste trabalho a classificação dos poluentes em apenas estado gasoso e estado particulado, incluindo neste último as gotículas e as partículas.

Outras formas de classificar os poluentes atmosféricos dizem respeito à sua composição química, que pode ser convenientemente agrupada em compostos sulfurados, compostos azotados, compostos carbonados e compostos radioactivos; à forma como afectam a saúde humana classificando-se em substâncias tóxicas ou substâncias cancerígenas; ou ainda, pelo modo como os poluentes atingem a atmosfera, sendo classificados como poluentes primários, quando provêm directamente da fonte de emissão, ou poluentes secundários, quando resultam de interacções químicas entre os elementos que compõem a atmosfera e os poluentes primários.

Na última década a comunidade internacional (através da Organização Mundial de Saúde, Organização das Nações Unidas, entre outras), tem vindo a dar cada vez mais atenção aos poluentes orgânicos persistentes (POPs). Estes poluentes, de origem essencialmente antropogénica, são substâncias químicas (substâncias presentes fundamentalmente em pesticidas e na combustão de resíduos) que se mantêm intactas no ambiente por longos períodos de tempo, e cuja toxicidade é um factor de risco para a saúde humana (<http://www.who.int/iomc/groups/pop/en/>).

As emissões de poluentes atmosféricos podem ser provocadas por fenómenos naturais, tais como as erupções vulcânicas (responsáveis pela emissão de partículas e gases poluentes, como por exemplo o dióxido de enxofre e o gás metano); incêndios florestais

(responsáveis pela emissão de poluentes sob a forma de fumos, monóxido de carbono, dióxido de carbono e óxidos de azoto); fenómenos decorrentes da actividade humana (antropogénicos), como por exemplo as actividades industriais (nomeadamente das indústrias extractivas e transformadoras) ou actividades relacionadas com os estilos de vida das pessoas, tais como o tráfego automóvel ou o aquecimento das casas.

Em Portugal, uma parte significativa das emissões de poluentes das actividades industriais é originada na queima de combustíveis fósseis (por exemplo em refinarias, siderurgias, cimenteiras ou centrais termoeléctricas). Neste processo estão envolvidas reacções químicas de combustão, onde são libertados para a atmosfera consideráveis quantidades de óxidos de enxofre ( $\text{SO}_2$ ), dióxido de carbono ( $\text{CO}_2$ ) monóxido de carbono ( $\text{CO}$ ), óxidos de azoto ( $\text{NO}_x$ ) e partículas.

Outras fontes de emissão, como a queima de resíduos urbanos, industriais ou agrícolas podem emitir, em situações não controladas, quantidades consideráveis de compostos perigosos, enquanto que as pedreiras e áreas de construção são uma causa importante da emissão de partículas.

Por outro lado, as fontes móveis de emissão, caracterizadas pelos veículos (rodoviários, ferroviários, marítimos ou aéreos) que utilizam motores (de 2 ou 4 tempos, ou a jacto) de combustão (diesel, gasolina ou carvão), emitem para a atmosfera uma grande variedade e quantidade de poluentes. Os transportes rodoviários, classificados por Boubel e seus co-autores (1994) como a fonte móvel de emissões predominante em todos os países industrializados, são os principais responsáveis pela emissão de gases de escape e evaporação de combustíveis, onde são libertados, entre outros,  $\text{CO}$ ,  $\text{NO}_x$ , partículas e  $\text{CO}_2$ .

As emissões de poluentes a partir das suas fontes, quer fixas quer móveis, tendem a expandir-se e a misturar-se na atmosfera, em função da natureza dos poluentes, das características da emissão e das condições meteorológicas que existem numa determinada vizinhança do emissor.

Boubel e seus co-autores (1994) definem várias vizinhanças (ou escalas) para análise dos problemas da poluição do ar: a escala local, que analisa os problemas de poluição atmosférica para distâncias inferiores a 5 quilómetros das fontes de emissão, a escala urbana que se estende até aos 50 quilómetros, a escala regional entre os 50 e 500 quilómetros, a escala continental entre os 500 e vários milhares de quilómetros e a

escala global que analisa o problema da poluição atmosférica à escala do planeta. Descrevem-se a seguir, de forma sumária, os problemas principais causados pela poluição atmosférica, identificados à escala urbana (que inclui a escala local), na medida em que as escalas restantes estão fora do âmbito da dissertação.

Boubel e seus co-autores (1994) identificam dois tipos diferentes de problemas em áreas urbanas que se distinguem na forma como os poluentes atingem a atmosfera, isto é, se são emitidos directamente da fonte de emissão ou se se formam por reacções fotoquímicas dos poluentes primários. No caso de poluentes primários, como o CO ou as partículas, os problemas colocam-se em zonas consideradas *hot spots*, ou seja, zonas onde se registam elevadas concentrações provocadas, por exemplo, pelos gases de escape e evaporação de combustíveis dos veículos motorizados. Os locais onde tipicamente se registam estas elevadas concentrações são os cruzamentos, particularmente em áreas cercadas por edifícios altos onde o volume de ar é limitado, condicionando a dispersão dos poluentes. Outro dos principais problemas observados nas áreas metropolitanas está relacionado com o fenómeno designado por *smog* fotoquímico, caracterizado pela formação de uma névoa alguns metros acima da superfície do solo, com uma percentagem significativa de ozono formado a partir de reacções que ocorrem, na presença de radiação solar, com hidrocarbonetos ou com NO<sub>2</sub>, NO e CO emitidos principalmente pelos automóveis, indústrias e aterros sanitários.

### **2.3 Os efeitos da poluição atmosférica na saúde pública**

A inalação de ar poluído induzida quer por exposição prolongada a concentrações baixas quer por exposição a episódios de elevadas concentrações, pode provocar efeitos nocivos na saúde pública, nomeadamente ao nível respiratório onde contribui para o agravamento ou até mesmo para o desenvolvimento de doenças como a asma, infecções pulmonares, enfisemas pulmonares ou cancro do pulmão.

No âmbito deste trabalho, analisam-se em particular associações entre a prevalência de asma e alergias em crianças e jovens e a exposição a concentrações de partículas inaláveis. Estas partículas são uma mistura de substâncias químicas e partículas sólidas e líquidas de dimensão variável com capacidade para penetrar em diferentes regiões do tracto respiratório (Boubel, Fox et al., 1994). A escolha das partículas inaláveis para este trabalho deve-se ao facto de existir um número significativo de estudos

epidemiológicos que apontam para uma forte associação entre a mortalidade e morbidade e a exposição a concentrações de partículas inaláveis (Ostro, 2004).

Sabe-se que a função principal do sistema respiratório é levar oxigênio ao sistema sanguíneo e remover o dióxido de carbono do organismo. O oxigênio contido no ar é inalado pelo nariz e/ou boca e é conduzido ao longo das vias respiratórias aéreas através da faringe, laringe, traqueia, brônquios e bronquíolos. Estes últimos, constituídos por uma extensa arborização dos brônquios desembocam nas unidades funcionais do sistema respiratório que se designam por alvéolos pulmonares. É nos alvéolos que se dão as trocas gasosas entre o ar inalado e os gases trazidos da corrente sanguínea, permitindo a passagem do oxigênio para a corrente sanguínea e a libertação do CO<sub>2</sub> produzido pelo metabolismo celular.

No caso da inalação das partículas e gases poluentes, o seu efeito no sistema respiratório é muito influenciado pela região do pulmão que atingem. Enquanto que as partículas dependem das suas características aerodinâmicas - dimensão, forma e densidade - para penetrar nos pulmões, os gases dependem da solubilidade das moléculas gasosas inaladas no revestimento das diferentes regiões do sistema respiratório (Boubel, Fox et al., 1994).

Na atmosfera, as dimensões das partículas vão desde as poeiras, fumos e resíduos de carbono não queimados, visíveis a olho nu, designadas por partículas grosseiras cujos diâmetros variam entre 2.5 µm e 40 µm, até dimensões microscópicas inferiores a 2.5 µm, designadas por partículas finas e aerossóis. Na Figura 2.1 apresenta-se a relação entre a dimensão aerodinâmica das partículas e as regiões onde estas se depositam.

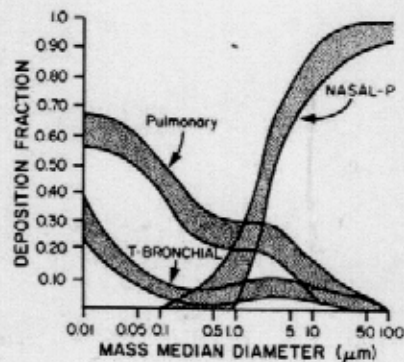


Figura 2.1- Deposição de partículas em função do seu diâmetro, em várias regiões dos pulmões (retirado de Boubel, Fox et al., 1994).

Pela análise do gráfico observa-se que à medida que as partículas se tornam mais finas, a fracção depositada nos mecanismos respiratórios mais profundos aumenta. As partículas de diâmetro superior a 10 µm depositam-se fundamentalmente no tracto superior do aparelho respiratório (onde se incluem a região nasal e a faringe- *Nasal-P*). Para dimensões inferiores a 10 µm, verifica-se um aumento da fracção da deposição das partículas nas regiões do tracto inferior do aparelho respiratório (composto pela traqueia, brônquios - *T-Bronchial* - e região pulmonar - *Pulmonary*). Actualmente considera-se que para partículas de dimensão inferior a 2.5 µm a penetração nos mecanismos respiratórios mais profundos contribui de modo dramático para a deterioração das condições de saúde pública.

Vários estudos epidemiológicos revelam ligações entre a poluição atmosférica em meio urbano e numerosos efeitos adversos na saúde. No âmbito de um relatório publicado em conjunto pela OMS e a Agência Europeia do Ambiente (AEA), Tamburini e Ehrenstein (2002) compilaram e publicaram um conjunto de artigos que apresentam evidências de relações significativas entre a qualidade do ar e a saúde das crianças.

Num estudo epidemiológico realizado em Leipzig, Alemanha (von Mutius, Sherrill et al., 1995), foram detectadas associações entre as concentrações elevadas de SO<sub>2</sub>, concentrações moderadas de PM<sub>10</sub> e NO<sub>x</sub> e a prevalência de doenças agudas nas vias aéreas respiratórias. Noutro estudo epidemiológico desenvolvido em Paris (Fauroux, Sampil et al., 2000), foram encontradas associações significativas entre as

concentrações de ozono e os ataques de asma nas crianças registados nas urgências hospitalares. Hirsch e seus co-autores (1999) elaboraram um estudo analítico em crianças, em Dresden (Alemanha), onde detectaram que o crescimento moderado da poluição atmosférica causada pelo tráfego automóvel está associado a um aumento da prevalência de tosse e bronquite. Ghering e Cyrus (2002) analisaram os efeitos das exposições de longa duração a partículas de diâmetro inferior a  $2.5 \mu\text{m}$  e  $\text{NO}_2$  em crianças, na cidade de Munique (Alemanha), aplicando uma metodologia baseada em ferramentas de sistemas de informação geográfica e na aplicação de regressão logística para analisar os efeitos das exposições com ajustamento das variáveis de confusão. Os resultados obtidos revelaram associações significativas entre a presença dos poluentes e ocorrências de tosse sem infecção e tosse seca nocturna.

## 2.4 Os mapas de incidência ou prevalência de doenças

Conforme foi referido no capítulo sobre os estudos ecológicos, a análise da distribuição geográfica da prevalência ou incidência de doenças é de importância considerável para os epidemiologistas. A aplicação de mapas de incidência ou ocorrência de doenças pode ser importante para descrever a distribuição espacial do fenómeno e formular hipóteses sobre a sua etiologia. É também importante para identificar variações não aleatórias das doenças e contribuir para um melhor planeamento na locação dos recursos nas áreas de maior risco ou ainda para formulação de hipóteses sobre a etiologia das doenças no âmbito de estudos ecológicos (Lawson, Biggeri et al., 2000; Pascutto, Wakefield et al., 2000; Lawson, 2001).

No primeiro caso, onde os mapas são usados para descrever a distribuição da doença no espaço, o objecto de análise é atribuir uma estimativa do verdadeiro risco relativo de incidência ou ocorrência de uma doença na área de estudo. No segundo caso, onde os mapas são usados para identificar variações não aleatórias (ou *clusters*) de doenças, o objecto de análise é saber onde existem *clusters* e onde se localizam. No terceiro caso, o dos estudos ecológicos, o objecto de análise é a distribuição da doença no espaço e sua relação com variáveis explicativas.

No âmbito deste trabalho, pretende-se abordar em detalhe os aspectos relativos à descrição da distribuição das doenças no espaço, para introduzir a formulação de hipóteses no âmbito de estudos ecológicos e fornecer o enquadramento teórico necessário para compreender a metodologia apresentada nos capítulos posteriores.

O primeiro aspecto a considerar na elaboração de um mapa de incidência de doenças, e que distingue claramente a abordagem estatística a seguir, está relacionado com o tipo de dados espaciais que o compõem. Os tipos de dados observados podem variar entre casos indexados à residência, onde cada caso ou ocorrência é referenciado no espaço individualmente, e a contagem de casos indexados a uma região ou área, onde os dados observados referem-se a contagens de incidência ou ocorrência observadas em áreas definidas (geralmente áreas administrativas ou áreas de influência dos centros de saúde). A incidência geográfica de uma doença com base em dados indexados ao local onde reside ou trabalha o indivíduo fornece dados importantes sobre a exposição a potenciais factores de risco ambientais. No entanto, os dados não estão geralmente disponíveis a esta escala por razões de confidencialidade ou por necessidade de sistematizar grandes volumes de informação.

A representação de um mapa com as contagens de ocorrências de doença em populações distribuídas no espaço de forma heterogénea é considerada por vários autores (Tsal e Wen, 1986; Lawson, Biggeri et al., 2000; Pascutto, Wakefield et al., 2000; Lawson, 2001; Kulldorff, 2003; Berke, 2004; Elliott e Wartenberg, 2004; Goovaerts e Jacquez, 2004) como não sendo adequada para a análise de eventuais padrões espaciais, dado que não considera as diferenças das estruturas das populações relativamente à sua dimensão, estrutura etária ou distribuição por sexo, em cada região. Para ultrapassar esta questão, estes autores sugerem métodos que incorporam as diferenças das estruturas das populações, de modo a torná-las comparáveis. Considerem-se as contagens de ocorrência de doenças,  $O_i$ , na população  $N_i$ , onde  $i = 1, \dots, I$  representa o índice da região observada. Nestes casos pode assumir-se que a contagem de ocorrências segue uma distribuição de Poisson, de média  $E_i \cdot \theta$ , ou seja,

$$O_i \sim \text{Poisson}(E_i \theta_i)$$

onde  $E_i$  representa o número esperado de ocorrências na população da região  $i$  e  $\theta$  representa o risco relativo de incidência da população da região  $i$ . O estimador de  $\theta$  mais frequente na área da epidemiologia, é o estimador de máxima verosimilhança (MV), designado por *Standard Incidence Ratio*<sup>1</sup> ( $\varphi$ ), ou razão padronizada de incidência,

---

<sup>1</sup> *Standard Incidence Ratio* no caso de estudo de incidência. No caso de estudo de mortalidade ou de morbilidade designa-se respectivamente por *Standard Mortality Ratio* ou *Standard Morbidity Ratio* (SMR).

que compara a incidência de uma doença entre uma população sujeita ao estudo e uma população padrão, definido por:

$$\varphi_i = \frac{O_i}{E_i}$$

onde  $O_i$  é o número observado de ocorrências na população  $i$  e  $E_i$  é o número esperado de ocorrências na mesma população (admitindo a independência das observações). Apesar de  $\varphi$  ser um estimador centrado, é apontado como um modelo que, em termos estatísticos, apresenta algumas desvantagens, na medida em que para populações de pequenas dimensões uma ligeira variação no número de ocorrências tende a causar grandes variações nos valores de  $\varphi$  e nos valores da variância estimada (Lawson, Biggeri et al., 2000; Pascutto, Wakefield et al., 2000; Elliott e Wartenberg, 2004). Lawson (2001) sugere que a utilização deste estimador para representar a incidência de uma doença, seja sempre acompanhada pelo mapa de variabilidade das estimativas obtidas.

Alternativamente, para ultrapassar os problemas derivados da atribuição frequencista de probabilidades de risco de incidência relativa, existem modelos caracterizados pela utilização de informação adicional para atenuar os valores de  $\theta_i$ , nomeadamente nas regiões de maior instabilidade (ou maior variabilidade, que coincidem com as regiões de populações de pequena dimensão). Os referidos modelos apontam para a utilização de modelos de alisamento segundo a abordagem bayesiana, que pondera as distribuições *à priori* de  $O_i$  e de  $\theta_i$  em função da precisão das estimativas em cada região, para obter uma distribuição *à posteriori*. Sankoh e seus co-autores (2002) utilizam o estimador empírico de Bayes num estudo sobre a mortalidade infantil no Burkina Faso para criar um mapa de risco de incidência. Berke (2004) aplica uma metodologia semelhante num estudo sobre o síndrome infantil da morte súbita (SIDS) na região da Carolina do Norte. Lawson (2001) apresenta dois estudos de caso sobre a incidência de cancro na laringe na região de Lancashire (Inglaterra) e a mortalidade causada por cancro nas vias respiratórias, na região de Falkirk (Escócia), onde ajusta alguns modelos bayesianos para representar os riscos de incidência nas regiões administrativas que compõem as áreas de estudo.

## **3 Sistemas de informação geográfica (SIG)**

### **3.1 Introdução**

Passaram cerca de 40 anos desde que R. F. Tomlinson propôs o conceito de sistema de informação geográfica (SIG) como um sistema concebido num ambiente computacional para armazenamento e organização de informação geográfica. Desde então, a evolução das ferramentas SIG processou-se de forma fulgurante, nomeadamente nas décadas de 80 e 90 do século passado, e tornou-se um ambiente de trabalho comum em universidades, instituições públicas e empresas, com aplicações nas mais variadas áreas do conhecimento desde as ciências exactas às ciências sociais (por exemplo, nas áreas da geodesia, da detecção remota, do ambiente, no planeamento do território ou da geografia). Influenciadas pelo contexto em que são referidas, várias definições de SIG aparecem na literatura (Dueker, 1979; Burrough, 1986; Aronoff, 1989), no entanto todas comungam da ideia de que um SIG representa um sistema que integra informação espacial.

No âmbito deste trabalho adopta-se a definição de Dueker, citada por Chrisman (2002) em que um SIG define-se como um sistema constituído por hardware, software, dados, pessoas, organizações ou um ambiente institucional, com o objectivo de recolher, armazenar, analisar e disseminar informações sobre áreas da Terra. As componentes de recolha e armazenamento de dados geográficos, permitem integrar as características geográficas e alfanuméricas dos dados num SIG, isto é, integrar num único sistema a localização relativa no espaço e os atributos dos dados que os caracterizam e identificam. A componente de análise de informação geográfica consiste na possibilidade de usar um conjunto de algoritmos implementados num SIG, que permitem integrar e transformar os dados geográficos de modo a produzir informação útil aos processos de tomada de decisão. Por fim, o SIG permite uma disseminação da informação produzida rápida e eficaz devido à diversidade e flexibilidade das suas aplicações, que permite tornar os resultados acessíveis a uma grande variedade de utilizadores, desde especialistas ou investigadores até aos decisores (frequentemente sem conhecimentos sobre ferramentas de SIG).

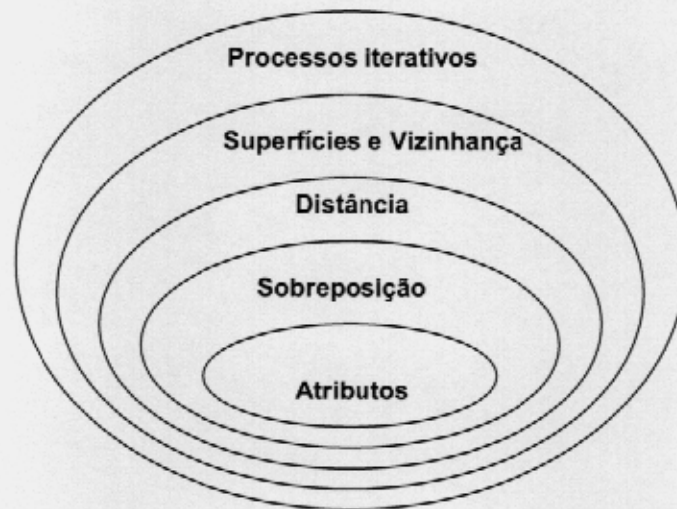
### **3.2 A integração de dados geográficos**

A representação dos dados geográficos num suporte lógico pode assumir várias formas, que se designam por modelos ou estruturas de dados (assim se designam

porque representam um conjunto complexo e organizado de *bits*). Vários autores (Aronoff, 1989; Matos, 2001; Chrisman, 2002) concordam que os modelos de dados dominantes para representação geométrica de dados geográficos em SIG são os vectoriais relacionais e os matriciais (raster). Os dados de estrutura vectorial, representados por pontos, linhas ou polígonos, são organizados em estruturas topológicas (ou de grafos) onde se armazenam informações sobre a posição relativa dos dados geográficos. A integração com uma base de dados relacional ou modelo relacional, assegura a associação entre os dados geográficos e os alfanuméricos, armazenados usualmente em tabelas compostas por atributos (colunas) e tuplos (linhas). Nos dados de estrutura matricial, a representação do espaço é constituída por uma discretização do espaço em células de dimensão definida e dispostas de forma regular, em que cada célula está associada a um único valor. Esta estrutura de dados é bastante mais simples que a vectorial e é particularmente adequada à modelação de fenómenos com distribuição contínua (Matos, 2001).

Actualmente é possível em grande parte dos suportes lógicos, a utilização conjunta de dados produzidos de acordo com ambos os modelos (vectorial e raster), o que permite integrar procedimentos mais complexos de operação ou transformação dos dados geográficos.

Segundo Chrisman (2002) a complexidade destas operações varia entre as que não usam a geometria dos dados geográficos (apenas usam a sua componente alfanumérica) e as que usam processos iterativos para estudar a propagação de um fenómeno ao longo de uma área. Na Figura 3.1 representam-se os vários tipos de operação implementados em SIG, classificados de acordo com Chisman (2002), segundo o grau de complexidade da sua implementação.



**Figura 3.1- Operações num SIG segundo o seu grau de complexidade de implementação.**

O primeiro conjunto de operações referido no esquema (Atributos), envolve a utilização dos atributos dos dados geográficos sem envolver a componente espacial ou geométrica dos dados. Neste grupo de operações incluem-se a manipulação de atributos, como por exemplo a classificação de dados geográficos ou a utilização de álgebra relacional (no caso do modelo vectorial) para combinar (ou relacionar) atributos entre fontes de informação geográfica distinta.

O conjunto de operações de sobreposição (ou *map overlay*) permite integrar informações geográficas de diferentes fontes de dados, ao nível das componentes geométrica e alfanumérica, e são usadas geralmente para comparar valores de atributos. Estas operações, implementadas de modo distinto nos modelos raster e vectorial, permitem combinar valores provenientes de diversas fontes de informação num ponto ou local. No caso de dados de estrutura raster, a sobreposição de dados é obtida através de um sistema de operações aplicadas a conjuntos, genericamente designado por álgebra booleana, enquanto que no modelo vectorial a sobreposição obriga ao processamento das intersecções geométricas geradas, construção de nova estrutura topológica (e geométrica) e actualização dos valores dos atributos.

As operações de distância consistem em algoritmos que permitem detectar relações de distância implícitas nas representações espaciais. Uma das aplicações mais comuns com operações de distância é para a definição de áreas de influência (como por exemplo a delimitação de áreas de exposição a factores de risco para a saúde pública),

através do cálculo de *buffers* (polígonos que delimitam uma área em redor de um ponto, linha ou polígono) ou de polígonos de Thiessen<sup>2</sup> (polígonos gerados a partir de um conjunto de pontos, e que delimitam a área mais próxima de um ponto em comparação a todos os outros).

As operações de distância pertencem a um conjunto mais vasto de algoritmos que se centram na vizinhança, ou seja, no contexto espacial que rodeia cada valor observado. A percepção das operações de vizinhança é mais evidente em superfícies contínuas (modelo de dados raster) do que em modelos vectoriais, razão pela qual Chrisman incluiu as operações de vizinhança e superfícies na mesma categoria. Matos (2001) apresenta de modo sistematizado as operações de vizinhança para modelos raster, e classifica-as de focais e zonais. As funções focais tomam em consideração os valores de uma vizinhança como argumento da função. Várias aplicações em SIG utilizam estas funções, como por exemplo para cálculo de declives (morfologia do terreno) ou definição de direcção de escoamento de recursos hídricos. As funções zonais são semelhantes às funções focais, mas com a diferença de a vizinhança ter uma forma fixa, podendo as zonas serem definidas por outra fonte de informação geográfica (como por exemplo, *buffers* ou polígonos de Thiessen).

Os processos iterativos são extensões das operações de vizinhança, em que o resultado obtido num ponto pode propagar-se de modo a influenciar o resultado noutros pontos. Neste grupo de operações, destacam-se como aplicações mais frequentes a determinação de bacias visuais a partir de um modelo digital do terreno (determinação da visibilidade potencial a partir de um ponto qualquer da superfície) ou as funções de custo-acumulação (usadas por exemplo para determinar acessibilidades num estudo de transportes ou para determinar a acumulação de escoamento em recursos hídricos).

### **3.3 Análise espacial em saúde pública**

A análise espacial no âmbito da saúde pública foca-se na construção de mapas de incidência ou prevalência de doenças e na correlação das suas distribuições no espaço, através de métodos gráficos e quantitativos para identificar características particulares sobre a dependência e heterogeneidade espacial das doenças. Segundo Douven e Scholten (1995) a definição de dependência espacial consiste na caracterização das

---

<sup>2</sup> Nas ciências matemáticas e de computação designam-se por Diagramas de Voronoi.

relações evidenciadas entre os dados geográficos, enquanto que a definição de heterogeneidade espacial consiste na caracterização dos diferentes efeitos provocados pela dependência espacial. Conforme referido em 2.4, pretende-se com esta abordagem contribuir para um melhor planeamento na locação de recursos nas áreas de maior risco, ou para formulação de hipóteses sobre a etiologia das doenças no âmbito de estudos ecológicos.

De modo a estruturar a análise espacial de dados em saúde pública em função das diferentes técnicas usadas e dos seus objectivos específicos, Gatrell e Bailey (1995) distinguem três patamares de análise designados por visualização, análise exploratória e modelação dos dados espaciais. A visualização envolve a integração dos dados para construção de mapas permitindo evidenciar padrões espaciais, associações entre variáveis, áreas com valores anómalos ou áreas de *clusters*; a análise exploratória envolve ainda a descrição das características locais e regionais das variáveis, através de medidas de localização (ou tendência central) e de dispersão, incluindo ou não técnicas e métodos para quantificar eventuais associações espaciais (neste caso designa-se por análise exploratória de dados espaciais); a modelação abrange um conjunto de ferramentas estatísticas espaciais e não espaciais para construção de modelos univariados (mais simplistas) ou multivariados (mais complexos).

Conforme referido no sub-capítulo 1.2, o princípio da análise exploratória de dados, ou abordagem *data-driven*, reside no facto de não serem impostas condições sobre modelos subjacentes aos dados, o que permite pôr a descoberto características dos dados até aí desconhecidas ou até mesmo sustentar a formulação de hipóteses sobre a sua etiologia. A utilização de um SIG num ambiente exploratório aplicado à pesquisa em saúde pública é representada na Figura 3.2.

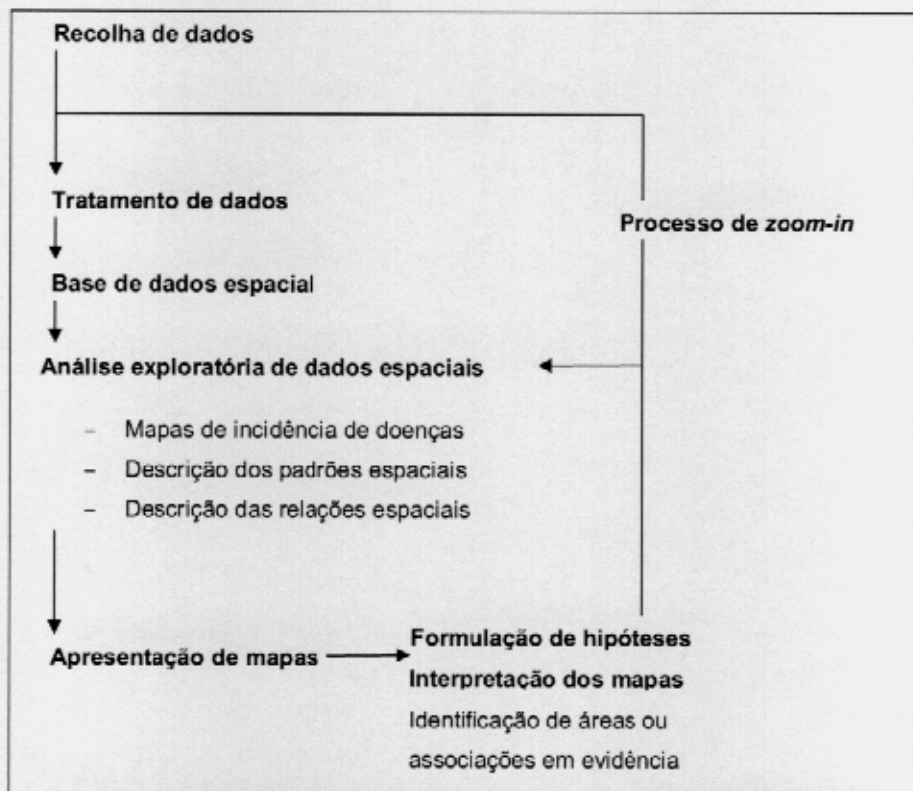


Figura 3.2- O enquadramento dos SIG na análise exploratória em saúde pública (adaptado de Douven e Scholten, 1995).

As tarefas desempenhadas por um SIG estão representadas no lado esquerdo da Figura 3.2, onde funcionam num sistema desenvolvido para armazenamento, processamento e visualização de dados espaciais.

Estas funcionalidades são fundamentais no processo de análise exploratória de dados espaciais em saúde pública, na medida em que fornecem indicações sobre áreas caracterizadas por valores extremos de prevalência ou incidência de doenças ou sobre relações ou associações entre variáveis da saúde e outros factores (sociais, demográficos, económicos, ambientais, entre outros). O processo exploratório é geralmente cíclico (representado no esquema por processo *zoom-in*), envolvendo a utilização de dados cada vez mais detalhados, ou seja, obtidos a escalas superiores (por exemplo passando de uma escala regional para uma local).

Neste processo, a combinação de um SIG, usado como ferramenta para armazenar, manipular e visualizar informação espacial, com métodos estatísticos e geoestatísticos

para análise exploratória de dados espaciais, permite descrever e sintetizar padrões espaciais de incidência ou prevalência de doenças e associações com outros factores, constituindo um instrumento valioso na área da investigação em saúde pública.

## 4 Dados e análise exploratória

### 4.1 Variáveis da saúde

#### 4.1.1 O International Study of Asthma and Allergies In Childhood (ISAAC)

Os dados sobre a incidência de asma e doenças alérgicas em crianças e jovens provêm de um inquérito realizado em 2002 a jovens de 13/14 anos e a pais de crianças com 6/7 anos, alunos em 143 escolas do ensino básico e secundário da região da Grande Lisboa, no âmbito de um estudo epidemiológico realizado a nível internacional sobre asma e outras doenças alérgicas na infância designado por International Study of Asthma and Allergies In Childhood (ISAAC). No âmbito deste trabalho consideram-se seis questões do inquérito apresentadas na terceira fase do projecto (as mesmas para ambos os grupos etários), de respostas do tipo binário (assume os valores 1 ou 0, para as respostas 'Sim' e 'Não' respectivamente) e que se apresentam na Tabela 4.1.

Questão do ISAAC
Alguma vez o médico lhe diagnosticou asma?
Alguma vez teve sintomas de pieira?
Alguma vez teve crises de espirros ou de pingos no nariz?
Alguma vez o médico lhe diagnosticou febre dos fenos?
Alguma vez teve lesões de pele com comichões?
Alguma vez o médico lhe diagnosticou um eczema?

Tabela 4.1- Questões do ISAAC consideradas no âmbito do estudo.

As variáveis apresentadas na Tabela 4.2, representam os resultados obtidos em cada uma das seis questões.

Questão do ISAAC	Variáveis por grupo de idades	
	6/7 anos	13/14 anos
Alguma vez o médico lhe diagnosticou asma?	Ash_6	Ash_13
Alguma vez teve sintomas de pieira?	Whez_6	Whez_13
Alguma vez teve crises de espirros ou de pingos no nariz?	Pnose_6	Pnose_13
Alguma vez o médico lhe diagnosticou febre dos fenos?	Hay_6	Hay_13
Alguma vez teve lesões de pele com comichões?	Rash_6	Rash_13
Alguma vez o médico lhe diagnosticou um eczema?	Ecz_6	Ecz_13

Tabela 4.2- Variáveis de saúde consideradas no âmbito do estudo.

Para obter a localização geográfica das escolas procedeu-se ao registo das coordenadas geográficas através da recolha de dados com apoio de um receptor GPS (Global Positioning System) e de uma base de dados fornecida pelos Correios de Portugal (CTT) num ficheiro ASCII, com códigos de escolas e coordenadas geográficas dos centroídes de entidades lineares associadas aos códigos postais dessas escolas. O resultado deste trabalho permitiu referenciar e integrar num SIG, 103 das 143 escolas envolvidas no inquérito. As escolas cuja localização não foi identificada foram retiradas deste estudo.

#### 4.1.2 A razão padronizada de incidência (SIR)

A análise da incidência da asma e doenças alérgicas não foi feita com base na contagem de casos observados em cada escola, dado que as diferentes dimensões (número de alunos) e estruturas etárias não permitiam a sua comparação. Neste contexto, considerou-se a utilização do estimador de máxima verosimilhança da razão padronizada de incidência ou *SIR* (*Standard Incidence Ratio*), já referido no sub-capítulo 2.4, que se define por:

$$\varphi_i = \frac{O_i}{E_i}$$

onde  $O_i$  é o número observado de ocorrências na escola  $i$  (assume-se que  $O_i$  segue uma distribuição de Poisson independente) e  $E_i$  é o número esperado de ocorrências nessa escola. O *SIR* permite a comparação entre as escolas, pois compara a incidência de uma dada doença de uma população padrão se esta estivesse submetida ao risco que a população sujeita ao estudo está. Um valor de *SIR* superior a 1 indica que a incidência observada na escola (população sujeita ao estudo) é superior ao esperado (com distribuição uniforme da probabilidade de incidência). Um valor de *SIR* inferior a 1 indica o oposto, ou seja, que a incidência observada na escola  $i$  é inferior ao esperado.

A variância deste estimador define-se por (Pascutto, Wakefield et al., 2000; Lawson, 2001; Bohning, Malzahn et al., 2002):

$$Var(\varphi_i) = \frac{\theta_i}{E_i} = \frac{O_i}{E_i^2}$$

onde  $E_i$  representa o número esperado de ocorrências na população da região  $i$  e  $\theta_i$  representa o estimador de máxima verosimilhança do risco relativo de incidência da população da região  $i$ .

### 4.1.3 A dimensão da amostra

No âmbito deste trabalho, apenas foram consideradas as escolas cuja dimensão da amostra (número de inquiridos ou observações por escola) é suficientemente representativa da escola para uma dada precisão.

Admite-se que, para cada grupo etário (6/7 e 13/14 anos) em cada escola, os alunos foram seleccionados por amostragem aleatória com probabilidades iguais de serem seleccionados para a amostra. A dimensão da amostra mínima necessária na escola  $i$  e grupo etário  $j$ ,  $n_{ij}$ , é calculada com base na expressão seguinte (Hansen, Hurwitz et al., 1953):

$$n_{ij} = \frac{k^2 N_{ij} S_j^2}{k^2 S_j^2 + d^2 N}$$

onde  $N_{ij}$  representa o número total de alunos da escola  $i$  e grupo etário  $j$ ,  $S_j^2$  representa a variância corrigida no grupo  $j$ ,  $d$  representa o limite da variação da medida de precisão relativa desejada e  $k$  representa o valor da função inversa da função de distribuição Normal (estandardizada) para um intervalo de confiança definido. O valores admitidos para  $d$  e  $k$  foram respectivamente 0.1 e 1.281 (valor da função para um intervalo de confiança de 80%). As variâncias corrigidas para os grupos etários foram obtidas a partir dos resultados publicados por Pinto e Almeida (Pinto e Almeida, 2003).

Questão do ISAAC	6/7 anos		13/14 anos	
	p	s <sup>2</sup>	p	s <sup>2</sup>
Alguma vez o médico lhe diagnosticou asma?	0.094	0.085	0.147	0.125
Alguma vez teve sintomas de pieira?	0.281	0.202	0.218	0.170
Alguma vez teve crises de espirros ou de pingos no nariz?	0.291	0.206	0.371	0.233
Alguma vez o médico lhe diagnosticou febre dos fenos?	0.042	0.040	0.075	0.069
Alguma vez teve lesões de pele com comichões?	0.210	0.166	0.133	0.115
Alguma vez o médico lhe diagnosticou um eczema?	0.141	0.121	0.127	0.111

Tabela 4.3- Proporção (p) e variância corrigida (s<sup>2</sup>) da amostra que declarou já ter tido pieira, asma, rinite ou eczema (adaptado de Pinto e Almeida, 2003).

Os resultados obtidos levaram ao abandono de 28 escolas, dado que nestas o valor de  $n_{ij}$  foi superior à dimensão da amostra recolhida, pelo que o estudo prosseguiu com as restantes 75 escolas do ensino Básico e Secundário.

### 4.1.4 Área de estudo

As escolas consideradas no estudo distribuem-se numa área com cerca de 1800 km<sup>2</sup> abrangendo 9 concelhos (Lisboa, Odivelas, Oeiras, Sintra, Amadora, Loures, Vila

Franca de Xira, Mafra e Sobral de Monte Agraço) com particular incidência nos concelhos de Sintra (19 escolas) e Loures (16). Localizam-se junto aos núcleos urbanos, semi-urbanos ou rurais, ao longo das vias de comunicação.



Figura 4.1-Localização das escolas consideradas no estudo.

A grande área abrangida pelo estudo, proporciona meios envolventes às escolas bastante diversificados, desde áreas urbanas com elevada densidade populacional, de tráfego e industrial, até às áreas rurais caracterizadas por baixa densidade populacional concentrada nas actividades agrícolas e na criação de gado.



a)



b)

Figura 4.2- Ambiente envolvente em duas das escolas inseridas no estudo a) escola em ambiente rural e b) escola em ambiente urbano.

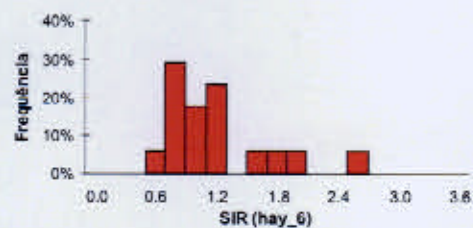
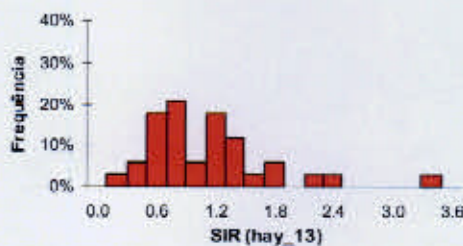
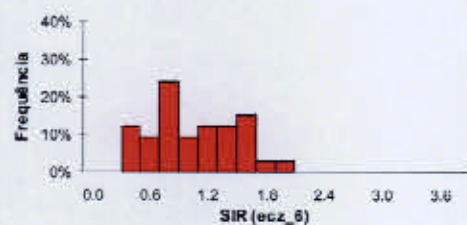
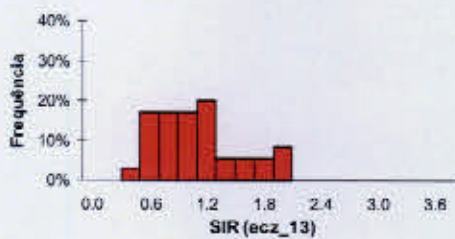
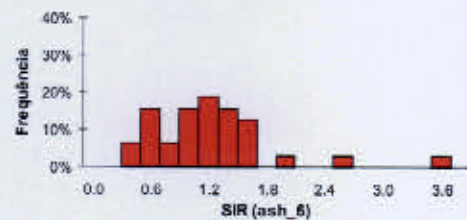
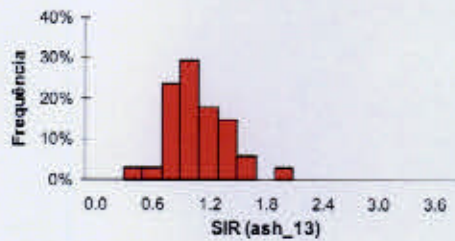
Nas regiões de Lisboa, Oeiras e Sintra caracterizadas por núcleos urbanos densamente povoados, a dimensão das escolas da amostra é variada, enquanto que no meio rural (concelhos de Mafra ou Sobral do Monte Agraço) há maior densidade de escolas do ensino básico, geralmente de dimensão reduzida.

#### 4.1.5 Análise descritiva de dados

A análise descritiva dos dados do ISAAC foi feita com base na elaboração de histogramas, no cálculo de parâmetros amostrais (descrição univariada) e na descrição da distribuição geográfica (descrição espacial) das razões padronizadas de incidência (cf. 4.1.2) por grupo etário, com o objectivo de caracterizar, em termos exploratórios, as colecções de dados recolhidas.

##### 4.1.5.1 Descrição univariada das variáveis da saúde

Da análise conjunta dos histogramas e dos valores das estatísticas descritivas das variáveis da saúde, evidenciam-se algumas diferenças na distribuição das frequências. Os histogramas apresentam formas assimétricas com peso relativo das caudas geralmente baixo e em alguns casos encontram-se valores anómalos.



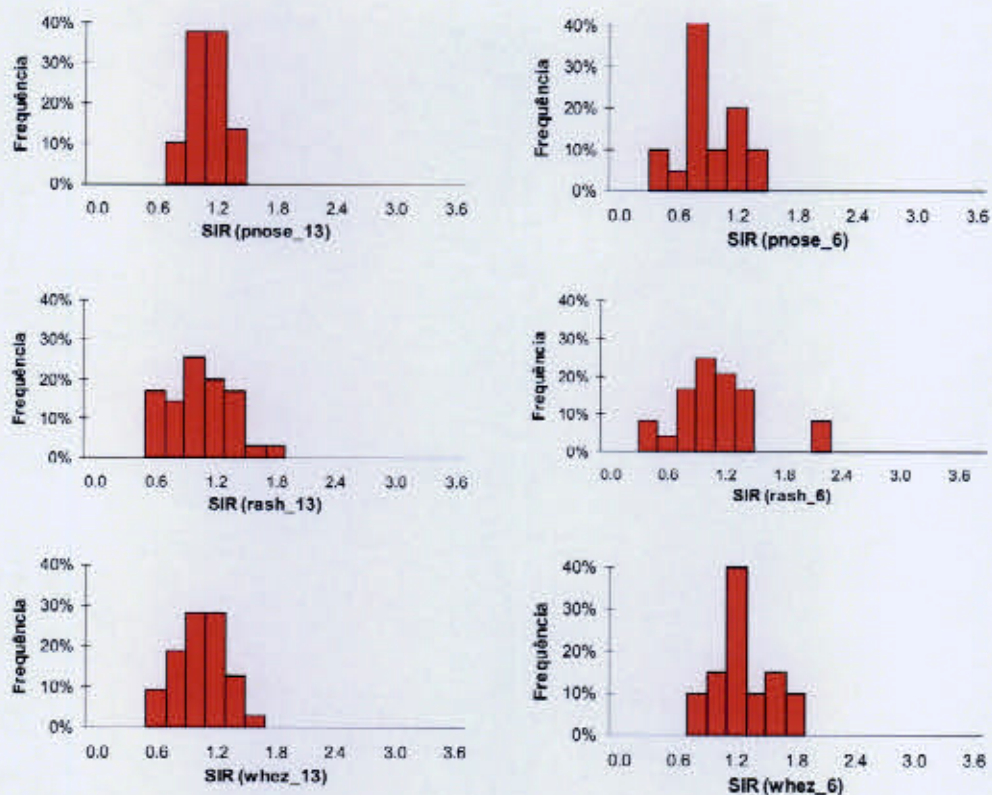


Figura 4.3- Histogramas dos valores das variáveis da saúde

Os respectivos parâmetros amostrais apresentam-se na tabela seguinte (Tabela 4.4).

Variável	Nº de obs.	Média	P.25	P.50	P.75	P.95	Variância	Min	Max	Assimetria
Ash_13	34	0.98	0.74	0.97	1.07	1.53	0.10	0.24	1.85	0.42
Ash_6	32	1.13	0.79	1.04	1.31	2.60	0.41	0.32	3.46	1.85
Ecz_13	35	1.00	0.61	0.94	1.23	1.85	0.18	0.39	1.97	0.69
Ecz_6	33	0.97	0.68	0.91	1.36	1.66	0.18	0.26	1.82	0.10
Hay_13	34	1.02	0.56	0.90	1.27	2.27	0.39	0.20	3.24	1.58
Hay_6	17	1.10	0.77	1.00	1.16	2.44	0.26	0.42	2.44	1.29
Rash_13	35	0.96	0.76	0.89	1.20	1.60	0.09	0.45	1.61	0.24
Rash_6	24	1.00	0.73	0.93	1.20	2.03	0.20	0.22	2.12	0.82
Prose_13	29	1.00	0.91	1.01	1.14	1.22	0.02	0.71	1.25	-0.18
Prose_6	20	0.82	0.84	0.79	1.08	1.24	0.08	0.29	1.26	-0.02
Whez_13	32	0.97	0.74	0.97	1.16	1.30	0.06	0.53	1.58	0.09
Whez_6	20	1.15	0.99	1.07	1.40	1.68	0.08	0.69	1.73	0.53

Tabela 4.4- Parâmetros amostrais das variáveis da saúde

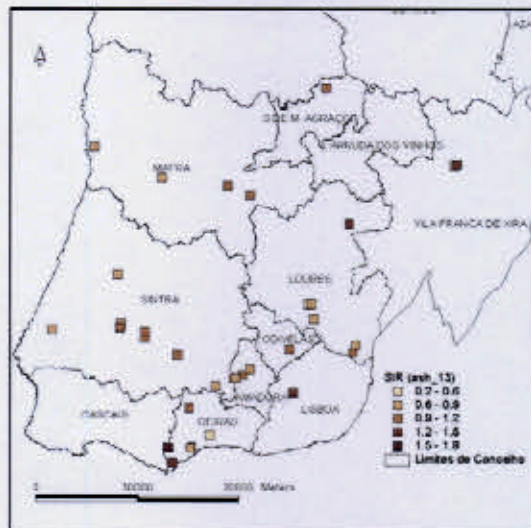
A elevada assimetria positiva evidenciada pelos histogramas das variáveis ash\_6, hay\_13 e hay\_6, quantificada pelos coeficientes de assimetria, dá indicações da existência de valores anómalos nestas variáveis. Verifica-se que estas são as variáveis com valores máximos mais elevados e com maiores amplitudes.

As variáveis *prose\_13* e *prose\_6* são as únicas que apresentam histogramas com forma assimétrica negativa, evidenciando uma tendência para frequências elevadas de valores acima da média.

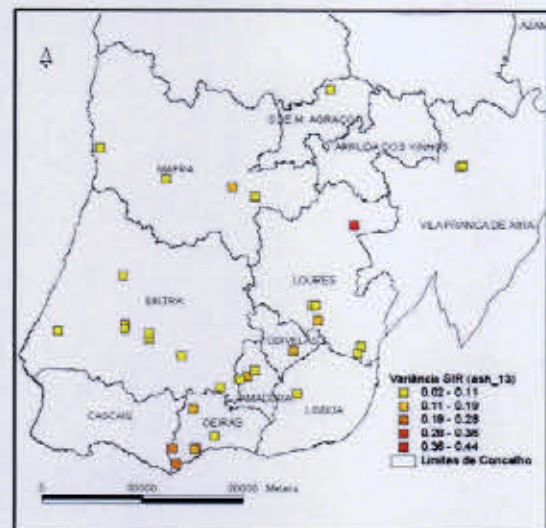
Quanto aos valores médios e variâncias observados são semelhantes entre si, com coeficientes de variação (definido pelo quociente entre a raiz quadrada da variância e a média da variável) a variar entre 0.15 (*prose\_13*) e 0.61 (*hay\_13*).

#### 4.1.5.2 Descrição espacial das variáveis da saúde

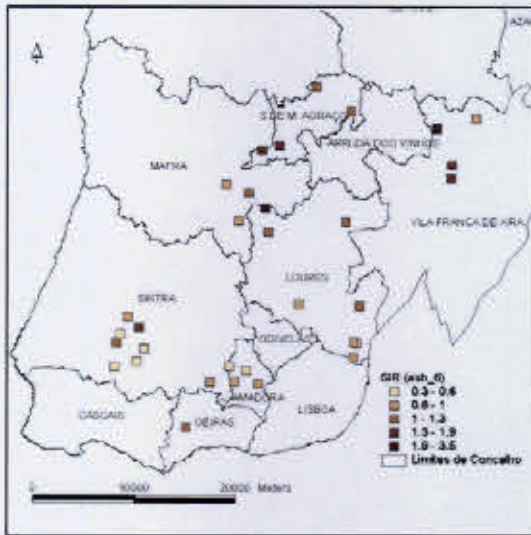
A distribuição das escolas na área de estudo (ou geometria das amostras) apresenta, em ambos os grupos etários, áreas de densidade muito variada caracterizadas por várias amostras dispostas em aglomerados. Apresentam-se na Figura 4.4 os mapas com representação dos SIR obtidos com o estimador de máxima verosimilhança, e respectivas variâncias.



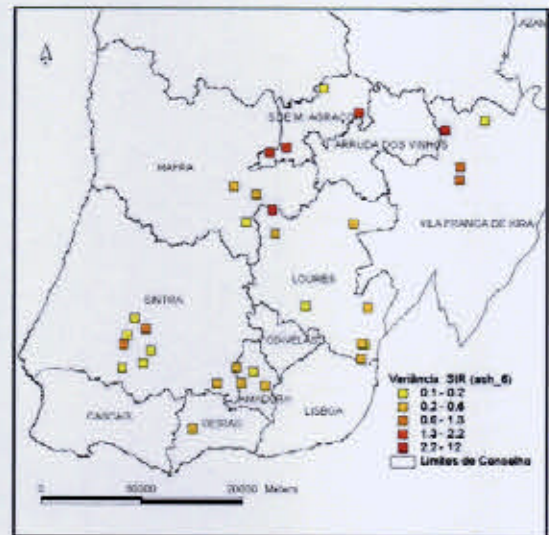
1.a



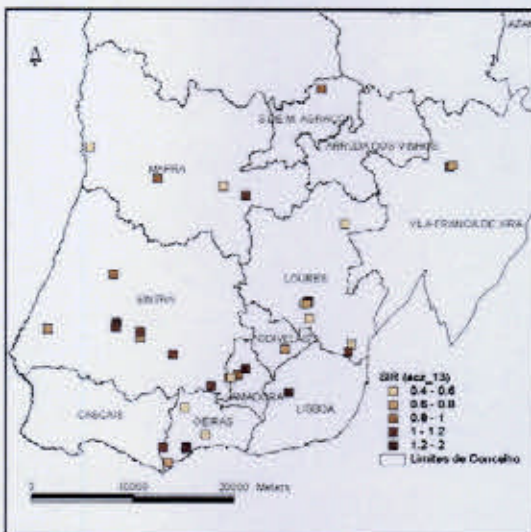
1.b



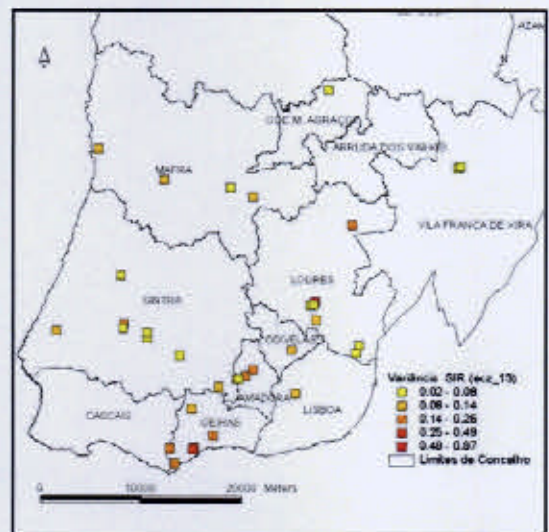
2.a



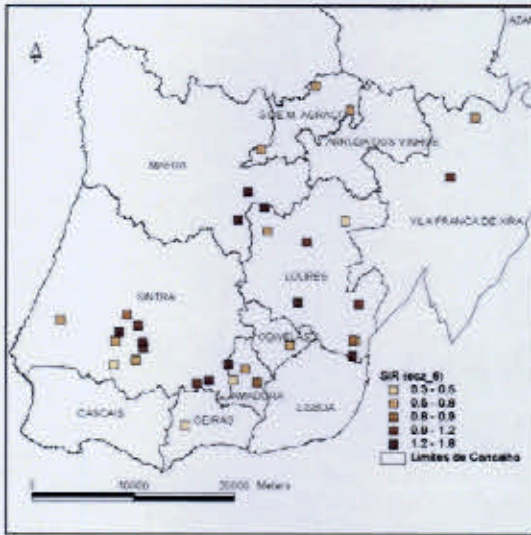
2.b



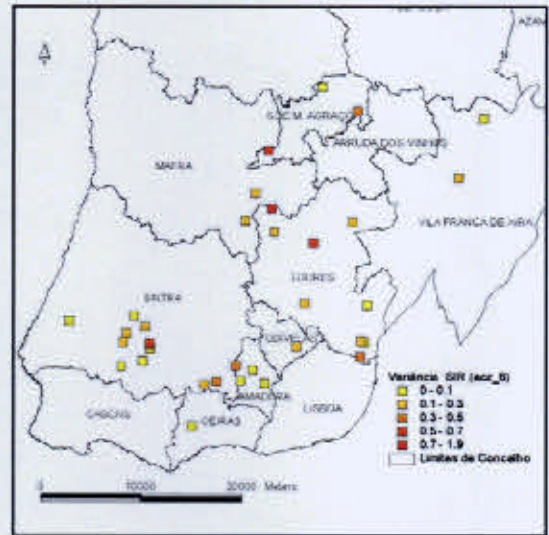
3.a



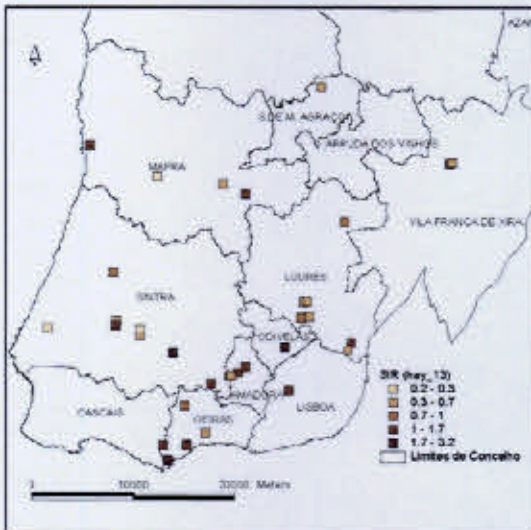
3.b



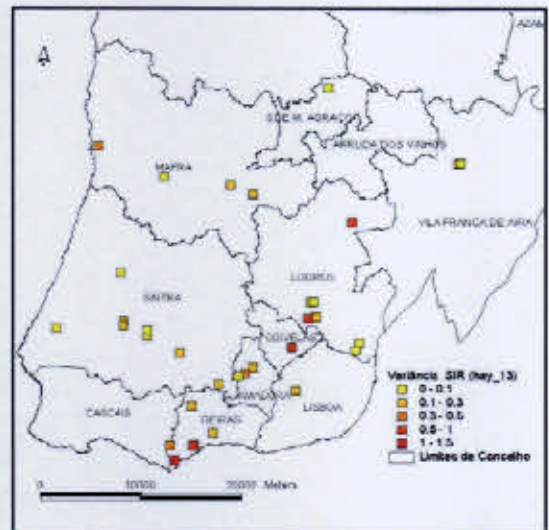
4.a



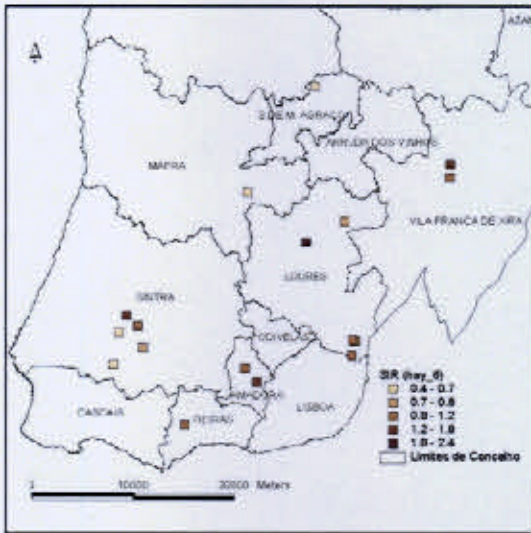
4.b



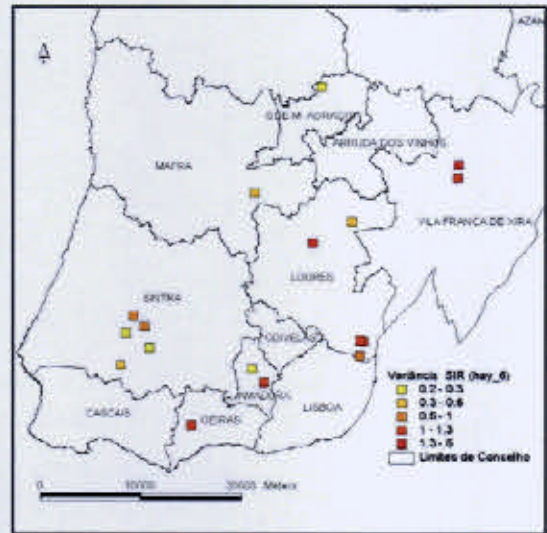
5.a



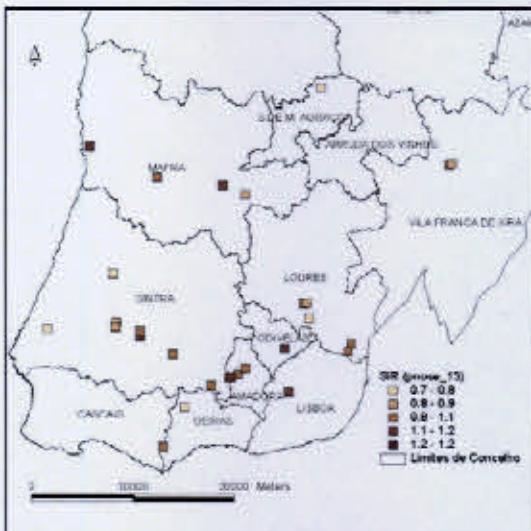
5.b



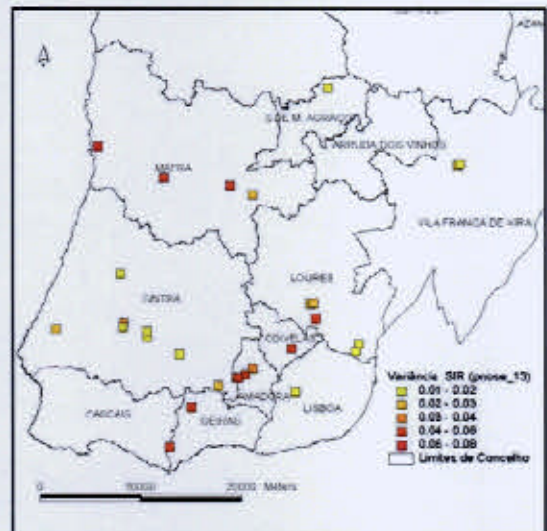
6.a



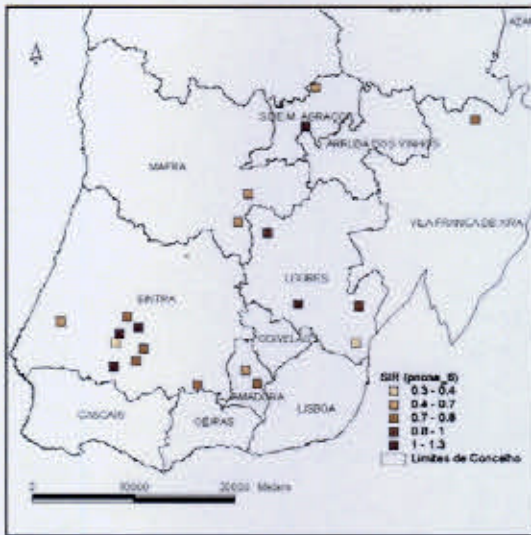
6.b



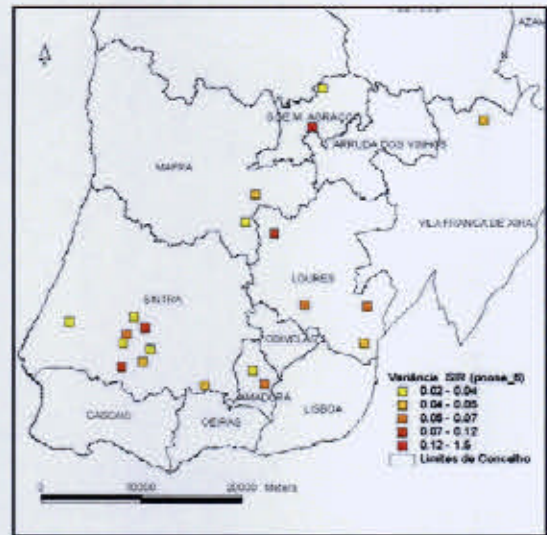
7.a



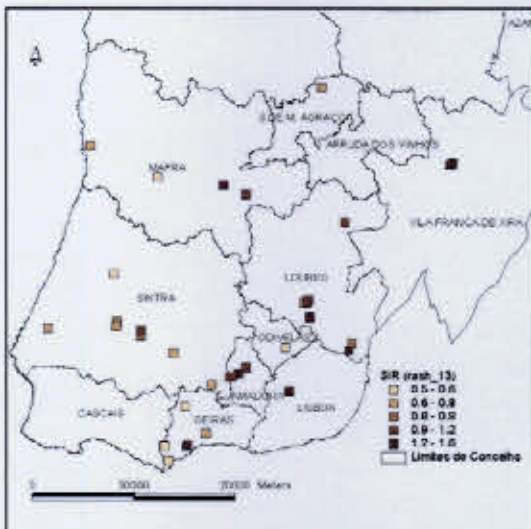
7.b



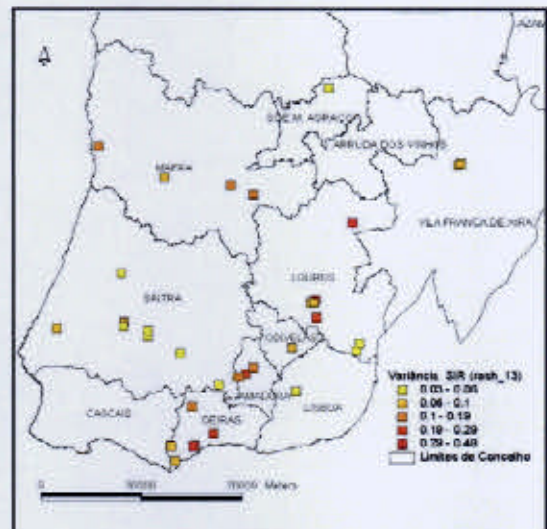
8.a



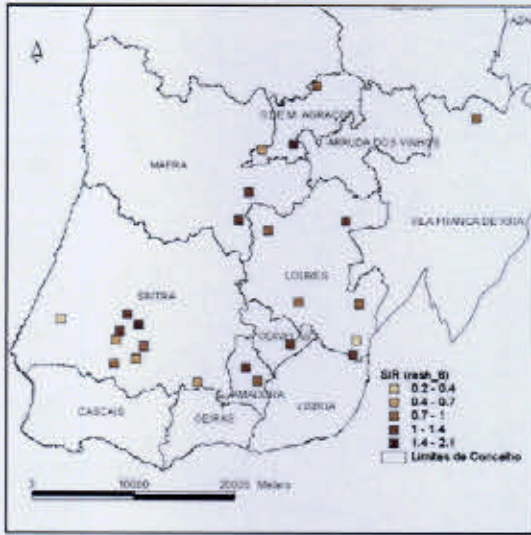
8.b



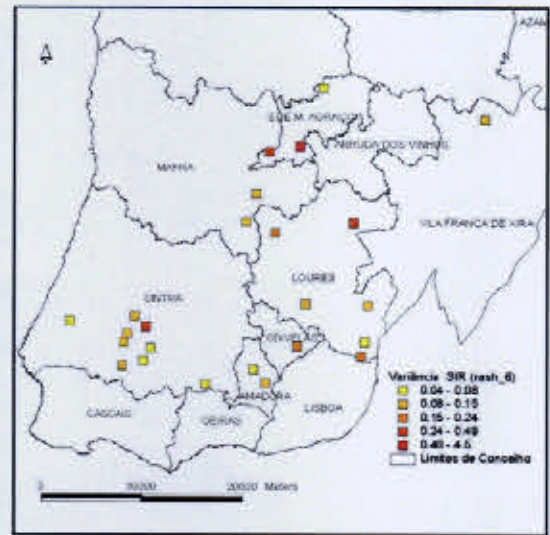
9.a



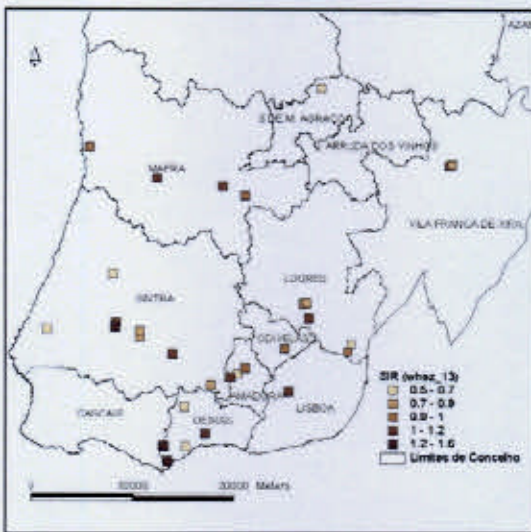
9.b



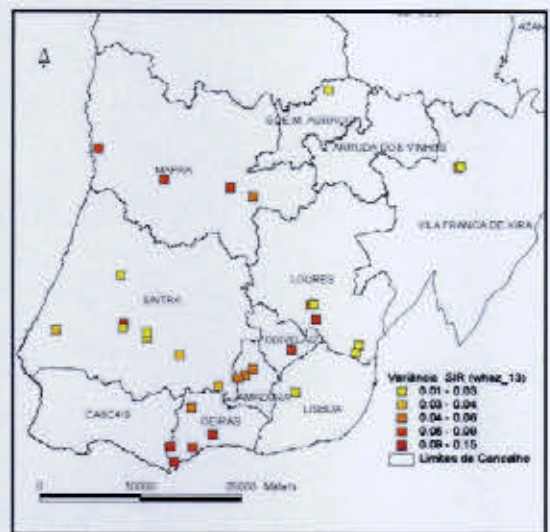
10.a



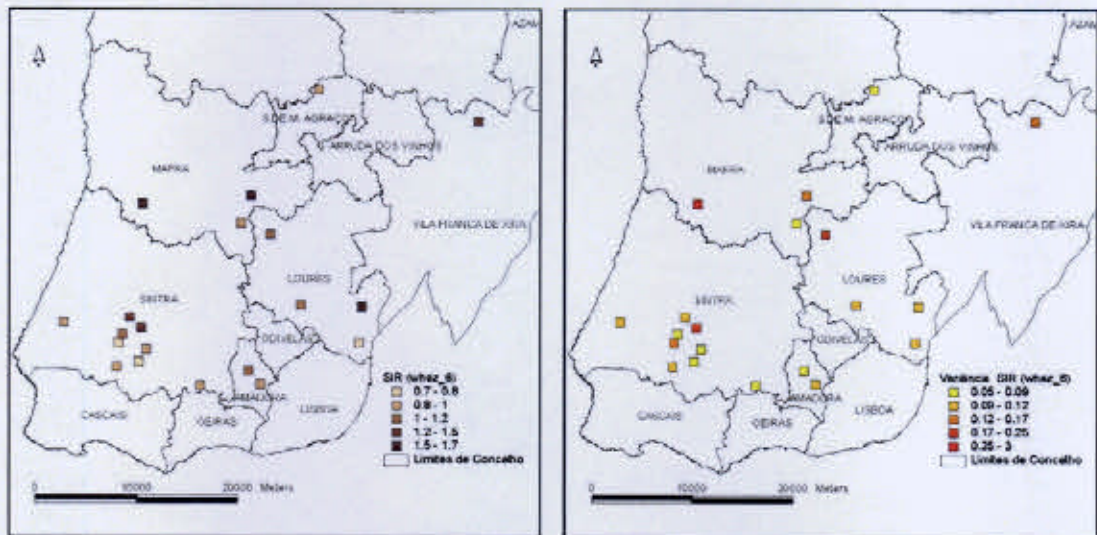
10.b



11.a



11.b



12.a

12.b

**Figura 4.4-** Distribuição espacial dos SIR (.a) e das respectivas variâncias (.b) das variáveis ash\_13 (1), ash\_6 (2), ecz\_13 (3), ecz\_6 (4), hay\_13 (5), hay\_6 (6), pnose\_13 (7), pnose\_6 (8), rash\_13 (9), rash\_6 (10), whez\_13 (11), whez\_6 (12).

Os valores de variância evidenciam uma tendência para serem mais elevados nas áreas onde os *SIR* são também mais elevados. Esta evidência resulta do facto da variância do estimador ser inversamente proporcional ao valor esperado (ver 4.1.2) e por isso, nas escolas com menos alunos (onde o valor esperado é mais baixo) há normalmente maior variabilidade amostral.

A inspecção visual da distribuição das variáveis no espaço permite identificar algumas características mais evidentes da estrutura espacial:

- A variável ash\_13 apresenta uma tendência para valores mais elevados nas regiões limítrofes da área de estudo, enquanto que variável ash\_6 evidencia uma tendência para os valores mais elevados se concentrarem na região Norte;
- Nas variáveis ecz\_13 e hay\_13 existe uma tendência para os valores mais elevados se localizarem na região Sul, enquanto que os valores mais baixos encontram-se dispersos por toda a área de estudo;
- Os valores mais elevados da variável rash\_13 localizam-se ao longo da direcção Nordeste-Sudoeste, enquanto que os valores mais baixos se concentram preferencialmente no quadrante Sudoeste.

Não se detectaram quaisquer padrões particulares na distribuição espacial das restantes variáveis -ecz\_6, hay\_6, rash\_6, pnose\_13, pnose\_6, whez\_13 e whez\_6.

## 4.2 Variável de qualidade do ar

### 4.2.1 Base de dados Qualar

Os dados relativos à caracterização espacial das concentrações de  $PM_{10}$  ( $\mu\text{m}/\text{m}^3$ ) foram fornecidos pela base de dados qualidade do ar do Instituto do Ambiente, disponível via *World Wide Web* ([www.qualar.org](http://www.qualar.org)), onde se recolheram as séries temporais de concentrações ( $\mu\text{m}/\text{m}^3$ ) horárias registadas, entre o dia 1 de Janeiro de 2001 e 31 de Dezembro de 2003, de oito estações localizadas na região de Lisboa, conforme Tabela 4.5.

Estação	Código da estação	Tipo de estação
Av. Da Liberdade	E1	Tráfego
Cascais (mercado)	E2	Tráfego
Entre Campos	E3	Tráfego
Loures	E4	Fundo
Mem Martins	E5	Fundo
Olivais	E6	Fundo
Reboleira	E7	Fundo
Qta. do Marquês	E8	Fundo

Tabela 4.5- Estações de medição de  $PM_{10}$  consideradas no âmbito do estudo

A localização geográfica das estações foi também fornecida através da base de dados referida, e inserida num ficheiro ASCII (American Standard Code for Information Interchange) com códigos das estações e respectivas coordenadas geográficas para referenciar e integrar num SIG.

### 4.2.2 Área de estudo

Os dados recolhidos abrangem uma área com cerca de  $600 \text{ km}^2$ , envolvendo os concelhos de Lisboa (E1, E3 e E6), Amadora (E8), Oeiras (E7), Cascais (E2), Sintra (E5) e Loures (E4).

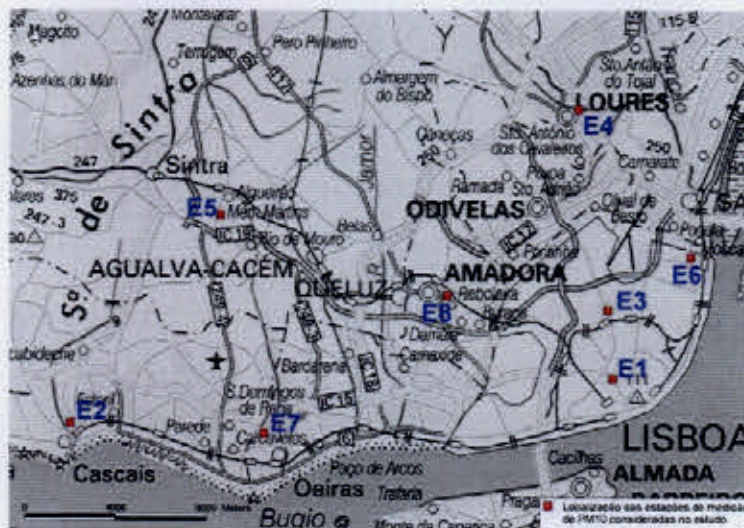


Figura 4.5- Localização das estações de medição consideradas no estudo

O meio ambiente que envolve as estações de medição de  $PM_{10}$  é caracterizado pelo ambiente urbano com diferentes níveis de exposição potencial da população à poluição atmosférica.



a)

b)

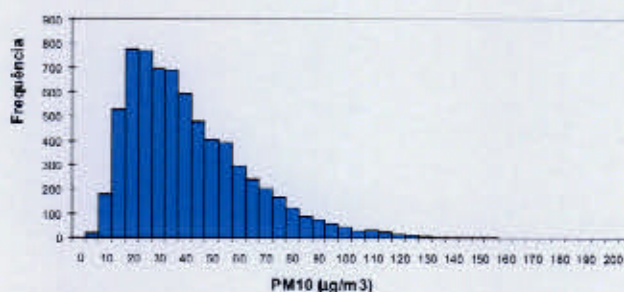
Figura 4.6- Localização de duas das estações de medição a) de fundo (E6) e b) de tráfego (E1)

As estações E1 a E3 localizam-se em áreas de circulação automóvel intensa e de passagem de pessoas (também designadas por estações de tráfego), enquanto que as estações E4 a E8 localizam-se em zonas mais afastadas das vias de tráfego principais ou de fontes industriais importantes (também designadas por estações de fundo).

## 4.2.3 Análise descritiva de dados

### 4.2.3.1 Estatística descritiva univariada

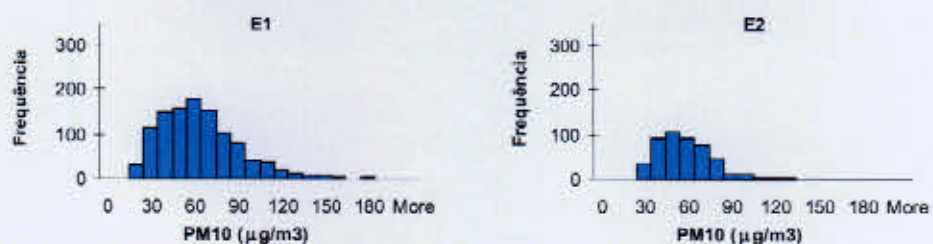
Na Figura 4.7 apresenta-se o histograma e os parâmetros amostrais do conjunto de observações de  $PM_{10}$ . A distribuição da concentração de  $PM_{10}$  na área de estudo evidencia, uma forma leptocúrtica de assimetria positiva cujo peso relativo das caudas é de  $1.39 \mu\text{g}^3/\text{m}^9$  e a concentração de valores junto do centro (coeficiente de achatamento) é de  $2.84 \mu\text{g}^4/\text{m}^{12}$ . A média estimada é de  $39.04 \mu\text{g}/\text{m}^3$  e o desvio-padrão é de  $23.56 \mu\text{g}/\text{m}^3$ .

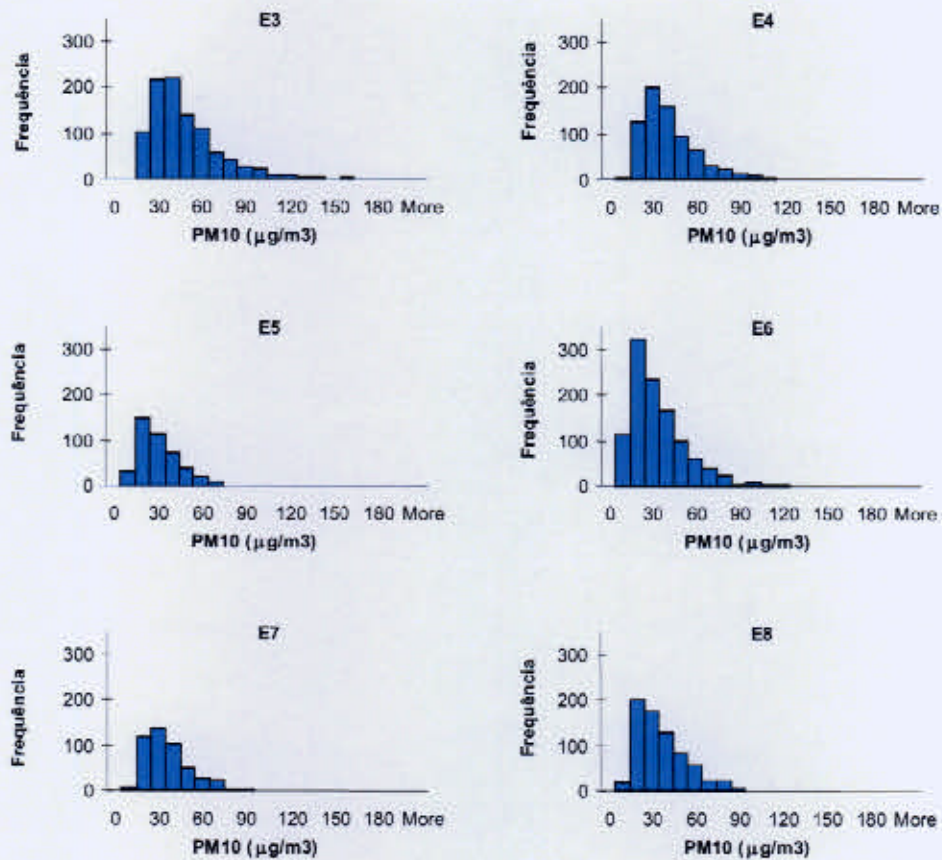


Nº de Obs.	Média	Desvio-Padrão	Assimetria	Achatamento	Mínimo	Mediana	Máximo
5970	39.74	23.78	1.34	2.67	1.9	34.39	184.55

Figura 4.7- Histograma e alguns parâmetros amostrais de concentrações médias diárias de  $PM_{10}$  observadas no conjunto das estações de medição.

Para caracterizar e avaliar as diferenças nas distribuições (localização e escala) de cada estação apresentam-se na Figura 4.8 os histogramas das concentrações médias diárias de  $PM_{10}$  observadas em cada uma delas para o período analisado.





**Figura 4.8- Histogramas das concentrações médias diárias de PM<sub>10</sub>**

Os histogramas evidenciam uma forma de assimetria positiva semelhante à do histograma global. As diferenças na forma dos histogramas centram-se no peso relativo das caudas e no seu achatamento, onde as estações E3, E4 e E6 evidenciam maior concentração de valores junto ao centro da distribuição. Os respectivos parâmetros amostrais apresentam-se na Tabela 4.6.

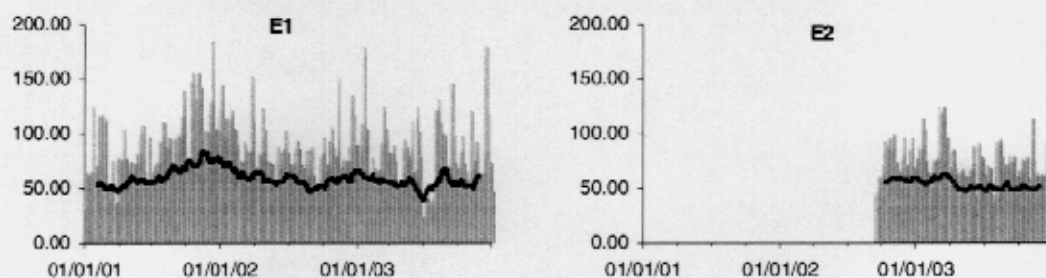
Est.	N	Média	P.25	P.50	P.75	P.95	D-padrão	Min	Máx	Assimetria
E1	1094	58.98	38.80	54.88	73.35	108.92	27.25	08.47	184.55	01.01
E2	476	53.01	39.20	50.29	63.10	86.87	18.20	21.42	124.42	00.88
E3	965	43.19	25.65	37.46	53.32	92.12	24.05	09.08	171.41	01.62
E4	726	35.97	22.71	31.59	45.70	71.67	18.35	08.01	152.53	01.39
E5	430	25.78	15.75	22.88	32.74	52.87	13.26	06.14	69.94	00.95
E6	1077	28.36	15.01	23.73	37.02	65.40	18.46	01.90	119.93	01.34
E7	476	31.14	19.71	27.76	38.93	63.25	15.46	08.17	99.68	01.16
E8	726	32.00	18.30	27.29	42.16	69.33	18.10	06.22	122.36	01.23

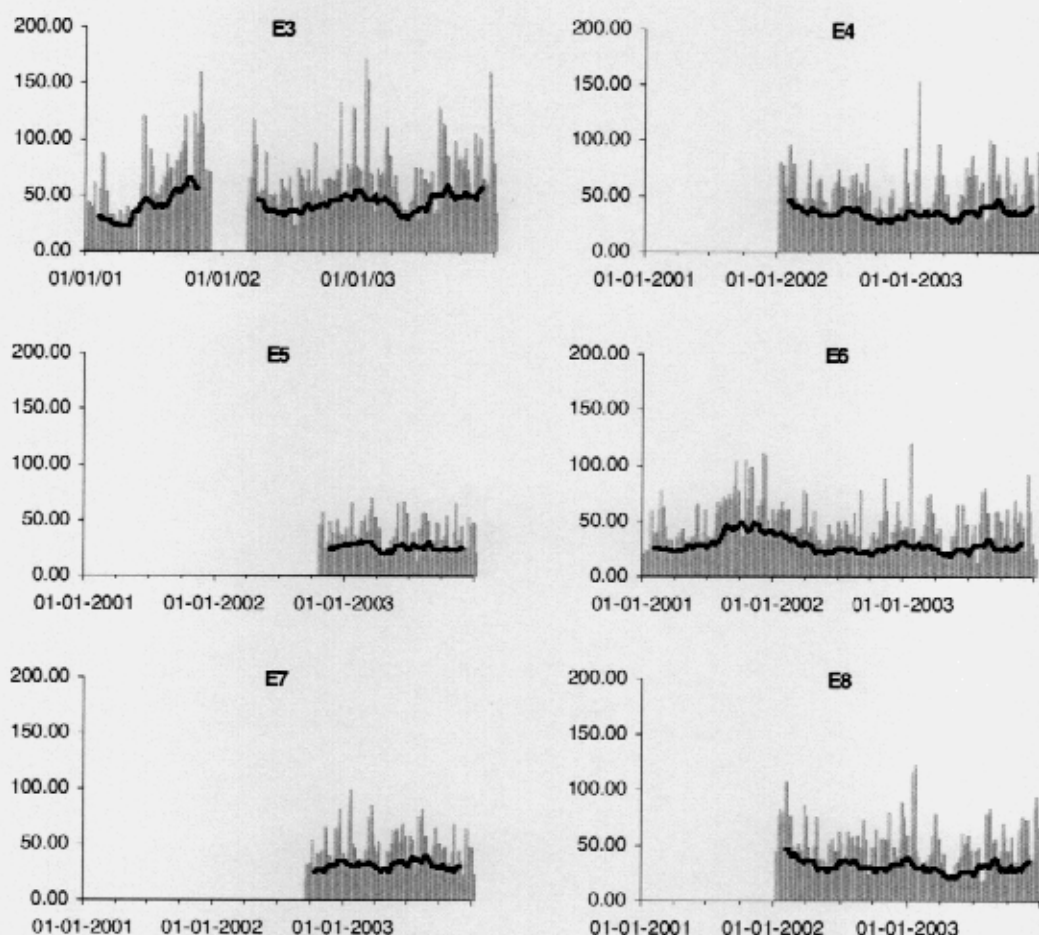
**Tabela 4.6- Medidas de localização e dispersão das concentrações médias diárias de PM<sub>10</sub>**

De forma global, verifica-se uma prevalência dos valores médios, máximos e desvios-padrão mais elevados nas estações de tráfego, enquanto que as maiores variabilidades relativas, quantificadas pelos coeficientes de assimetria são exibidas pelas estações E3, E4, E6 e E8. Nos gráficos seguintes (Figura 4.9) apresentam-se as séries temporais das concentrações médias diárias de PM<sub>10</sub> ( $\mu\text{g}/\text{m}^3$ ) das estações de monitorização.

Apresentam-se ainda os resultados do alisamento das séries, obtidos pelo método das médias móveis, com intenção de reduzir o impacto das altas frequências observadas e tornar mais evidentes eventuais padrões subjacentes nas sucessões cronológicas. Após consideradas várias hipóteses, o período escolhido para o alisamento das séries foi de 60 dias.

As descontinuidades que aparecem nos gráficos são derivadas do facto de existirem valores omissos nas medições. No caso das estações E2, E4, E5, E7 e E8, só têm registos a partir de 2002, altura em que entraram em funcionamento.





**Figura 4.9- Séries temporais das concentrações médias diárias de  $PM_{10}$**

De uma forma geral, as estações de medição evidenciam oscilações mais ou menos regulares ao longo do ano, com tendência para os valores mais elevados se localizarem nos períodos de Dezembro a Março. No caso das estações de tráfego, observa-se ainda uma tendência para valores mais baixos localizarem-se nos períodos de Junho a Agosto. Estas observações sugerem que as variações na concentração de  $PM_{10}$  estão relacionadas com a quantidade de veículos em circulação, na medida que em se pensar em termos meteorológicos a situação deveria ser contrária (concentrações mais elevadas no Verão e menores no Inverno, devido ao efeito da precipitação que contribui para a redução de concentrações de  $PM_{10}$  no ambiente), se exceptuarmos os dias de Inverno em que ocorrem inversões térmicas.

A observação das sucessões cronológicas e dos resultados obtidos por alisamento das médias móveis permitiu identificar flutuações dos valores em relação às médias o que

indicia a existência de diferentes valores médios ao longo do tempo, ou seja uma variabilidade temporal não detectada pelas estatísticas globais.

#### 4.2.3.2 Estatística descritiva bivariada

A descrição bivariada dos valores observados nas estações de medição auxilia a percepção de associações entre valores observados em diferentes estações, considerando a variabilidade conjunta das séries.

Neste sentido, apresentam-se na Tabela 4.7 os coeficientes de correlação de Pearson (em vez das covariâncias, para eliminar os efeitos de escala observados em 4.2.3.1) calculados para todos os pares de estações de medição, onde na maioria dos casos é evidente um elevado grau de associação linear.

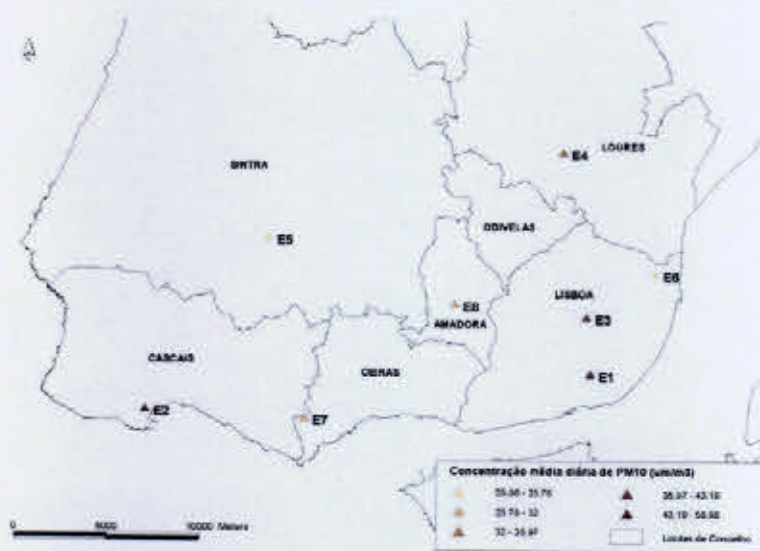
	E1	E2	E3	E4	E5	E6	E7	E8
E1	1.000	0.669	0.846	0.799	0.709	0.853	0.808	0.841
E2	0.669	1.000	0.721	0.694	0.749	0.767	0.771	0.728
E3	0.846	0.721	1.000	0.881	0.746	0.942	0.823	0.907
E4	0.799	0.694	0.881	1.000	0.819	0.918	0.874	0.865
E5	0.709	0.749	0.746	0.819	1.000	0.826	0.876	0.804
E6	0.853	0.767	0.942	0.918	0.826	1.000	0.881	0.901
E7	0.808	0.771	0.823	0.874	0.876	0.881	1.000	0.887
E8	0.841	0.728	0.907	0.865	0.804	0.901	0.887	1.000

Tabela 4.7- Coeficientes de correlação de Pearson das concentrações médias diárias de PM<sub>10</sub>.

Os resultados obtidos evidenciam correlação linear elevada entre as séries temporais. No Anexo A apresentam-se os diagramas de dispersão e as rectas de regressão linear (calculadas pelo critério dos mínimos quadrados) sobrepostas, para visualizar graficamente as associações evidenciadas entre os dados observados.

#### 4.2.3.3 Descrição espacial univariada e bivariada

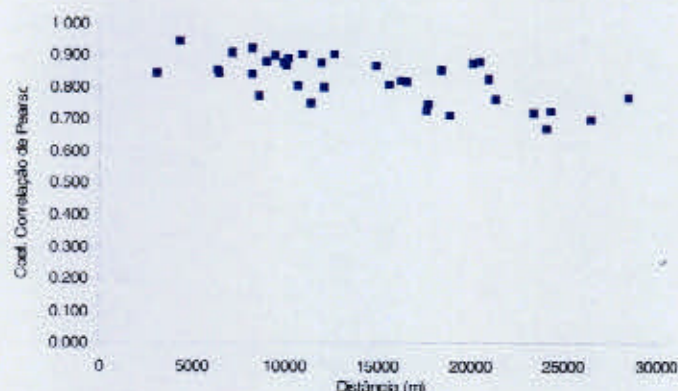
A distribuição espacial das concentrações médias de PM<sub>10</sub> observadas nas estações é caracterizada por uma dispersão elevada das estações, conforme mapa da Figura 4.10.



**Figura 4.10- Distribuição geográfica e concentrações médias de PM<sub>10</sub>**

Verifica-se que a distribuição geográfica das estações de medição apresenta médias locais heterogéneas e uma dispersão de valores extremos, o que indicia a presença de heterocedasticidade. De facto, pode observar-se que valores obtidos em estações próximas entre si (E3 e E6 ou E1 e E6) são bastante discordantes, o que pode ser justificado pelo facto de E1 e E3 estarem localizadas em pontos de circulação automóvel intensa e de passagem de pessoas, enquanto que E6 localiza-se num ponto mais afastado das vias de tráfego principais (e de fontes de poluição industrial).

Para compreender como é que os valores observados nas estações se correlacionam no espaço, apresenta-se na Figura 4.11 o correlograma respectivo onde se representam os valores dos coeficientes de correlação calculados em função da distância euclidiana entre observações.

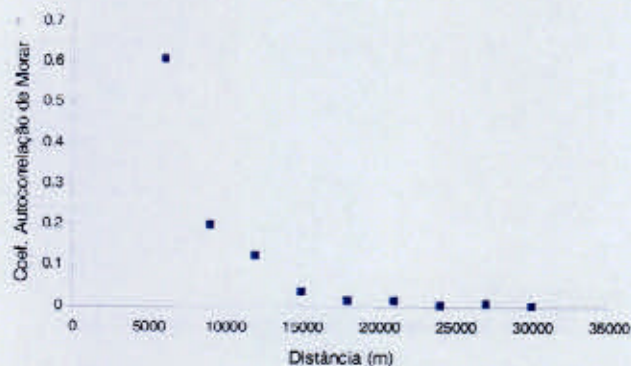


**Figura 4.11- Correlograma (correlação linear de Pearson) das concentrações médias diárias de PM<sub>10</sub>**

O diagrama de dispersão apresenta uma tendência para as correlações entre amostras decrescerem à medida que aumentam as distâncias entre si. No entanto, essa tendência é ténue, e não é clara a presença de uma estrutura de continuidade no espaço. Num cenário em que as observações são realizações de um mesmo processo aleatório, logo com a mesma distribuição de probabilidades, a função de correlação pode ser substituída pelo coeficiente de autocorrelação de Moran  $I$  onde se admite que a média é constante no domínio espacial e a variância depende apenas da distância (euclidiana) entre amostras. O valor de  $I$  para cada classe de distância  $h$  é definida por:

$$I(h) = n \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \cdot [Z(x_i, t) - m(x)] \cdot [Z(x_j, t) - m(x)]}{s^2(x) \cdot \sum_{i=1}^n \sum_{j=1}^n w_{ij}}$$

onde  $Z(x_i, t)$  e  $Z(x_j, t)$  representam os valores de PM<sub>10</sub> medidos nas estações  $i$  e  $j$  no instante  $t$ ,  $m(x)$  e  $s^2(x)$  representam respectivamente a média e a variância de todas as observações de PM<sub>10</sub>, de dimensão  $n$ , e  $w_{ij}$  representa o ponderador da distância entre  $i$  e  $j$  (definido neste caso pelo inverso do quadrado da distância). Os coeficientes de autocorrelação de Moran  $I$ , apresentam-se no correlograma da Figura 4.12.



**Figura 4.12- Correlograma (autocorrelação de Moran I) das concentrações médias diárias de  $PM_{10}$**

Este novo resultado evidencia a presença de uma estrutura de continuidade espacial, que decai rapidamente entre os 5000 e 8500 metros de distância.

### 4.3 Comentários

Nos sub-capítulos 4.1 e 4.2 foram estimadas algumas medidas de localização e de dispersão dos atributos e analisaram-se as suas distribuições no espaço e, no caso das  $PM_{10}$ , também no tempo.

Estimaram-se parâmetros amostrais e construíram-se histogramas das variáveis, e procedeu-se à inspeção visual da sua distribuição espacial. No caso das  $PM_{10}$  calcularam-se os correlogramas e construíram-se ainda os gráficos das séries temporais onde se sugere que as variações na concentração de  $PM_{10}$  estão relacionadas com a quantidade de veículos em circulação.

Os histogramas obtidos para as variáveis da saúde apresentaram formas assimétricas e achatadas e a sua distribuição espacial evidenciou a existência de vários aglomerados de escolas. Para a qualidade do ar, o histograma apresentou uma forma acentuadamente assimétrica, em que a distribuição das estações é escassa no espaço e abundante no tempo.

Verificou-se ainda que os domínios espacial e temporal das variáveis (de qualidade do ar e da saúde) não são comuns. A distribuição espacial das variáveis da saúde abrange

uma área mais vasta do que a variável da qualidade do ar e as séries temporais apenas foram recolhidas para as  $PM_{10}$ .

## 5 Modelação geoestatística

### 5.1 Introdução ao formalismo geoestatístico

Os fundamentos teóricos sobre a modelação geoestatística, desenvolvida originalmente para as ciências da Terra são aplicados na área da epidemiologia segundo duas grandes perspectivas, a da construção de mapas descritivos, onde o objectivo da modelação geoestatística é atribuir uma estimativa do verdadeiro risco relativo de incidência, e a da construção de mapas para formulação de hipóteses sobre a etiologia das doenças no âmbito de estudos ecológicos (Elliott, Cuzick et al., 1992; Lawson, Biggeri et al., 2000; Pascutto, Wakefield et al., 2000; Lawson, 2001). Neste último caso, o objectivo da modelação geoestatística é fornecer resultados que permitam gerar hipóteses sobre relações causais entre as doenças e variáveis explicativas.

Em qualquer dos casos, a abordagem geoestatística implica a modelação de um processo espacial onde existe um certo grau de incerteza sobre o fenómeno entre os locais das amostras. O formalismo geoestatístico baseia-se no conceito de função aleatória, aonde o conjunto de valores desconhecidos constitui um conjunto de variáveis aleatórias espacialmente dependentes. Desta forma a geoestatística reconhece a incerteza do fenómeno que pretende modelar e permite a estimação de valores em locais não amostrados, considerando algumas hipóteses sobre as características estatísticas do fenómeno.

Uma variável aleatória é uma variável que pode assumir uma série de valores de acordo com uma distribuição de probabilidade. Considere-se o estudo de uma variável aleatória<sup>3</sup> contínua onde  $s$  representa a localização de  $Z$  no espaço. A distribuição de probabilidade de  $Z(s)$  é definida pela função de distribuição:

$$F(z) = \Pr[Z(s) \leq z] = \int_{-\infty}^z f(z) dz$$

onde  $f(z)$  representa a função de densidade de probabilidade (f.d.p) de  $Z(s)$ . A caracterização da distribuição de probabilidade de  $Z(s)$  é dada pelos seus momentos onde o primeiro momento em relação à origem é a média ou valor esperado:

---

<sup>3</sup> Por convenção, denota-se a variável aleatória por uma letra maiúscula, e as observações ou realizações da variável pela letra minúscula.

$$E[Z(s)] = \int_{-\infty}^{+\infty} z \cdot f(z) dz$$

Para avaliar a dispersão da distribuição de  $Z(s)$  recorre-se a momentos de ordem superior em relação à média, nomeadamente o segundo momento, ou variância:

$$V[Z(s)] = \int_{-\infty}^{+\infty} (z - m)^2 \cdot f(z) dz$$

onde  $m$  representa o valor esperado de  $Z(s)$ .

A função aleatória, FA, é definida como um conjunto de variáveis aleatórias  $Z(s)$  normalmente dependentes, uma para cada localização  $s$  numa área de estudo  $A$ . (componente aleatória das variáveis) e de correlações existentes entre as variáveis definidas nos pontos da região de interesse. A distribuição de probabilidade da função da FA define-se pela função de distribuição  $k$ -variada:

$$F(z_1, \dots, z_k) = \int_{-\infty}^{z_1} \dots \int_{-\infty}^{z_k} f(z_1, \dots, z_k) dz_1 \dots dz_k$$

onde  $f(z_1, \dots, z_k)$  representa a função de densidade conjunta de  $k$  variáveis aleatórias,  $Z(s)$ , dependentes entre si. Esta função descreve a incerteza conjunta sobre as  $k$  realizações  $z_k$ . Na prática, a análise da incerteza conjunta envolve o conhecimento das funções de distribuição de duas variáveis aleatórias de cada vez, e dos seus primeiros momentos, normalmente até ao segundo. Para os conhecer é necessário conhecer a distribuição de frequências em cada ponto  $s$ , o que implica a necessidade de se realizarem várias campanhas de amostragem em cada ponto. No entanto, geralmente apenas se conhece uma realização da cada variável<sup>4</sup> em alguns pontos  $s$  amostrados, pelo que é necessário assumirem-se alguns pressupostos (ou restrições) sobre o comportamento da função aleatória.

A solução proposta pelo formalismo geoestatístico passa por inferir a função de distribuição a partir da distribuição dos dados recolhidos na área de interesse, e assumir a estacionaridade dos dois primeiros momentos da função aleatória, ou seja, que a média,  $m$ , é constante no domínio espacial:

$$E[Z(s)] = E[Z(s+h)] = m$$

<sup>4</sup> Quando o conjunto de repetições consiste em medições registadas no mesmo local, em momentos diferentes a função aleatória é definida no domínio espaço-temporal e a variável denota-se por  $Z(s, t)$

e que a covariância espacial  $C(h)$  entre duas variáveis depende apenas da distância entre elas e não da sua localização:

$$C(h) = E[Z(s) \cdot Z(s+h)] - E[Z(s)] \cdot E[Z(s+h)]$$

onde  $h$  representa a distância e direcção entre os pontos  $s$  e  $s+h$

Existem outros parâmetros da função aleatória, como o correlograma, definido por

$$\rho(h) = \frac{C(h)}{C(0)}$$

ou o variograma, definido por

$$2\gamma(h) = E\left\{[Z(s) - Z(s+h)]^2\right\}$$

que são usados frequentemente para caracterizar e medir o modo como as variáveis aleatórias se correlacionam no espaço. O variograma é geralmente usado em substituição da função de covariância, dado que a sua definição é menos exigente, em termos de estacionaridade, porque não impõe a estacionaridade de segunda ordem da função aleatória, mas sim a estacionaridade de segunda ordem dos acréscimos espaciais da função aleatória (designada por estacionaridade intrínseca).

## 5.2 Análise da continuidade espacial e temporal

### 5.2.1 Introdução

A inferência dos parâmetros da função aleatória permite caracterizar e quantificar o modo como as variáveis aleatórias se correlacionam no espaço (e no tempo), e serve de base aos processos de inferência espacial como a estimação ou a simulação. Os parâmetros são estimados a partir dos valores observados, onde a estrutura de continuidade espacial é caracterizada pelo estimador<sup>5</sup> do variograma (Eq. 5.1) ou da covariância (Eq. 5.2):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \cdot \sum_{\alpha=1}^{N(h)} [z(s_{\alpha}) - z(s_{\alpha} + h)]^2$$

Eq. 5.1

<sup>5</sup> Por convenção, para distinguir de forma inequívoca um estimador amostral de um parâmetro da população, adiciona-se um ^ ao estimador.

$$\hat{C}(h) = \frac{1}{N(h)} \cdot \sum_{\alpha=1}^{N(h)} z(s_{\alpha}) \cdot z(s_{\alpha} + h) - \hat{m}_{\cdot h} \cdot \hat{m}_{+h}$$

Eq. 5.2

onde  $N(h)$  representa o número de pares de observações à distância  $h$ ,  $z(s_{\alpha})$  e  $z(s_{\alpha}+h)$  representam os valores das realizações do processo espacial localizadas nos pontos  $s_{\alpha}$  e  $s_{\alpha}+h$  (ou, seja os valores das amostras em  $s_{\alpha}$  e  $s_{\alpha}+h$ ), e  $\hat{m}_{\cdot h}$  e  $\hat{m}_{+h}$  representam os estimadores da média das observações localizadas à distância  $|h|$  de um dado ponto da região de estudo.

Conforme referido no sub-capítulo 5.1, o estimador do variograma é geralmente preferido ao estimador da covariância porque em processos aleatórios não-estacionários (com deriva) a continuidade espacial não é detectada pela covariância (visto que a variância não é finita) mas sim pela distância estrutural que é função dos acréscimos espaciais da função aleatória. Perante isto, o estimador do variograma assume-se como a ferramenta geoestatística para caracterizar e quantificar a estrutura de continuidade espacial.

Para estimar os valores experimentais do variograma,  $\hat{\gamma}(h)$ , são definidas direcções, tolerâncias angulares e classes de distância para o vector  $h$ , tomando em consideração que deve haver um compromisso entre a direcção e tolerância angular do vector, de forma a não anular evidências de anisotropia e entre o comprimento e tolerância angular do vector, de forma a existirem pares de observações suficientes em cada classe de distâncias, dotando-as de significado estatístico.

Os valores experimentais obtidos no variograma (Eq. 5.1), servem de referência para o ajustamento de uma curva média definida por uma função contínua que permite deduzir valores do variograma para qualquer distância  $h$ .

A modelação do variograma está limitada a um conjunto restrito de funções teóricas (ou combinações lineares dessas funções teóricas) que verificam a condição de positividade (função de covariância definida positiva). É esta condição que garante não só que existe solução para o sistema de krigagem (estimador geoestatístico) e que esta é única, mas também que a solução do sistema é estável. A prática geoestatística tem mostrado que existe um número relativamente pequeno de funções definidas positivas

que cobrem a generalidade das situações de dispersão de fenómenos espaciais nas ciências da Terra e do Ambiente (Soares, 2000): o modelo de efeito-pepita (Eq. 5.3), o modelo esférico (Eq. 5.4), o modelo exponencial (Eq. 5.5), o modelo gaussiano (Eq. 5.6) e o modelo de potência (Eq. 5.7):

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 & h \neq 0 \end{cases}$$

**Eq. 5.3**

$$\gamma(h) = \begin{cases} C_0 + C_1 \cdot \left[ 1.5 \cdot \frac{h}{a} - 0.5 \cdot \left( \frac{h}{a} \right)^3 \right] & h \leq a \\ C_0 + C_1 & h > a \end{cases}$$

**Eq. 5.4**

$$\gamma(h) = C_0 + C_1 \cdot (1 - e^{-3h/a})$$

**Eq. 5.5**

$$\gamma(h) = C_0 + C_1 \cdot (1 - e^{-3h^2/a^2})$$

**Eq. 5.6**

$$\gamma(h) = C_0 + C_1 \cdot h^w \quad 0 < w < 2$$

**Eq. 5.7**

onde  $h$  representa a distância absoluta,  $a$  representa a amplitude (valor da distância a partir da qual os pares de observações deixam de estar correlacionados entre si),  $C_0$  representa o efeito-pepita (descontinuidade na origem do modelo causada por factores como erros de amostragem ou variabilidade à pequena escala), e  $C_0 + C_1$  representa o patamar (limite superior para o qual tendem os valores do variograma com o aumento dos valores de  $h$ ).

Estes modelos estão expressos na sua forma isotrópica (assumem a mesma forma em todas as direcções) e podem ser classificados de acordo com o seu comportamento na origem ou no modo como se comportam quando  $h$  tende para infinito (com ou sem limite superior). Na origem, os modelos podem ter um comportamento parabólico (modelo gaussiano) ou linear (modelos exponencial ou esférico) ou ainda descontínuo (modelo de efeito-pepita), enquanto que no modelo de potência o comportamento junto da origem, é função do valor do expoente  $w$  admitido (andamento linear quando  $w$  é

igual a 1, e um comportamento parabólico quando  $w$  tende para 2). Relativamente ao modo como se comportam quando  $h$  tende para infinito, o modelo de potência não atinge um patamar (não tem variância finita), enquanto que os restantes modelos (Eq. 5.3 a Eq. 5.6) são limitados superiormente (também designados por modelos de transição), embora no caso exponencial e no gaussiano, atinjam o patamar de modo assintótico.

No caso da estrutura de continuidade espacial variar com a direcção (fenómeno anisótropo) é necessário modelar a anisotropia utilizando funções que dependem da direcção do vector  $h$  (para discussão detalhada sobre os modelos de anisotropia consultar Isaaks e Srivastava, 1989).

Para estruturas de continuidade onde uma função não é suficiente para modelar os valores experimentais, é possível embricar as funções apresentadas, em  $L+1$  combinações lineares, beneficiando da propriedade dos modelos definidos positivos em que qualquer combinação linear de modelos definidos com coeficientes positivos é ainda um modelo definido positivo (Goovaerts, 1997):

$$\gamma(h) = \sum_{i=0}^L C_i \cdot \gamma_i(h) \quad C_i \geq 0$$

onde  $C_i$  representa o patamar ou coeficiente positivo da função  $\gamma_i(h)$ .

## 5.2.2 Variáveis da saúde

### 5.2.2.1 Análise da continuidade espacial

Na análise descritiva das variáveis da saúde detectaram-se alguns sinais de anisotropia. Neste sentido foram experimentadas várias classes de ângulos e de distância para recolha das amostras a considerar na análise, nomeadamente direcções com intervalos de 15°, distâncias entre os 1000 e 6000 metros (com intervalos de 500 metros) e tolerâncias angulares até 60°. Foram também estimados variogramas omnidireccionais, com classes de distância a variar entre os 1000 e 8000 metros. No entanto não foi possível modelar nenhuma anisotropia por causa da disposição irregular das amostras na área de estudo, caracterizada por um número reduzido de observações em várias direcções e por amostras dispostas em aglomerados, onde amostras próximas entre si

têm valores de incidência muito diferentes. Assim, apenas foi possível interpretar os variogramas assumindo que a estrutura de continuidade espacial é igual em todas as direcções (ou isotropa).

Durante o processo experimental procedeu-se à análise da variabilidade local dos dados ao longo da área de estudo, com o objectivo de detectar a presença de uma forma particular de heterocedasticidade (designação dada a uma situação em que a variância depende da localização), designada por efeito proporcional, onde a variância local varia de modo proporcional à média local. Segundo vários autores (Isaaks e Srivastava, 1989; Goovaerts, 1997; Soares, 2000), a presença de um efeito proporcional, quando combinado com amostras dispostas em aglomerados de valores elevados pode enviesar o estimador do variograma (Goovaerts, 1997), dado que os valores da média e da variância função da distância,  $|h|$ , tendem a decrescer com o aumento de  $|h|$ , o que num caso limite pode gerar um variograma representativo de um fenómeno não estruturado no espaço. Isaaks e Srivastava (1989) e Goovaerts (1997) abordam nestes casos, algumas alternativas ao estimador do variograma clássico para detectar estruturas de continuidade subjacentes ao fenómeno, como por exemplo o cálculo de variogramas relativos (locais, globais ou emparelhados) ou a utilização de outros estimadores do variograma, como o variograma mediano.

A presença do efeito proporcional foi analisada com diagramas de dispersão das variâncias e médias locais, por aplicação do método das janelas móveis, que consiste na divisão da área de estudo em janelas de dimensões iguais, onde são estimadas a média e variância para cada janela. Experimentaram-se várias dimensões de janela, de modo a garantir um número mínimo de observações admitidas para estimar as médias e variâncias locais, e optou-se pela definição de janelas com 20 Kms de lado, a cada 5 Kms (com sobreposição das janelas), em que cada uma incluiu entre 8 a 30 observações. Os resultados obtidos e representados nos gráficos do Anexo B, evidenciam a presença de um efeito proporcional nas variáveis ash\_13, ash\_6, ecz\_13, hay\_13, rash\_6, pnose\_13 e whez\_13, onde os parâmetros estimados de  $\rho$  e  $\rho_{\text{rank}}$  (apresentados nos gráficos das alienas .a do anexo), representam respectivamente o coeficiente de correlação linear de Pearson e o de Spearman (o último é menos sensível a valores extremos). Dado o pequeno número de pontos dos diagramas de dispersão de hay\_6 e whez\_6, os resultados destas variáveis não são conclusivos.

Calcularam-se os diagramas de dispersão dos estimadores da média e da variância em função da distância,  $|h|$ , que se apresentam no mesmo anexo (gráficos das alienas .b do Anexo B). Para permitir a comparabilidade dos gráficos, as estatísticas de cada variável foram estandardizadas respectivamente pela sua média e variância global.

Observa-se que as estimativas da média e da variância ao longo de  $|h|$  flutuam em torno da unidade (de modo errático no caso da variância), pelo que se considerou que o efeito proporcional observado não provoca um enviesamento dos estimadores do variograma. Como resultado desta análise, optou-se por avançar para a inferência do estimador de variograma clássico para caracterização da continuidade espacial das variáveis da saúde.

### 5.2.2.1.1 Asma

A observação da distribuição geográfica das amostras permitiu identificar, em ambos os grupos de idade, áreas onde aglomerados de escolas apresentam valores bastante discordantes entre si. Para a variável ash\_13 (Figura 5.1) identificou-se uma estrutura de continuidade espacial que foi ajustada um modelo esférico com uma amplitude de 6800 metros (os parâmetros do modelo encontram-se do lado direito da figura). Isto significa que se admite que, em média, abaixo dos 6800 metros de distância entre pontos amostrados, os valores observados estão correlacionados entre si. O valor assumido pelo patamar ( $C_0+C_1$ ) refere-se à variância da amostra, e é representado na figura seguinte por uma linha vermelha paralela ao eixo das abcissas.

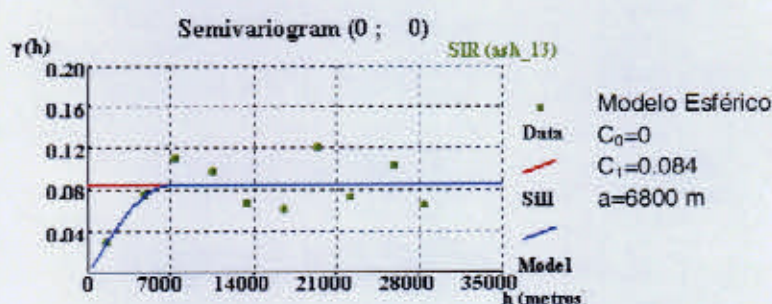


Figura 5.1- Variograma experimental e ajustamento de modelo esférico (ash\_13)

Na variável ash\_6 observa-se que a continuidade espacial contém duas estruturas de continuidade imbricadas, detectadas a diferentes escalas. Assim, optou-se por ajustar

uma combinação linear de dois modelos esféricos sem efeito pepita, conforme Figura 5.2.

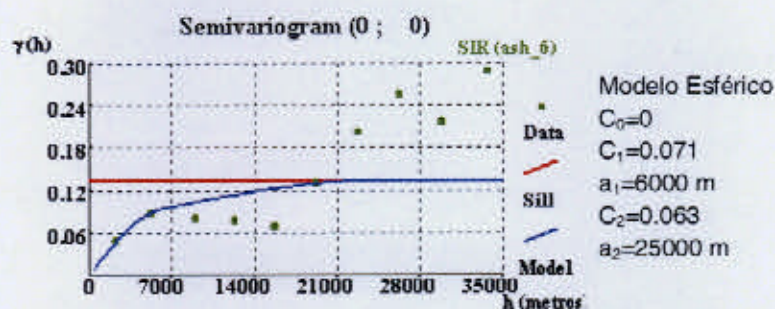


Figura 5.2- Variograma experimental e combinação linear de modelos esféricos (ash\_6)

À primeira estrutura foi ajustado um modelo com uma amplitude de 6000 metros (equivalente ao modelo do variograma ajustado para a variável ash\_13) e uma contribuição de 53% da variância total e, à segunda estrutura ajustou-se um modelo com uma amplitude de 25000 metros e uma contribuição de 47% da variância total.

### 5.2.2.1.2 Febre dos fenos

Os resultados obtidos para hay\_13 evidenciaram uma estrutura de continuidade espacial, à qual foi ajustada uma função exponencial com uma amplitude de 4700 metros e sem efeito pepita.

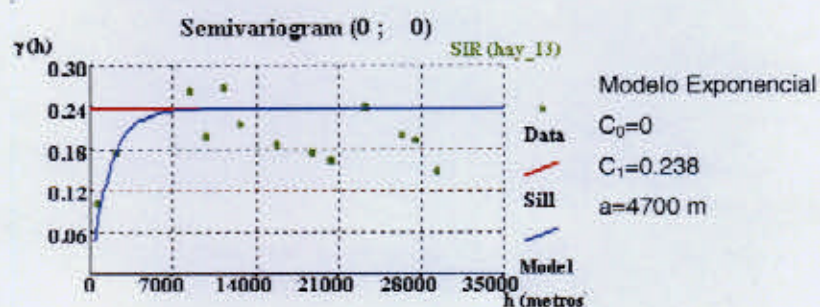


Figura 5.3- Variograma experimental e ajustamento de modelo esférico (hay\_13)

A variável hay\_6, composta por um número reduzido de observações (17) dispostas de modo irregular e em aglomerados com valores de incidência muito diferentes, não foi possível detectar uma estrutura de continuidade espacial.

### 5.2.2.1.3 Comichões com lesões de pele

Os resultados da análise exploratória da variografia omnidireccional da variável rash\_6 não revelaram qualquer estrutura de continuidade espacial. No caso da variável rash\_13, foram removidos dois pares de pontos, que originavam valores anômalos no variograma experimental em classes de distância próximas da origem. Após este procedimento, os resultados do variograma experimental permitiram identificar uma estrutura de continuidade espacial que foi ajustada um modelo exponencial sem efeito pepita e com uma amplitude de 4200 metros, conforme representado na Figura 5.4.

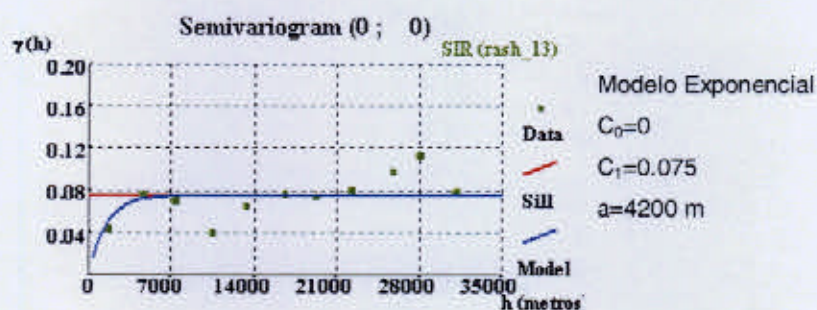


Figura 5.4- Variograma experimental e ajustamento de modelo exponencial (rash\_13)

### 5.2.2.1.4 Crises de espirros ou de pingos no nariz

Os resultados da análise exploratória da variografia omnidireccional da variável pnose\_6 não revelaram qualquer estrutura de continuidade espacial. No caso da variável pnose\_13, os resultados do variograma experimental permitiram identificar uma estrutura de continuidade espacial que foi ajustada por um modelo esférico sem efeito pepita e com uma amplitude de 3500 metros, conforme representado na Figura 5.5.

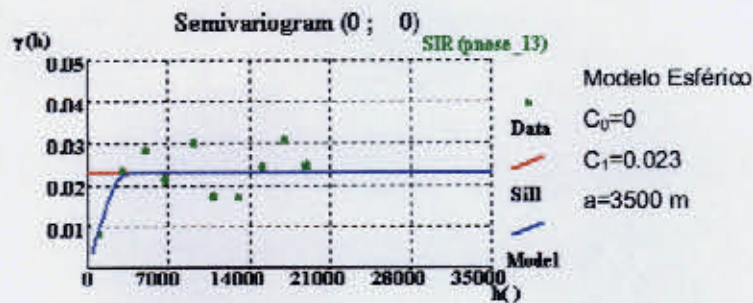


Figura 5.5- Variograma experimental e ajustamento de modelo esférico (pnose\_13)

### 5.2.2.2 Síntese dos resultados obtidos

Para além das variáveis hay\_6, rash\_6 e pnose\_6, também os resultados dos variogramas experimentais da pieira (whez\_13 e whez\_6) e eczema (ecz\_13 e ecz\_6) não revelam qualquer evidência de estrutura espacial. Perante isto, optou-se por continuar o trabalho apenas com as variáveis que apresentaram boas estruturas de continuidade espacial e às quais foi possível ajustar um modelo teórico de variograma - ash\_13, ash\_6, hay\_13, rash\_13 e pnose\_13 -, permitindo avançar para os processos de estimação e simulação. Na Tabela 5.1 estão identificados os parâmetros dos modelos ajustados considerados.

Variável	Modelo	$C_0$	$C_1$	$a_1$	$C_2$	$a_2$
Ash_13	Esférico	0	0.084	6800	-	-
Ash_6	Esférico	0	0.071	6000	0.063	25000
Hay_13	Exponencial	0	0.238	4700	-	-
Hay_6	-	-	-	-	-	-
Rash_13	Exponencial	0	0.075	4200	-	-
Rash_6	-	-	-	-	-	-
Pnose_13	Esférico	0	0.023	3500	-	-
Pnose_6	-	-	-	-	-	-
Whez_13	-	-	-	-	-	-
Whez_6	-	-	-	-	-	-
Ecz_13	-	-	-	-	-	-
Ecz_6	-	-	-	-	-	-

Tabela 5.1- Modelos de variograma e respectivos parâmetros, ajustados às variáveis de saúde

### 5.2.3 Variável da qualidade do ar

O conjunto de dados recolhidos sobre  $PM_{10}$  é composto por um número reduzido de observações no espaço (8) e um número abundante de observações no tempo (5970).

Este facto é importante para a análise da estrutura de continuidade espacial, na medida em que se admite uma forte dependência da localização espacial e temporal dos valores observados. No entanto, no âmbito deste estudo, as continuidades espacial e temporal são analisadas pela decomposição das covariâncias espacio-temporais em componentes espaciais e temporais isoladas.

Neste estudo admite-se que a estrutura de continuidade espacial é igual em todas as direcções (ou omnidireccional), dado que o número reduzido de observações (no espaço) é inadequado para a detecção de anisotropia.

### 5.2.3.1 Análise da continuidade temporal das PM<sub>10</sub>

Conforme referido em 5.2.1, a estrutura de continuidade temporal das PM<sub>10</sub> é caracterizada, pelo estimador do variograma. Este estimador,  $\gamma_i(t)$ , é definido por um comportamento médio que correlaciona os valores medidos para cada estação em momentos desfasados no tempo ( $t$ ), definido pela expressão seguinte.

$$\gamma_i(t) = \frac{1}{2N(t)} \cdot \sum_{j=1}^{N(t)} [z(s, t_j) - z(s, t_j + t)]^2$$

onde  $z(s, t_j)$  representa o valor de PM<sub>10</sub> medido na estação  $s$  no instante  $t_j$  e  $N(t)$  representa o número de pares de pontos separados por um período de tempo  $t$ .

Para comparar os resultados optou-se por eliminar os efeitos de localização e de escala, transformando os valores de cada estação em valores standardizados,  $z(s, t)^*$ , de média igual a 0 e desvio padrão igual a 1:

$$z(s, t)^* = \frac{z(s, t) - m(s)}{\sigma(s)}$$

onde  $z(s, t)$  representa o valor de PM<sub>10</sub> medido na estação  $s$  num instante  $t$ , e  $m(s)$  e  $\sigma(s)$  representam respectivamente a média e desvio padrão dos valores medidos na estação  $s$ . Na Figura 5.6 apresentam-se os modelos exponenciais standardizados (patamar igual à unidade) obtidos. No Anexo C apresenta-se o ajustamento dos modelos aos variogramas e respectivos parâmetros estimados para cada estação.

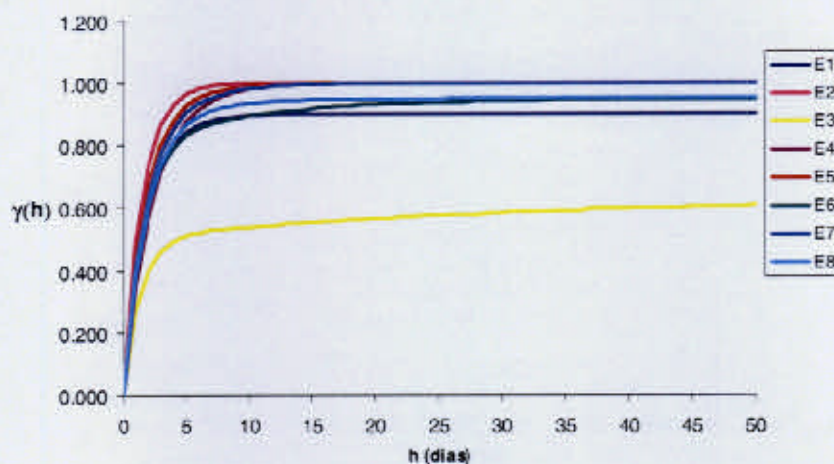


Figura 5.6- Modelos temporais ajustados, das concentrações médias diárias de PM<sub>10</sub>

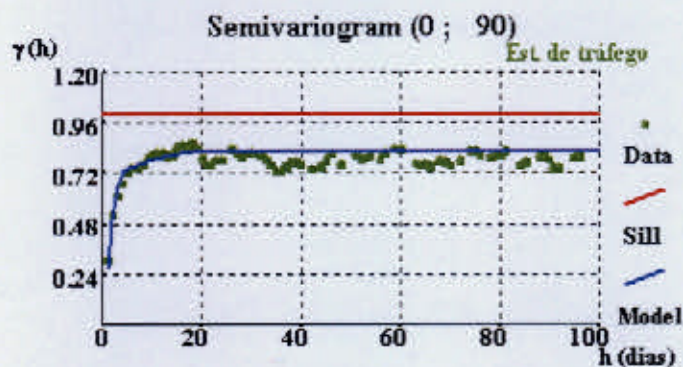
Os resultados evidenciam em todas as estações, elevadas correlações que se caracterizam pela presença de duas estruturas a duas escalas: uma primeira estrutura à pequena escala, com amplitudes que variam entre 4 a 6 dias, e uma segunda estrutura de menor influência, cuja amplitude varia entre 10 dias (E2) e 16 dias (E8). As estações E1 e E3 não atingem, em média, a variância dos dados, durante o período representado (desfasamentos até 50 dias). Verifica-se na Figura 4.9 que estas estações apresentam um elevado número de altas frequências, que têm grande influência no estimador da variância.

Os resultados sugerem diferentes comportamentos das estruturas de continuidade temporal nas estações de tráfego e nas estações de fundo. Para quantificar as diferenças, considerou-se em seguida a estrutura de continuidade temporal das estações de tráfego e de fundo separadamente, através do cálculo da média dos variogramas temporais ( $\gamma$ ) de cada um dos tipos de estação (tráfego/fundo), definida por:

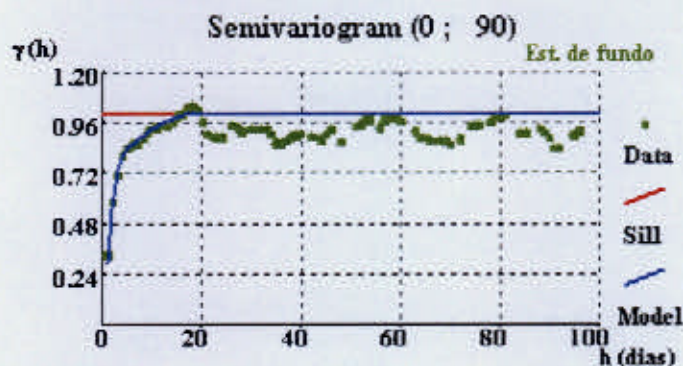
$$\gamma(t) = \frac{1}{2Nt(t)Ns} \cdot \sum_{i=1}^{Ns} \sum_{j=1}^{Nt(t)} [z(s_i, t_j) - z(s_i, t_j + t)]^2$$

onde  $z(s_i, t_j)$  representa o valor de PM<sub>10</sub> medido na estação  $s_i$  no instante  $t_j$  e  $Ns$  representa o número de estações de medição consideradas no cálculo. Os variogramas experimentais obtidos para ambos os tipos de estação apresentam-se na Figura 5.7. Para permitir a sua comparabilidade procedeu-se à transformação de dados originais

em dados estandardizados e optou-se por representar os modelos de ajustamento, com o objectivo de quantificar as diferenças encontradas.



1. Modelo esférico,  $c_0=0$ ;  $c_1=0.69$ ;  $a_1=4$ ;  $c_2=0.13$ ;  $a_2=19$



2. Modelo esférico,  $c_0=0$ ;  $c_1=0.73$ ;  $a_1=4$ ;  $c_2=0.27$ ;  $a_2=18$

Figura 5.7- Variogramas temporais das concentrações médias diárias de  $PM_{10}$  nas estações de tráfego (1.) e de fundo (2.)

Os variogramas obtidos apresentam duas estruturas a duas escalas e foram modelados sem efeito pepita, por combinação linear de dois modelos esféricos, cujos valores dos respectivos parâmetros se encontram indicados por baixo dos gráficos. Verificou-se que os modelos apresentam estruturas muito semelhantes em termos de continuidade temporal e apenas diferem na contribuição da variância para cada uma das estruturas.

### 5.2.3.2 Análise da continuidade espacial e temporal das $PM_{10}$

Na descrição espacial das  $PM_{10}$  (4.2.3.3) identificam-se indícios de heterocedasticidade espacial devido à dispersão de valores extremos na área de estudo. No entanto essa evidência baseou-se apenas na análise dos valores médios observados em cada uma

das estações, e não considerou as correlações espaciais entre as estações, medidas ao longo do tempo para cada instante de  $t$ . Para as considerar analisou-se a continuidade espacial, definida de forma análoga à da continuidade temporal, onde o variograma médio espacial ( $\gamma_s$ ) é calculado através da média dos variogramas espaciais em cada instante  $t$  do domínio temporal:

$$\gamma_s(h) = \frac{1}{2N_s(h)N_t} * \sum_{j=1}^{N_t} \sum_{i=1}^{N_s(h)} [Z(x_i, t_j) - Z(x_i + h, t_j)]^2$$

onde  $N_s(h)$  representa o número de pares de estações separadas pela distância  $h$  e  $N_t$  representa o número de observações no tempo consideradas. Apresentam-se na Figura 5.8 os resultados obtidos com este estimador.

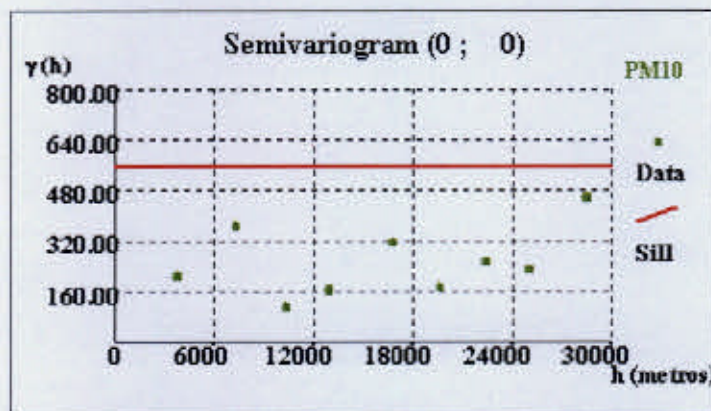


Figura 5.8- Variograma espacial das concentrações médias diárias de PM10

O variograma obtido a partir do conjunto global das 5970 observações, reflecte a presença de um fenómeno não estruturado no espaço, caracterizado por um conjunto de observações recolhidas num domínio espacial onde as covariâncias (espaciais) entre estações dependem da localização, violando desta forma a hipótese das covariâncias entre estações dependerem apenas da distância entre si.

Outras razões podem ser apontadas para justificar este resultado, como o facto do variograma experimental ser produto de uma média de variogramas espaciais calculados ao longo do tempo, ignorando a hipótese de existirem períodos do domínio temporal homocedásticos no espaço (em que as covariâncias espaciais dependem apenas da distância entre si); ou pelo facto de existirem valores anómalos dispersos na área de estudo, que provocam variâncias locais elevadas; ou simplesmente por não existir correlação espacial (o que é difícil de explicar num fenómeno desta natureza).

Outra hipótese ainda é considerar que a análise exploratória com base no variograma dos valores originais pode não ser o melhor instrumento para a detecção da continuidade espacial deste fenómeno. Enquanto que na descrição espacial das  $PM_{10}$  se sublinhou a existência de médias distintas na área de estudo, a descrição bivariada das  $PM_{10}$  deu indícios da existência de continuidade espacial ao longo do tempo (ver Figura 4.11 e Figura 4.12), o que sugere a presença de uma estrutura de continuidade subjacente ao fenómeno, não detectada pelo variograma global da Figura 5.8. Serpa (2003) exemplifica um caso típico, onde variáveis originais caracterizadas por médias heterogéneas e com elevadas correlações no tempo, podem resultar em variogramas sem estrutura de continuidade.

Conforme já referido no sub-capítulo 5.2.2.1, Isaaks e Srivastava (1989) sugerem nestes casos outras alternativas como os estimadores de variogramas relativos (locais, globais ou emparelhados), ou de variogramas medianos, ou ainda a transformação de variáveis com vista a extrair informação sobre a estrutura espacial subjacente ao fenómeno. No âmbito deste estudo foram avaliadas as alternativas propostas por Isaaks e Srivastava, e optou-se por transformar a variável  $PM_{10}$  em cada estação, numa nova variável,  $z(s,t)$ , de média igual a zero e desvio padrão igual a um:

$$z(s,t) = \frac{z(s,t) - m(s)}{\sigma(s)}$$

onde  $z(s,t)$  representa o valor de  $PM_{10}$  medido na estação  $s$  num instante  $t$ , e  $m(s)$  e  $\sigma(s)$  representam respectivamente a média e desvio padrão dos valores medidos na estação  $s$ . O variograma espacial da variável transformada apresenta-se na figura seguinte (Figura 5.9).

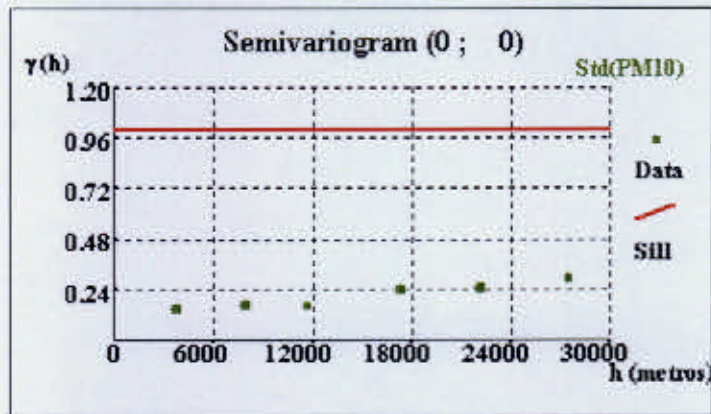


Figura 5.9- Variograma espacial das concentrações médias diárias de  $PM_{10}$  estandardizadas

Ao retirar o efeito da média e variância locais, o resultado exhibe a presença de continuidade espacial de larga escala. Esta evidência é, no entanto, atenuada pelo facto do variograma experimental ser produto de uma média de variogramas espaciais calculados ao longo do tempo, o que provoca um alisamento do variograma experimental.

Para detectar eventuais diferenças nas estruturas de continuidade espacial provocadas pelas oscilações dos valores observados ao longo do tempo nas estações de medição (referidas no sub-capítulo 4.2.3.1), representam-se no gráfico da Figura 5.10 os coeficientes de correlação linear de Pearson entre as estações de medição, calculados para janelas de 30 dias ( $t=30$ ), no período de Fevereiro de 2002 a Dezembro de 2003. O número mínimo de pares de observações admitido por janela é 15, razão pela qual foram retiradas as observações de 2001 caracterizadas por muitos valores omissos em várias estações. As colunas estão ordenadas de forma ascendente, da esquerda para a direita, em função da distância entre as estações de medição, de modo a tornar mais evidente os períodos do domínio temporal onde as covariâncias espaciais entre estações são mais dependentes da distância entre si.

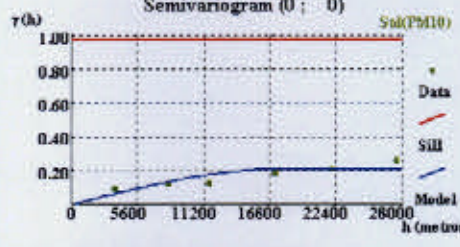
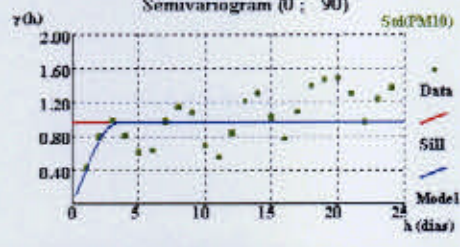
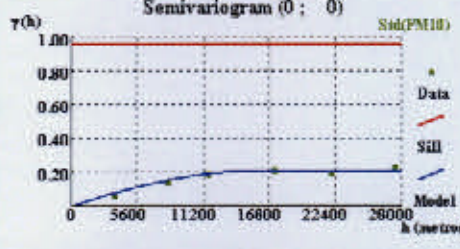
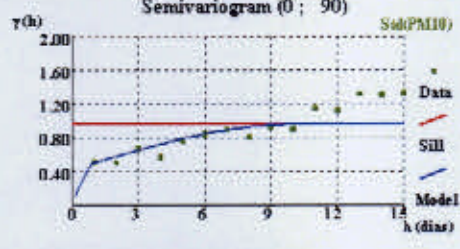
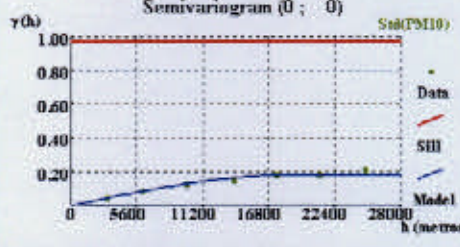
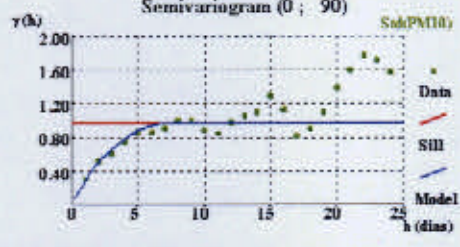
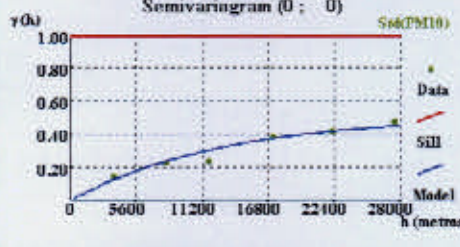
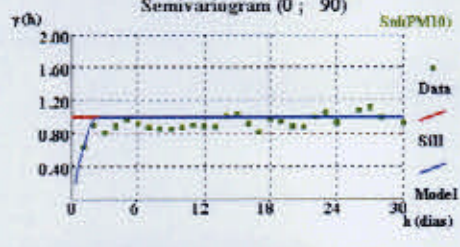


**Figura 5.10- Coeficientes de correlação linear de Pearson entre estações de medição ao longo do tempo por períodos: A – de 15/08/2003 a 31/12/2003; B – 30/04/2003 a 15/08/2003; C – 11/03/2003 a 30/04/2003; D - 09/02/2003 a 11/03/2003; E – de 17/01/2003 a 09/02/2003; F – de 17/12/2002 a 17/01/2003; G – de 02/09/2002 a 17/12/2002; H – de 14/07/2002 a 02/09/2002; I – de 25/04/2002 a 14/07/2002; J – de 24/02/2002 a 25/04/2002**

As letras A, B, C, D, E, F, G, H, I e J representadas no gráfico identificam intervalos ou períodos do tempo onde se considera que as correlações lineares não sofrem oscilações e por conseguinte o padrão de continuidade espacial ou ausência dele se mantém possivelmente constante. Para quantificar as estruturas dos intervalos de tempo identificados, calcularam-se os variogramas (experimentais) espaciais e temporais obtidos para vários períodos entre Fevereiro de 2002 a Dezembro de 2003 (Figura 5.11). Os modelos escolhidos para o ajustamento dos variogramas espaciais e

temporais, encontram-se sobrepostos aos variogramas experimentais com indicação dos respectivos parâmetros considerados. Para os casos em que não se detectaram evidências da presença de uma estrutura espacial, apresentam-se apenas os variogramas experimentais, sem ajustamento de modelo teórico (espacial e temporal).

Período	Variograma espacial	Variograma temporal
<b>A</b> De 15/08/03 a 31/12/03	<p>Semivariogram (0 ; 0)</p>	<p>Semivariogram (0 ; 90)</p>
	<b>1.a. Modelo esférico embricado:</b> $c_0=0; c_1=0.2; a_1=20000; c_2=0.8; a_2=1E+20$	<b>1.b. Modelo esférico embricado: <math>c_0=0;</math></b> $c_1=0.2; a_1=4; c_2=0.8; a_2=4$
<b>B</b> De 30/04/03 a 15/08/03	<p>Semivariogram (0 ; 0)</p>	<p>Semivariogram (0 ; 90)</p>
	<b>2.a. Período não estruturado no espaço</b>	<b>2.b. Período não estruturado no espaço</b>
<b>C</b> De 11/03/03 a 30/04/03	<p>Semivariogram (0 ; 0)</p>	<p>Semivariogram (0 ; 90)</p>
	<b>3.a. Modelo esférico embricado:</b> $c_0=0; c_1=0.1; a_1=7500; c_2=0.6; a_2=1E+20;$ $c_3=0.3; a_3=1E+20$	<b>3.b. Modelo esférico embricado:</b> $c_0=0; c_1=0.1; a_1=2.5; c_2=0.6; a_2=2.5;$ $c_3=0.3; a_3=23$

<p><b>D</b></p> <p>De 09/02/03 a 11/03/03</p>	<p>Semivariogram (0 ; 0)</p> 	<p>Semivariogram (0 ; 90)</p> 
	<p>4.a. Modelo esférico embricado: <math>c_0=0; c_1=0.21; a_1=18000; c_2=0.79; a_2=1E+20</math></p>	<p>4.b. Modelo esférico embricado: <math>c_0=0;</math> <math>c_1=0.21; a_1=4; c_2=0.79; a_2=4</math></p>
<p><b>E</b></p> <p>De 17/01/03 a 09/02/03</p>	<p>Semivariogram (0 ; 0)</p> 	<p>Semivariogram (0 ; 90)</p> 
	<p>5.a. Modelo esférico embricado: <math>c_0=0;</math> <math>c_1=0.2; a_1=14500; c_2=0.22; a_2=1E+20;</math> <math>c_3=0.58; a_3=1E+20</math></p>	<p>5.b. Modelo esférico embricado: <math>c_0=0; c_1=0.2; a_1=1; c_2=0.22; a_2=1;</math> <math>c_3=0.58; a_3=10</math></p>
<p><b>F</b></p> <p>De 17/12/02 a 17/01/03</p>	<p>Semivariogram (0 ; 0)</p> 	<p>Semivariogram (0 ; 90)</p> 
	<p>6.a. Modelo esférico embricado: <math>c_0=0;</math> <math>c_1=0.18; a_1=19000; c_2=0.82; a_2=1E+20</math></p>	<p>6.b. Modelo esférico embricado: <math>c_0=0; c_1=0.18; a_1=2; c_2=0.82; a_2=7</math></p>
<p><b>G</b></p> <p>De 02/09/02 a 17/12/02</p>	<p>Semivariogram (0 ; 0)</p> 	<p>Semivariogram (0 ; 90)</p> 
	<p>7.a. Modelo esférico embricado: <math>c_0=0; c_1=0.5; a_1=38470; c_2=0.5; a_2=1E+20</math></p>	<p>7.b. Modelo esférico embricado: <math>c_0=0; c_1=0.5; a_1=2.5; c_2=0.5; a_2=2.5</math></p>

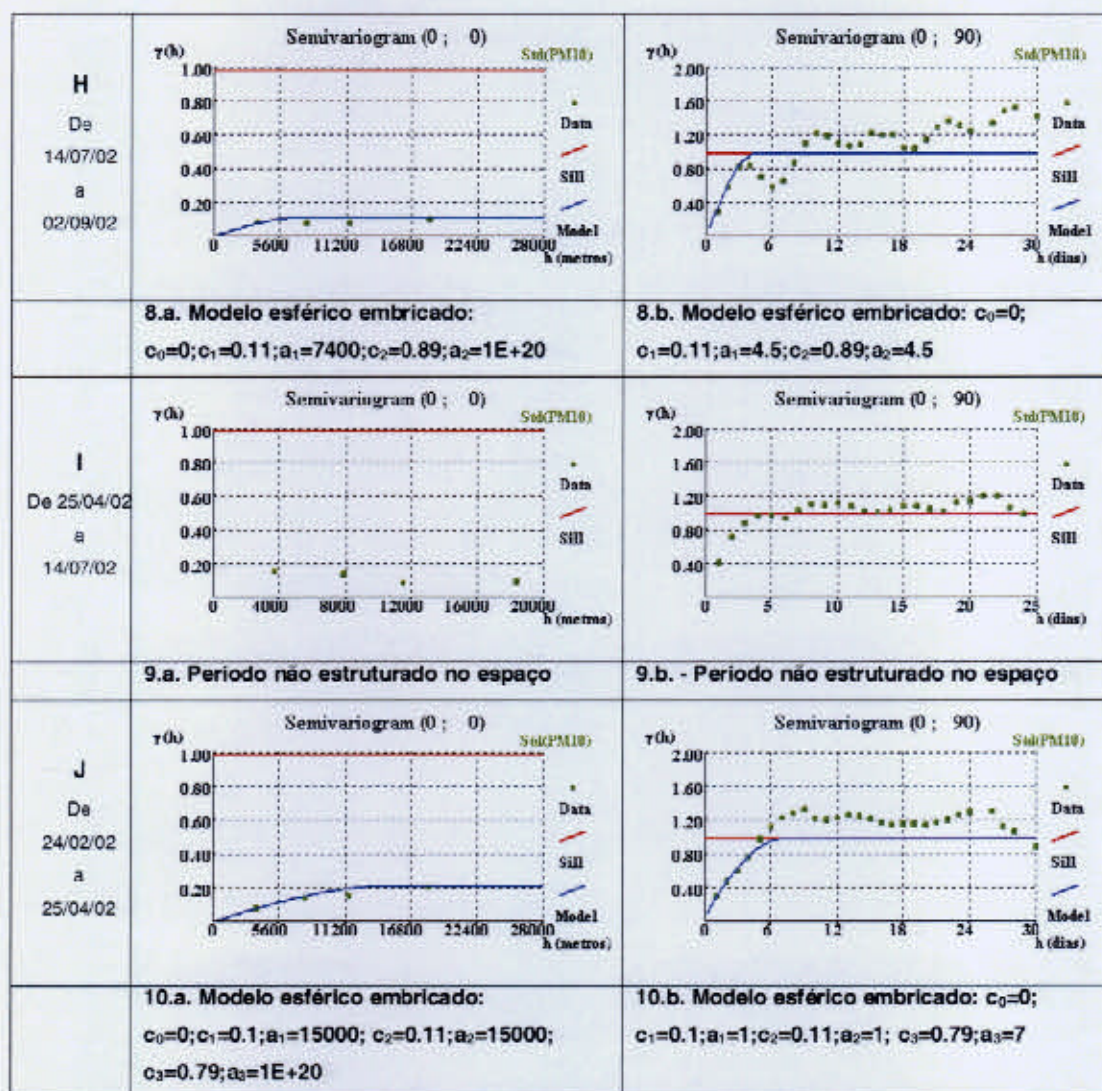


Figura 5.11- Variogramas espaciais (a) e temporais (b) das concentrações médias diárias de  $PM_{10}$  estandardizadas, ao longo do tempo por períodos: A – de 15/08/2003 a 31/12/2003; B – 30/04/2003 a 15/08/2003; C – 11/03/2003 a 30/04/2003; D - 09/02/2003 a 11/03/2003; E – de 17/01/2003 a 09/02/2003; F – de 17/12/2002 a 17/01/2003; G – de 02/09/2002 a 17/12/2002; H – de 14/07/2002 a 02/09/2002; I – de 25/04/2002 a 14/07/2002; J – de 24/02/2002 a 25/04/2002

Os variogramas experimentais dos períodos A, C, D, E, F, G, H e J exibem a presença de continuidade espacial de larga escala (não atingem o patamar no domínio espacial finito definido). Ajustaram-se modelos esféricos aos variogramas experimentais e ajustaram-se o mesmo número de estruturas em ambos os domínios (espacial e temporal) de modo a cumprir a imposição para a solução do sistema de Krigagem para estimação ou simulação de fenômenos anisotrópicos, visto que se identifica a presença de anisotropia zonal nos períodos considerados.

### 5.2.3.3 Síntese dos resultados obtidos

De um modo geral, a variografia temporal das  $PM_{10}$  evidenciou uma primeira estrutura à pequena escala, com amplitudes de 4 a 6 dias, e uma segunda estrutura, de menor influência com amplitudes de 10 a 16 dias. Não se encontraram diferenças significativas entre as estruturas temporais das estações de fundo e de tráfego.

Os resultados obtidos na variografia espacial evidenciaram estruturas de continuidade, quando retirados os efeitos da média e da variância local, para intervalos ou períodos do tempo homogêneos no espaço. Na Figura 5.12 representam-se os modelos esféricos standardizados (patamar igual à unidade) ajustados para os períodos de tempo A, C, D, E, F, G, H e J, que exibiram a presença de continuidade espacial de larga escala.

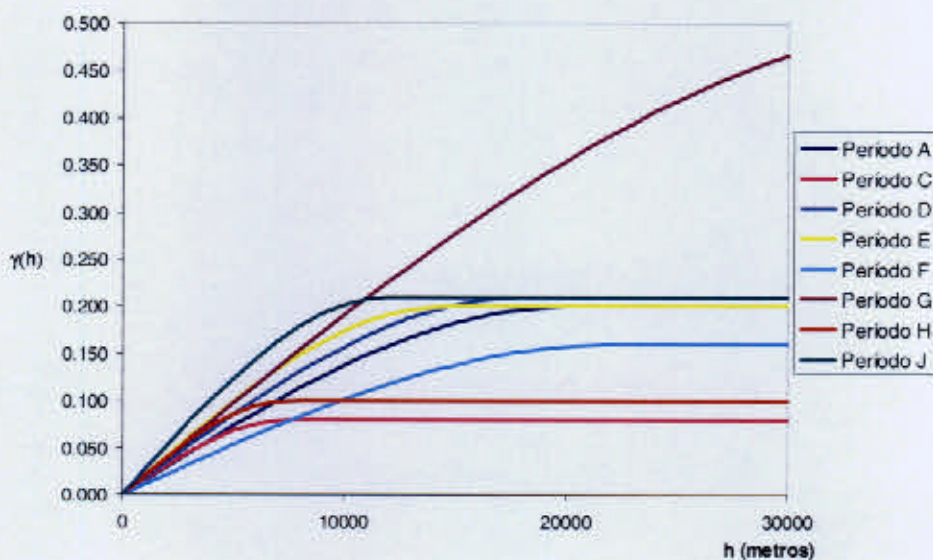


Figura 5.12- Modelos esféricos obtidos para caracterizar a variografia espacial

Os modelos de variogramas espacio-temporais apresentam uma anisotropia zonal, dado que a componente espacial se apresenta muito contínua, ao contrário da componente temporal, isto é, os patamares dos variogramas temporais atingem os valores das variâncias *a priori* das amostras em amplitudes relativamente pequenas (10 a 16 dias), enquanto que os patamares dos variogramas espaciais nunca alcançam os valores das variâncias *a priori* das amostras no domínio espacial em estudo.

Perante os resultados obtidos, optou-se por prosseguir para os processos de e simulação considerando apenas os períodos que exibem a presença de um padrão de continuidade espacial representada nos modelos ajustados, conforme Figura 5.12. Assume-se que, neste caso, a variografia obtida com recurso à transformação da variável é um bom estimador da estrutura espacial subjacente ao fenómeno.

### 5.3 Inferência espacial

#### 5.3.1 Introdução aos métodos de estimação geoestatísticos

Os modelos de dependência espacial definidos no sub-capítulo 5.2, caracterizam e quantificam o modo como cada uma das variáveis da saúde e da qualidade do ar se correlaciona em função do espaço e do tempo. Para estimar o valor das variáveis em qualquer ponto do espaço não observado, utilizou-se o estimador linear de Krigagem,  $Z^*(s)$ , definido por:

$$Z^*(s) - m(s) = \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}(s) \cdot [Z(s_{\alpha}) - m(s_{\alpha})]$$

onde  $\lambda_{\alpha}(s)$  representa o ponderador do peso atribuído a  $z(s_{\alpha})$ , realização da variável aleatória  $Z(s_{\alpha})$ ,  $m(s)$  e  $m(s_{\alpha})$  representam os valores esperados das variáveis aleatórias  $Z^*(s)$  e  $Z(s_{\alpha})$ , e  $n(s)$  representa o número de valores observados envolvidos na estimativa. Este estimador é um estimador de variância mínima, isto é, é escolhido de forma a minimizar o erro (ou variância) de estimação ( $\sigma_{\varepsilon}^2$ ) entre  $Z^*(s)$  e  $Z(s)$ ,

$$\sigma_{\varepsilon}^2(s) = \text{Var}[Z^*(s) - Z(s)]$$

sujeito à restrição de não enviesamento de  $Z^*(s)$ , isto é:

$$E[Z^*(s) - Z(s)] = 0$$

As variantes dos algoritmos do estimador linear de Krigagem dependem do modelo adoptado para a função aleatória  $Z(s)$ . Neste trabalho, o modelo considerado consiste em decompor a função numa componente residual,  $R(s)$  e numa componente de tendência ou deriva,  $m(s)$ :

$$Z(s) = R(s) + m(s)$$

onde se assume que  $R(s)$  é função aleatória estacionária de média zero e covariância a depender apenas do vector  $h$ , e o valor da componente de tendência é o valor esperado da variável aleatória  $Z(s)$ ,  $E\{Z(s)\} = m(s)$ . Neste cenário, as variantes do algoritmo do estimador linear geoestatístico, podem assumir três formas diferentes (Goovaerts, 1997), que dependem apenas do modelo considerado para a componente da tendência. Um primeiro cenário passa por admitir que a média é conhecida e constante na área de estudo, onde  $m(s)$  é uma componente estacionária e de valor conhecido,  $m(s) = m$ . O estimador linear de Krigagem pode rescrever-se da seguinte forma:

$$\begin{aligned} Z_{SK}^*(s) &= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{SK}(s) \cdot [Z(s_{\alpha}) - m] + m \\ &= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{SK}(s) \cdot [Z(s_{\alpha})] - \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{SK}(s) \cdot m + m \\ &= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{SK}(s) \cdot [Z(s_{\alpha})] + m \cdot \left[ 1 - \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{SK}(s) \right] \end{aligned}$$

e que se designa por estimador linear de Krigagem Simples. Para minimizar o erro de estimação do estimador de Krigagem Simples, não é necessário impor a restrição de não enviesamento, na medida em que este estimador já é centrado:

$$E[Z_{SK}^*(s) - Z(s)] = m - m = 0$$

A minimização do erro de variância passa então por determinar os  $n(s)$  ponderadores  $\lambda_{\alpha}^{SK}(s)$ , igualando a zero as  $n$  derivadas parciais da expressão da variância de estimação em ordem a  $\lambda_{\alpha}^{SK}(s)$ .

Outro cenário, passa por admitir que a média local varia ao longo da área de estudo e é desconhecida, mas é constante numa vizinhança admitida em torno de  $s$ . O estimador linear de Krigagem passa ser a definido por uma combinação linear de variáveis aleatórias mais uma média local constante,  $m(s)$ , e designa-se por estimador linear de Krigagem Normal:

$$\begin{aligned}
Z_{OK}^*(s) &= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{OK}(s) \cdot [Z(s_{\alpha}) - m(s)] + m(s) \\
&= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{OK}(s) \cdot [Z(s_{\alpha})] - \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{OK}(s) \cdot m(s) + m(s) \\
&= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{OK}(s) \cdot [Z(s_{\alpha})] + m(s) \cdot \left[ 1 - \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{OK}(s) \right] \\
&= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{OK}(s) \cdot [Z(s_{\alpha})] \quad \text{com} \quad \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{OK}(s) = 1
\end{aligned}$$

A restrição de não enviesamento deste estimador é assegurada desde que

$\sum_{\alpha=1}^{n(s)} \lambda_{\alpha}^{OK}(s) = 1$ , e a minimização do erro de variância passa por determinar os  $n(s)$

ponderadores  $\lambda_{\alpha}^{OK}(s)$ , igualando a zero as  $n(s) + 1$  derivadas parciais da expressão da variância de estimação em ordem aos  $\lambda_{\alpha}^{OK}(s)$  e a  $\mu$ , que representa o parâmetro de Lagrange introduzido no sistema de equações para eliminar o problema de optimização imposto pela restrição de não enviesamento do estimador.

Por último, pode considerar-se que a média local varia ao longo da área de estudo e é desconhecida, e varia de forma suave numa vizinhança admitida em torno de  $s$ . Neste caso, a componente de tendência é descrita por um modelo definido por uma combinação linear de funções conhecidas  $f_k(s')$  das coordenadas:

$$m(s') = \sum_{k=0}^K a_k(s') \cdot f_k(s')$$

onde os coeficientes  $a_k(s')$  representam constantes desconhecidas na vizinhança do ponto  $s$ . Nestas circunstâncias o estimador linear de Krigagem com modelo de deriva é definido por:

$$\begin{aligned}
Z^*(s) &= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}(s) \cdot [Z(s_{\alpha}) - m(s')] + m(s') \\
&= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}(s) \cdot \left[ Z(s_{\alpha}) - \sum_{k=0}^K a_k(s) \cdot f_k(s_{\alpha}) \right] + \sum_{k=0}^K a_k(s) \cdot f_k(s) \\
&= \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}(s) \cdot Z(s_{\alpha}) + \sum_{k=0}^K a_k(s) \cdot \left[ f_k(s) - \sum_{\alpha=1}^{n(s)} \lambda_{\alpha}(s) \cdot f_k(s_{\alpha}) \right]
\end{aligned}$$

O não enviesamento deste estimador é assegurado pelas  $K+1$  restrições seguintes:

$$\sum_{\alpha=1}^{n(s)} \lambda_{\alpha}(s) \cdot f_k(s_{\alpha}) = f_k(s) \quad k = 0, \dots, K$$

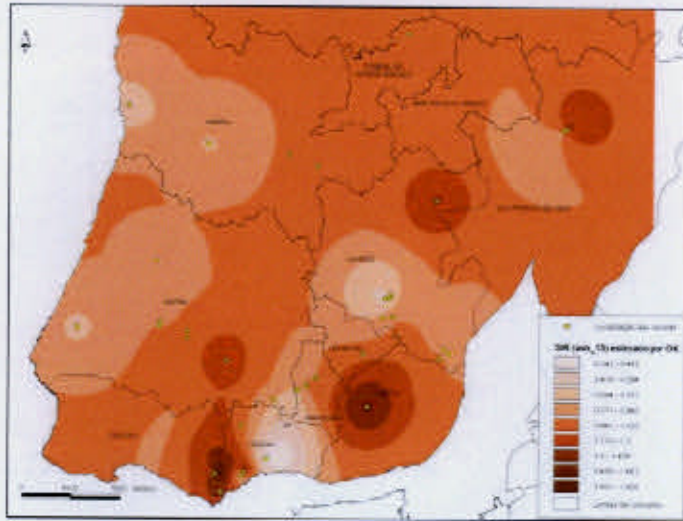
e a minimização do erro de variância passa por determinar os  $n(s)$  ponderadores  $\lambda_{\alpha}(s)$ , igualando a zero as  $n(s) + K + 1$  derivadas parciais da expressão da variância de estimação em ordem aos  $n(s)$   $\lambda_{\alpha}(s)$  e aos  $K+1$   $\mu_k(s)$  que representam os parâmetros de Lagrange introduzidos no sistema de equações para eliminar o problema de otimização imposto pela restrição de não enviesamento do estimador.

Soares (2000) e Goovaerts (1997) referem que este modelo apresenta grandes inconvenientes por ser necessário calcular o variograma dos resíduos da variável (desconhecidos) e definir *a priori* a forma das  $k$  funções de tendência  $f_k(s)$ . Os mesmos autores sugerem alguns métodos práticos para a solução destes problemas, que conduzem à utilização do estimador linear de Krigagem Normal para a inferência espacial.

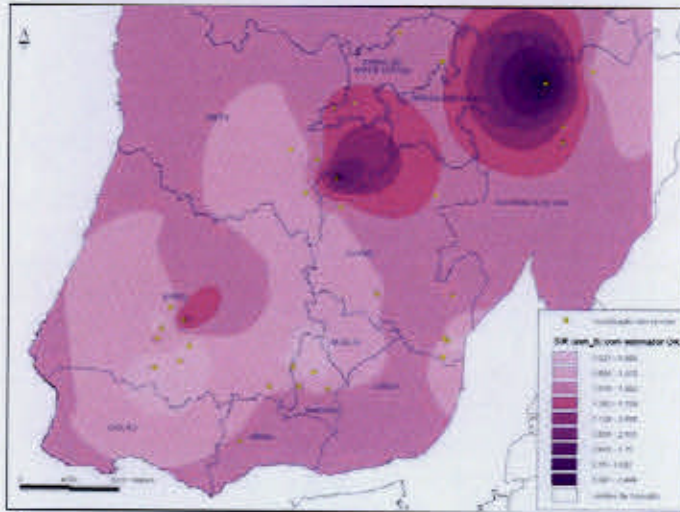
Existem ainda outros modelos de estimação linear da família da Krigagem que se aplicam a fenómenos não estacionários ou correlacionados com outras variáveis (Krigagem com deriva externa e Co-Krigagem), mas cujo enquadramento teórico ultrapassa o âmbito deste trabalho.

### 5.3.2 Variáveis da saúde

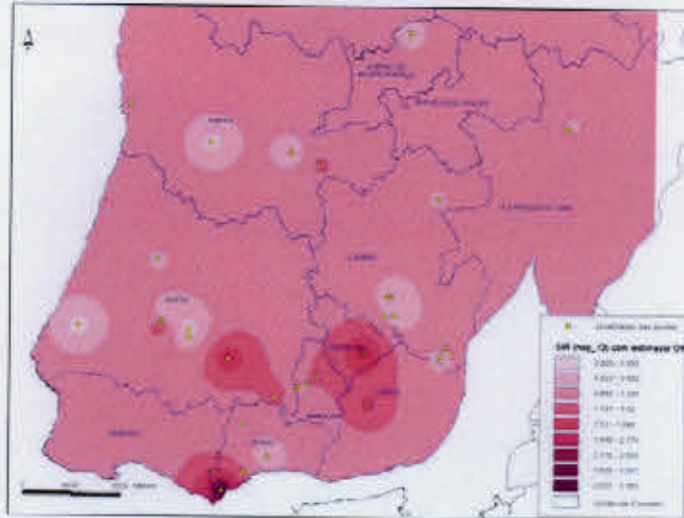
Para a inferência espacial das variáveis da saúde, incorporaram-se as estruturas de continuidade espacial estimadas no sub-capítulo 5.2.2.1. O modelo considerado para a tendência tomou em linha de conta a análise das médias locais abordada no mesmo sub-capítulo e ilustrada no Anexo B, onde se considerou a existência de variações nas médias locais, ao longo da área de estudo, para as variáveis consideradas (ash\_13, ash\_6, hay\_13, rash\_13 e prose\_13). Os mapas seguintes ilustram a distribuição espacial estimada da razão padronizada de incidência das variáveis da saúde obtidas pelo o estimador linear de Krigagem Normal.



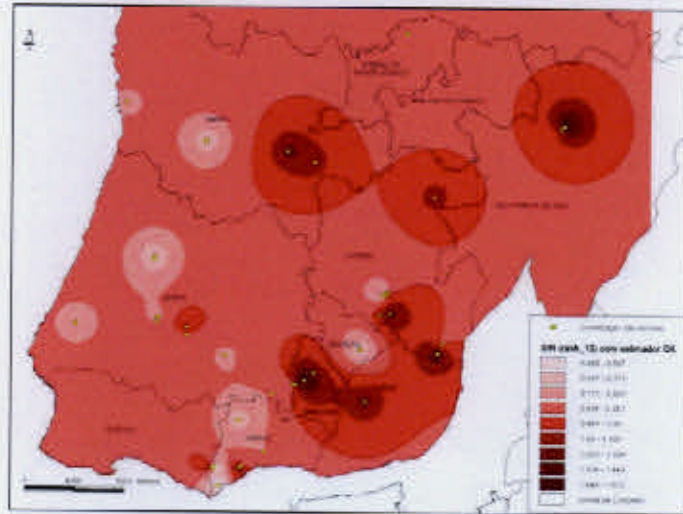
a.



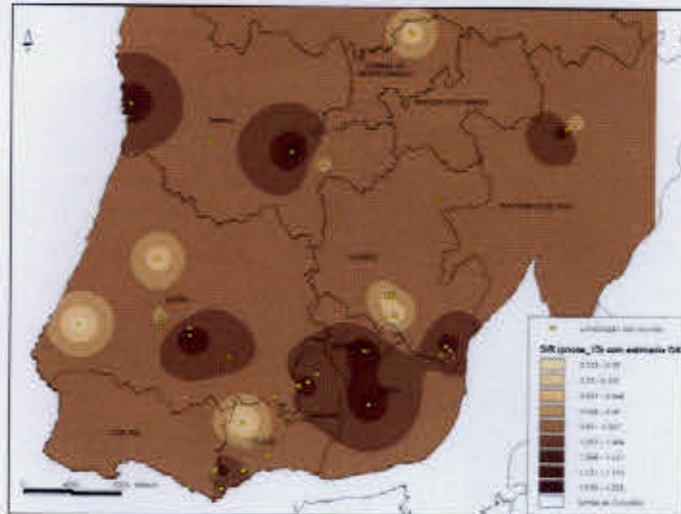
b.



**c.**



**d.**

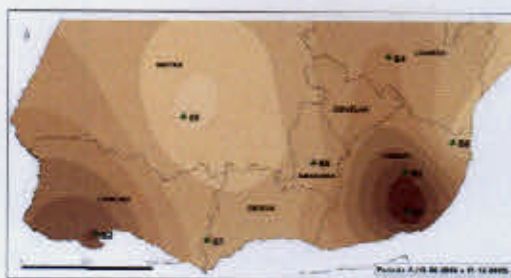


e.

Figura 5.13- Mapas das estimativas da razão padronizada de incidência de ash\_13 (a.) ash\_6 (b.) hay\_13 (c.) rash\_13 (d.) e pnose\_13 (e.) obtidos com o estimador linear de Krigagem Normal.

### 5.3.3 Variável de qualidade do ar

Para a inferência espacial das concentrações médias de  $PM_{10}$  ao longo do tempo, incorporaram-se as estruturas de continuidade espacial obtidas nos sub-capítulos 5.2.3.2 e 5.2.3.3. Os mapas seguintes foram obtidos com o estimador linear de Krigagem Normal e resultam da média aritmética das estimativas obtidas nos períodos de tempo considerados.



a.



b.

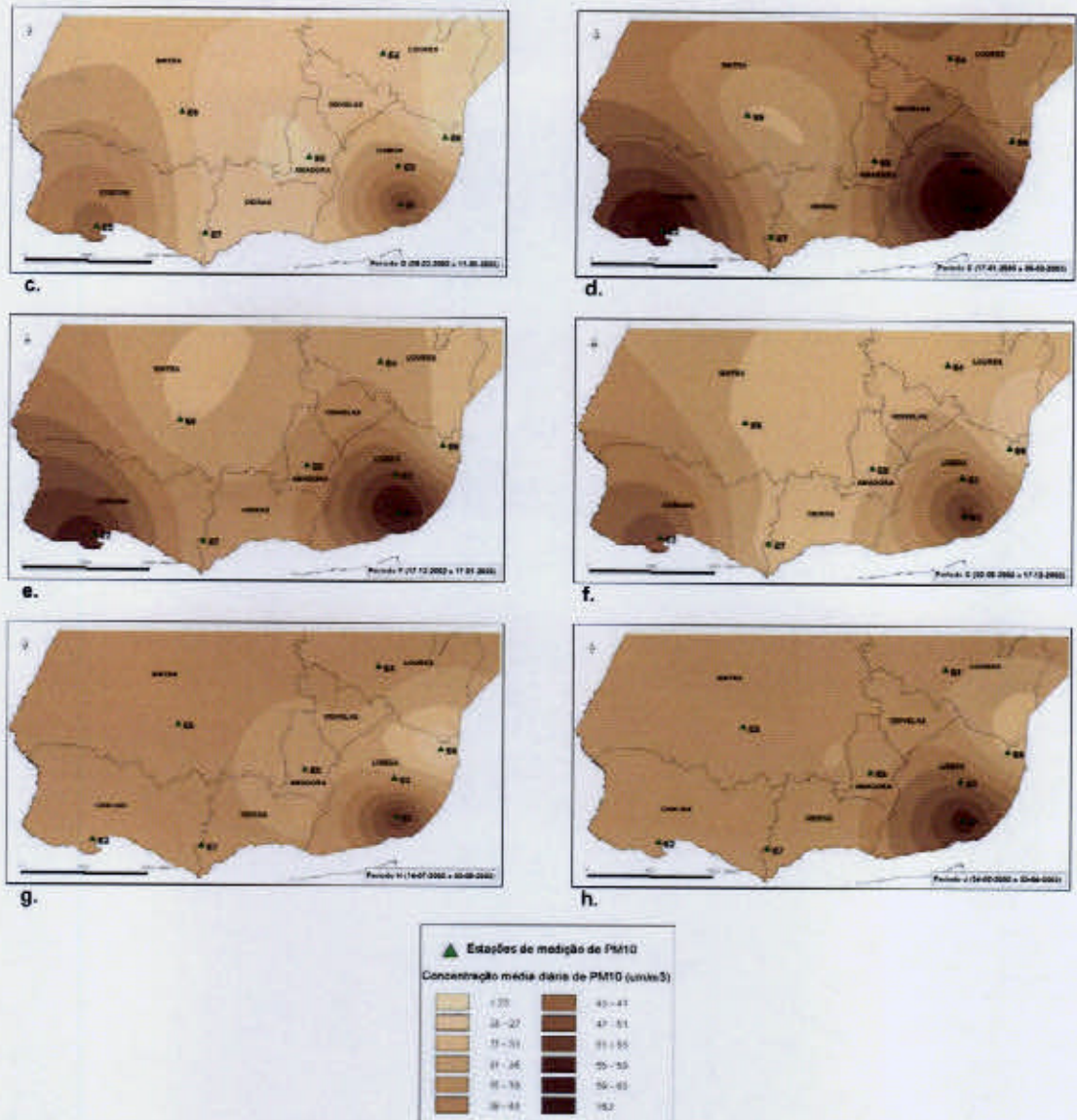


Figura 5.14- Mapas de concentrações médias diárias de PM<sub>10</sub> obtidos com o estimador linear de Krigagem Normal, para os períodos de tempo A (a.), C (b.), D (c.), E (d.), F (e.), G (f.), H (g.) e J (h.).

Note-se que nos períodos H e J, não existem observações para as estações E2, E5 e E7 pelo que as estimativas para essas áreas tendem para os valores médios das estações que lhes estão mais próximas.

## 5.4 Avaliação da incerteza espacial

### 5.4.1 Introdução à simulação estocástica geoestatística

Os modelos de estimação abordados no sub-capítulo 5.3 têm como principal objectivo inferir valores das variáveis geradas por um processo espacial estocástico, em locais onde os valores são desconhecidos. Estas estimativas obtidas para qualquer ponto de valor desconhecido, são consideradas as melhores na perspectiva da minimização do erro quadrático médio. No entanto, para avaliar a incerteza resultante do comportamento simultâneo do conjunto de amostras, este critério de qualidade do estimador (variância mínima) não é adequado, na medida em que os mapas resultantes apresentam um alisamento das eventuais variações locais existentes e um alisamento ainda mais pronunciado em áreas sub amostradas. Por estes motivos, para reproduzir a variabilidade espacial dos valores experimentais, a utilização de modelos de simulação estocástica geoestatística é mais adequada (Goovaerts, 1997; Soares, 2000).

Com estes modelos pretendem-se reproduzir a realização dos processos estocásticos, nomeadamente a sua lei de distribuição conjunta e estrutura de continuidade espacial, condicionada aos valores experimentais, ou seja, para um conjunto de  $k$  variáveis aleatórias,  $\{Z(s_\alpha), \alpha = 1, \dots, k\}$ , definidas em  $k$  pontos  $s_\alpha$  de uma área de estudo pretende-se gerar  $L$  realizações conjuntas  $\{z^l(s_\alpha), \alpha = 1, \dots, k\}$  com  $l=1, \dots, L$ , condicionadas pelos valores experimentais reproduzindo o modelo estrutural dos variogramas e histogramas de valores experimentais.

O modelo de simulação usado neste trabalho enquadra-se na família dos modelos de simulação sequencial (para variáveis contínuas), cujos algoritmos baseiam-se na aplicação recursiva do axioma de Bayes. Este axioma enuncia que uma função de distribuição condicional pode ser expressa pelo produto de duas funções de distribuição condicionais univariadas, pelo que a sua aplicação em passos sucessivos permite gerar uma função de distribuição conjunta  $k$ -variada (realização da função de distribuição) através do produto de  $k$  funções de distribuição univariadas condicionais:

$$F(s_1, \dots, s_\alpha, \dots, s_k; z_1, \dots, z_\alpha, \dots, z_k | (n)) = F(s_k; z_k | (n+k-1)) \cdot F(s_{k-1}; z_{k-1} | (n+k-2)) \cdot \dots \cdot F(s_2; z_2 | (n+1)) \cdot F(s_1; z_1 | (n))$$

onde  $F(s_k; z_k | (n+k-1))$  é a função de distribuição de  $Z(s_k)$  condicionada pelo conjunto de  $n$  dados observados (valores experimentais) e das  $k-1$  realizações anteriores de  $z(s_\alpha), \alpha = 1, \dots, k-1$  simuladas. Esta decomposição em  $k$  funções de distribuição condicionais, permite gerar uma realização do processo estocástico, ao visitar de modo sequencial todos os pontos da área de estudo a simular o que, em termos teóricos, requer o conhecimento dos momentos da função de distribuição condicional em cada passo da simulação.

Neste trabalho o método de simulação sequencial usado é o proposto por Soares (2001), que se designa por Simulação Sequencial Directa (SSD), onde a lei de distribuição condicional em cada passo é obtida por estimação da média e da variância condicional e amostragem posterior da lei de distribuição marginal. O processo de simulação usado desenvolve-se nos seguintes passos:

- Escolha aleatória de um trajecto que visita todos os nós a serem simulados numa malha definida;
- Em cada nó da malha  $s_i, i=1, \dots, k$ 
  - i. Estimar a média e variância local de  $Z(s_i), z_{SK}(s_i)^*$  e  $\sigma_{SK}^2(s_i)$ , com o estimador de Krigagem Simples, condicionadas pelo conjunto de valores experimentais e  $k-1$  valores previamente simulados;
  - ii. Definir um intervalo da função de distribuição global (histograma dos valores experimentais), centrada em  $z_{SK}(s_i)^*$  e de variância  $\sigma_{SK}^2(s_i)$ , e proceder a uma transformação gaussiana,  $\varphi$ , da variável  $z_{SK}(s_i)^*$ ,

$$G(y(s_i)^*, \sigma_{SK}^2(s_i)), \text{ onde } y(s_i)^* = \varphi(z_{SK}(s_i)^*)$$

- iii. Gerar um número aleatório  $p$  de distribuição uniforme,  $p \sim U(0,1)$
- iv. Gerar um valor  $y^s$  a partir da lei de distribuição  $G(y(s_i)^*, \sigma_{SK}^2(s_i))$ :

$$y^s = G^{-1}(y(s_i)^*, \sigma_{SK}^2(s_i), p)$$

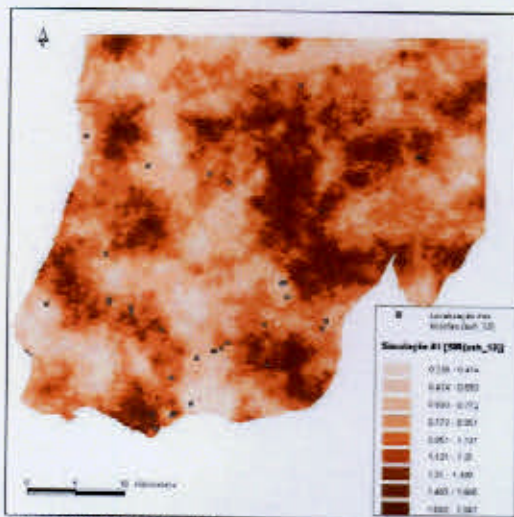
- v. Transformar o valor simulado obtido nas unidades originais:

$$z^s(s_i) = \varphi^{-1}(y^s)$$

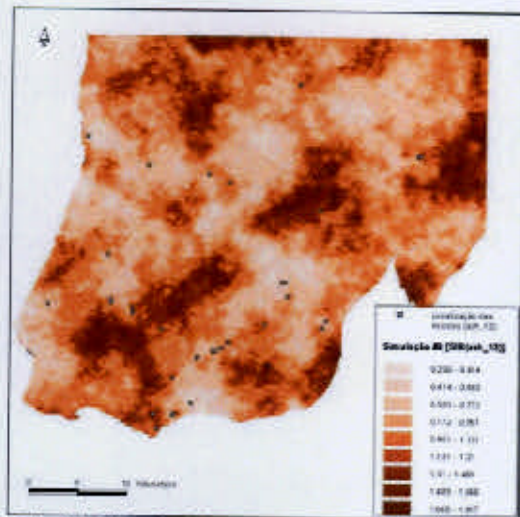
- vi. Voltar ao passo i. até simular todos os  $k$  pontos terem sido visitados e simulados.

### 5.4.2 Variáveis da saúde

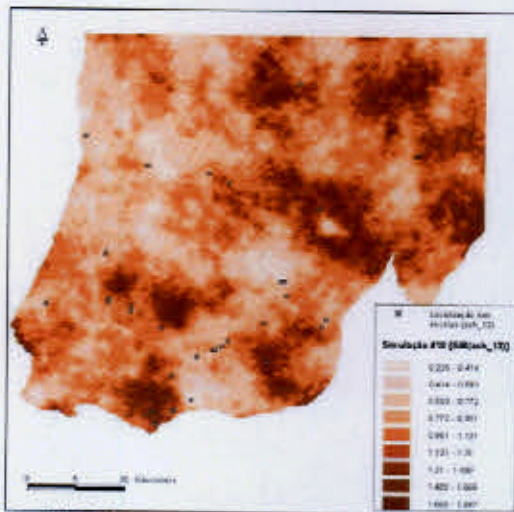
Através da aplicação do processo de simulação apresentado no ponto anterior, foram geradas 30 simulações para cada variável com o objectivo de avaliar a incerteza espacial, considerando a distribuição conjunta dos valores experimentais. Apresentam-se na Figura 5.15 três exemplos de realizações equiprováveis simuladas pelo método DSS, para cada uma das 5 variáveis.



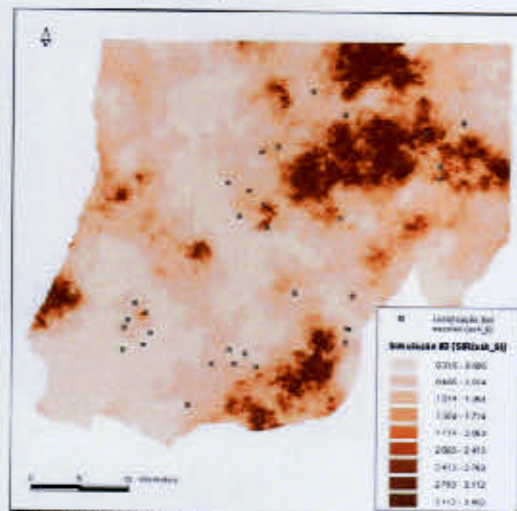
1.a. #1 SIR(ash\_13)



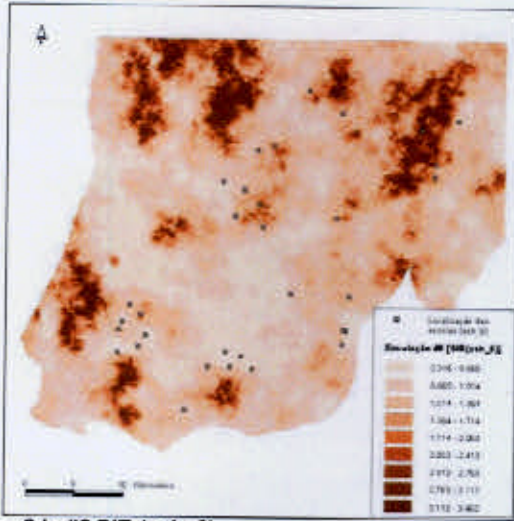
1.b. #9 SIR(ash\_13)



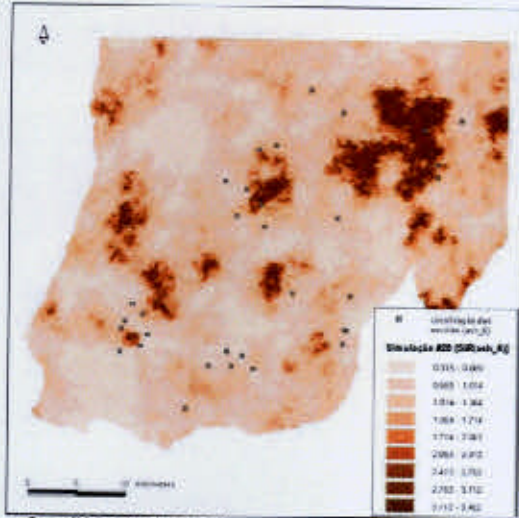
1.c. #18 SIR(ash\_13)



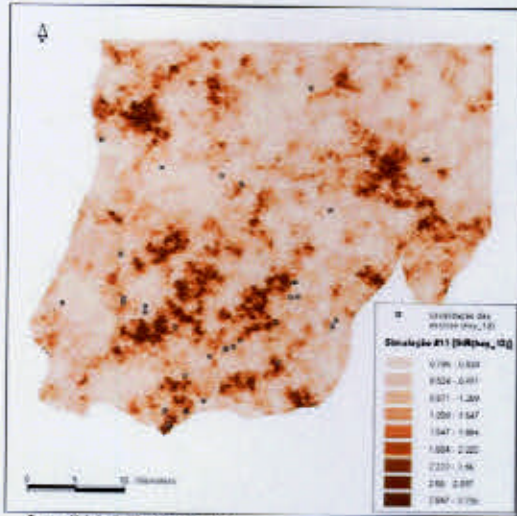
2.a. #3 SIR(ash\_6)



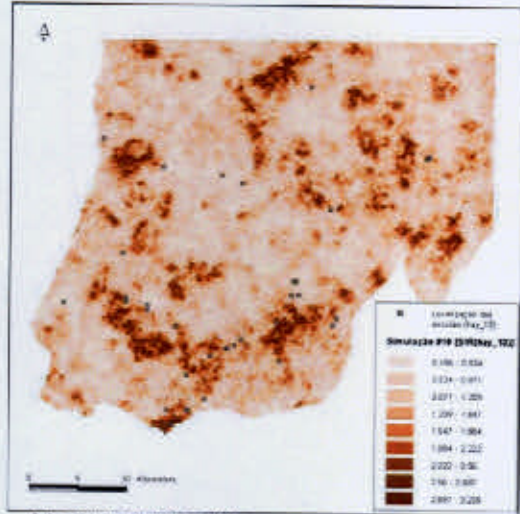
2.b. #6 SIR (ash\_6)



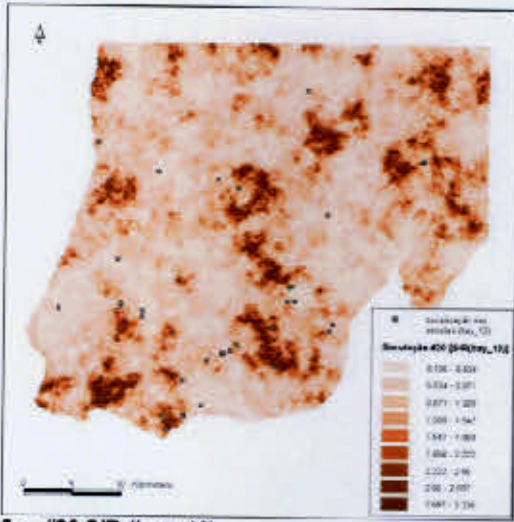
2.c. #20 SIR (ash\_6)



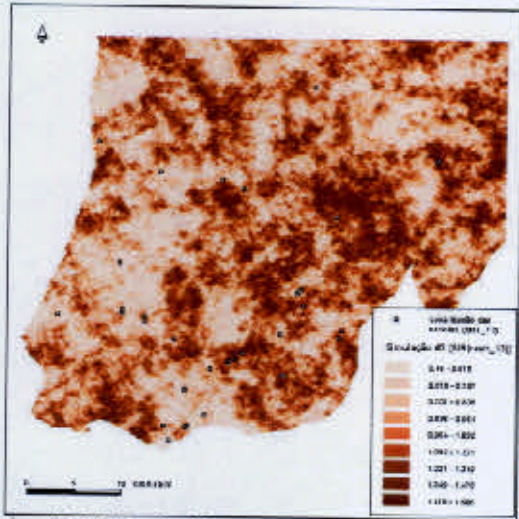
3.a. #11 SIR (hay\_13)



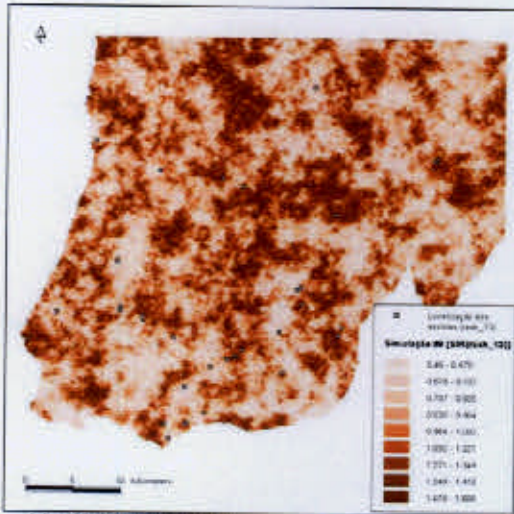
3.b. #16 SIR (hay\_13)



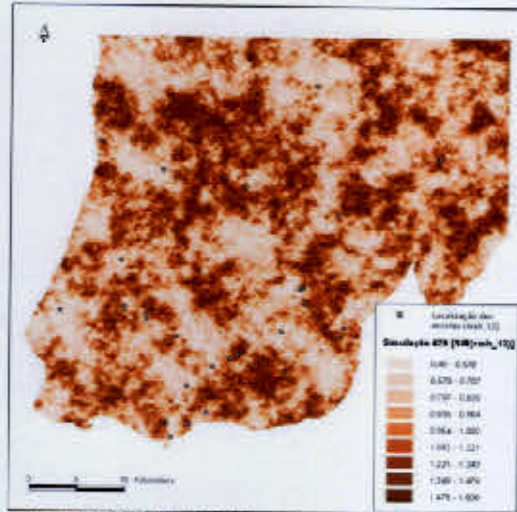
3.c. #20 SIR (hay\_13)



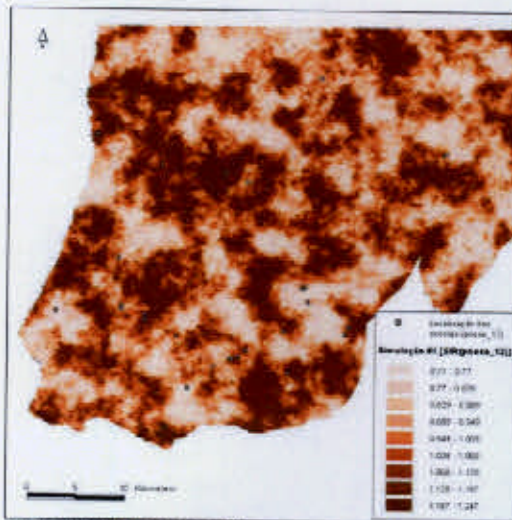
4.a. #3 SIR (rash\_13)



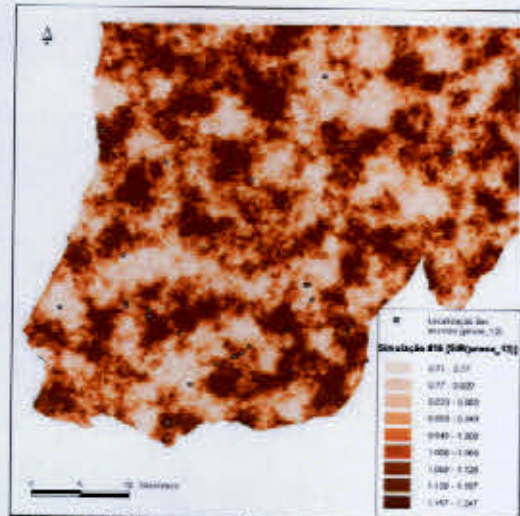
4.b. #9 SIR (rash\_13)



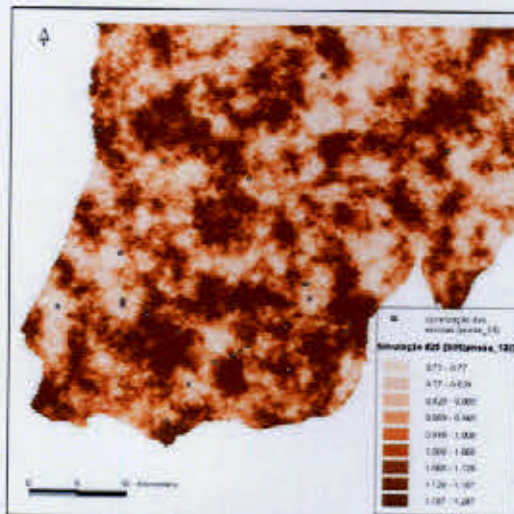
4.c. #29 SIR (rash\_13)



5.a. #1 SIR(pnose\_13)



5.b. #19 SIR(pnose\_13)



5.c. #25 SIR(pnose\_25)

**Figura 5.15- Exemplos de realizações equiprováveis simuladas para as variáveis da saúde**

Os variogramas e histogramas de todas as simulações estão representados respectivamente nos Anexos D e E. A comparação entre os variogramas e histogramas dos mapas simulados e dos valores experimentais foi realizada para todas as simulações, onde se concluiu que as imagens simuladas apresentam padrões de continuidade espacial e formas de distribuição semelhantes às estruturas de continuidade espacial e histogramas dos valores experimentais.

Para a análise conjunta dos mapas simulados em cada variável, calcularam-se para cada ponto as médias das 30 realizações simuladas,  $m(s)$ , para obter mapas de médias:

$$m(s) = \frac{\sum_{l=1}^L z'(s)}{L}$$

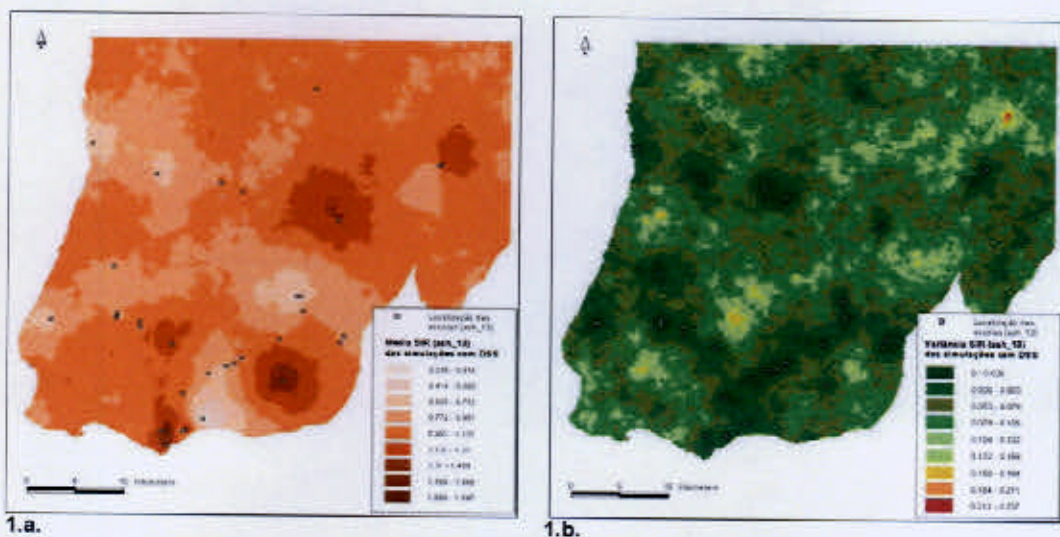
onde  $z'(s)$  representa o valor obtido no ponto  $s$  na simulação  $l$ ,  $l=1, \dots, L$ . Para identificar áreas onde os riscos relativos de incidência apresentam maior variabilidade, calcularam-se para cada ponto, as variâncias das 30 realizações simuladas,  $\sigma^2(s)$ :

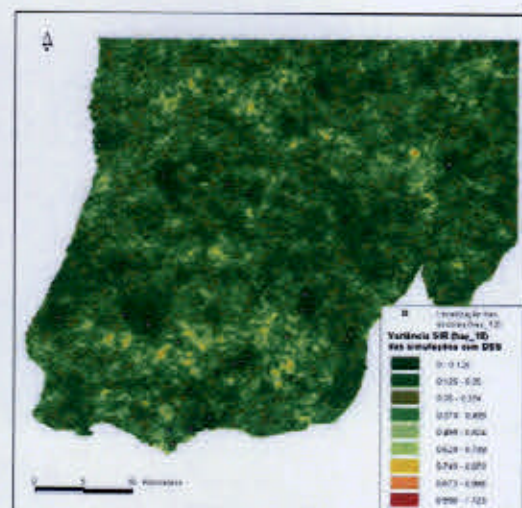
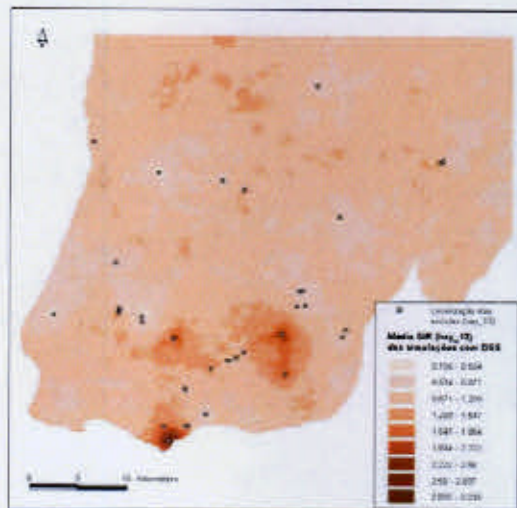
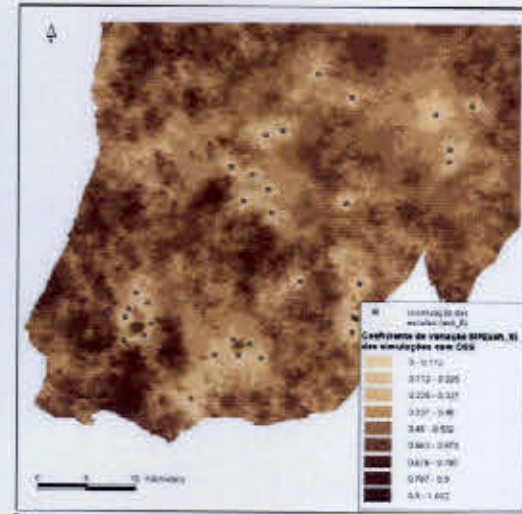
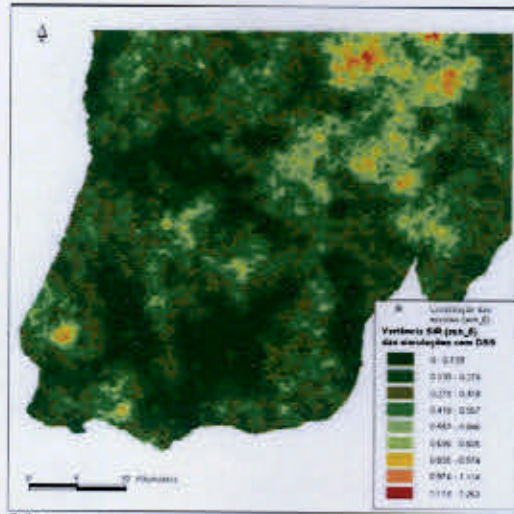
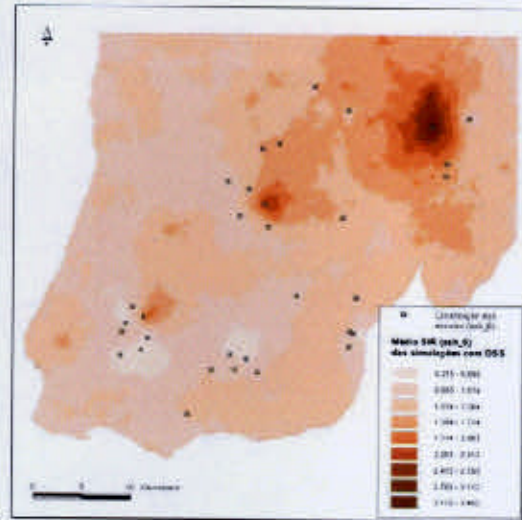
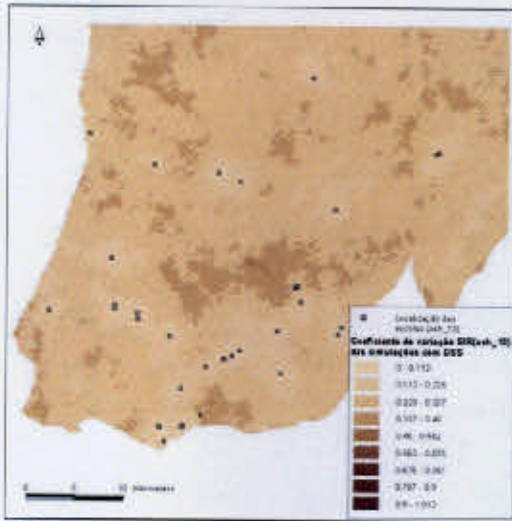
$$\sigma^2(s) = \frac{\sum_{l=1}^L [z'(s) - m(s)]^2}{L}$$

Para eliminar os efeitos de escala e comparar a variabilidade obtida entre as variáveis consideradas, calcularam-se os mapas de coeficientes de variação, onde para cada ponto, o coeficiente de variação,  $CV(s)$ , é definido por:

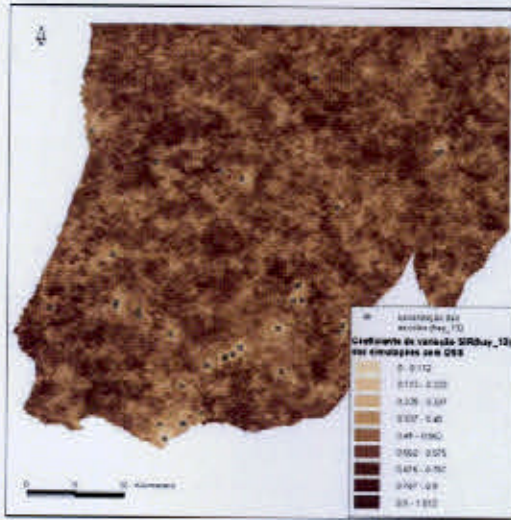
$$CV(s) = \frac{m(s)}{\sigma^2(s)}$$

Os resultados obtidos apresentam-se na Figura 5.16.

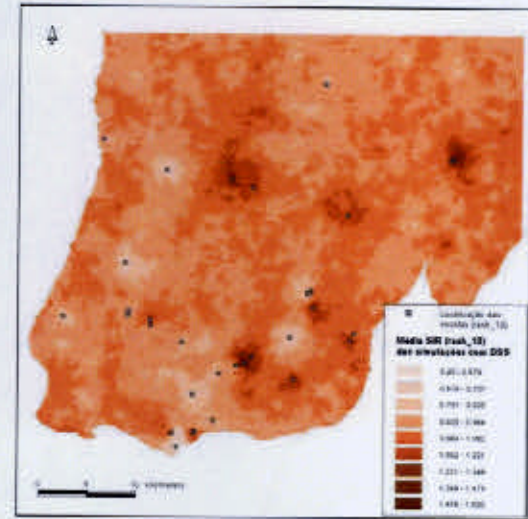




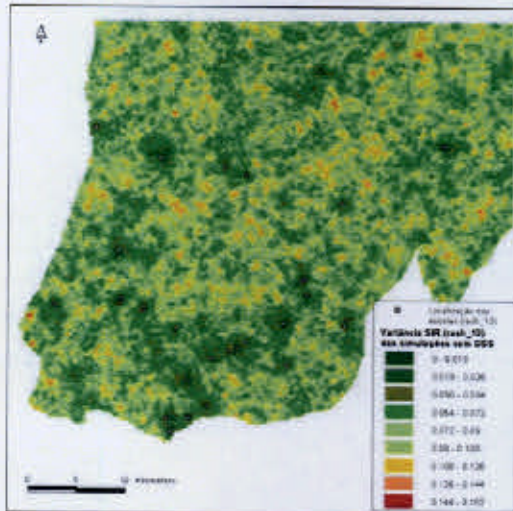
3.a.



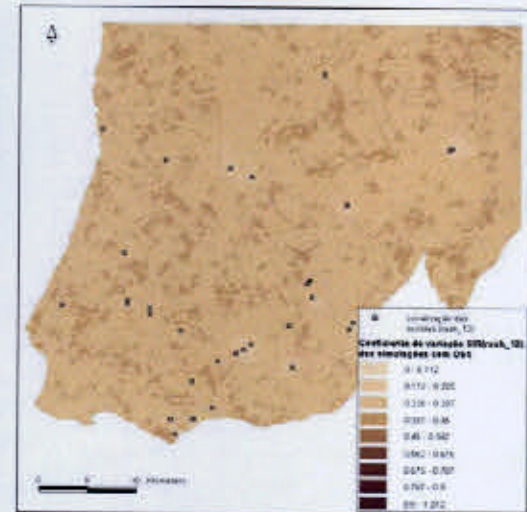
3.b.



3.c.

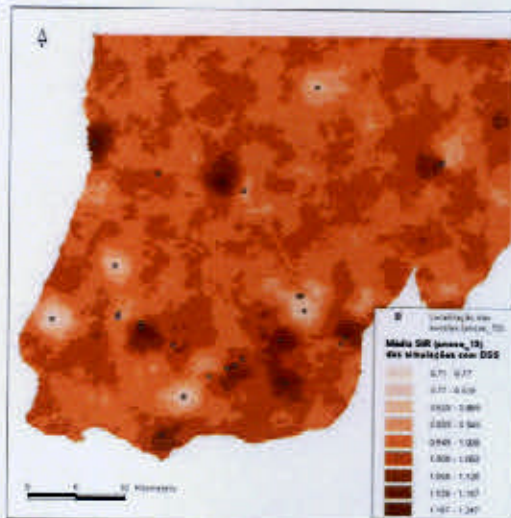


4.a.

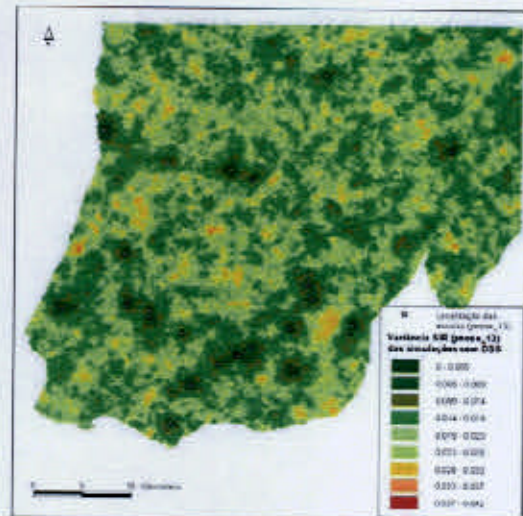


4.b.

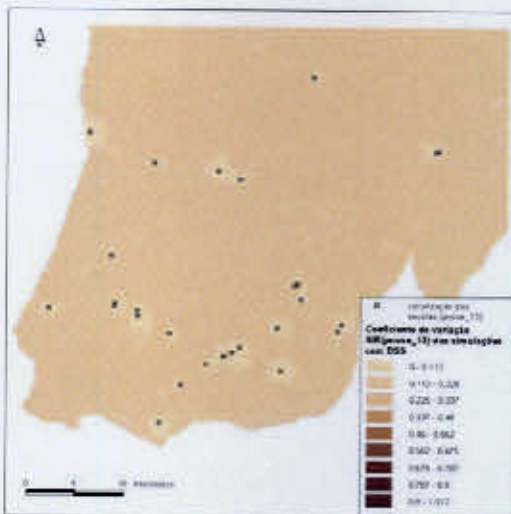
4.c.



5.a.



5.b.



5.c.

**Figura 5.16- Mapas de médias (a.), variâncias (b.) e coeficientes de variação (c.) obtidos pelo método de simulação sequencial directa: ash\_13 (1.); ash\_6 (2.); hay\_13 (3.); rash\_13 (4.) e prose\_13 (5.)**

Os mapas de médias representam um alisamento das variações (em relação à média) observadas nos vários mapas simulados e evidenciam semelhanças aos mapas estimados no sub-capítulo 5.3.2. Os mapas de variâncias representam a variabilidade das simulações em cada ponto, e permitem identificar as regiões onde a incerteza sobre o valor do SIR é maior. Nos casos das variáveis ash\_13 e ash\_6 (alíneas 1.b, 2.b da Figura 5.16) verificou-se que os mapas apresentam grandes manchas de baixa variabilidade nos valores simulados (áreas representadas a verde escuro). Nas restantes variáveis (hay\_13, rash\_13 e prose\_13) verificou-se que os mapas apresentam maior variabilidade à pequena escala (fruto dos padrões de estrutura de

continuidade impostos pelos modelos de variograma, caracterizados por baixas amplitudes), que se evidencia ao longo de toda a área de estudo. Verificou-se ainda que em termos relativos, a incerteza associada à distribuição espacial dos valores de SIR nas variáveis ash\_6 e hay\_13 é superior à das restantes variáveis.

### 5.4.3 Variável da qualidade do ar

Para simular as concentrações de PM<sub>10</sub>, foram incorporados os modelos espaciais e temporais conforme Figura 5.11 e geraram-se 30 simulações para cada período de tempo considerado (A, C, D, E, F, G, H, J) com DSS, de acordo com uma malha cúbica definida no referencial (X, Y, Z) conforme tabela seguinte.

Elxo	Mínimo	Máximo	Nº Células (ou nós)	Resolução da malha	Unidade
X	80 000	120 000	81	500	Metros
Y	185 000	210 000	51	500	Metros
Z	di	df	$\lambda$	1	Dias

Tabela 5.2- Definição das dimensões X, Y e Z do domínio a simular

Os parâmetros di, df e  $\lambda$  representam o dia de início, de fim e número de dias ( $\lambda = df - di$ ) de cada período, respectivamente. Na Tabela 5.3 apresentam-se os valores dos parâmetros considerados para a dimensão Z.

Período	di	df	$\lambda$
A	957	1095	139
C	800	850	51
D	770	800	31
E	747	770	24
F	716	747	32
G	610	715	106
H	560	609	50
J	420	479	60

Tabela 5.3- Definição da dimensão Z para cada período de tempo

Para cada período de tempo, o domínio espacial e temporal em cada simulação ficou definido por uma malha cúbica discretizada em  $81 \times 51 \times \lambda$  células. As simulações obtidas para cada período foram comparadas com os histogramas, variogramas e estatísticas descritivas dos dados experimentais, para avaliar as simulações como instrumento de avaliação da incerteza.

Apresenta-se em seguida, como exemplo, o resultado da aplicação da DSS a três simulações de três dias consecutivos obtidos para o período A (dias 27, 28 e 29 de Setembro de 2003).

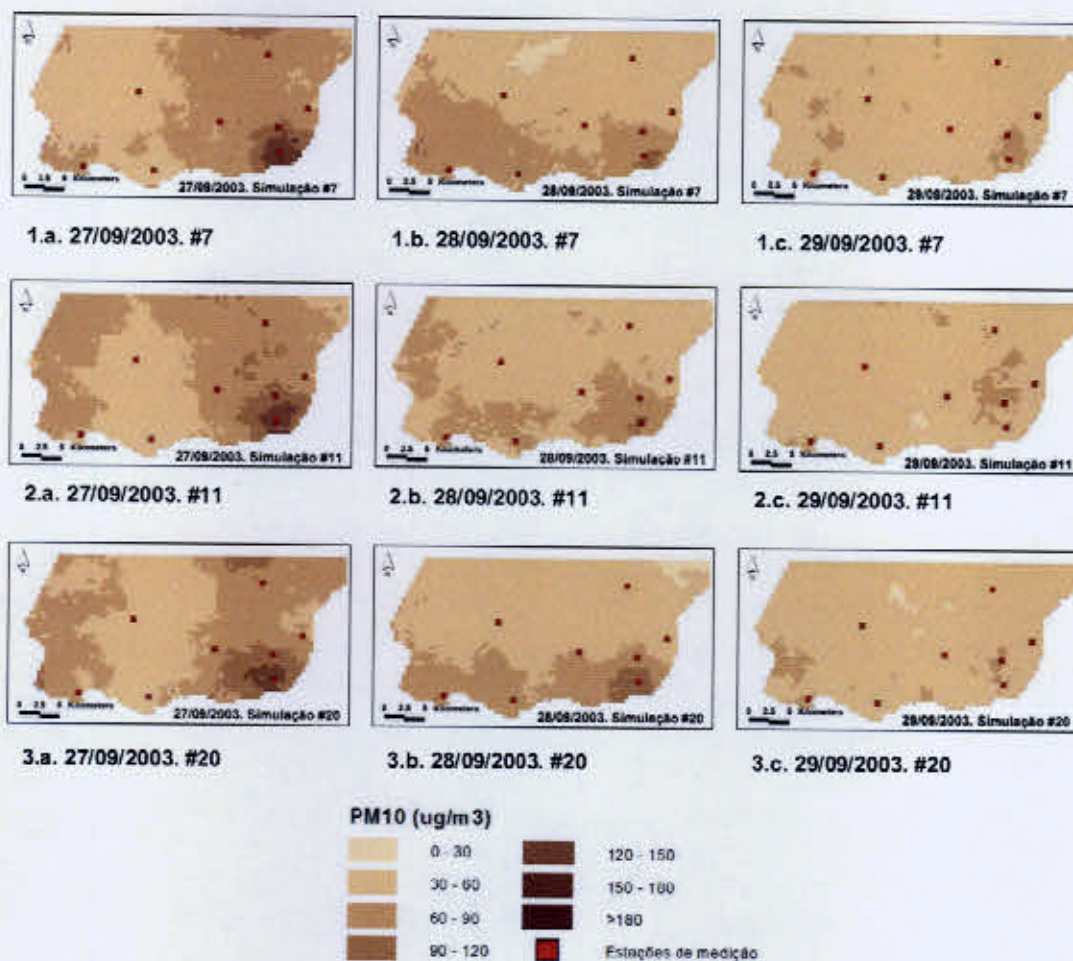
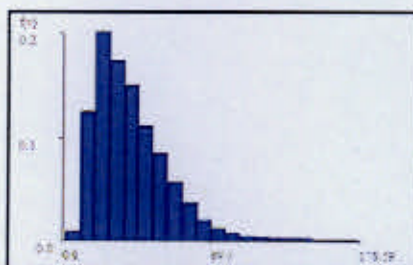


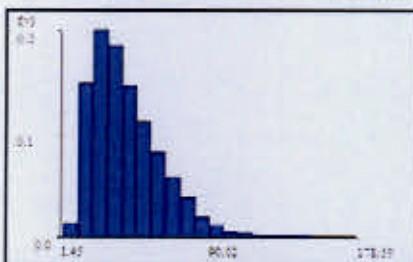
Figura 5.17- Três realizações dos dias 27, 28 e 29 de Setembro de 2003

Com base nas realizações obtidas em todas as simulações para todos os períodos, compararam-se as formas dos histogramas e as estatísticas descritivas com as dos dados experimentais, para avaliar a qualidade da reprodução do fenómeno. Apresentam-se na Figura 5.18 os resultados obtidos para as três realizações apresentadas na Figura 5.17.



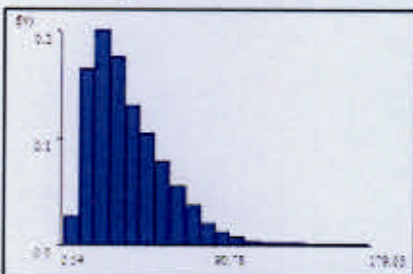
<b>Média</b>	39.06
<b>D.Padrão</b>	20.93
<b>Mínimo</b>	0
<b>Mediana</b>	34.95
<b>P.95</b>	78.42
<b>Máximo</b>	178.59

1.a. Histograma e estatísticas descritivas da simulação #7



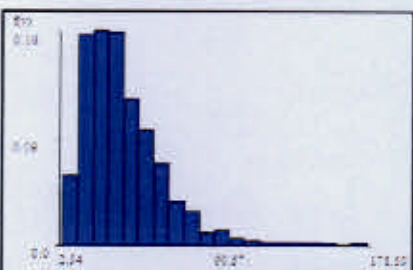
<b>Média</b>	39.01
<b>D.Padrão</b>	20.43
<b>Mínimo</b>	1.45
<b>Mediana</b>	34.86
<b>P.95</b>	77.45
<b>Máximo</b>	178.59

2.a. Histograma e estatísticas descritivas da simulação #11



<b>Média</b>	39.04
<b>D.Padrão</b>	20.82
<b>Mínimo</b>	2.54
<b>Mediana</b>	34.39
<b>P.95</b>	78.72
<b>Máximo</b>	179.03

3.a. Histograma e estatísticas descritivas da simulação #20



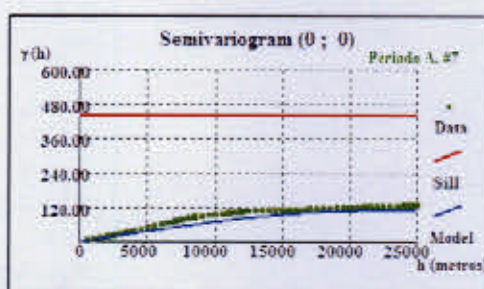
<b>Média</b>	37.96
<b>D.Padrão</b>	22.77
<b>Mínimo</b>	2.54
<b>Mediana</b>	33.14
<b>P.95</b>	78.50
<b>Máximo</b>	178.59

4.a. Histograma e estatísticas descritivas dos dados experimentais

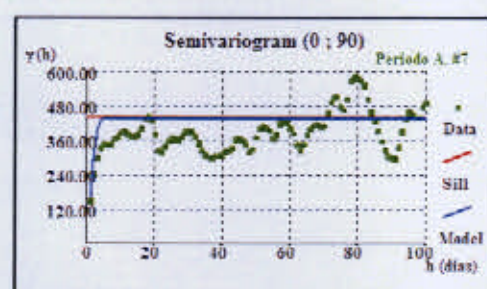
Figura 5.18- Histogramas e estatísticas descritivas dos dados experimentais e das simulações #7, #11 e #20 do período A.

De forma geral, os resultados obtidos para o período A foram satisfatórios, ou seja, consideraram-se semelhantes as formas dos histogramas das simulações com o do histograma dos dados experimentais. Por outro lado, consideraram-se também que os parâmetros estimados nas simulações apresentaram valores semelhantes aos parâmetros estatísticos dos dados experimentais.

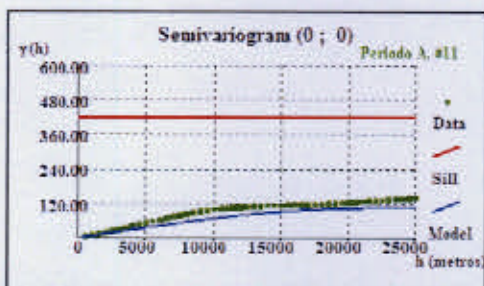
Compararam-se os resultados da variografia espacial e temporal de cada simulação, com a variografia das variáveis estandardizadas, o que serviu para avaliar a reprodução dos padrões (espaciais e temporais) imposta pelo modelo teórico dos dados transformados. Apresentam-se os resultados obtidos para as três simulações apresentadas na Figura 5.17, com sobreposição do modelo teórico ajustado à variável transformada.



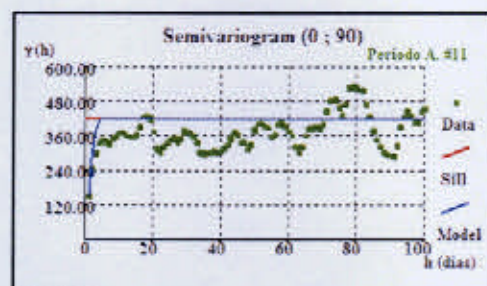
1.a. #7



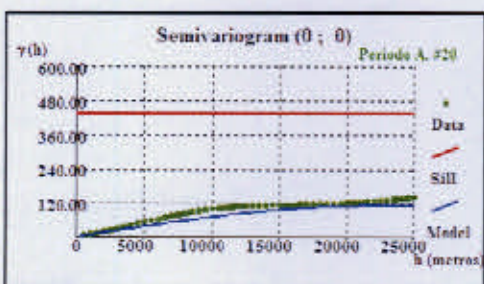
1.b. #7



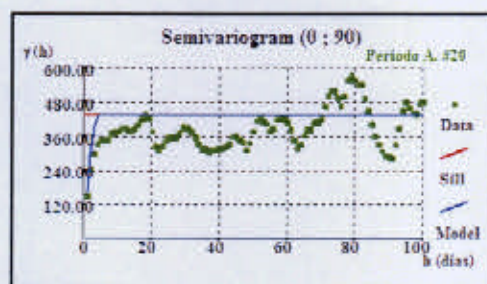
2.a. #11



2.b. #11



3.a. #20



3.b. #20

Figura 5.19- Variogramas espaciais (a.) e temporais (b.) das simulações #7, #11 e #20 do período A.

Nos três exemplos apresentados, considera-se que as amplitudes da primeira estrutura foram reproduzidas. Na reprodução da segunda estrutura as realizações tenderam a apresentar maior continuidade do que fora imposto pelo modelo.

Em termos globais, os resultados das simulações do período A foram satisfatórias, na medida em que as estruturas espacial e temporal do fenómeno foram reproduzidas com a DSS, assim como os histogramas e parâmetros estatísticos.

Para os restantes períodos de tempo, os resultados obtidos revelaram-se bons apenas para o período E. De facto, na maioria dos casos as estruturas de continuidade espacial e temporal modeladas a partir das variáveis transformadas, bem como os histogramas e estatísticas descritivas, não foram reproduzidas com a DSS. Apresenta-se em seguida um exemplo dos resultados insatisfatórios obtidos, com base em três simulações de três dias consecutivos obtidas para o período C (dias 25, 26 e 27 de Março de 2003).

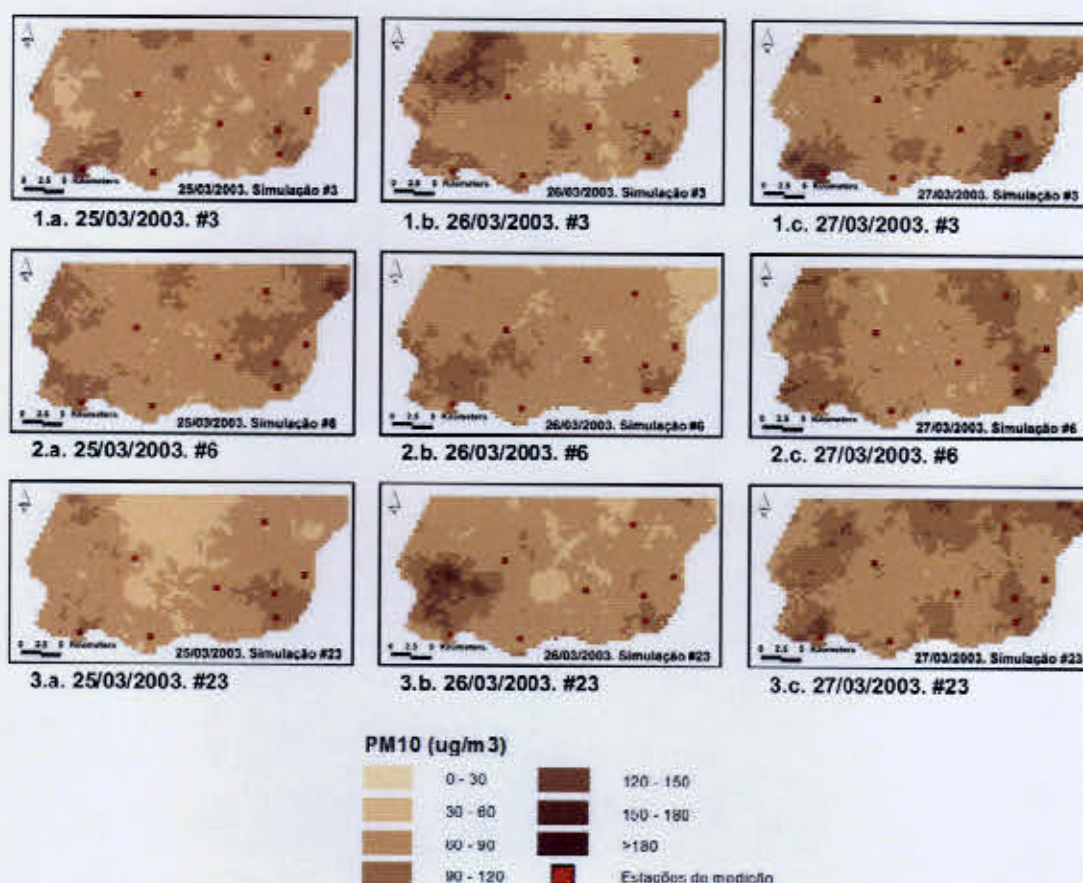
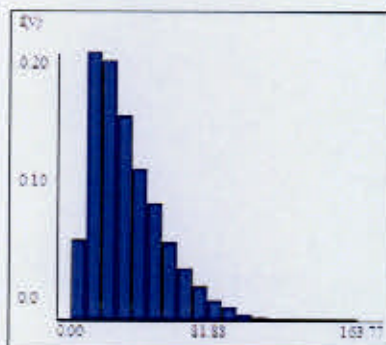


Figura 5.20- Três realizações dos dias 25, 26 e 27 de Março de 2003

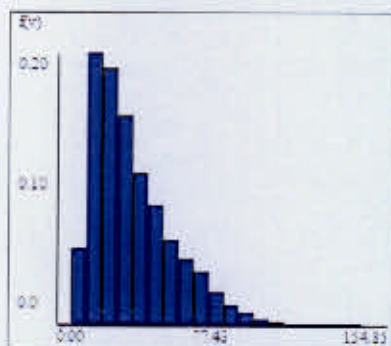
Os histogramas e estatísticas descritivas dos dados originais evidenciam diferenças significativas relativamente aos das simulações apresentadas (Figura 5.21). A assimetria dos histogramas é semelhante, mas o achatamento do histograma dos

dados experimentais é mais acentuado (veja-se o peso da cauda direita do histograma, relativamente ao das simulações). Também os parâmetros estatísticos estimados apresentam diferenças acentuadas, nomeadamente os valores máximos e valores médios obtidos.



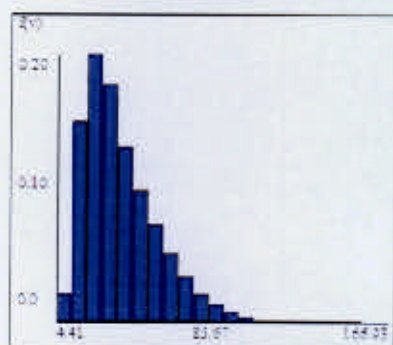
Média	39.04
D.Padrão	19.98
Mínimo	0
Mediana	34.37
P.95	78.11
Máximo	163.77

1.a. Histograma e estatísticas descritivas da simulação #3



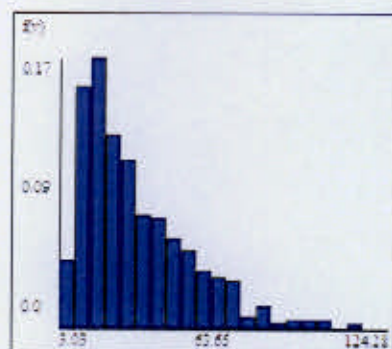
Média	39.04
D.Padrão	20.57
Mínimo	0.00
Mediana	33.81
P.95	79.96
Máximo	154.85

2.a. Histograma e estatísticas descritivas da simulação #6



Média	39.03
D.Padrão	20.22
Mínimo	4.41
Mediana	34.29
P.95	78.39
Máximo	166.93

3.a. Histograma e estatísticas descritivas da simulação #23

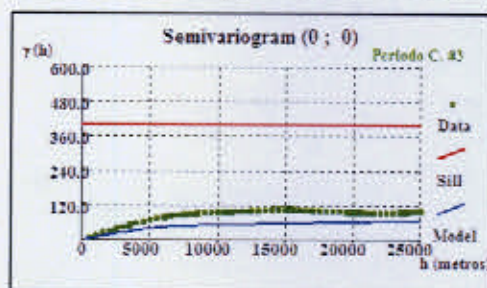


Média	35.70
D.Padrão	22.01
Mínimo	3.03
Mediana	30.02
P.95	81.08
Máximo	124.28

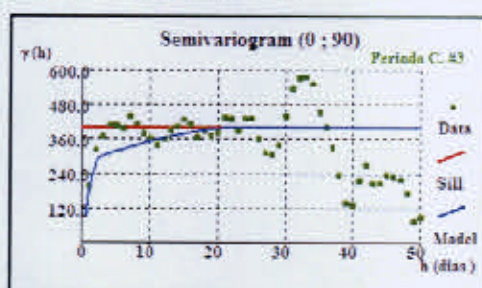
4.a. Histograma e estatísticas descritivas dos dados experimentais

Figura 5.21- Histogramas e estatísticas descritivas dos dados experimentais e das simulações #3, #6 e #23 do período C.

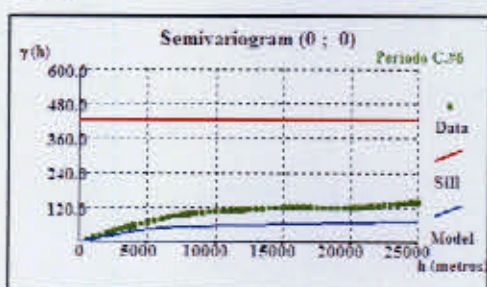
As estruturas de continuidade espacial e temporal apresentam também diferenças relativamente aos modelos teóricos ajustados (representados a azul).



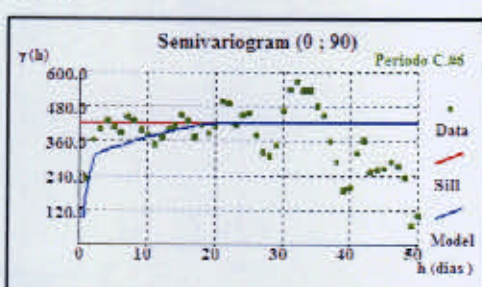
1.a. #3



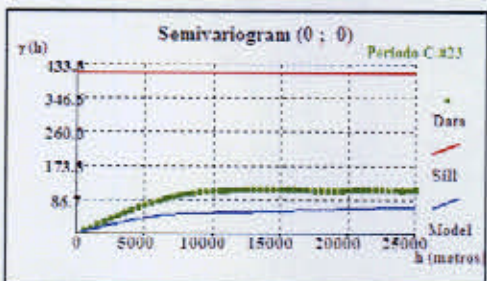
1.b. #3



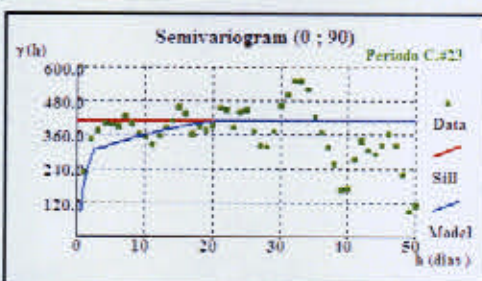
2.a. #6



2.b. #6



3.a. #23



3.b. #23

Figura 5.22- Variogramas espaciais (a.) e temporais (b.) das simulações #3, #6 e #23 do período C.

A estrutura de continuidade espacial e temporal do período C não foi reproduzida no processo de simulação (Figura 5.22). As diferenças evidenciaram-se na amplitude da primeira estrutura espacial e temporal (é superior nas simulações) e no caso da variografia temporal, na amplitude da segunda estrutura (é inferior nas simulações).

Numa análise global aos resultados considerou-se que não foram satisfatórios, porque o processo de simulação não cumpriu o objectivo de reproduzir o fenómeno, nomeadamente a variabilidade dos dados experimentais, a sua lei de distribuição e a estrutura de continuidade espacial e temporal. No entanto, os resultados obtidos tiveram o mérito de revelar a fragilidade dos pressupostos assumidos no capítulo 5.2.3.3, onde se considerou a estrutura espacial subjacente ao fenómeno (e não a estrutura espacial do próprio fenómeno) como modelo para os processos de inferência e simulação do fenómeno original.

Perante estes resultados, optou-se por repetir o processo de simulação, de modo a garantir a reprodução das características dos dados originais, através da aplicação da metodologia de simulação por DSS com a variável transformada,  $z(s,t)'$ , e posterior transformação inversa,  $z(s,t)''$ , conforme se segue (Serpa, 2003):

$$z(s,t)'' = z(s,t)' \cdot \sigma(s) + m(s)$$

onde  $z(s,t)''$  representa o valor de  $PM_{10}$  na escala original,  $z(s,t)'$  representa o valor simulado da variável transformada no ponto  $s$  num instante  $t$  da malha cúbica, e  $m(s)$  e  $\sigma(s)$  representam respectivamente a média e desvio padrão dos valores medidos no ponto  $s$ .

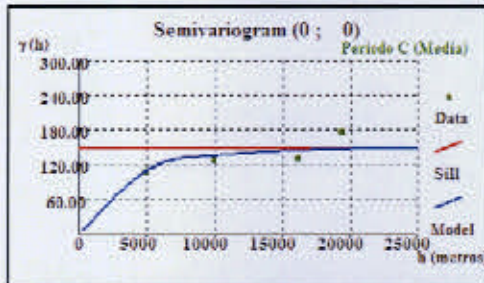
A dificuldade da aplicação deste método reside no facto de não existirem os parâmetros estatísticos para cada ponto da malha simulada, sendo por isso necessário estimá-los. Optou-se por estimar as médias e variâncias para cada ponto da malha, com o estimador de Krigagem Normal (Goovaerts, 1997; Serpa, 2003), a partir dos valores médios e variâncias observadas nos dados experimentais, para cada período. Os valores conhecidos das médias e desvios padrão em cada estação, para os vários períodos apresentam-se na Tabela 5.4.

Estação	E1		E2		E3		E4	
Período	Média	DP	Média	DP	Média	DP	Média	DP
A	55.8	29.5	49.1	15.8	50.7	25.8	37.0	18.5
C	54.2	25.2	56.1	21.2	40.3	23.1	32.2	19.1
D	52.3	19.3	60.6	22.5	43.8	16.8	31.5	12.3
E	59.6	26.6	51.7	19.3	43.3	28.3	33.4	28.3
F	65.5	35.3	59.3	18.7	53.4	32.7	35.2	20
G	55.1	24.0	54.7	17.8	44.5	19.9	28.7	12.9
H	56.9	20.2	nd	nd	41.0	19.9	39.2	16.8
J	61.4	29.1	nd	nd	48.3	24.0	36.1	16.4
Estação	E5		E6		E7		E8	
Período	Média	DP	Média	DP	Média	DP	Média	DP
A	24.0	12.2	26.1	17.7	29.7	13.8	30.7	17.9
C	23.7	15.7	23.9	17.5	29.6	17.7	26.5	16.3
D	31.1	13.3	25.1	14.3	29.6	11.2	24.6	11.4
E	24.5	9.35	25.8	23.0	28.7	12.3	26.2	12.4
F	30.2	12.1	31.3	21.6	36.2	19.4	41.0	25.9
G	25.0	11.1	22.7	13.6	27.8	12.1	28.9	14.6
H	nd	nd	26.8	16.8	nd	nd	37.9	17.2
J	nd	nd	28.9	18.2	nd	nd	35.8	19.0

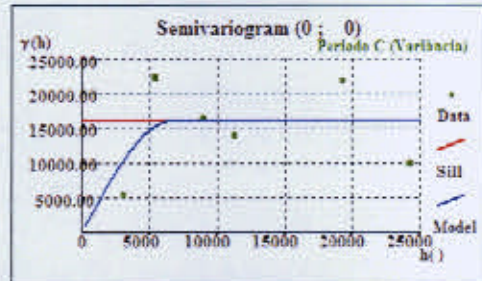
DP=Desvio Padrão. nd= dados não disponíveis.

**Tabela 5.4- Valores conhecidos das médias em cada estação, para os vários períodos.**

As estruturas de continuidade necessárias para a estimação foram modeladas para cada período. No entanto, a análise de continuidade espacial da variável original, abordada no capítulo 5.2.3.2, revelou a inexistência de um fenómeno estruturado no espaço. Desta forma, o cálculo dos variogramas só foi possível por eliminação de um ou dois pares de pontos (consoante o período analisado), responsáveis pelo ruído provocado na análise de continuidade da variável original. Apesar de não ser o procedimento óptimo revelou-se o único caminho possível para ajustar modelos às estruturas de continuidade espacial e produzir as estimativas necessárias da média e do desvio padrão para cada ponto da malha. Esta metodologia foi aplicada para toda a malha, em todos os períodos considerados nesta fase do estudo - A, C, D, E, F, G, H e J. Descrevem-se em seguida, de modo detalhado, os resultados obtidos para o período C.



a. Variograma espacial da média para o período C. Modelo esférico embricado:  $c_0=0; c_1=117; a_1=7400; c_2=31.5; a_2=23200$



b. Variograma espacial da variância para o período C. Modelo esférico:  $c_0=0; c_1=16050; a_1=6600$

Figura 5.23- Variografia espacial das médias e das variâncias para o período C.

No ajustamento dos modelos teóricos (linhas a azul na Figura 5.23) admitiu-se que a estrutura de continuidade espacial à pequena escala não tem efeito pepita. O coeficiente de correlação entre as médias e as variâncias é elevado ( $\rho^2=0.85$ ) o que, em termos de ajustamento do modelo justifica em parte, as semelhanças obtidas nos valores das amplitudes. Apresentam-se a seguir os mapas estimados das médias e dos desvios padrão locais obtidos com o estimador de Krigagem Normal.



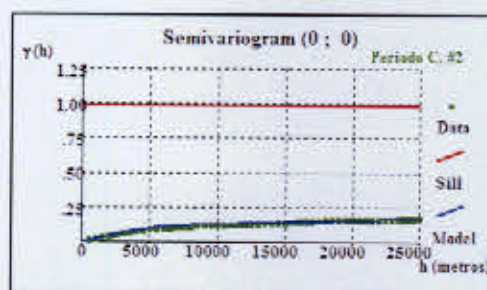
a. Mapa das médias locais para o período C



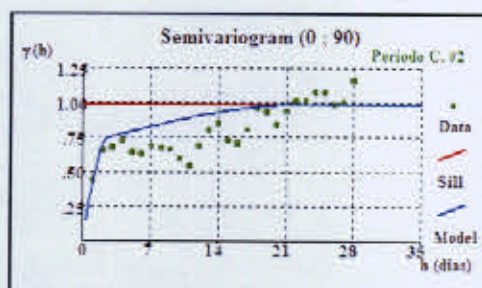
b. Mapa dos desvio-padrão locais para o período C.

Figura 5.24- Mapas das médias e desvios-padrão locais estimados para o período C

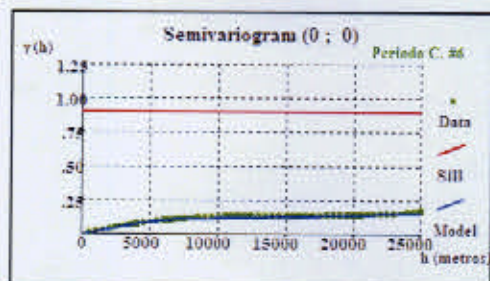
Nos mapas resultantes obtiveram-se valores médios e desvios-padrão de concentrações médias para qualquer ponto da área de estudo, isto é, a estimação do primeiro e segundo momentos. Procedeu-se então à repetição do processo de simulação, desta vez com a variável transformada de média igual a zero e variância igual a um. Apresentam-se os resultados obtidos em algumas das simulações obtidas.



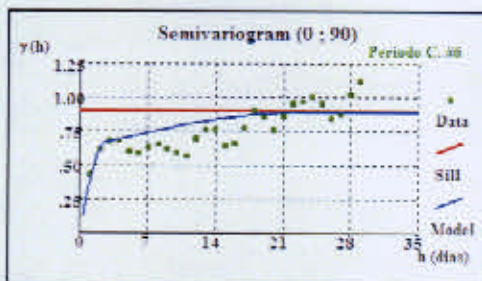
1.a. #2



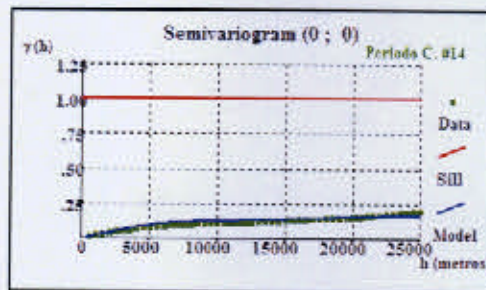
1.b. #2



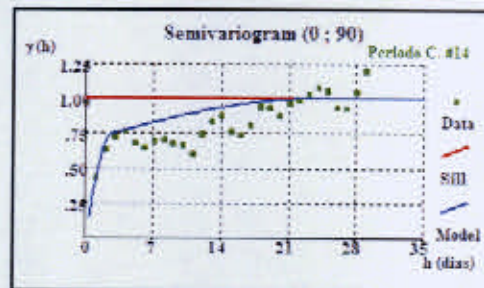
2.a. #6



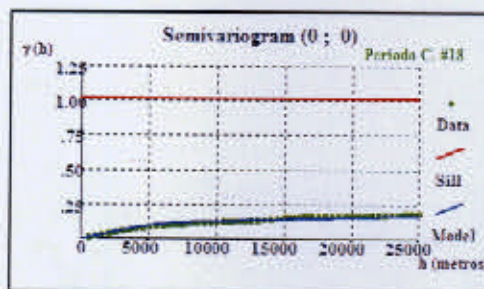
2.b. #6



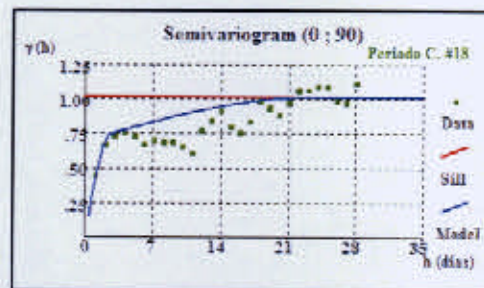
3.a. #14



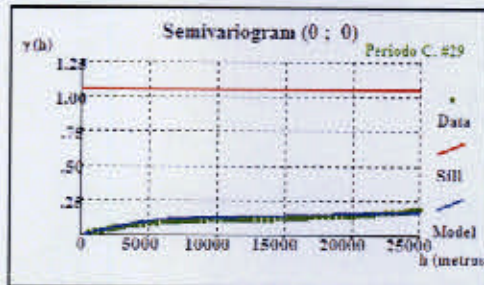
3.b. #14



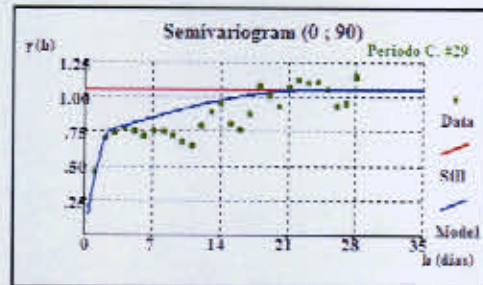
4.a. #18



4.b. #18



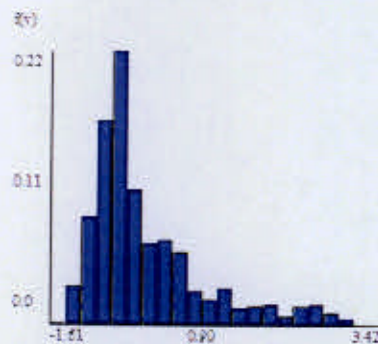
5.a. #29



5.b. #29

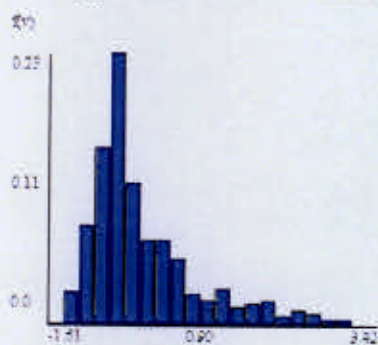
Figura 5.25- Variogramas espaciais e temporais de 5 das 30 simulações da variável transformada obtidas para o período C.

A simulação com DSS reproduziu as estruturas de continuidade espacial e temporal da variável transformada, conforme apresentadas nas alíneas 3.a. e 3.b. da Figura 5.11. Quanto aos histogramas e estatísticas descritivas, considera-se que os resultados também foram muito satisfatórios.



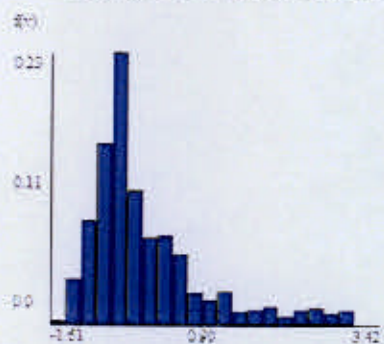
Média	0.00
D.Padrão	0.99
Mínimo	-1.61
Mediana	-0.36
P.95	2.39
Máximo	3.41

1.a. Histograma e estatísticas descritivas da simulação #2



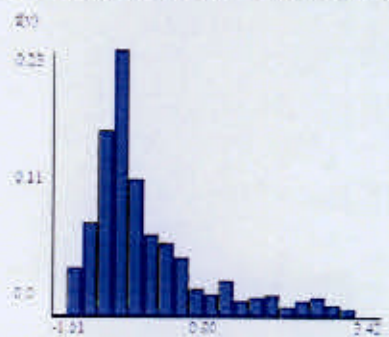
Média	-0.01
D.Padrão	0.95
Mínimo	-1.61
Mediana	-0.34
P.95	2.07
Máximo	3.42

2.a. Histograma e estatísticas descritivas da simulação #6



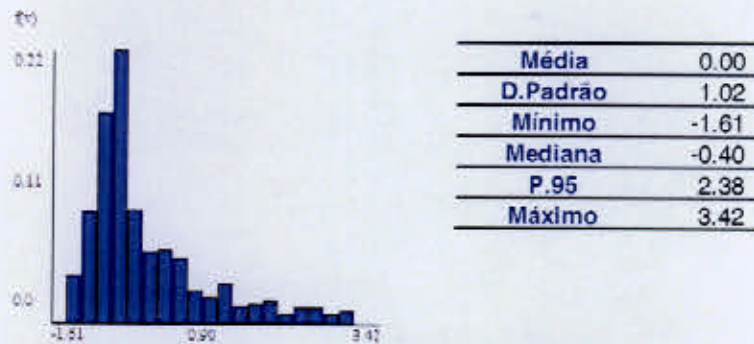
Média	0.00
D.Padrão	1.03
Mínimo	-1.61
Mediana	-0.36
P.95	2.46
Máximo	3.41

3.a. Histograma e estatísticas descritivas da simulação #14

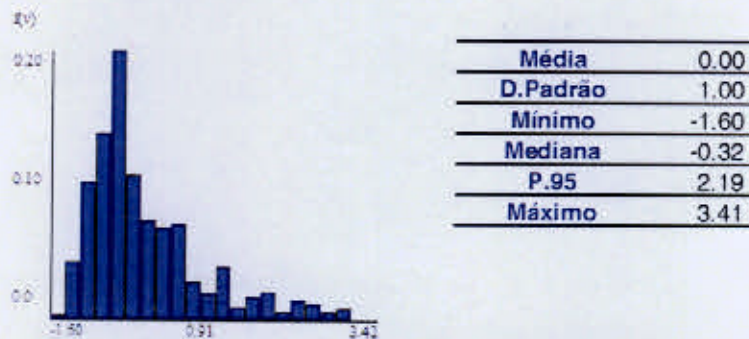


Média	0.00
D.Padrão	1.00
Mínimo	-1.61
Mediana	-0.37
P.95	2.41
Máximo	3.42

4.a. Histograma e estatísticas descritivas da simulação #18



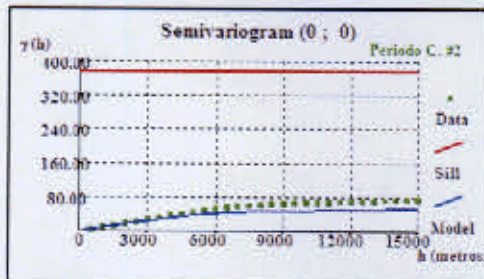
5.a. Histograma e estatísticas descritivas da simulação #29



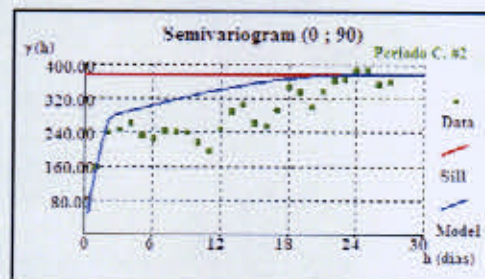
6.a. Histograma e estatísticas descritivas dos dados experimentais (transformados)

Figura 5.26- Histogramas e estatísticas descritivas dos dados experimentais (transformados) e de 5 simulações, para o período C.

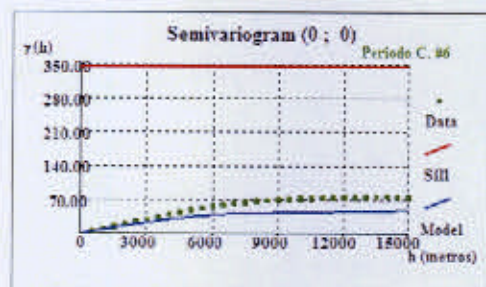
Perante os resultados obtidos a metodologia aplicada prosseguiu com a transformação inversa dos valores do cubo simulado, usando para esse efeito os parâmetros estimados das médias e desvios-padrão locais (cf. Figura 5.24). Os resultados da transformação apresentam-se na Figura 5.27.



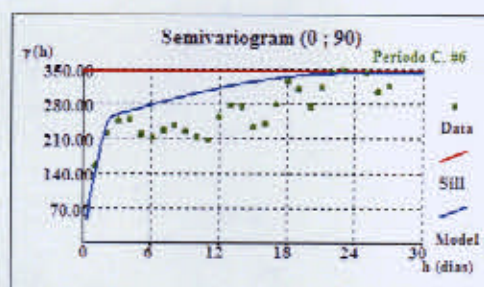
1.a.



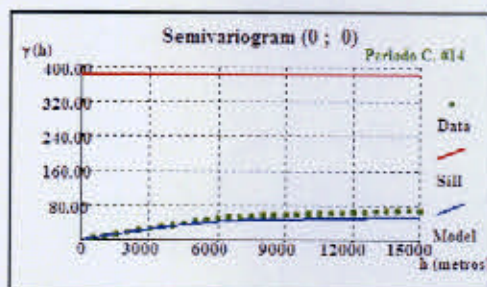
1.b.



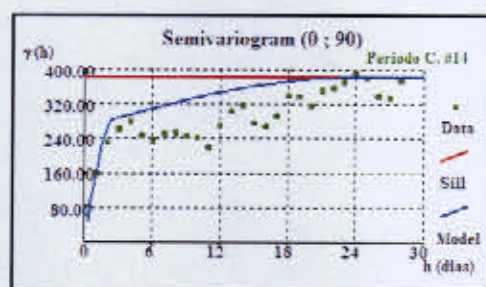
2.a.



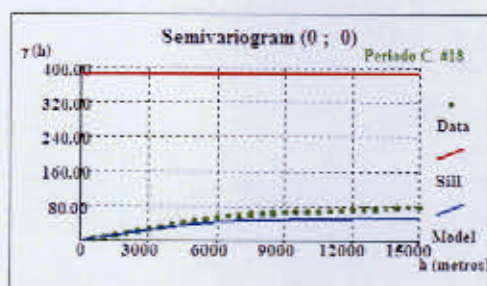
2.b.



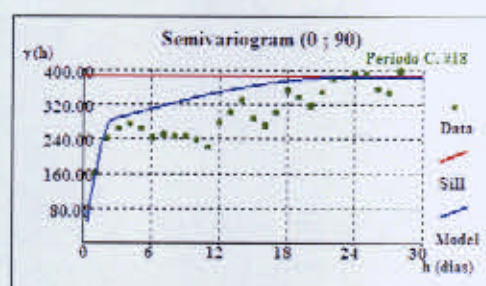
3.a.



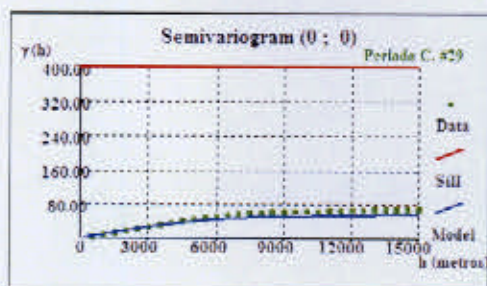
3.b.



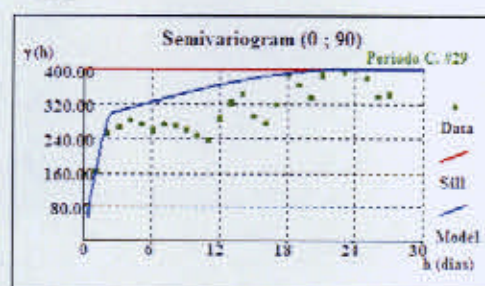
4.a.



4.b.



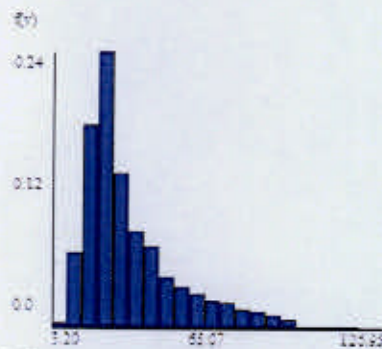
5.a.



5.b.

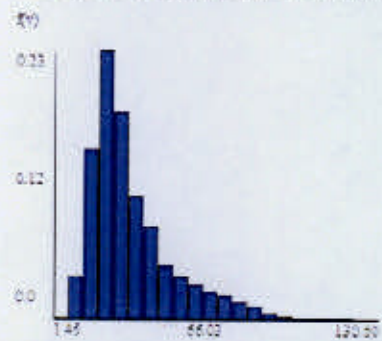
Figura 5.27- Variogramas espaciais e temporais de 5 das 30 simulações da variável original obtida por transformação inversa, para o período C.

Pelo ajustamento dos modelos espacial e temporal é visível a reprodução das estruturas de continuidade, em particular na primeira estrutura, com cerca de 7500 metros e 2.5 dias, respectivamente. Os histogramas e estatísticas descritivas que completam esta análise apresentam-se na Figura 5.28.



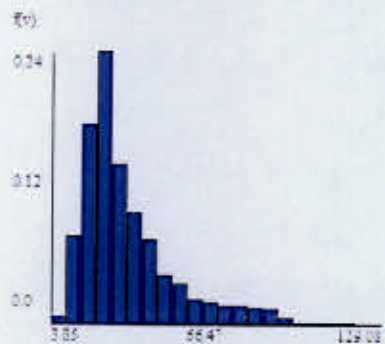
<b>Média</b>	35.39
<b>D.Padrão</b>	19.34
<b>Mínimo</b>	3.85
<b>Mediana</b>	28.91
<b>P.95</b>	80.22
<b>Máximo</b>	127.68

**1. Histograma e estatísticas descritivas da simulação #2**



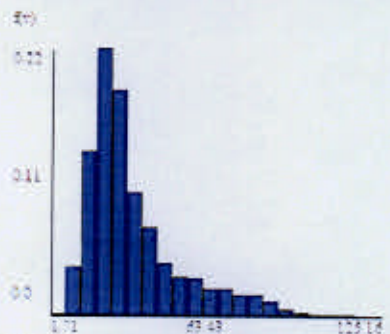
<b>Média</b>	35.30
<b>D.Padrão</b>	18.54
<b>Mínimo</b>	1.45
<b>Mediana</b>	29.41
<b>P.95</b>	75.94
<b>Máximo</b>	130.60

**2. Histograma e estatísticas descritivas da simulação #6**



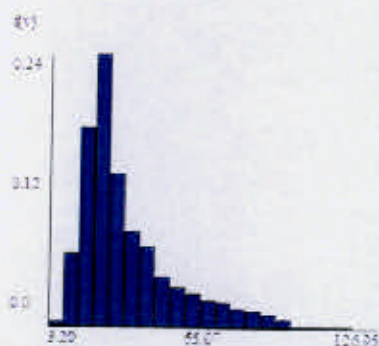
<b>Média</b>	35.31
<b>D.Padrão</b>	19.53
<b>Mínimo</b>	3.85
<b>Mediana</b>	29.01
<b>P.95</b>	80.69
<b>Máximo</b>	129.08

**3. Histograma e estatísticas descritivas da simulação #14**



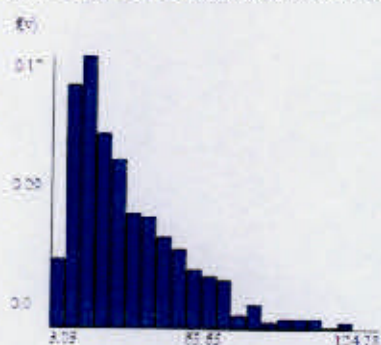
<b>Média</b>	35.30
<b>D.Padrão</b>	19.59
<b>Mínimo</b>	1.70
<b>Mediana</b>	28.89
<b>P.95</b>	79.95
<b>Máximo</b>	125.15

**4. Histograma e estatísticas descritivas da simulação #18**



<b>Média</b>	35.36
<b>D.Padrão</b>	19.96
<b>Mínimo</b>	3.19
<b>Mediana</b>	28.29
<b>P.95</b>	79.99
<b>Máximo</b>	126.94

5. Histograma e estatísticas descritivas da simulação #29



<b>Média</b>	35.70
<b>D.Padrão</b>	22.01
<b>Mínimo</b>	3.03
<b>Mediana</b>	30.02
<b>P.95</b>	81.08
<b>Máximo</b>	124.28

6. Histograma e estatísticas descritivas dos dados experimentais

Figura 5.28- Histogramas e estatísticas descritivas dos dados experimentais e de 5 simulações (após transformação inversa), para o período C

As médias das simulações são tendencialmente mais elevadas e os desvios-padrão mais baixos, enquanto que a forma dos histogramas é mais suave, nomeadamente na cauda direita das distribuições. Considera-se que, para o período analisado, o processo de simulação usado gerou imagens equiprováveis de concentrações locais de  $PM_{10}$ .

Perante os resultados obtidos no período C, procedeu-se à aplicação da metodologia acima descrita para cada ponto da malha cúbica simulada, para cada período, excepto para os períodos H e J, dado que o número de observações disponíveis nestes períodos (5 e 6 respectivamente) revelou-se manifestamente insuficiente para estimar os mapas de médias e de desvios-padrão (ver Tabela 5.4). Os resultados satisfatórios estenderam-se aos restantes períodos analisados – A, D, E, F e G – onde o processo de simulação cumpriu o objectivo de reproduzir o fenómeno, nomeadamente a variabilidade dos dados experimentais, a sua lei de distribuição e as estruturas de continuidade espacial e temporal.

Após a abordagem acima descrita que envolveu 1) a simulação do domínio espacial com a variável transformada e 2) a estimação das médias e desvios padrão locais para aplicação da transformação inversa e reprodução do fenómeno à escala original, obtiveram-se 30 simulações (para cada dia e) para cada período de tempo.

Este conjunto de cenários equiprováveis foi usado para quantificar comportamentos extremos do fenómeno (e a incerteza associada), onde para cada ponto da área de estudo calculou-se a proporção de realizações que ultrapassaram um determinado valor-limite (também designado por valor de corte) acima do qual se admitiu que existe risco para a saúde pública. Considere-se uma função indicatriz  $I(s)$ , definida por:

$$I(s) = \begin{cases} 1 & Z^l(s,t) > z \\ 0 & Z^l(s,t) \leq z \end{cases}$$

onde  $z$  representa um valor de corte definido,  $Z^l(s,t)$  representa o valor da variável  $Z$  no local  $s$  (pertencente à área de estudo), no momento  $t$  ( $t=1, \dots, T$ ) referente à simulação  $l$  ( $l=1, \dots, L$ ). A probabilidade da concentração de  $PM_{10}$  ultrapassar o valor de corte  $z$  no local  $s$ , no cenário  $l$  é dada por:

$$P[Z^l(s,t) > z] = \frac{\sum_{t=1}^T I^l(s,t)}{T}$$

Em termos legais, os valores-limite definidos para as  $PM_{10}$  estão definidos desde 1999, de acordo com a 1ª Directiva-Filha 1999/30/CE da Directiva-Quadro 1996/62/CE. Estes valores disponíveis para vários poluentes são definidos como "o nível fixado com base em conhecimentos científicos com o intuito de evitar, prevenir ou reduzir os efeitos nocivos sobre a saúde humana e/ou o ambiente na sua globalidade, susceptível de ser atingido num prazo determinado e que, quando atingido, não deverá ser excedido" (Artº 2, da Directiva-Quadro 1996/62/CE). Apresentam-se na Tabela 5.5 os valores-limite estabelecidos no Anexo III da 1ª Directiva-Filha, referentes aos valores-limite de  $PM_{10}$ .

Poluente	Valor-limite	Periodo amostragem	Margem de tolerância
PM <sub>10</sub> (1ª fase, até 2005)	50 µg/m <sup>3</sup> PM <sub>10</sub> a não exceder mais de 35 vezes em cada ano civil	24 horas	25 µg/m <sup>3</sup> (50%) com reduções sucessivas até atingir os 0% em 1 de Janeiro 2005
PM <sub>10</sub> (1ª fase, até 2005)	40 µg/m <sup>3</sup> PM <sub>10</sub>	Ano civil	8 µg/m <sup>3</sup> (20%) com reduções sucessivas até atingir os 0% em 1 de Janeiro 2005
PM <sub>10</sub> (2ª fase, de 2005 a 2010)	50 µg/m <sup>3</sup> PM <sub>10</sub> a não exceder mais de 7 vezes em cada ano civil	24 horas	A calcular em função dos dados, de modo a ser equivalente ao valor limite da 1ª fase até atingir os 0% em 1 de Janeiro de 2010
PM <sub>10</sub> (2ª fase, de 2005 a 2010)	20 µg/m <sup>3</sup> PM <sub>10</sub>	Ano civil	10 µg/m <sup>3</sup> (50%) com reduções sucessivas até atingir os 0% em 1 de Janeiro 2010

Tabela 5.5- Valores-limite estabelecidos pela 1ª Directiva-Filha 1999/30/CE para as PM<sub>10</sub>

No âmbito deste trabalho, definiram-se como valores de corte as concentrações médias de 40 µg/m<sup>3</sup> e 20 µg/m<sup>3</sup>, que correspondem aos limites anuais para a protecção da saúde pública, a cumprir em 2005 e 2010, respectivamente (cf. Tabela 5.5).

Admitindo, por exemplo, que se pretendiam conhecer as probabilidades de excedência do valor de corte de 40 µg/m<sup>3</sup>, calcularam-se os mapas de probabilidades dos vários períodos. Apresenta-se um dos mapas obtidos para o período A, onde são visíveis as diferenças nas probabilidades de excedência, superiores nas áreas das estações com concentrações observadas mais elevadas (E1, E2 e E3).

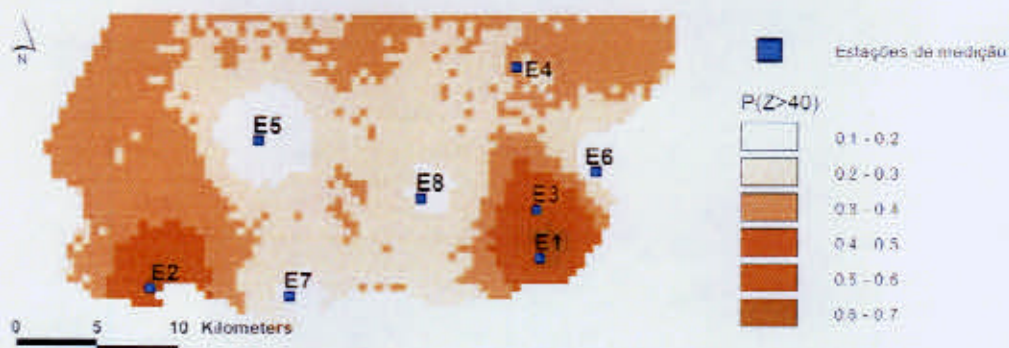


Figura 5.29- Um dos mapas de probabilidades para um valor de corte=40 µg/m<sup>3</sup>/dia (Período A)

Para permitir uma análise global dos resultados procedeu-se a uma síntese da informação contida nos mapas. Calculou-se o mapa de probabilidades médio,  $m(s)$  para cada período, onde cada valor da malha espacial é obtido por:

$$m(s) = \frac{\sum_{l=1}^L P[Z^l(s) > 40]}{L}$$

onde  $P[Z^l(s) > 40]$  representa a probabilidade estimada da variável  $Z$  exceder o valor de corte no ponto  $s$  na simulação  $l$ ,  $l=1, \dots, L$ . Dado que  $m(s)$  é uma representação atenuada de cada um dos mapas de probabilidades e não chega para caracterizar a sua distribuição, calculou-se para cada período, o mapa de variâncias,  $\sigma^2(s)$ , para identificar as áreas de maior incerteza ou variabilidade sobre a probabilidade de excedência estimada. O valor de  $\sigma^2(s)$  para um local qualquer da malha espacial é obtido por:

$$\sigma^2(s) = \frac{\sum_{l=1}^L [P[Z^l(s) > 40] - m(s)]^2}{L}$$

Para eliminar os efeitos de escala entre os mapas de probabilidades e de variâncias dos vários períodos, calculou-se ainda o mapa de coeficientes de variação. No Anexo F apresentam-se os mapas de probabilidades, de variâncias e de coeficientes de variação calculados. As distribuições de probabilidade obtidas apresentam formas aproximadamente gaussianas, mas centradas em valores diferentes. Neste contexto podem discriminar-se os resultados em três grupos: 1) os períodos A e D, que apresentaram valores médios semelhantes entre si, 2) os períodos C, E e G com valores médios e quartis semelhantes e, 3) o período F com uma probabilidade média de excedência muito elevada em relação aos restantes períodos. Apresentam-se na tabela seguinte as estatísticas descritivas dos mapas.

Período	Média	P.50	P.75	P.95	D. Padrão	Min.	Máx.
A	0.35	0.35	0.37	0.49	0.07	0.12	0.69
C	0.29	0.28	0.31	0.45	0.08	0.14	0.77
D	0.33	0.31	0.34	0.61	0.12	0.10	0.82
E	0.29	0.28	0.31	0.41	0.08	0.04	0.79
F	0.43	0.41	0.47	0.68	0.11	0.15	0.88
G	0.29	0.28	0.31	0.51	0.10	0.09	0.81

**Tabela 5.6-** Estatísticas descritivas dos mapas de probabilidade para um valor de corte de 40  $\mu\text{g}/\text{m}^3$

É junto às estações E1, E2 e E3 (estações de tráfego) que os mapas evidenciam maiores probabilidades de exceder o valor de corte, sendo este resultado coerente com os valores médios dos dados experimentais (ver Tabela 5.4).

#### **5.4.4 Síntese dos resultados obtidos**

Os mapas obtidos com recurso aos processos de simulação estocástica (para as variáveis da saúde e ambiente) permitiram quantificar a incerteza dos fenómenos observados, através de cenários extremos obtidos com simulações equiprováveis, dando origem a distribuições que, por exemplo, permitiram estimar probabilidades de exceder um valor de corte.

As simulações das variáveis da saúde reproduziram os padrões de continuidade espacial e parâmetros do risco relativo de incidência e permitiram identificar regiões onde a incerteza sobre o valor obtido é superior.

## **6 Análise das relações entre a qualidade do ar e a saúde**

### **6.1 Introdução**

A relação entre a qualidade do ar e a saúde foi analisada segundo uma perspectiva exploratória centrada nas características estatísticas e espaciais dos dados. Neste processo, o resultado da análise geoestatística sobre a qual se debruçou grande parte deste trabalho foi integrado em ferramentas de Sistemas de Informação Geográfica

(SIG) onde se executaram operações de análise espacial para processamento, visualização e modelação das correlações locais entre as variáveis estudadas. Para este efeito considerou-se uma área de estudo diferente da que se apresentou nos capítulos anteriores devido ao facto dos domínios espacial e temporal das variáveis (de qualidade do ar e da saúde) não serem totalmente coincidentes (conforme referido no sub-capítulo 4.3). Por este motivo, para analisar as relações num domínio comum a ambas as variáveis, o domínio espacial passou a ser definido por uma área mais limitada conforme a Figura 6.1, que abrange 47 escolas e 8 estações de medição. No domínio temporal foram considerados os dados referentes a 2002 que abrangem o período durante o qual foram realizados os inquéritos do ISAAC.



**Figura 6.1-Domínio espacial considerado na análise**

A utilização de uma ferramenta SIG com capacidades de armazenamento, manipulação e visualização de dados geográficos combinada com técnicas de análise exploratória, permitiu descrever e sintetizar padrões e relações espaciais evidenciadas nos dados. Compilaram-se os dados num SIG com capacidade para integrar e operar com modelos de dados vectorial relacional e raster (ou matricial).

A utilização conjunta de dados produzidos em ambos os modelos, flexibilizou os modos de representação dos dados geográficos e ampliou o leque de operações de análise espacial disponíveis. Os dados integrados foram referenciados de acordo com o sistema de coordenadas rectangulares planas referidas ao Ponto Fictício (SW do Cabo

de S. Vicente). Na tabela 6.1 sintetizam-se as principais características dos dados considerados nesta fase.

Tema	Descrição	Modelo de dados	Atributo(s)
Escolas	Dados do ISAAC indexados às escolas	Vectorial	- Coordenadas X,Y - Valor de SIR para cada variável
Escolas	Realizações de processos estocásticos obtidos com simulação sequencial directa	Matricial	- Valor de SIR para cada variável
PM <sub>10</sub>	Dados do Qualar referentes às estações da qualidade do ar da rede nacional	Vectorial	- Coordenadas X,Y - Valor médio de concentração de PM <sub>10</sub>
PM <sub>10</sub>	Realizações de processos estocásticos obtidos com simulação sequencial directa	Matricial	- Probabilidade de excedência de um valor de corte definido

**Tabela 6.1- Dados considerados na análise das relações entre as variáveis da saúde e qualidade do ar**

A utilização dos SIG incidiu fundamentalmente na inspecção visual e análise exploratória dos dados, com o objectivo de formular hipóteses sobre as relações entre as concentrações de PM<sub>10</sub> e as variáveis da saúde. Neste âmbito foram usadas operações de análise espacial, recorrendo às funções de distância ou de vizinhança para modelos de dados vectorial e de dados raster. Ambas as classes de funções, actualmente disponíveis na maioria das aplicações de SIG distribuídas no mercado, permitiram definir respectivamente áreas de influência em função de uma distância admitida, e atribuir um grau de exposição à poluição atmosférica com recurso a funções estatísticas. Os resultados obtidos permitiram calcular coeficientes de correlação entre ambas as variáveis e produzir mapas de correlações locais.

## 6.2 Análise e apresentação de resultados

### 6.2.1 Mapas de probabilidades para as variáveis da saúde e da qualidade do ar

A relação entre as variáveis foi analisada a partir dos mapas simulados, gerados no capítulo 5. Definiram-se valores de corte para ambas as variáveis, partindo do pressuposto de que nos cenários extremos a probabilidade de se evidenciarem relações com significado (estatístico) é superior. Optou-se por analisar 4 combinações de cenários (da saúde/qualidade do ar), com valores de corte iguais aos percentis 50 e 75 para as razões padronizadas de incidência, e 20 µg/m<sup>3</sup> e 40 µg/m<sup>3</sup> para a variável da qualidade do ar, conforme sintetizado na tabela seguinte.

Valores de corte	P.50	P.75
20 $\mu\text{g}/\text{m}^3$	Cenário 1	Cenário 2
40 $\mu\text{g}/\text{m}^3$	Cenário 3	Cenário 4

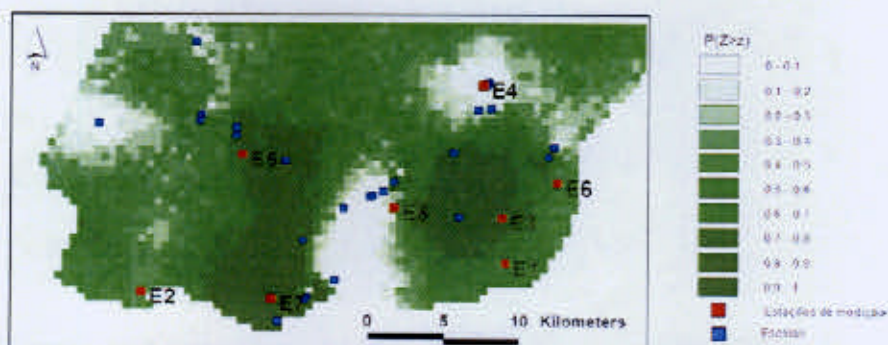
Tabela 6.2- Cenários escolhidos para análise das relações entre as variáveis da saúde e qualidade do ar

Os valores de corte definidos para as variáveis da saúde, foram obtidos a partir das distribuições dos dados experimentais (Tabela 6.3). Para as  $\text{PM}_{10}$  consideraram-se os valores de corte acima referidos.

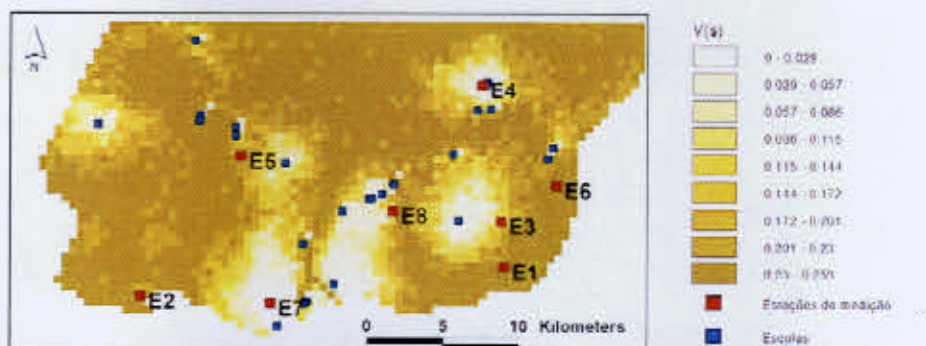
Variável	P.50	P.75
Ash_13	0.97	1.07
Ash_6	1.04	1.31
Hay_13	0.90	1.27
Rash_13	0.89	1.20
Prose_13	1.01	1.14

Tabela 6.3- Valores de corte definidos para as variáveis da saúde

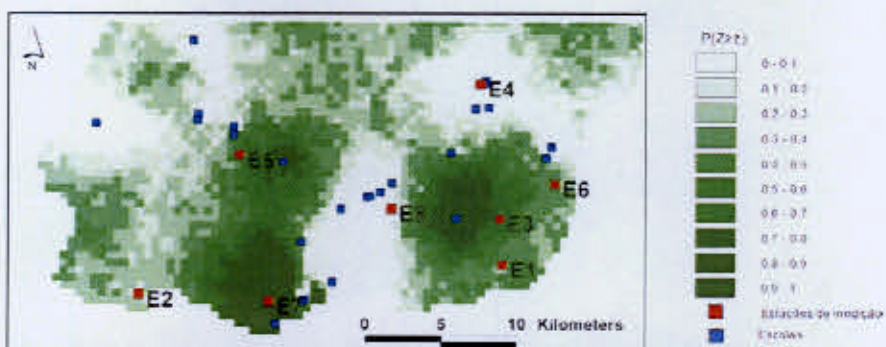
Os mapas da saúde simulados no capítulo 5.4.2 (ver Figura 5.15) permitiram estimar distribuições em cada ponto da malha espacial e obter os mapas de probabilidades de excedência dos valores de corte, para cada uma das variáveis. Com base nos 30 mapas de SIR simulados calcularam-se os mapas de probabilidade de excedência de um dado valor de corte, para cada variável da saúde, assim como os respectivos mapas de variância (Anexo G). Apresentam-se como exemplo, os resultados obtidos para a variável ash\_13, para ambos os valores de corte.



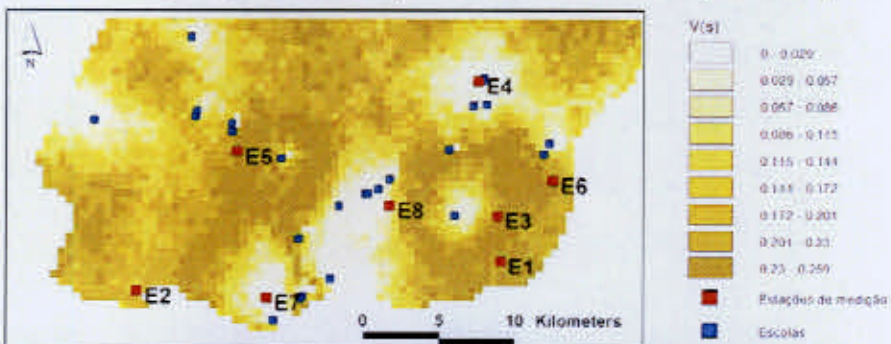
a. Ash\_13- Mapa de probabilidades para um valor de corte = 0.97 (percentil 50)



b. Ash\_13- Mapa de variâncias para um valor de corte = 0.97 (percentil 50)



c. Ash\_13- Mapa de probabilidades para um valor de corte = 1.07 (percentil 75)



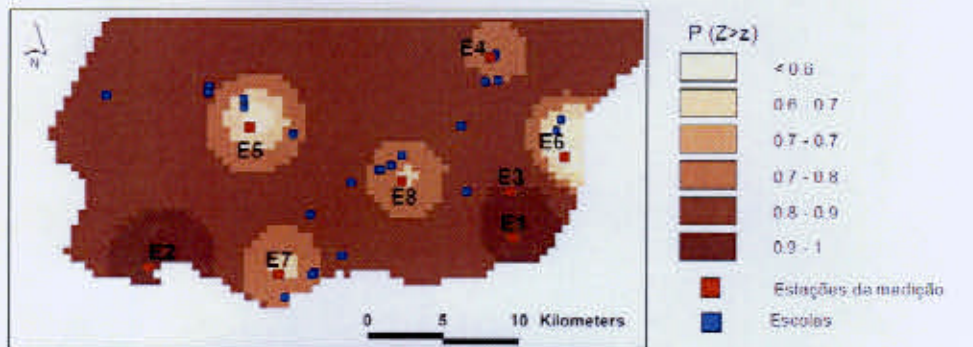
d. Ash\_13- Mapa de variâncias para um valor de corte = 1.07 (percentil 75)

**Figura 6.2- Mapas de probabilidades e de variâncias da variável Ash\_13 para os valores de corte definidos**

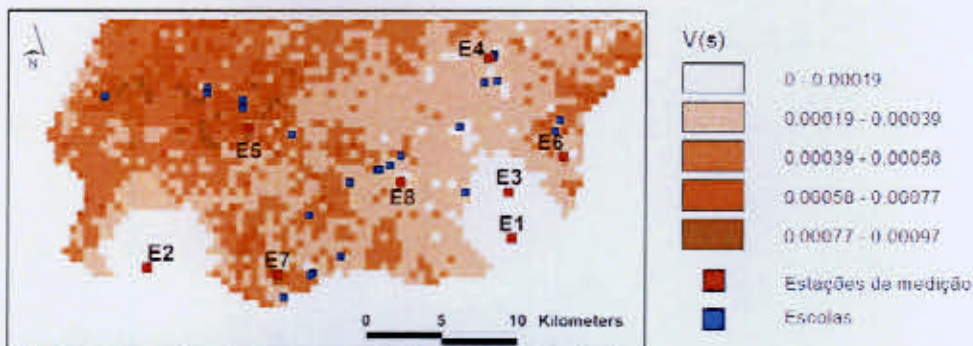
Na Figura 6.2 observa-se que as regiões mais prováveis de excederem os valores de corte localizam-se nas regiões onde a razão padronizada de incidência é, em média, superior. Verifica-se ainda uma clara tendência para um aumento de variabilidade em áreas sub-amostradas.

Os mapas de probabilidades da qualidade do ar utilizados neste capítulo referem-se ao período de Setembro de 2002 a Dezembro de 2002 (período G), pois coincidem com o período em que foram realizados os inquéritos do ISAAC (2002) e foram calculados

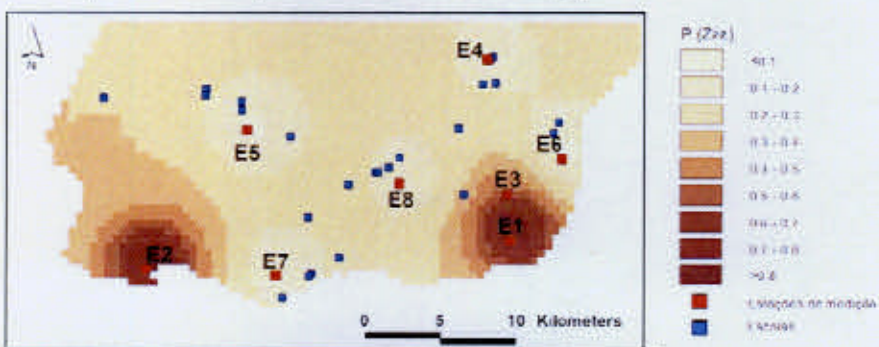
conforme metodologia descrita no capítulo 5.4.3. Apresentam-se a seguir os mapas de probabilidades e de variâncias obtidos para os valores de corte definidos.



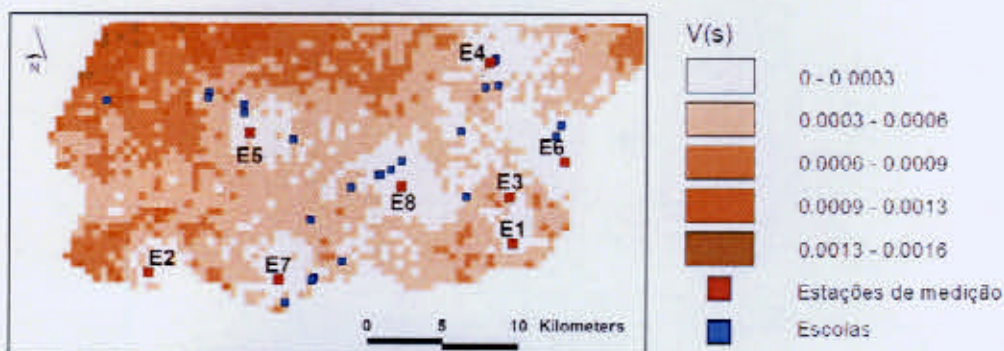
a. Mapa de probabilidades para um valor de corte  $z= 20 \mu\text{g}/\text{m}^3$



b. Mapa de variâncias para um valor de corte  $z= 20 \mu\text{g}/\text{m}^3$



c. Mapa de probabilidades para um valor de corte  $z= 40 \mu\text{g}/\text{m}^3$



d. Mapa de variâncias para um valor de corte  $z= 40 \mu\text{g}/\text{m}^3$

**Figura 6.3- Mapas de probabilidades e de variâncias de  $\text{PM}_{10}$  para os valores de corte definidos (Período G)**

As regiões mais prováveis de exceder a concentração de  $20 \mu\text{g}/\text{m}^3$  localizam-se nas áreas das estações de tráfego (E1, E2 e E3). Verifica-se também que, em grande parte da área de estudo (cerca de 93% da região) as suas probabilidades são superiores a 80%. O mapa de probabilidades para o valor de corte igual a  $40 \mu\text{g}/\text{m}^3$  apresenta uma distribuição semelhante, com as regiões mais prováveis de exceder o valor de corte, localizadas nas áreas das estações de tráfego. No entanto, devido à diferença entre os valores-limite definidos, a probabilidade da excedência ser superior a 80% apenas se verifica em 0,1% da área analisada. Note-se que a variabilidade local é muito baixa em ambos os valores de corte, o que significa que os mapas de probabilidades obtidos a partir das simulações foram semelhantes entre si para um dado valor de corte.

### 6.2.2 Mapas de correlações locais

Os mapas de correlações locais foram obtidos a partir dos mapas de probabilidades das variáveis da saúde e da qualidade do ar. Usaram-se as funções de distância disponíveis no SIG para definir áreas concêntricas em cada ponto dos mapas com um raio de 5000 metros, e para cálculo das correlações locais. Dado que na revisão crítica da literatura não se encontraram valores consensuais para os *buffers* a considerar, a escolha do raio de 5000 metros baseou-se numa análise prévia e qualitativa, de resultados observados para *buffers* de 2500, 5000 e 10000 metros. Remete-se para anexo (Anexo H) uma exposição sobre o modo como se abordou esta questão.

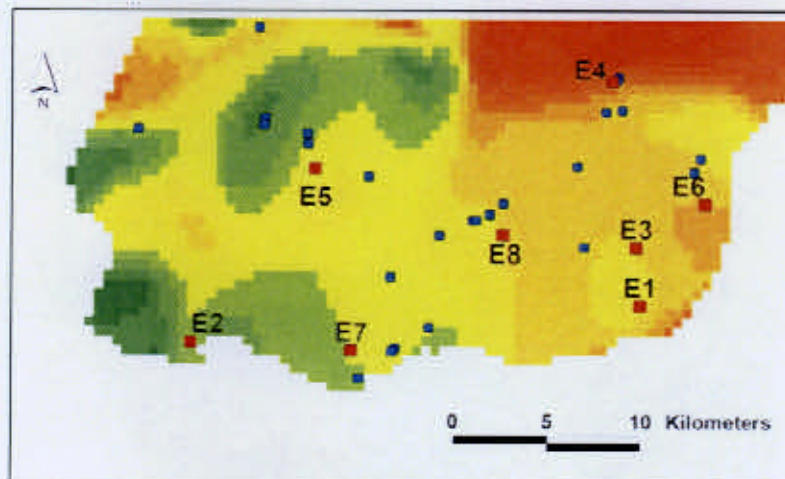
As correlações locais foram obtidas para cada ponto da região pelo coeficiente de correlação de Pearson ( $r$ ):

$$r = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

onde  $Cov(X,Y)$  representa a covariância e  $\sigma_x$  e  $\sigma_y$  representam os desvios padrão das variáveis X e Y.

Este coeficiente, é uma quantidade que indica a qualidade do ajustamento dos dados (independente da origem e da escala), segundo o critério dos mínimos quadrados, permitindo neste caso, pôr em evidência relações entre as variáveis da saúde e o factor de risco analisado.

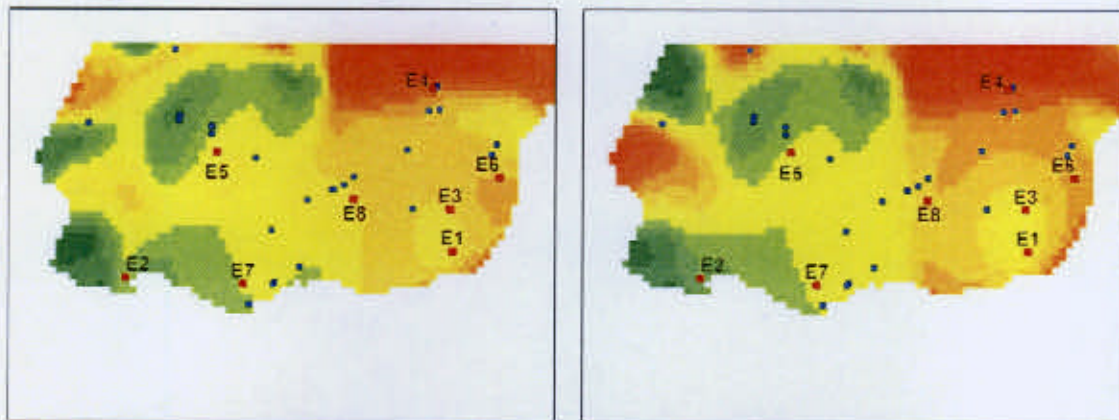
O mapa de correlações locais representado na Figura 6.4 é um exemplo ilustrativo dos resultados obtidos. Em cada ponto do mapa está representado o valor do coeficiente de correlação entre os mapas de excedência da razão padronizada de incidência de ash<sub>13</sub> (valor de corte=1.07) e da concentração de PM<sub>10</sub> (valor de corte=20 µg/m<sup>3</sup>), para um buffer de 5000 metros.





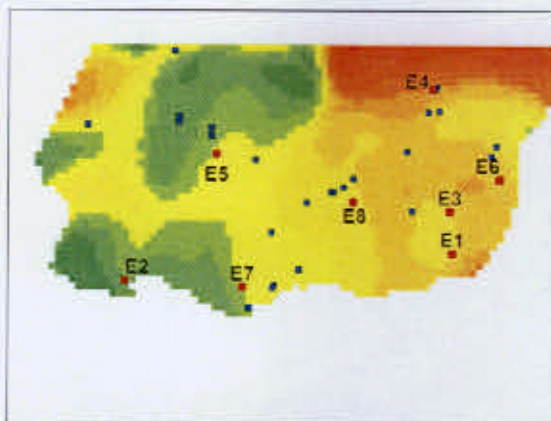
**Figura 6.4- Mapa de correlações locais entre a excedência da razão padronizada de incidência de ash\_13 = 1.07 e da concentração de PM<sub>10</sub> = 20 µg/m<sup>3</sup>**

Os mapas de correlações locais foram calculados para todas as variáveis e valores de corte considerados, e para cada variável da saúde obtiveram-se quatro mapas de correlações locais, conforme se apresenta na figura seguinte.

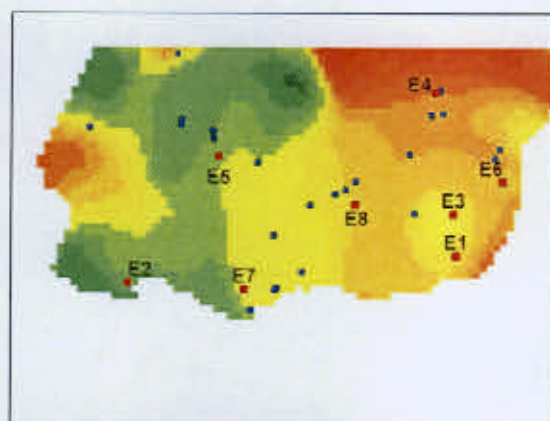


**a. Ash\_13 (P75) e PM10 (20 µg/m<sup>3</sup>)**

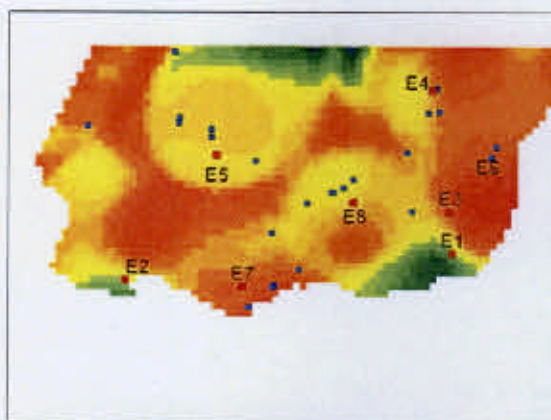
**b. Ash\_13 (P75) e PM10 (40 µg/m<sup>3</sup>)**



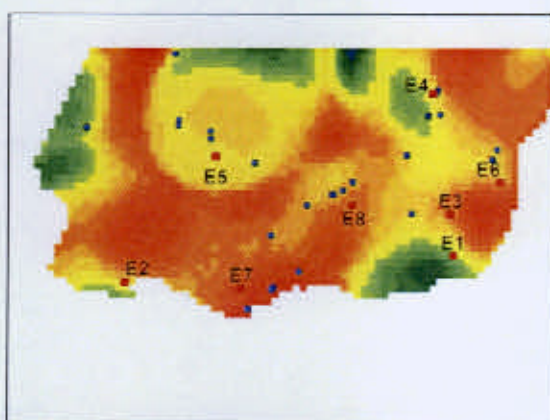
c. Ash\_13 (P50) e PM10 ( $20 \mu\text{g}/\text{m}^3$ )



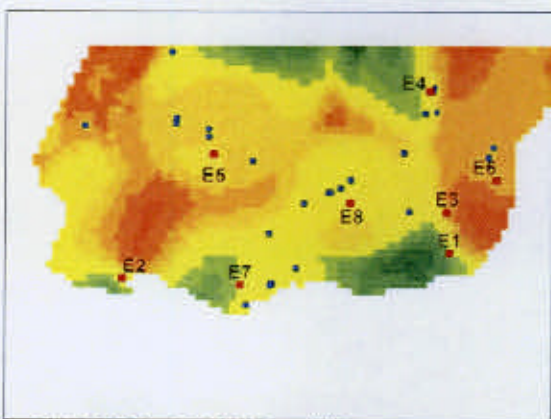
d. Ash\_13 (P50) e PM10 ( $40 \mu\text{g}/\text{m}^3$ )



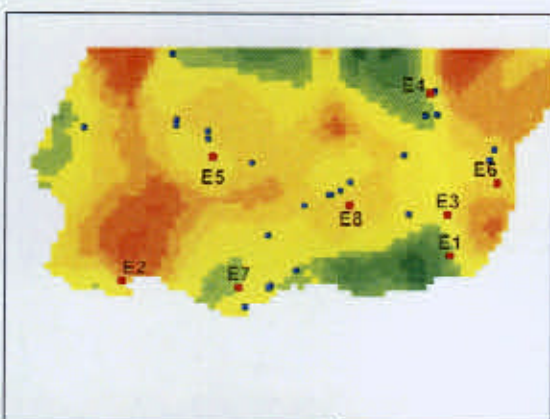
e. Ash\_6 (P75) e PM10 ( $20 \mu\text{g}/\text{m}^3$ )



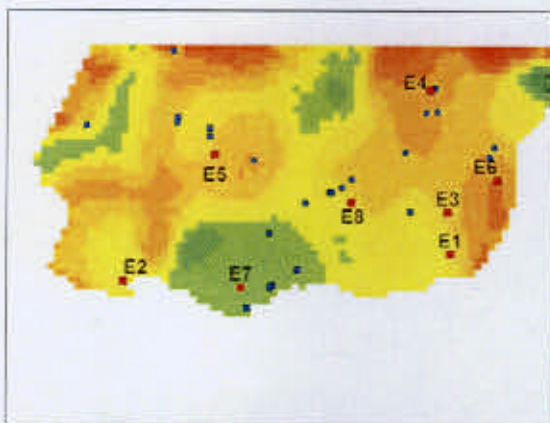
f. Ash\_6 (P75) e PM10 ( $40 \mu\text{g}/\text{m}^3$ )



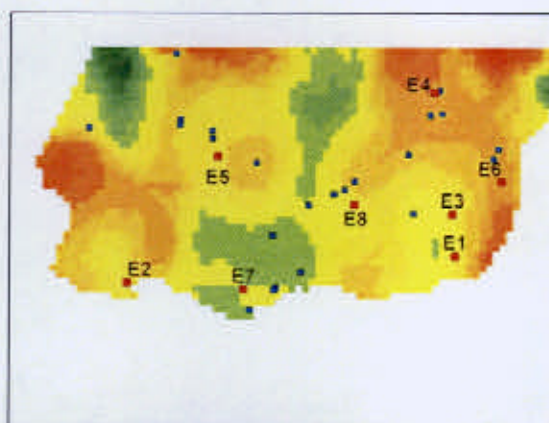
g. Ash\_6 (P50) e PM10 ( $20 \mu\text{g}/\text{m}^3$ )



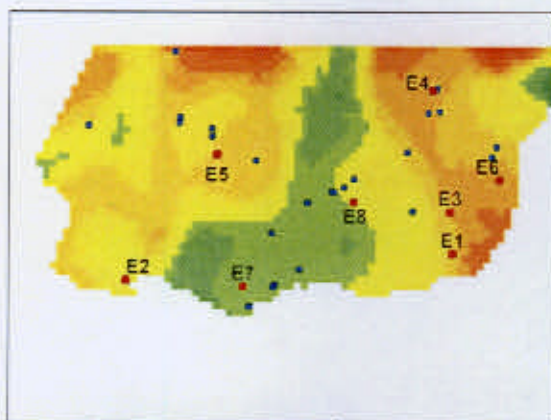
h. Ash\_6 (P50) e PM10 ( $40 \mu\text{g}/\text{m}^3$ )



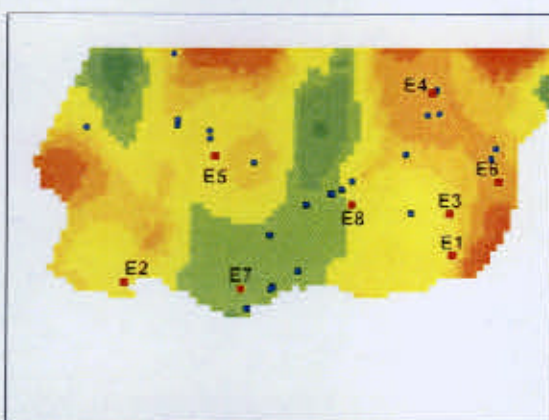
i. Hay\_13 (P75) e PM10 (20 µg/m<sup>3</sup>)



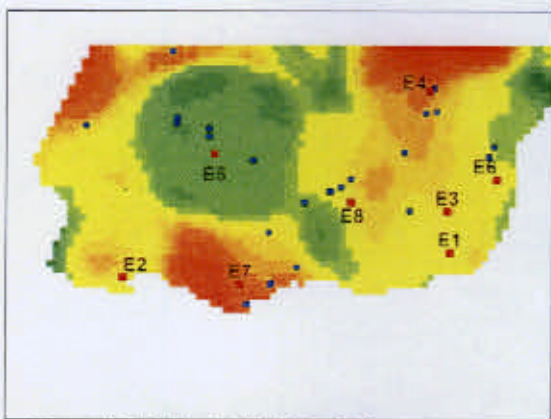
j. Hay\_13 (P75) e PM10 (40 µg/m<sup>3</sup>)



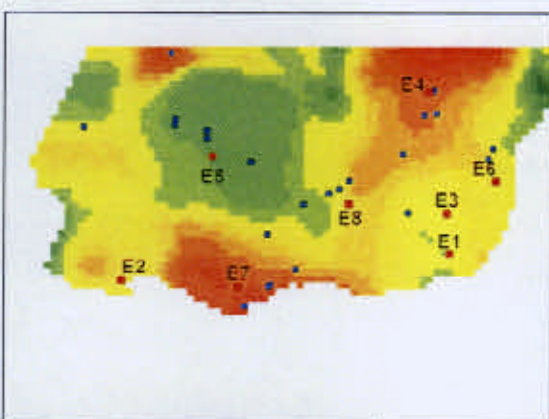
l. Hay\_13 (P50) e PM10 (20 µg/m<sup>3</sup>)



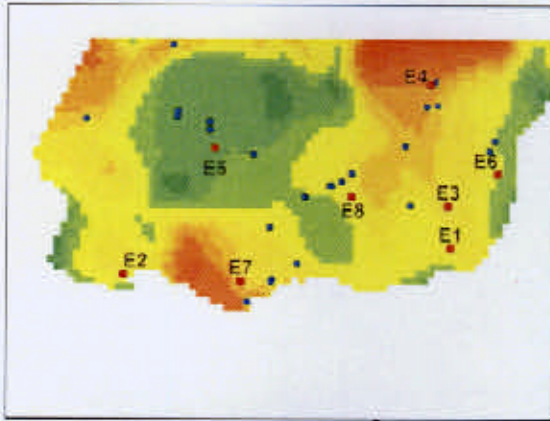
m. Hay\_13 (P50) e PM10 (40 µg/m<sup>3</sup>)



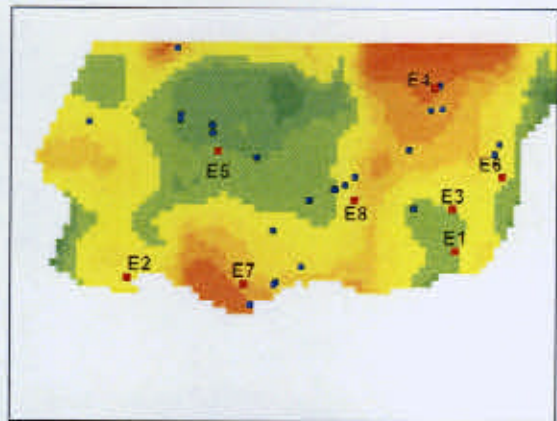
n. Pnose\_13 (P75) e PM10 (20 µg/m<sup>3</sup>)



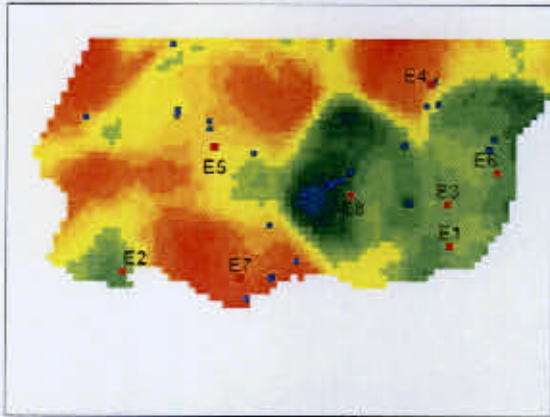
o. Pnose\_13 (P75) e PM10 (40 µg/m<sup>3</sup>)



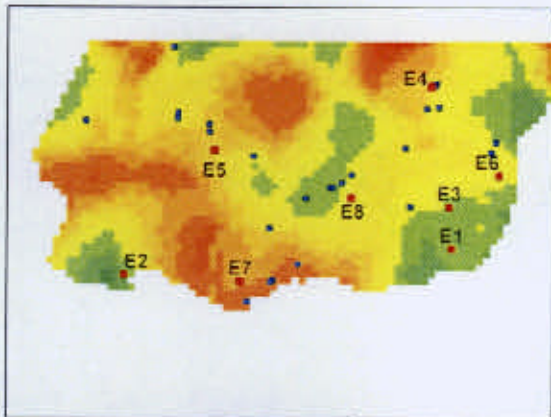
p. Pnose\_13 (P50) e PM10 ( $20 \mu\text{g}/\text{m}^3$ )



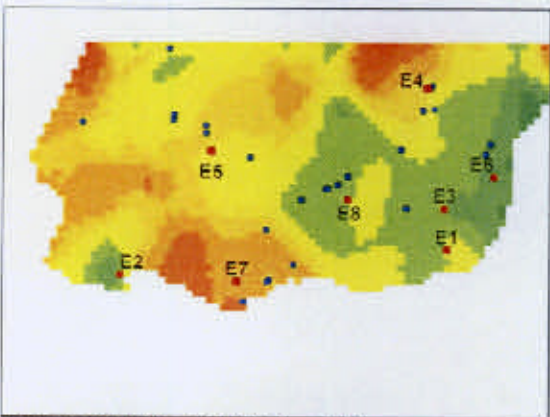
q. Pnose\_13 (P50) e PM10 ( $40 \mu\text{g}/\text{m}^3$ )



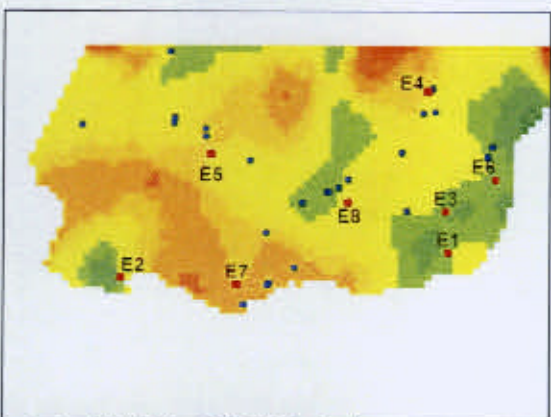
r. Rash\_13 (P75) e PM10 ( $20 \mu\text{g}/\text{m}^3$ )



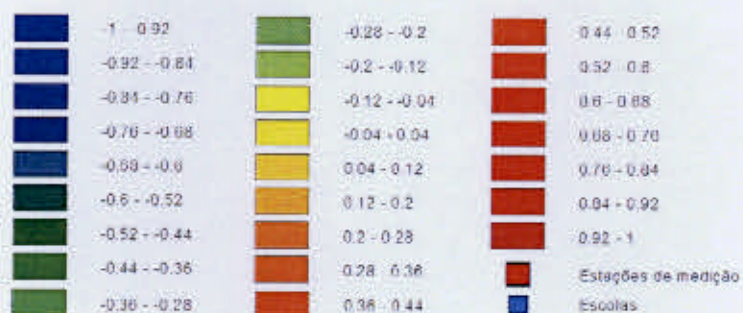
s. Rash\_13 (P75) e PM10 ( $40 \mu\text{g}/\text{m}^3$ )



t. Rash\_13 (P50) e PM10 ( $20 \mu\text{g}/\text{m}^3$ )



u. Rash\_13 (P50) e PM10 ( $40 \mu\text{g}/\text{m}^3$ )



**Figura 6.5- Mapas de correlações locais obtidos a. Ash\_13 (P75) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), b. Ash\_13 (P75) e PM10 (40  $\mu\text{g}/\text{m}^3$ ), c. Ash\_13 (P50) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), d. Ash\_13 (P50) e PM10 (40  $\mu\text{g}/\text{m}^3$ ), e. Ash\_6 (P75) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), f. Ash\_6 (P75) e PM10 (40  $\mu\text{g}/\text{m}^3$ ), g. Ash\_6 (P50) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), h. Ash\_6 (P50) e PM10 (40  $\mu\text{g}/\text{m}^3$ ), i. Hay\_13 (P75) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), j. Hay\_13 (P75) e PM10 (40  $\mu\text{g}/\text{m}^3$ ), l. Hay\_13 (P50) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), m. Hay\_13 (P50) e PM10 (40  $\mu\text{g}/\text{m}^3$ ), n. Pnose\_13 (P75) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), o. Pnose\_13 (P75) e PM10 (40  $\mu\text{g}/\text{m}^3$ ), p. Pnose\_13 (P50) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), q. Pnose\_13 (P50) e PM10 (40  $\mu\text{g}/\text{m}^3$ ), r. Rash\_13 (P75) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), s. Rash\_13 (P75) e PM10 (40  $\mu\text{g}/\text{m}^3$ ), t. Rash\_13 (P50) e PM10 (20  $\mu\text{g}/\text{m}^3$ ), u. Rash\_13 (P50) e PM10 (40  $\mu\text{g}/\text{m}^3$ ).**

Para a variável ash\_13 não se evidenciaram diferenças significativas entre a variação da razão padronizada de incidência e a do factor de risco para os diferentes valores de corte definidos. Todos os mapas da variável sugerem uma correlação entre a asma e o factor de risco na região oriental da área de estudo.

Para a variável ash\_6 verificaram-se diferentes resultados em função da variação dos valores de corte dos percentis da razão de incidência padronizada. As correlações locais foram mais evidentes nos mapas com probabilidades de exceder o percentil 75 da razão de incidência padronizada, ao longo dos limites Sul e Este da região de estudo.

Na variável hay\_13 não se evidenciaram diferenças significativas para os diferentes valores de corte definidos. De uma forma geral os mapas apresentam correlações mais elevadas na região oriental da área de estudo, em particular numa faixa localizada nas proximidades das estações de medição E1, E3, E4 e E6.

Nos mapas da variável pnose\_13 evidenciaram-se algumas diferenças entre as probabilidades de exceder os percentis 75 e 50 da razão de incidência padronizada. Essas diferenças foram mais evidentes na região de Oeiras (onde se situa a estação

E7), com coeficientes de correlação mais elevados para o percentil 75 (valor de corte de *prose\_13*).

A variável *rash\_13* evidenciou um comportamento semelhante ao da variável *prose\_13*, onde a variação dos valores de corte dos percentis da razão de incidência padronizada conduziu a mapas com correlações mais elevadas, nomeadamente na região de Oeiras.

Em termos globais, os mapas de correlação locais revelaram-se um instrumento que permitiu por em evidência variações locais sobre a relação entre a razão de incidência padronizada e a concentração de  $PM_{10}$ . Os diferentes resultados obtidos foram mais evidentes fazendo variar os valores de corte da SIR do que os valores de corte das concentrações de partículas. O pressuposto de que nos cenários extremos a probabilidade de se evidenciarem relações com significado (estatístico) é superior também se verificou neste exercício, na medida em que as correlações locais foram sistematicamente superiores (e positivas) para o valor de corte de SIR igual a 75.

## 7 Conclusões

O principal objectivo deste trabalho foi apresentar uma metodologia transversal às áreas do ambiente, da saúde e da geoestatística, suportadas em tecnologias de informação geográfica, para descrever a distribuição espacial de fenómenos da saúde e de factores de risco ambientais que possam contribuir para a formulação de hipóteses sobre a sua etiologia. Abordou-se em particular a aplicação de uma metodologia assente na análise da relação entre as estruturas de continuidade espacial da incidência (rácio padronizado) de uma doença e a de um factor de risco ambiental.

A metodologia usada centrou-se na utilização de modelos lineares de Krigagem para a estimação e processos de simulação estocástica geoestatística. Para relacionar as variáveis da saúde e da qualidade do ar, optou-se por um modelo de simulação estocástica proposto por Soares (2001). Calcularam-se correlações locais de valores de corte definidos para um raio de 5000 metros e obtiveram-se mapas de correlações locais que apresentaram, em alguns casos e em algumas regiões, correlações significativas com o factor de risco.

Os resultados mais significativos deste trabalho foram os seguintes:

- i. A distribuição espacial das escolas e das estações de monitorização colocou algumas dificuldades à aplicação da metodologia. As escolas caracterizaram-se por uma disposição em aglomerados com vastas áreas sub-amostradas, e incidências (rácios) díspares muito próximas entre si; as estações de monitorização caracterizaram-se por fornecer grande quantidade de dados no tempo mas pouca no espaço,
- ii. A utilização de um método de simulação estocástica foi de grande utilidade porque permitiu simular o comportamento das características espaciais das variáveis e gerar cenários equiprováveis. Permitiu ainda desenvolver uma análise probabilista onde se estimaram probabilidades locais de se excederem os valores de corte definidos,
- iii. A utilização de tecnologias de informação geográfica para integração dos dados e obtenção dos mapas revelou-se bastante eficiente dada a possibilidade de se analisar uma área vasta com relativa facilidade e

rapidez, permitindo ainda estabelecer cenários e condições alternativas, de resposta quase imediata,

- iv. A aplicação da metodologia desenvolvida ao caso de estudo apresentado, permitiu identificar correlações lineares locais elevadas nas regiões orientais e sul da área de estudo, nomeadamente para probabilidades de SIR extremas altas, superiores ao percentil 75 (em particular no caso da variável ash\_6);

## 8 Referências bibliográficas

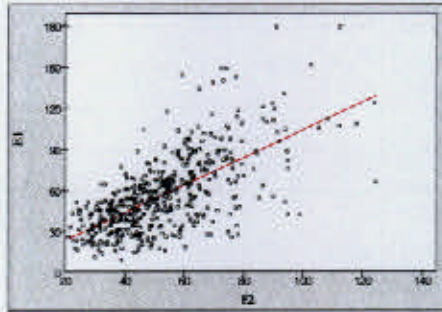
- Aronoff, S. (1989). Geographic information systems: a management perspective. Ottawa, WDL Publications.
- Bailey, T. C. e A. C. Gatrell (1995). Interactive Spatial Data Analysis. Essex, Longman.
- Berke, O. (2004). "Exploratory disease mapping: kriging the spatial risk function from regional count data." International Journal of Health Geographics **3**(1): 18.
- Beyea, J. e M. Hatch (1999). "Geographic Exposure Modeling: A Valuable Extension of Geographic Information Systems for Use in Environmental Epidemiology." Environmental Health Perspectives Supplements **107**(S1): 181-190.
- Bohning, D., U. Malzahn, et al. (2002). "Some general points in estimating heterogeneity variance with the DerSimonian-Laird estimator." Biostat **3**(4): 445-457.
- Boubel, R., D. Fox, et al. (1994). Fundamentals of Air Pollution. Londres, Academic Press, Inc.
- Burrough, P. (1986). Principles of GIS. Utrecht, University of Utrecht Press.
- Carrat, F. e A. Valleron (1992). "Epidemiologic mapping using the "kriging" method: application to an influenza-like illness epidemic in France." Am. J. Epidemiol. **135**(11): 1293-1300.
- Chrisman, N. (2002). Exploring Geographic Information Systems, John Wiley & Sons, Inc.
- Collins, S. (1998). Modelling Spatial Variations in Air Quality using GIS. GIS and Health. A. Gatrell e M. Loytonen. London, Taylor & Francis. **6**: 81-95.
- Douven, W. e H. Scholten (1995). Spatial analysis in health research. The Added Value of Geographical Information Systems in Public and Environmental Health. M. J. C. d. Lepper, H. Scholten e R. Stern. Dordrecht, Kluwer Academic Publishers. **24**: 117-133.
- Dueker, K. J. (1979). "Land resource information systems: A review of fifteen years experience." Geo-Processing **1**: 105-128.
- Elliott, P., J. Cuzick, et al. (1992). Geographical & Environmental Epidemiology. New York, Oxford University Press.
- Elliott, P. e D. Wartenberg (2004). "Spatial Epidemiology: Current Approaches and Future Challenges." Environmental Health Perspectives **112**(9): 998-1006.

- Fauroux, B., M. Sampil, et al. (2000). "Ozone: A trigger for hospital pediatric asthma emergency room visits." Pediatric Pulmonology **30**(1): 41.
- Gatrell, A. e M. Loytonen, Eds. (1998). GIS and Health. GISDATA. London, Taylor & Francis.
- Gehring, U., J. Cyrus, et al. (2002). "Traffic-related air pollution and respiratory health during the first 2 yrs of life." Eur Respir J **19**(4): 690-698.
- Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. New York, Oxford University Press Inc.
- Goovaerts, P. e G. Jacquez (2004). "Accounting for regional background and population size in the detection of spatial clusters and outliers using geostatistical filtering and spatial neutral models: the case of lung cancer in Long Island, New York." International Journal of Health Geographics **3**(1): 14.
- Haining, R. (1998). Spatial Statistics and the Analysis of Health Data. GIS and Health. A. Gatrell e M. Loytonen. London, Taylor & Francis. **6**: 29-47.
- Hansen, Hurwitz, et al. (1953). Sample Survey Methods and Theory. New York, John Wiley & Sons, Inc.
- Hertz-Picciotto, I. e B. Brunekreef (2001). "Environmental Epidemiology: Where We've Been and Where We're Going." Epidemiology **12**: 479-481.
- Hirsch, T., S. Weiland, et al. (1999). "Inner city air pollution and respiratory health and atopy in children." Eur Respir J **14**(3): 669-677.
- Isaaks, E. e R. M. Srivastava (1989). Applied Geostatistics. New York, Oxford University Press, Inc.
- Jacquez, G. M. (2000). "Spatial analysis in epidemiology : Nascent science or a failure of GIS ?" Journal of Geographical Systems **2**: 91-97.
- Kulldorff, M. (2003). "Spatial Epidemiology: Methods and Applications. P. Elliott, J. C. Wakefield, N. G. Best and D. J. Briggs (eds), Oxford University Press, Oxford, 2000. No. of pages: xvii + 475. ISBN 0-19-262941-7." Statistics in Medicine **22**(14): 2381.
- Last, J. M. (2001). "A DICTIONARY OF EPIDEMIOLOGY." Am. J. Epidemiol. **154**(4): 389-.
- Lawson, A., A. Biggeri, et al. (2000). "Disease mapping models: an empirical evaluation." Statistics in Medicine **19**: 2217-2241.
- Lawson, A. B. (2001). "Disease map reconstruction." Statistics in Medicine **20**(14): 2183-2204.
- Matos, J. (2001). Fundamentos de Informação Geográfica. Lisboa, Lidel.

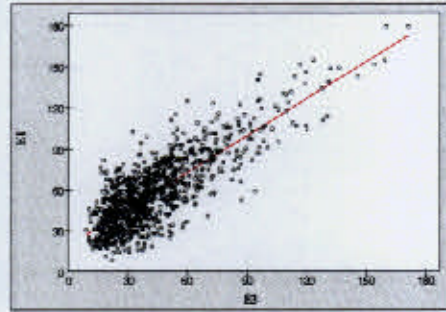
- Mausner, J. e S. Kramer (1984). Introdução à Epidemiologia. Lisboa, Fundação Calouste Gulbenkian.
- Nuckols, J., M. Ward, et al. (2003). "Using Geographic Information Systems for Exposure Assessment in Environmental Epidemiology Studies." Environmental Health Perspectives **112**(9): 1007-1015.
- Ostro, B. (2004). Outdoor air pollution: assessing the environmental burden of disease at national and local levels. Geneve, World Health Organization.
- Pascutto, C., J. C. Wakefield, et al. (2000). "Statistical issues in the analysis of disease mapping data." Statistics in Medicine **19**(17-18): 2493.
- Pinto, J. R. e M. M. d. Almeida (2003). A Criança Asmática no Mundo da Alergia. Lisboa, Euromédice.
- Sankoh, O. A., O. Berke, et al. (2002). Bayesian and GIS Mapping of Childhood Mortality in Rural Burkina Faso. Heidelberg, University of Heidelberg: 16.
- Seinfeld, J. H. (1986). Atmospheric chemistry and physics of air pollution, John Wiley & Sons, Inc.
- Serpa, C. (2003). Modelos Estocásticos Espaço-Temporais. Instituto Superior Técnico. Lisboa, Universidade Técnica de Lisboa: 226.
- Soares, A. (2000). Geoestatística para as Ciências da Terra e do Ambiente. Lisboa, IST Press.
- Soares, A. (2001). "Direct Sequential Simulation and Cosimulation." Mathematical Geology **33**(8): 911-926.
- Tamburini, G., O. Ehrenstein, et al. (2002). Children's health and environment: A review of evidence. Copenhagen, European Environment Agency (EEA) & World Health Organization (WHO): 223.
- Tsai, S. e C. Wen (1986). "A review of methodological issues of the standardized mortality ratio (SMR) in occupational cohort studies." Int. J. Epidemiol. **15**(1): 8-21.
- von Mutius, E., D. Sherrill, et al. (1995). "Air pollution and upper respiratory symptoms in children from East Germany." Eur Respir J **8**(5): 723-728.
- Westlake, A. (1995). Strategies for the use of geography in epidemiological analysis. The added value of geographical information systems in public and environmental health. M. J. C. d. Lepper, H. Scholten e R. Stern. Amsterdam, Kluwer Academic Publishers on behalf of the World Health Organization Regional Office for Europe. **24**: 135-144.

**ANEXO A**

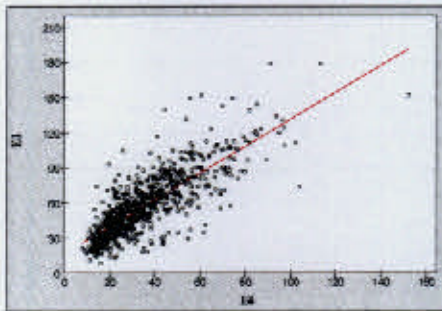
**Diagramas de dispersão das PM<sub>10</sub> - Período de 01/01/2001 a 31/12/2003**



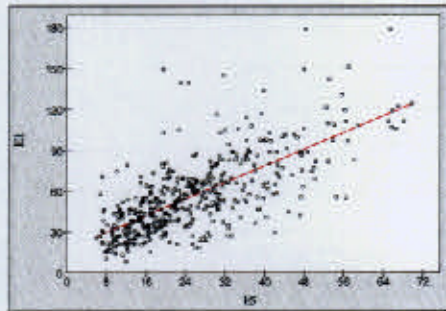
1. E1 vs E2



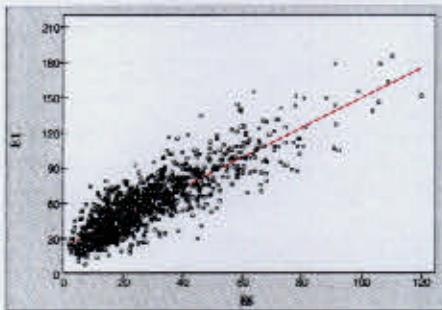
2. E1 vs E3



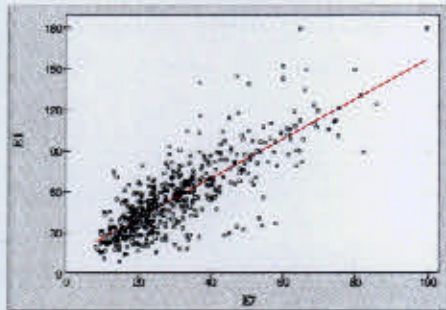
3. E1 vs E4



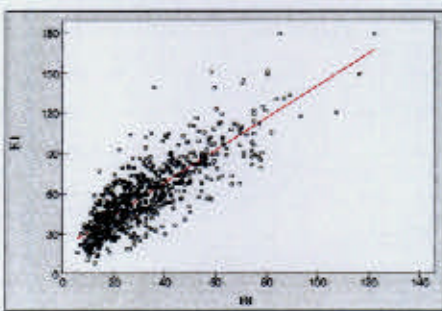
4. E1 vs E5



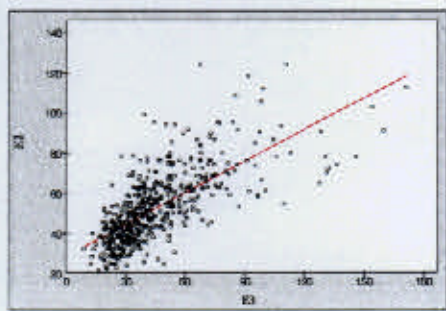
5. E1 vs E6



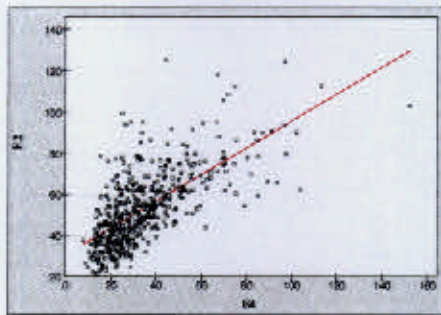
6. E1 vs E7



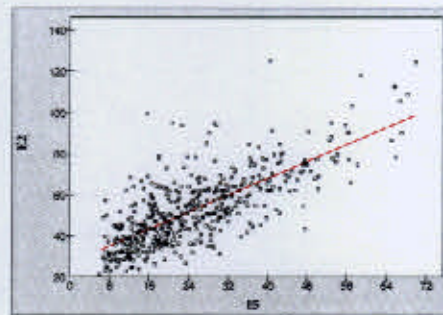
7. E1 vs E8



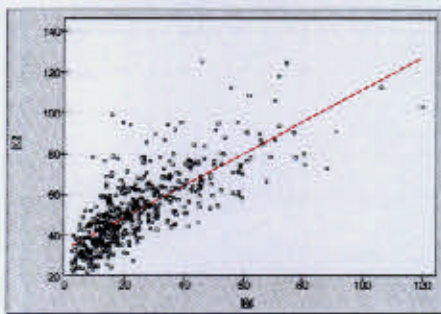
8. E2 vs E3



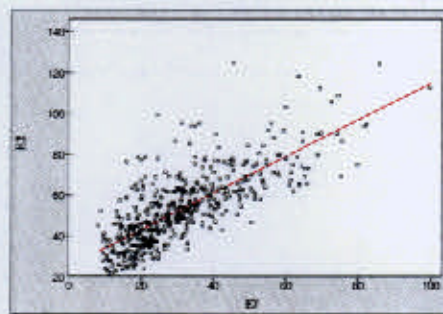
9. E2 vs E4



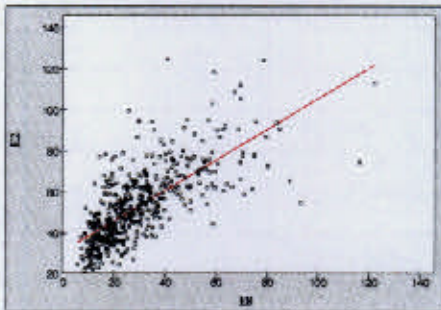
10. E2 vs E5



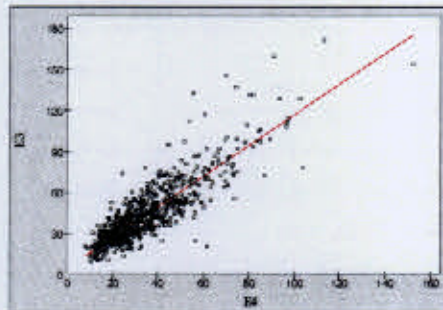
11. E2 vs E6



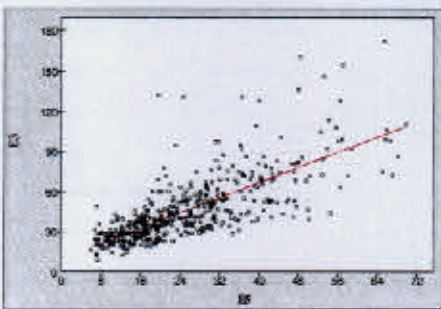
12. E2 vs E7



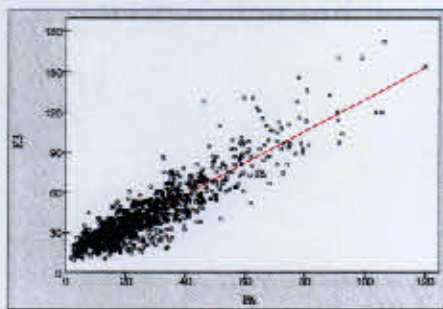
13. E2 vs E8



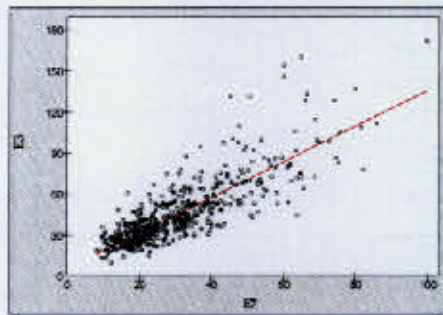
14. E3 vs E4



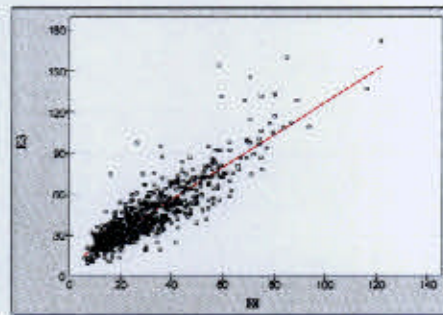
15. E3 vs E5



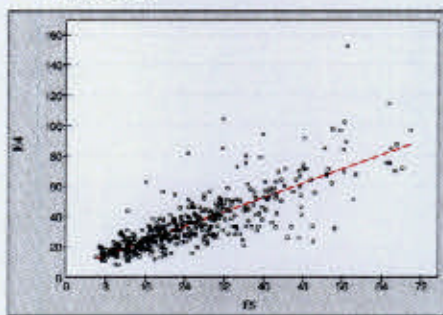
16. E3 vs E6



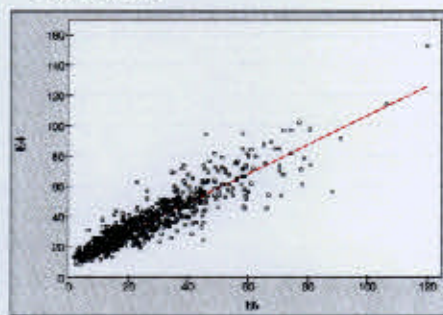
17. E3 vs E7



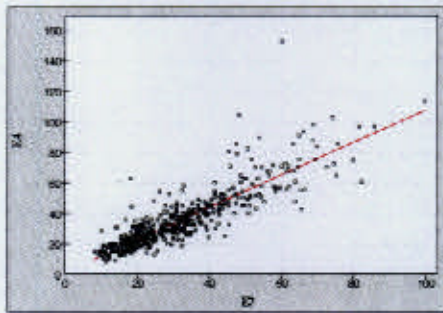
18. E3 vs E8



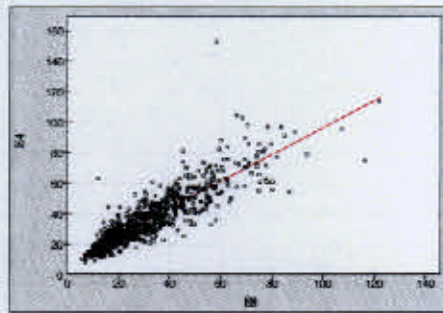
19. E4 vs E5



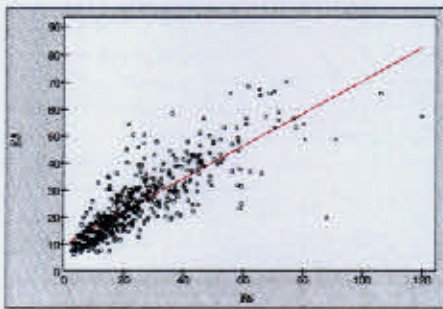
20. E4 vs E6



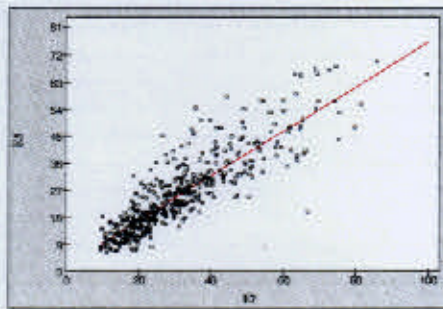
21. E4 vs E7



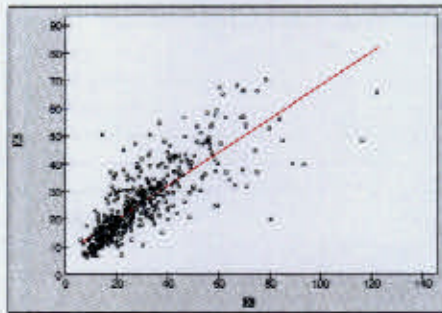
22. E4 vs E8



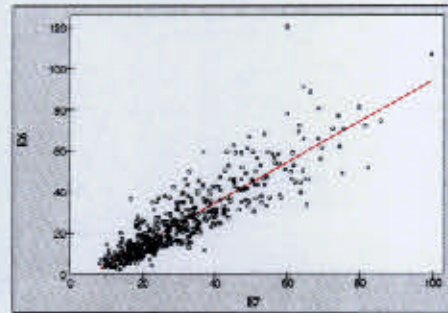
23. E5 vs E6



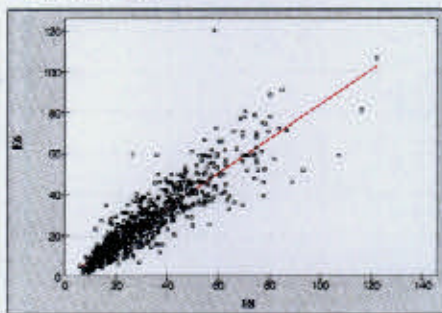
24. E5 vs E7



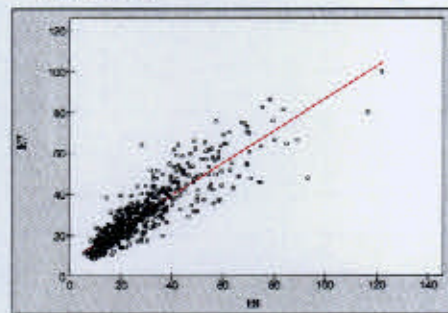
25. E5 vs E8



26. E6 vs E7



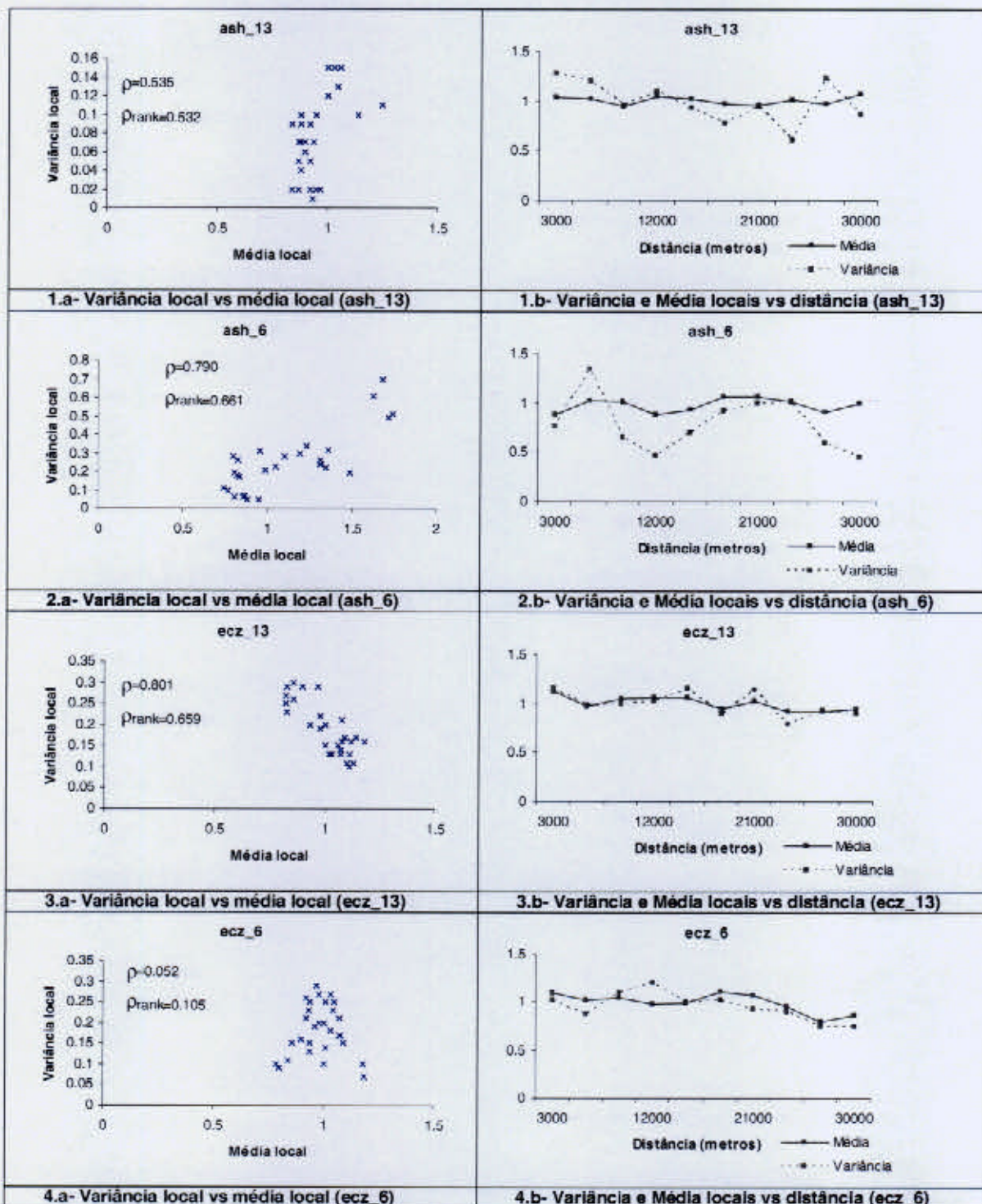
27. E6 vs E8

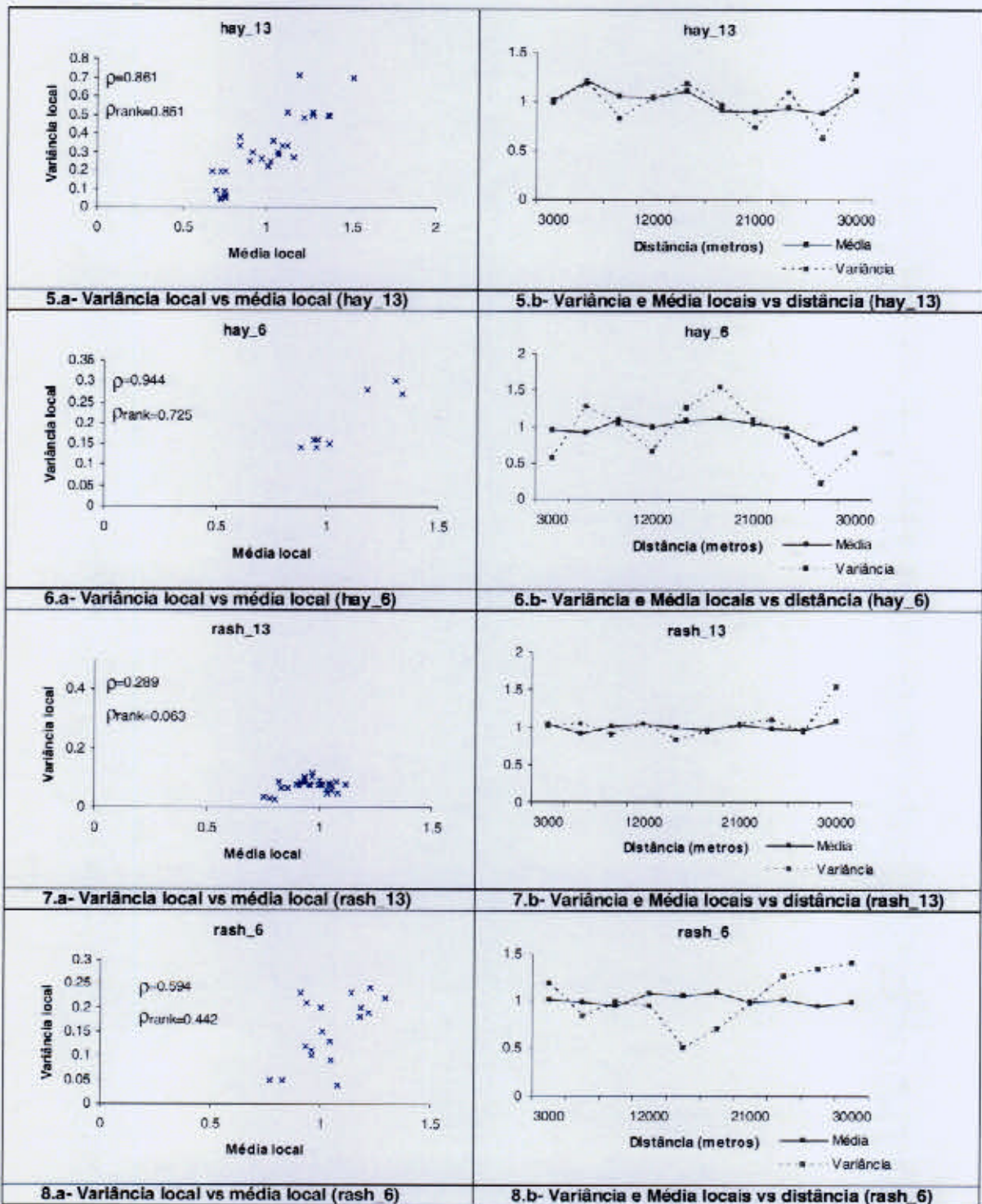


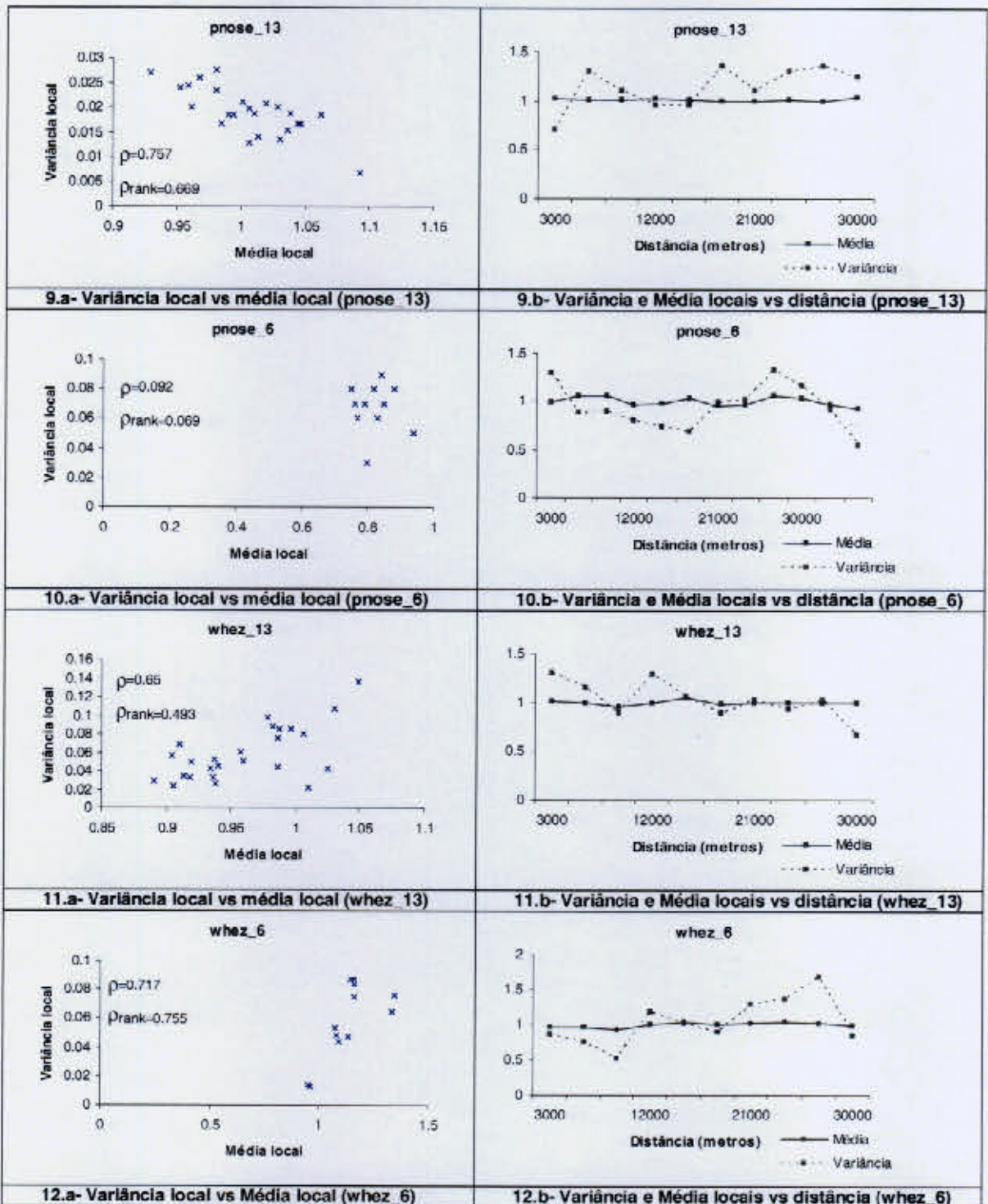
28. E7 vs E8

## **ANEXO B**

### **Análise do efeito proporcional nas variáveis da saúde**

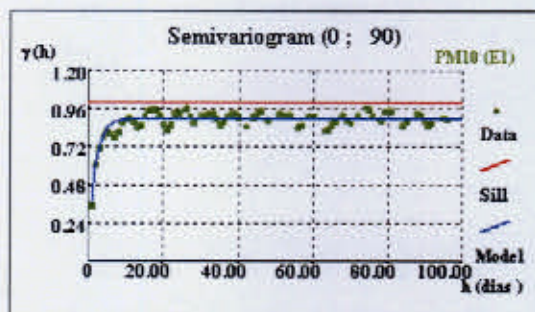




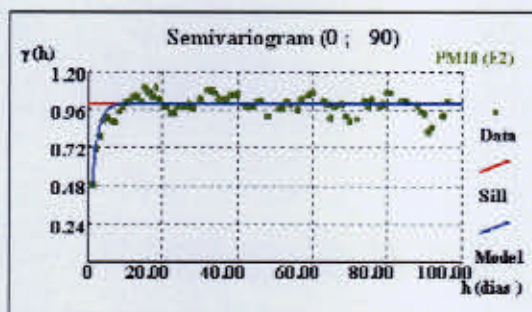


**ANEXO C**

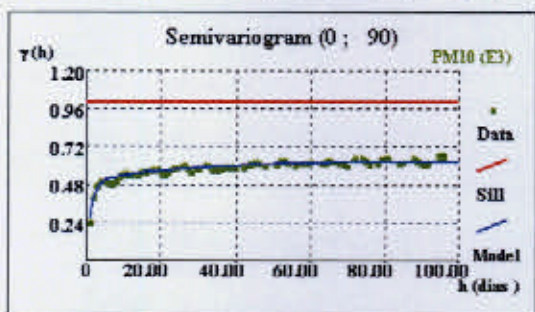
**Varlogramas temporais das  $PM_{10}$ - Período de 01/01/2001 a 31/12/2003**



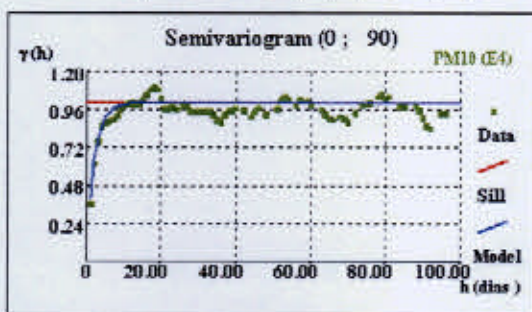
1. Modelo exponencial:  $c_0=0$ ;  $c_1=0.9$ ;  $a_1=5.5$  (E1)



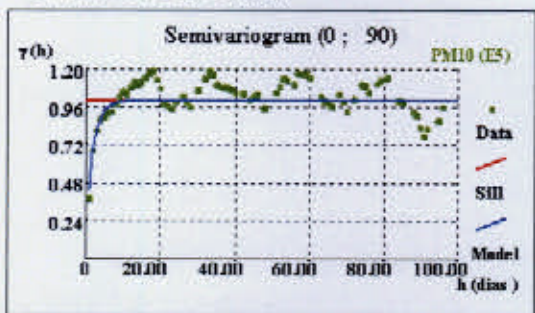
2. Modelo exponencial:  $c_0=0$ ;  $c_1=1$ ;  $a_1=4.5$  (E2)



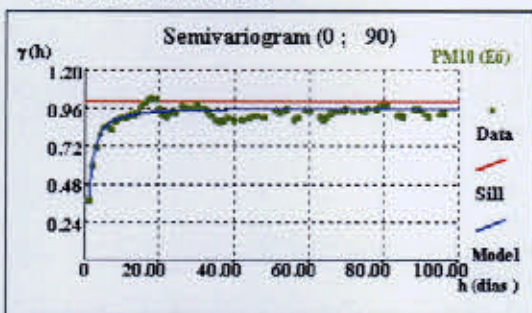
3. Modelo exponencial embricado:  $c_0=0$ ;  $c_1=0.5$ ;  $a_1=4$ ;  $c_2=0.13$ ;  $a_2=83$  (E3)



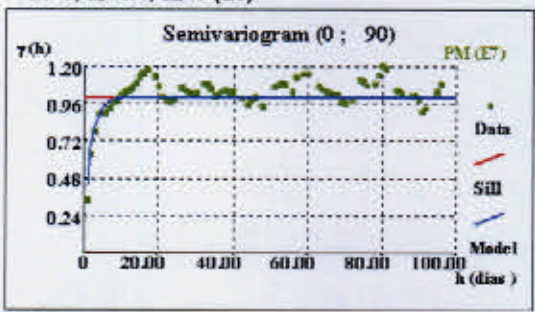
4. Modelo exponencial embricado:  $c_0=0$ ;  $c_1=0.8$ ;  $a_1=6.5$ ;  $c_2=0.2$ ;  $a_2=4$  (E4)



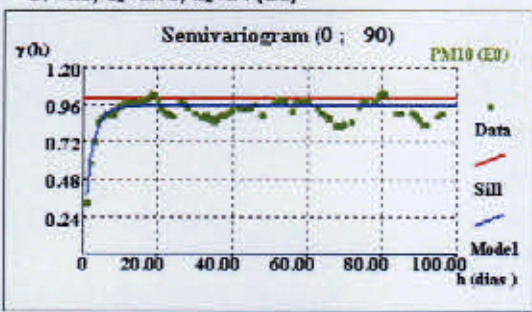
5. Modelo exponencial embricado:  $c_0=0$ ;  $c_1=0.9$ ;  $a_1=5$ ;  $c_2=0.1$ ;  $a_2=7$  (E5)



6. Modelo exponencial embricado:  $c_0=0$ ;  $c_1=0.8$ ;  $a_1=4.5$ ;  $c_2=0.15$ ;  $a_2=24$  (E6)



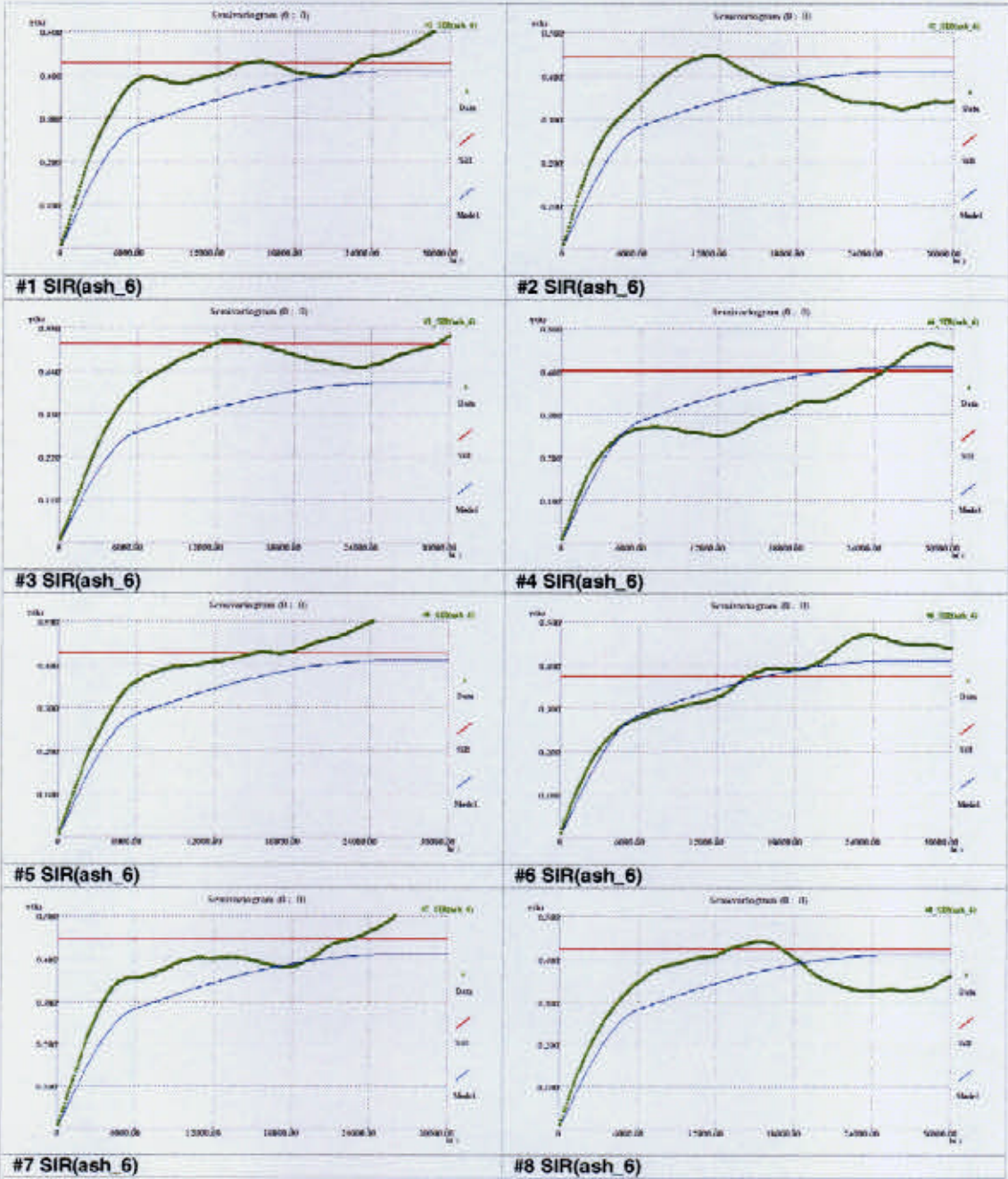
7. Modelo exponencial embricado:  $c_0=0$ ;  $c_1=0.8$ ;  $a_1=5$ ;  $c_2=0.2$ ;  $a_2=7$  (E7)

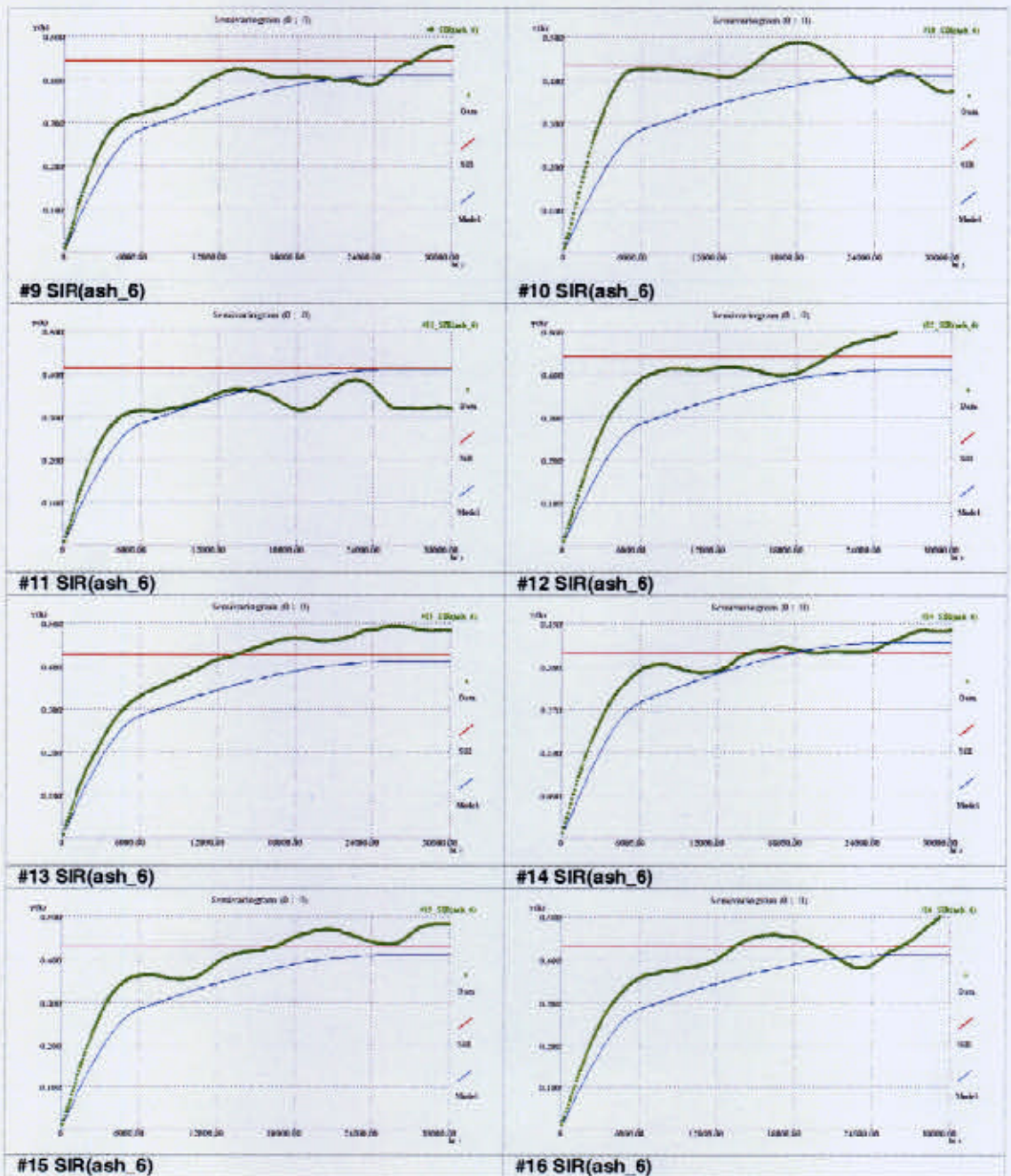


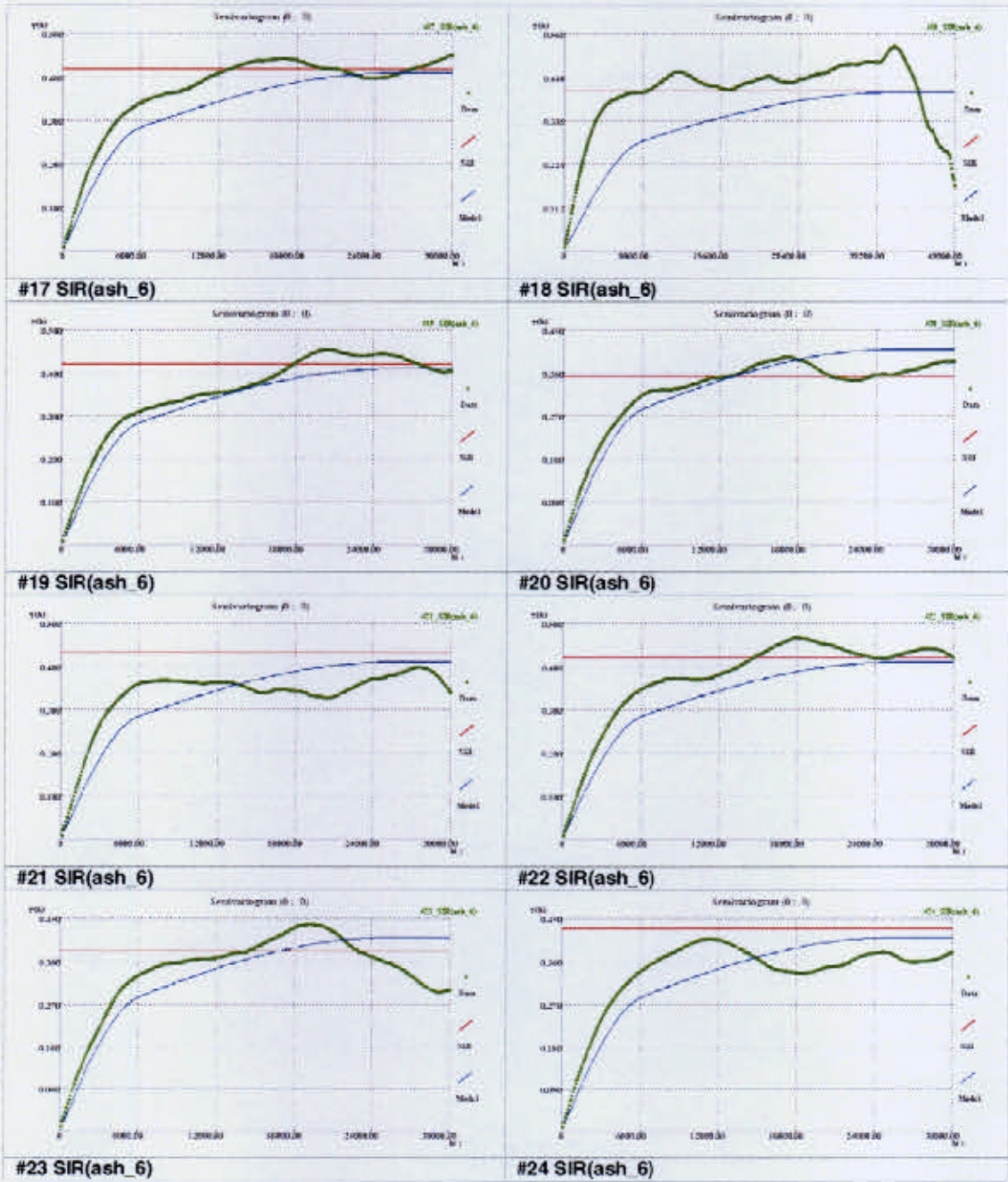
8. Modelo exponencial embricado:  $c_0=0$ ;  $c_1=0.88$ ;  $a_1=5.5$ ;  $c_2=0.07$ ;  $a_2=11$  (E8)

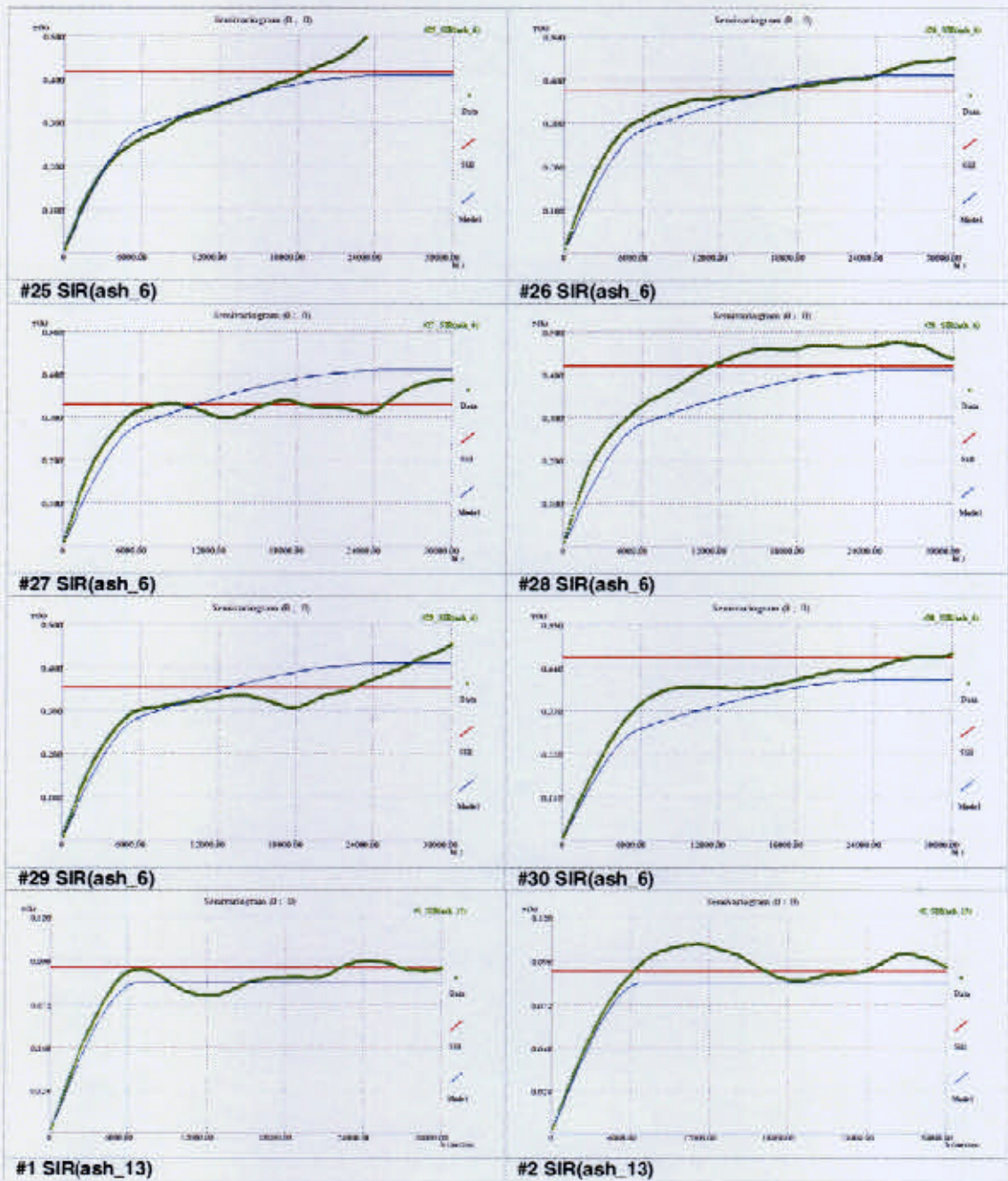
## **ANEXO D**

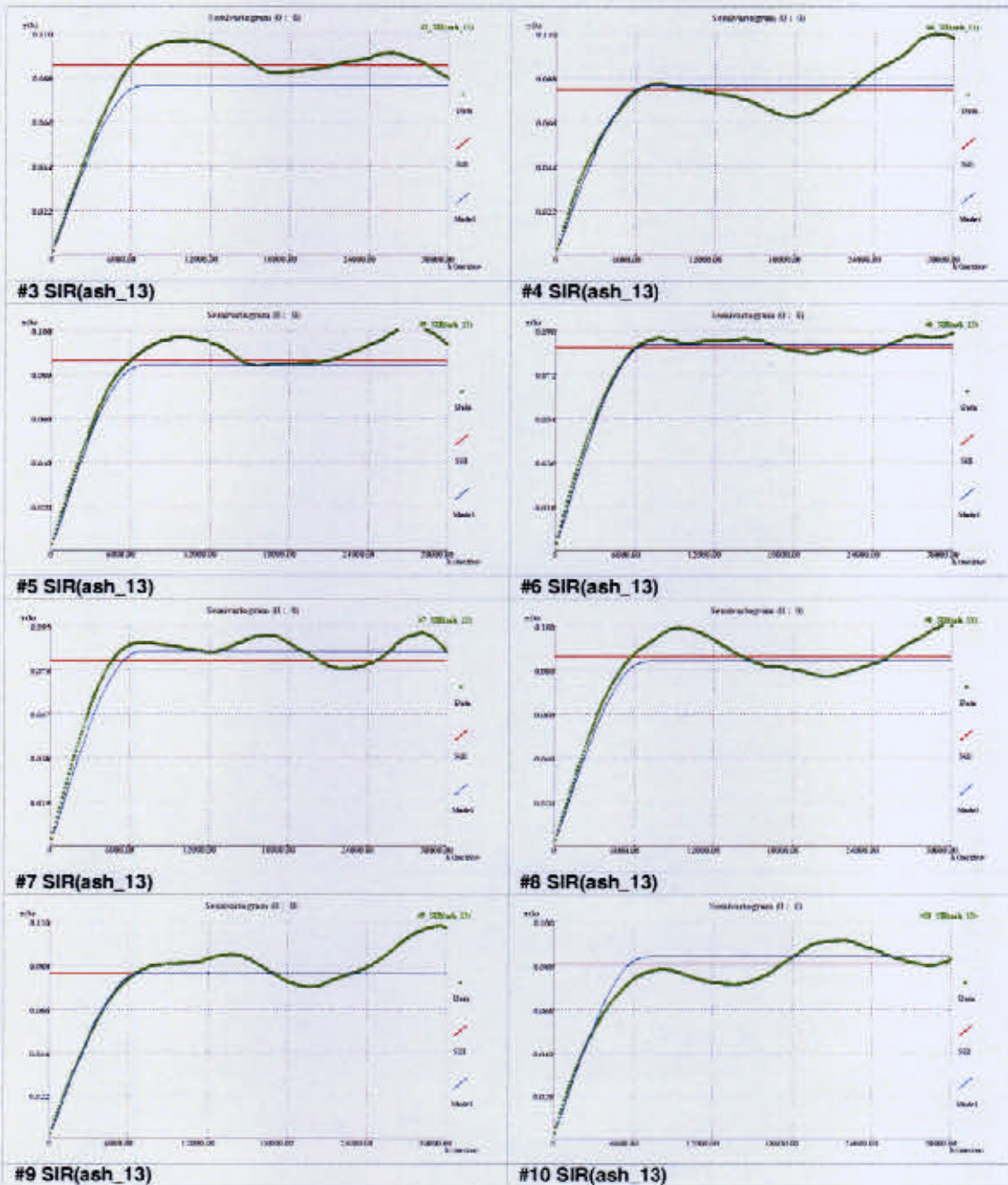
### **Variogramas das simulações das variáveis da saúde**

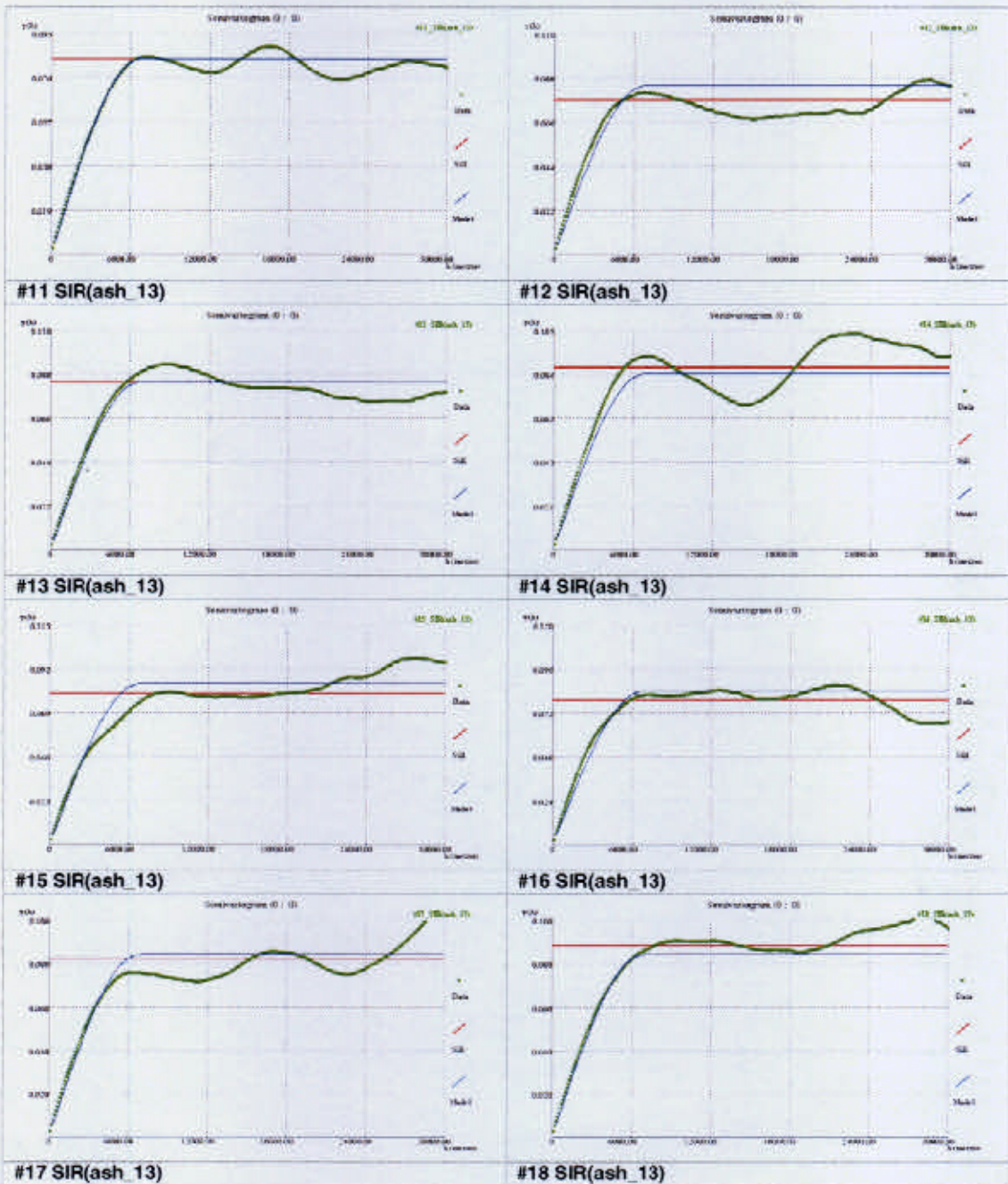


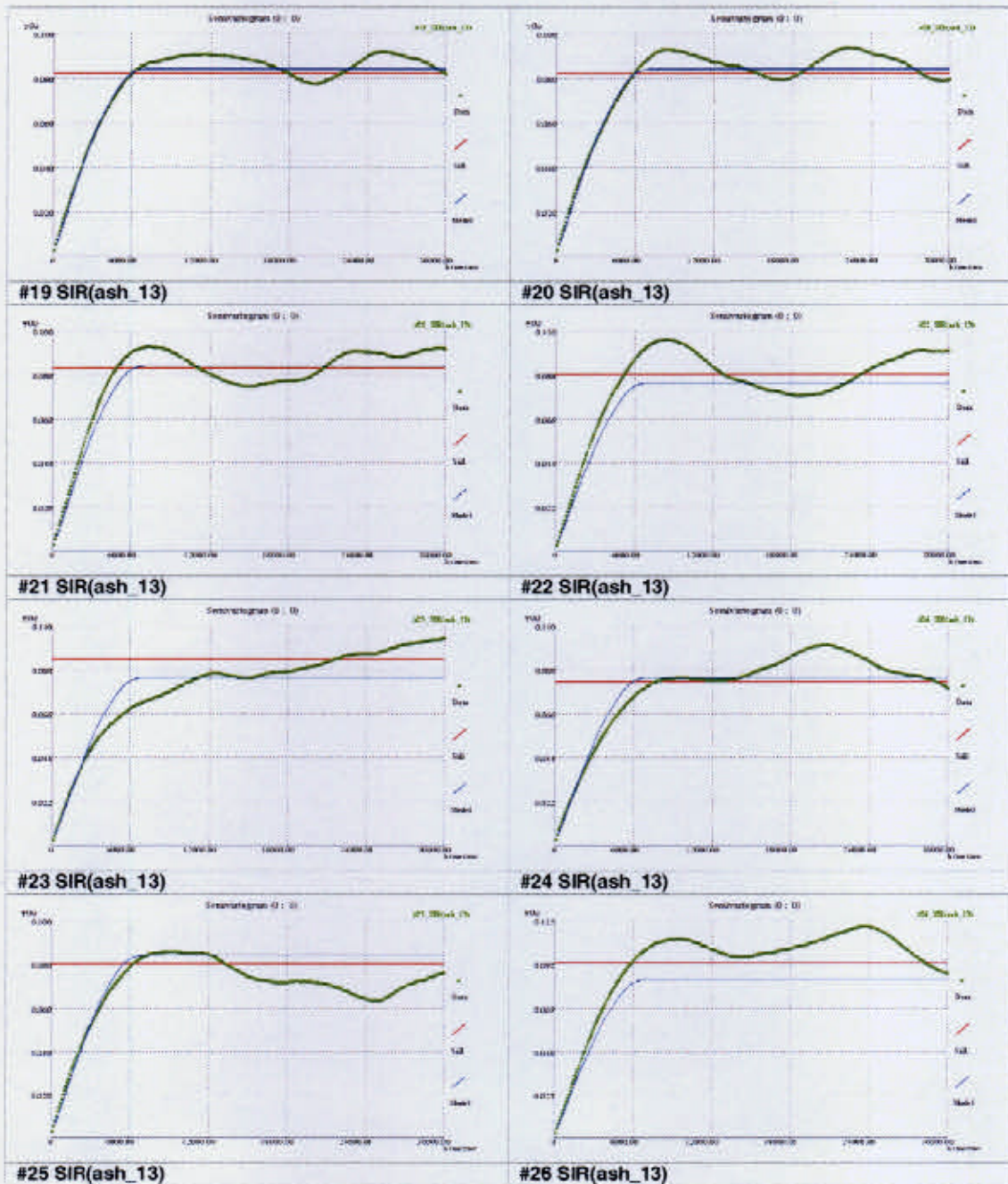


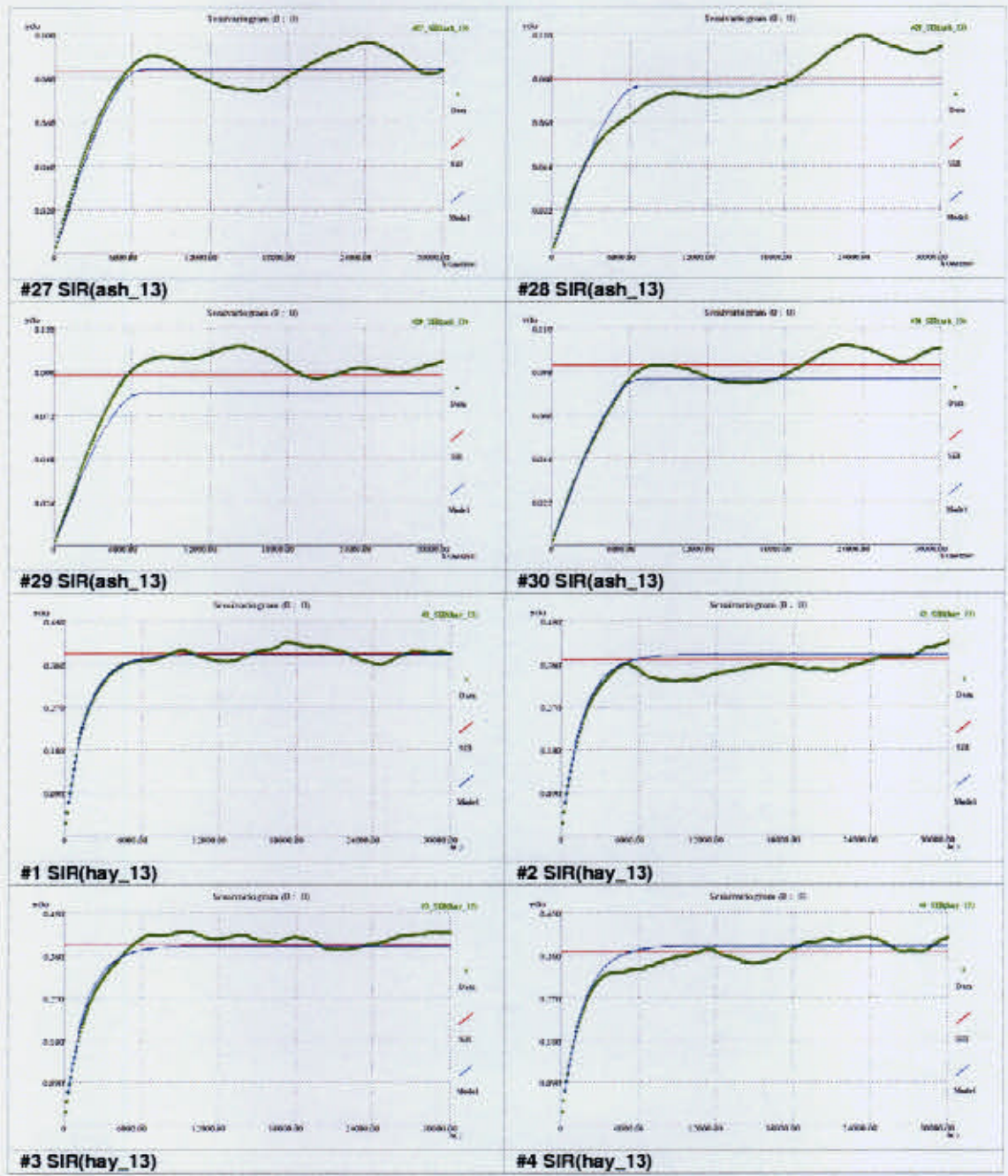


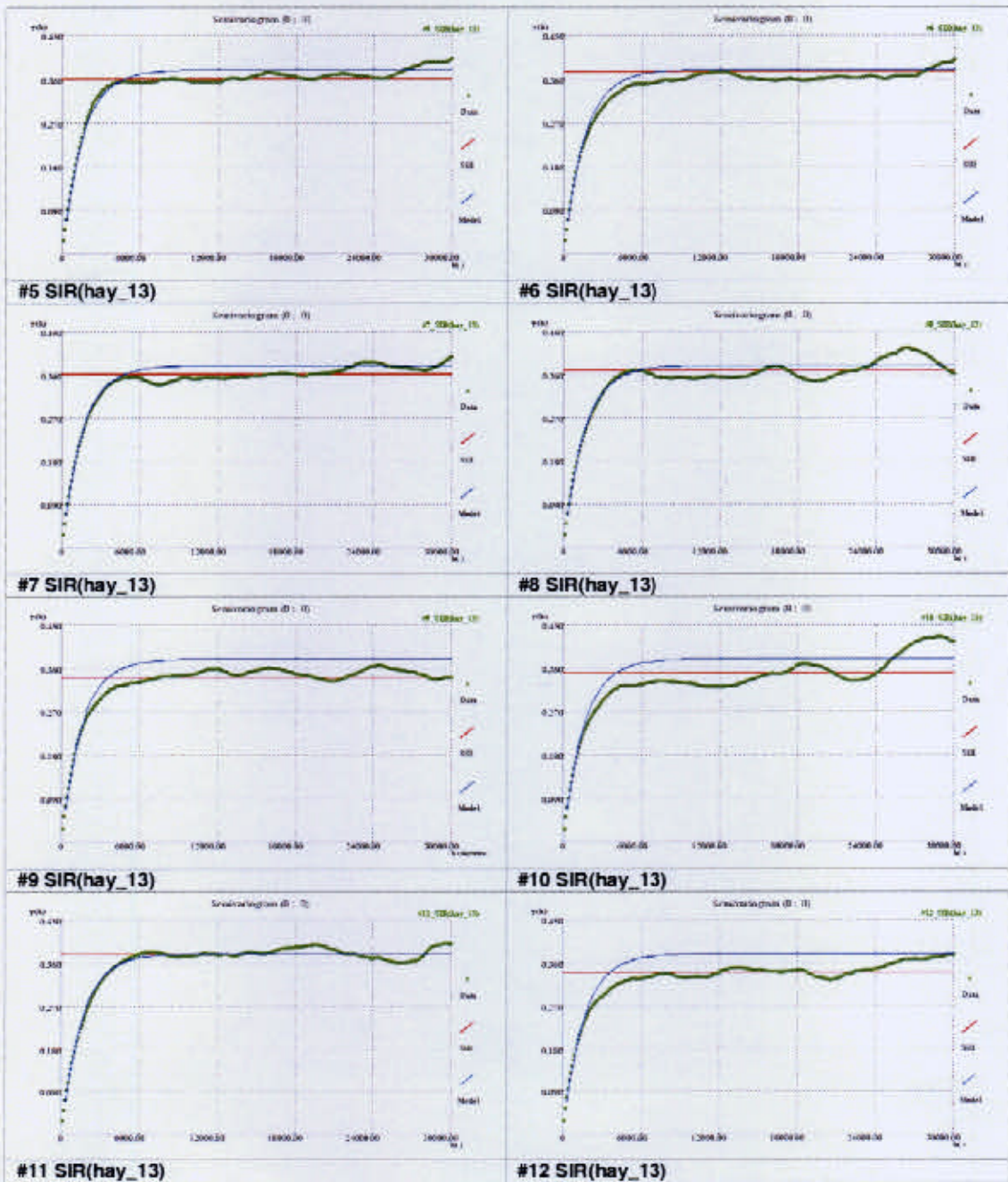


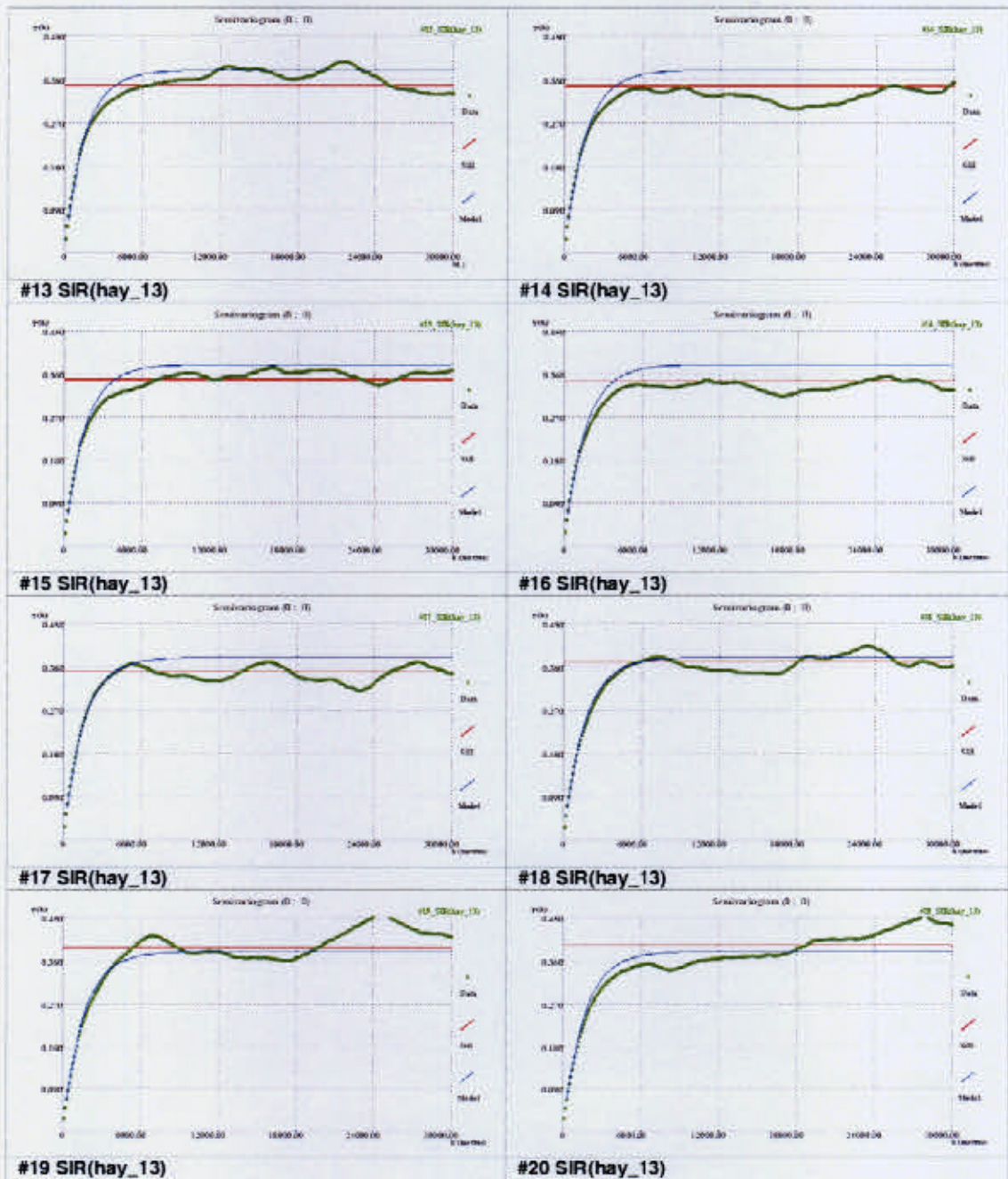


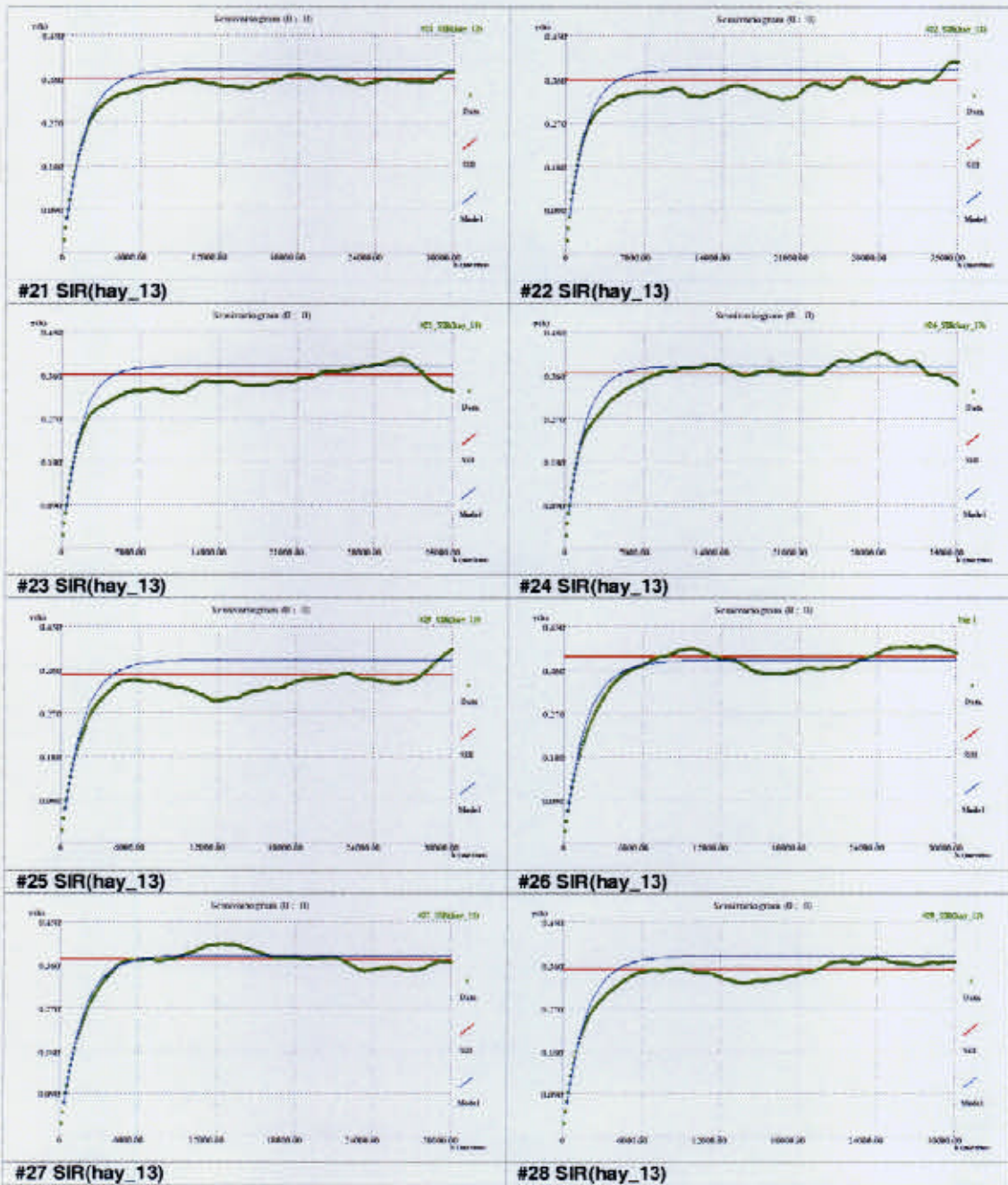


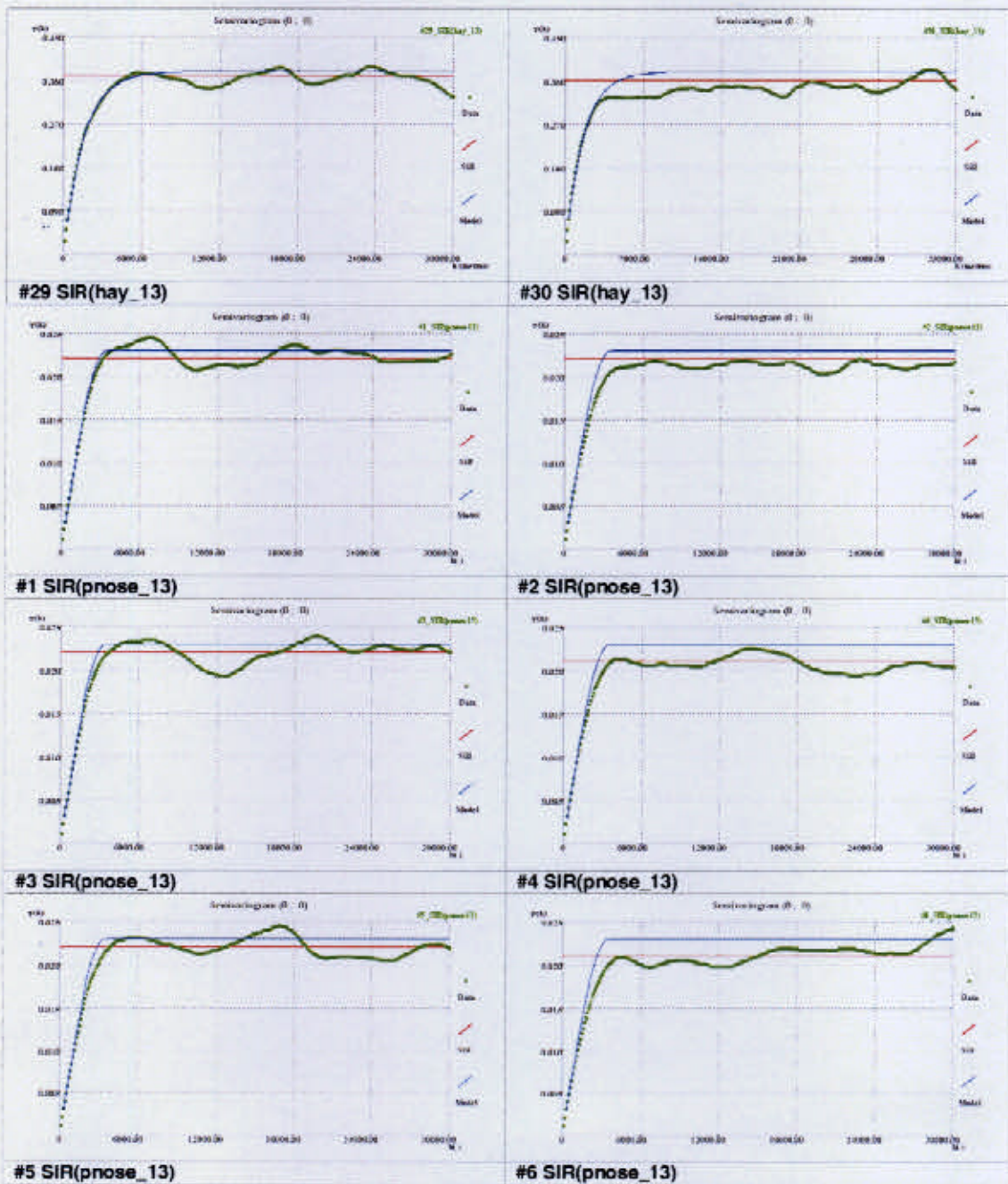


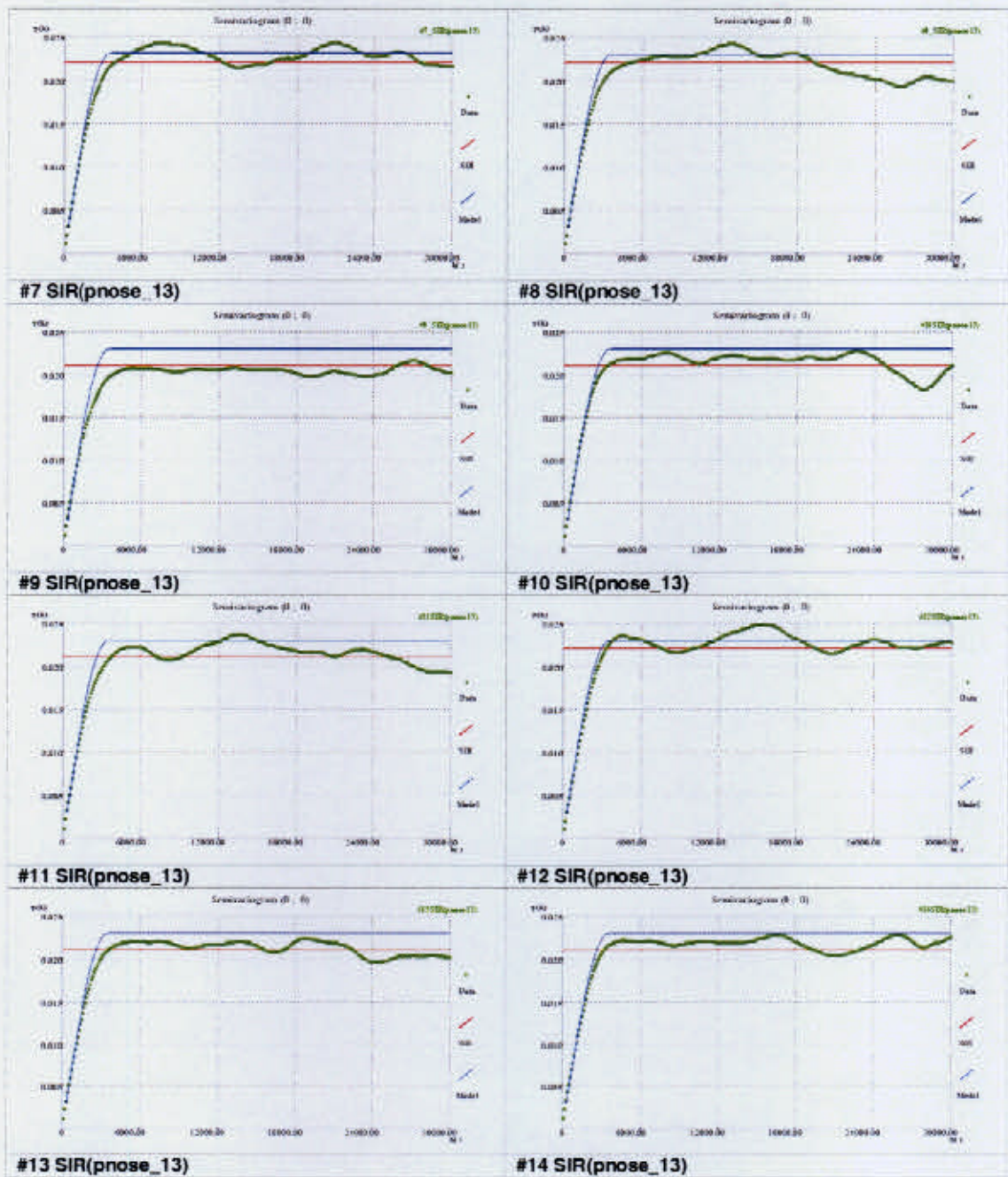


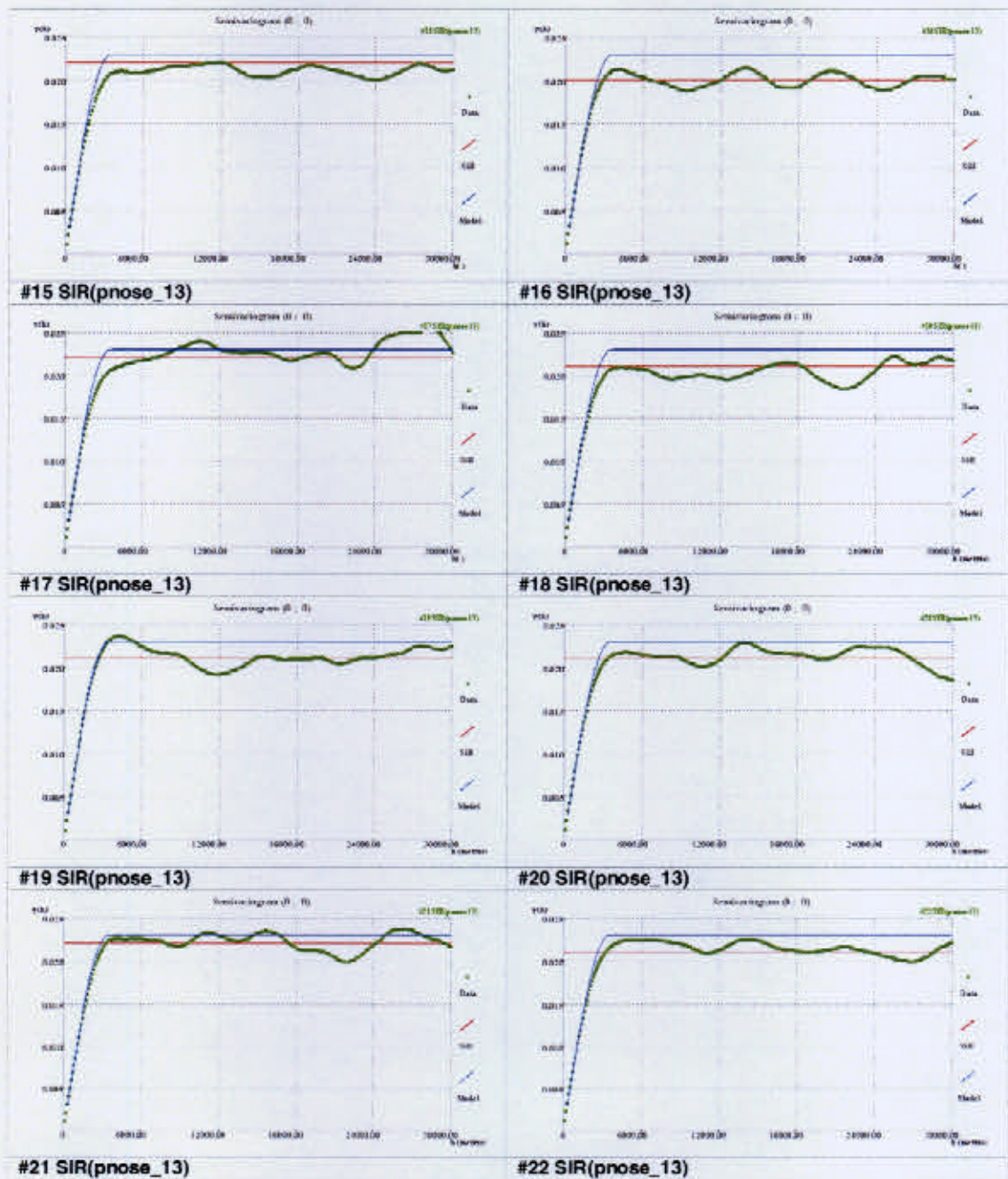


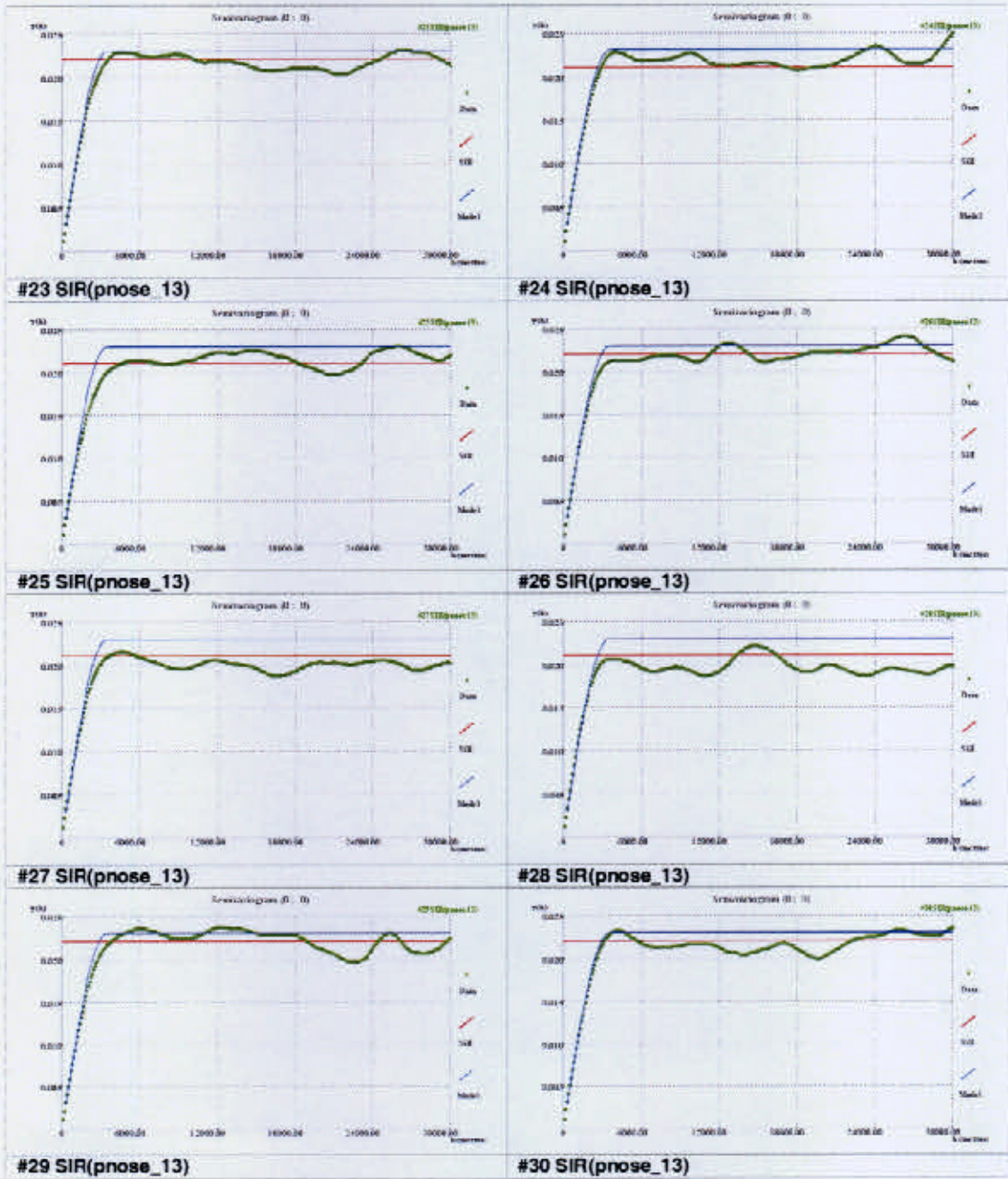


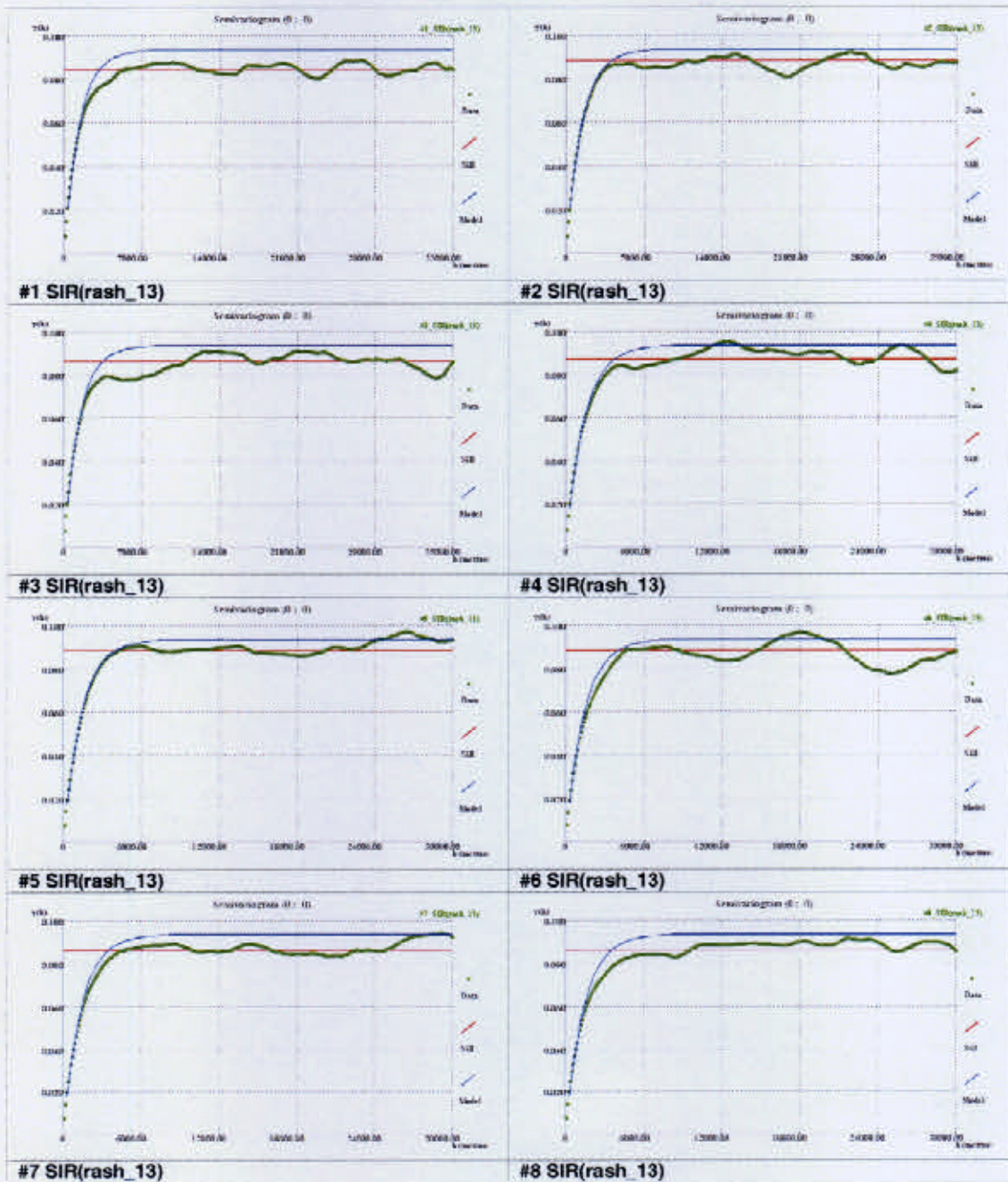


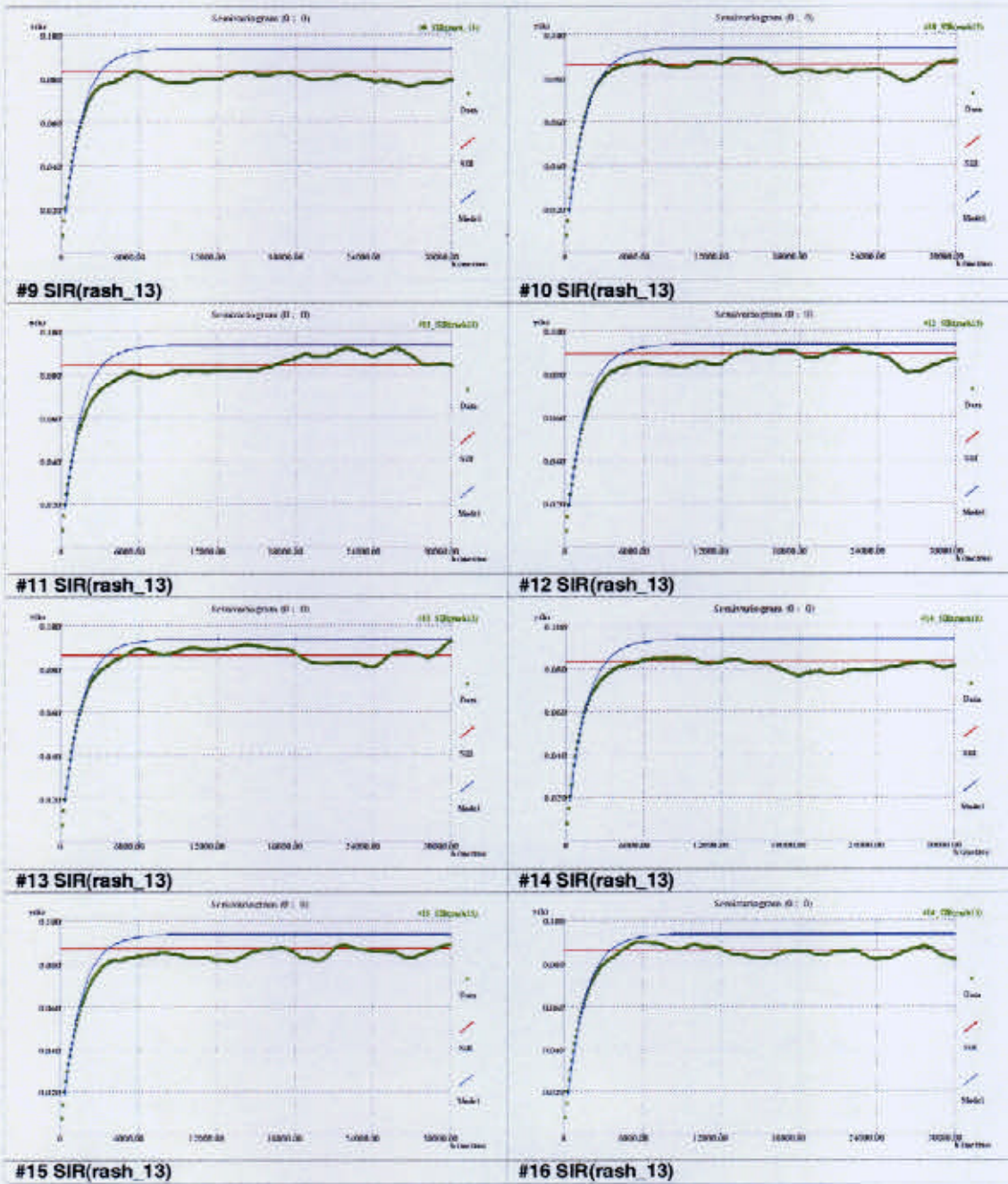


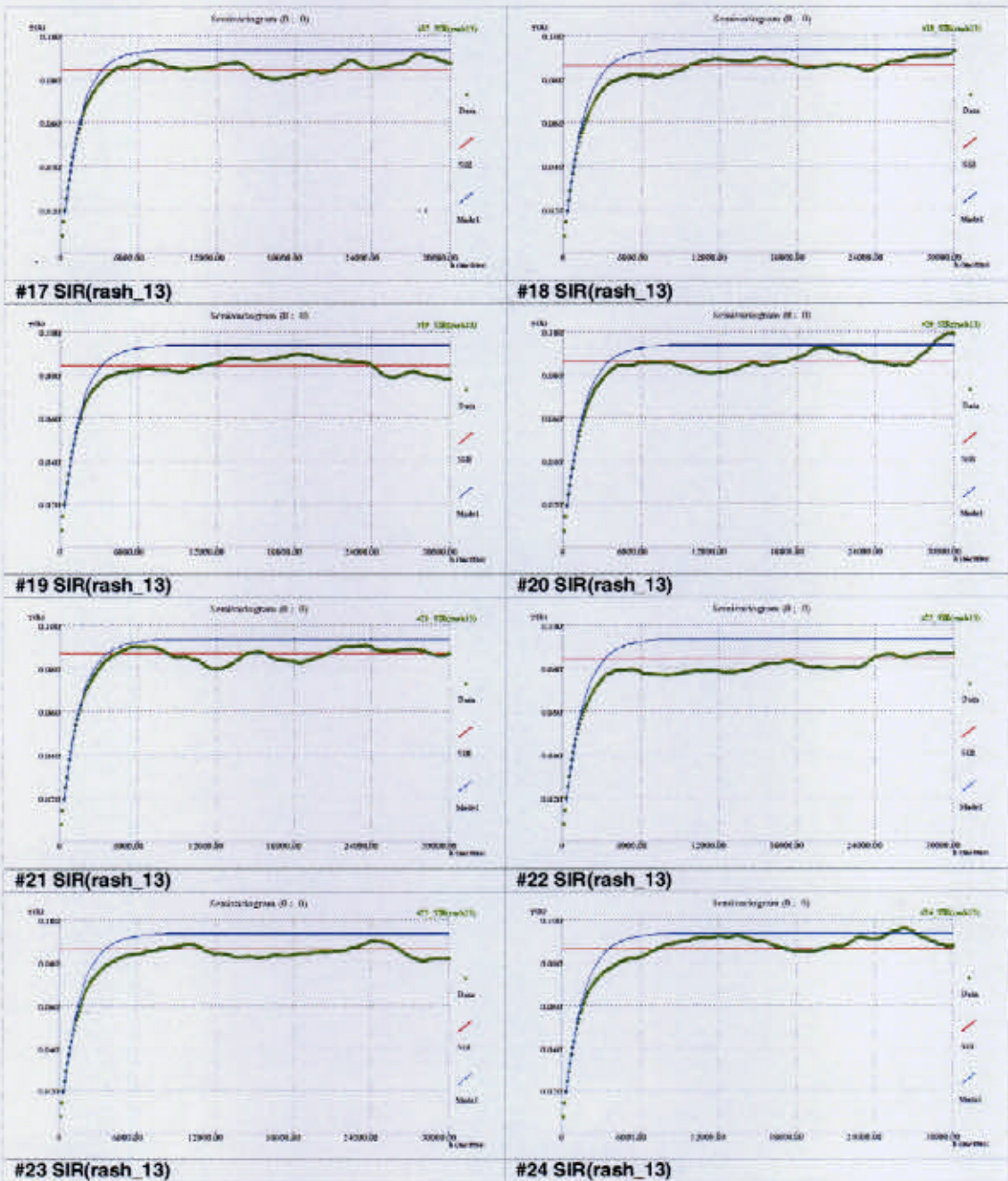








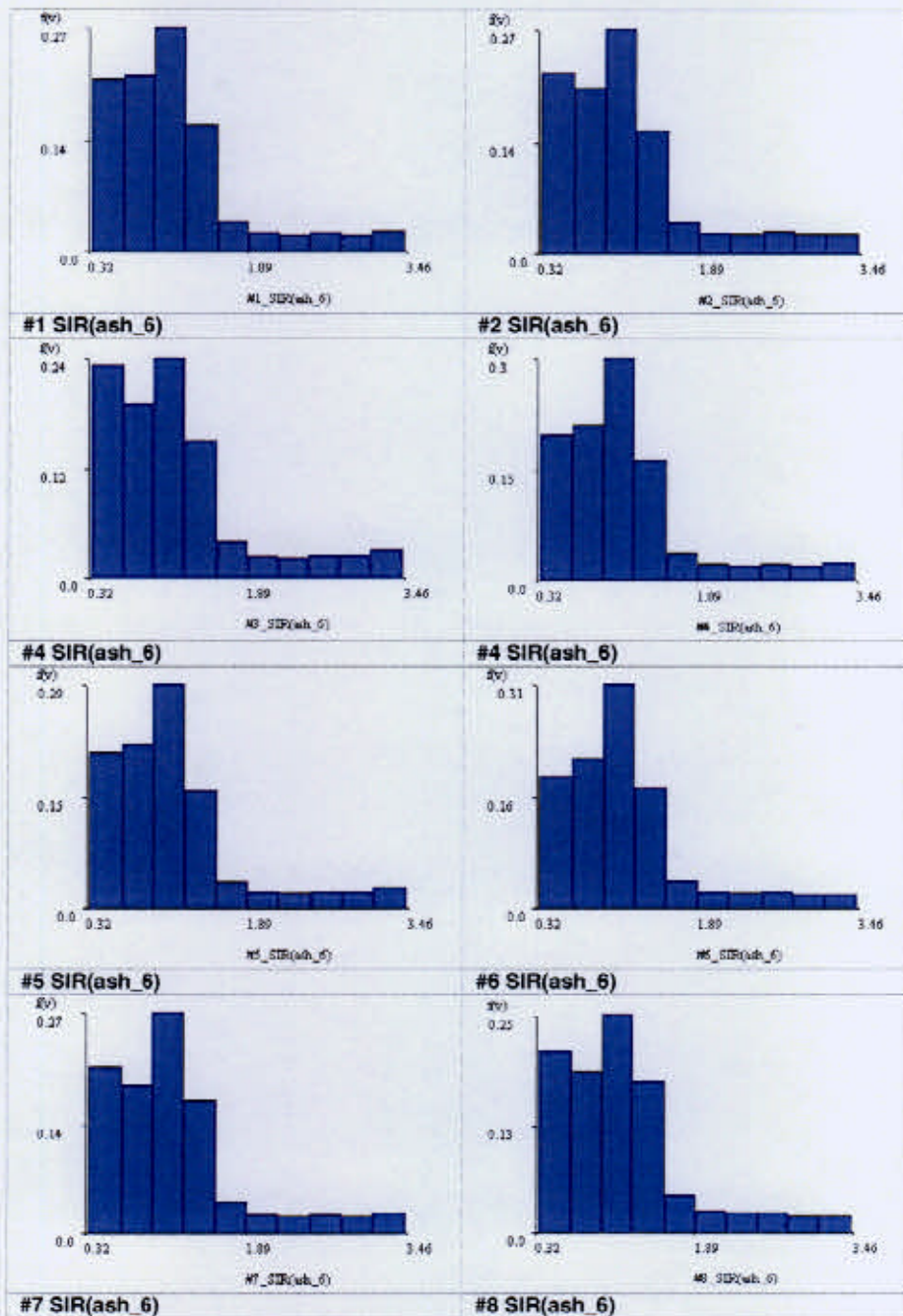


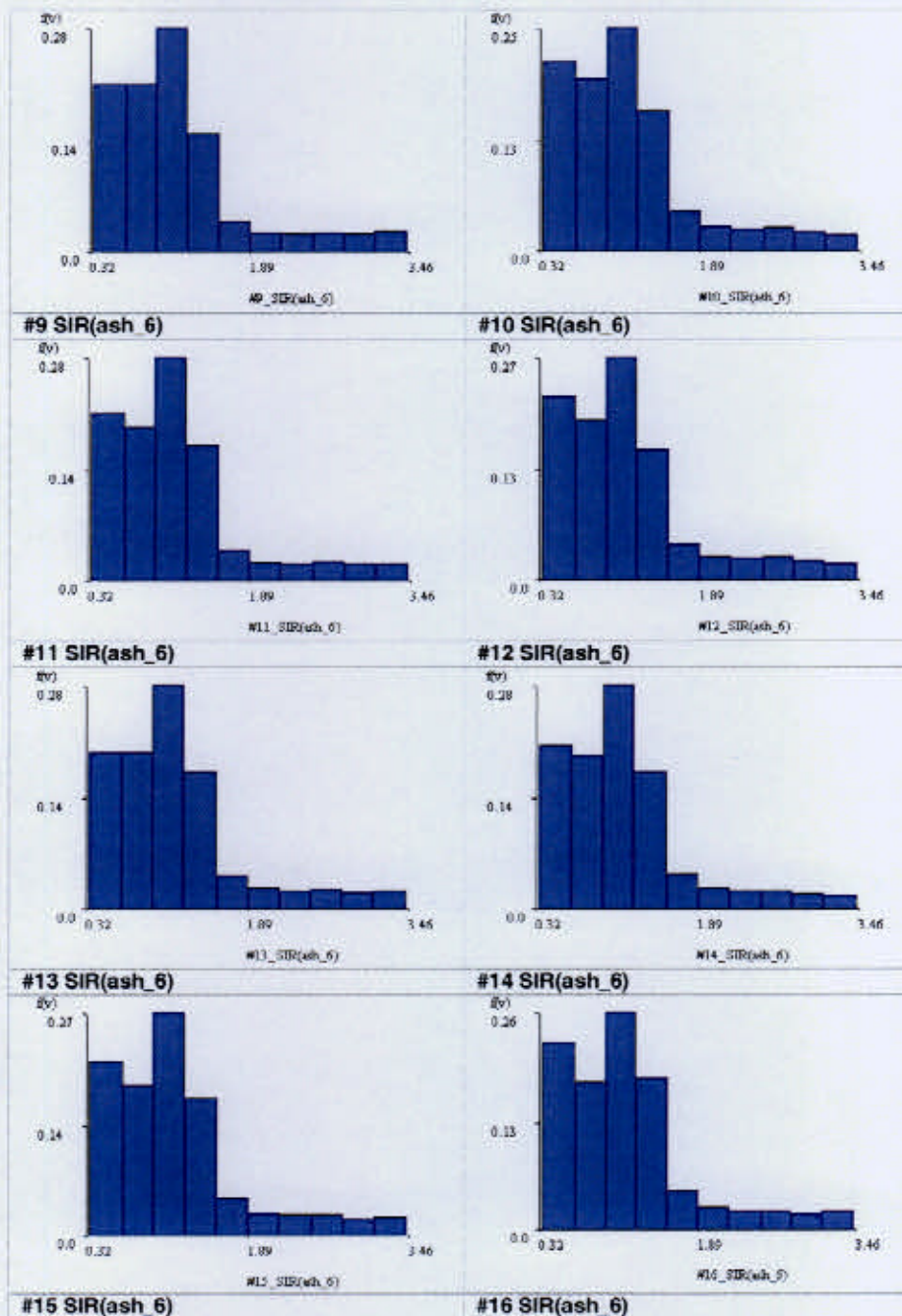


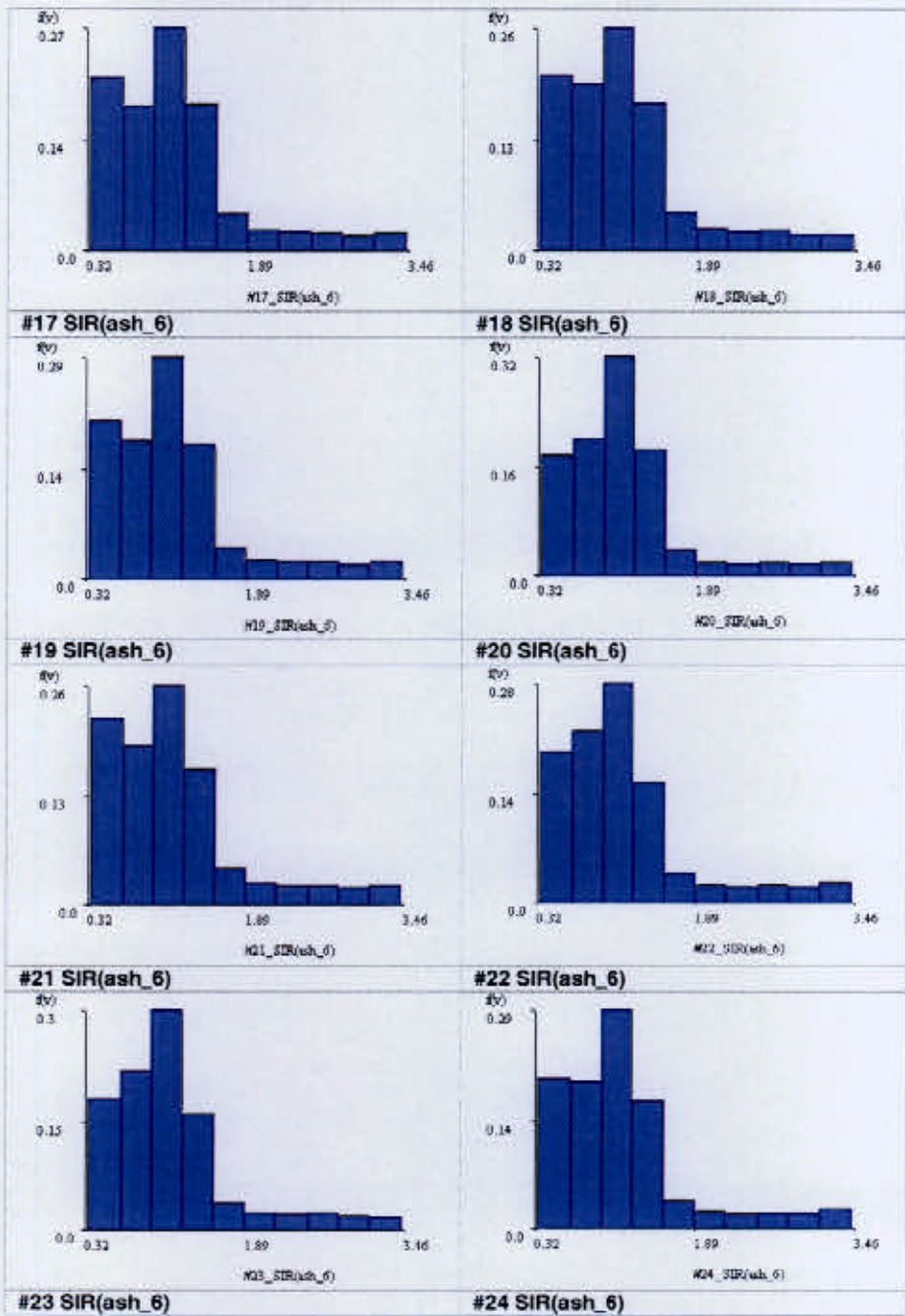


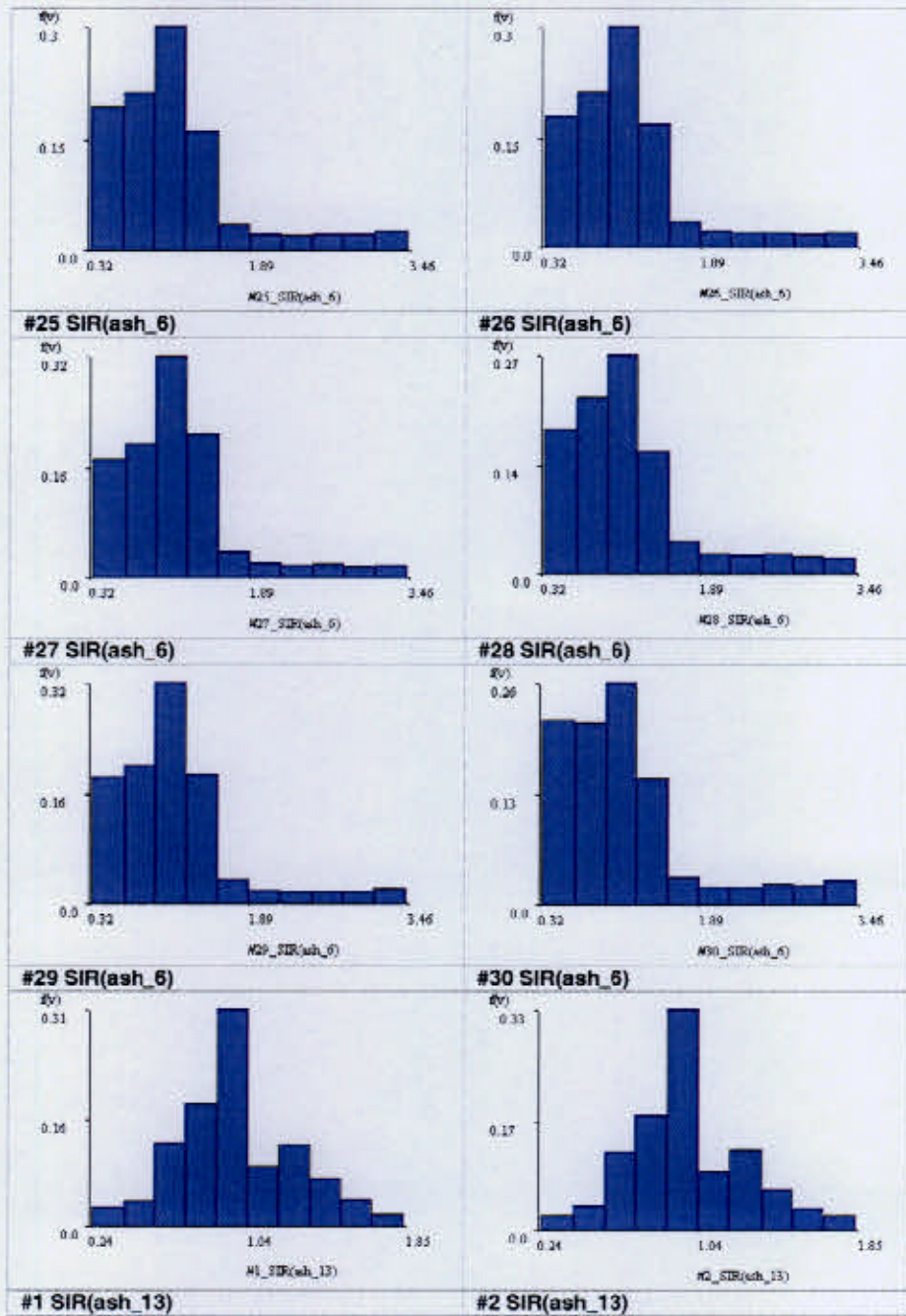
## **ANEXO E**

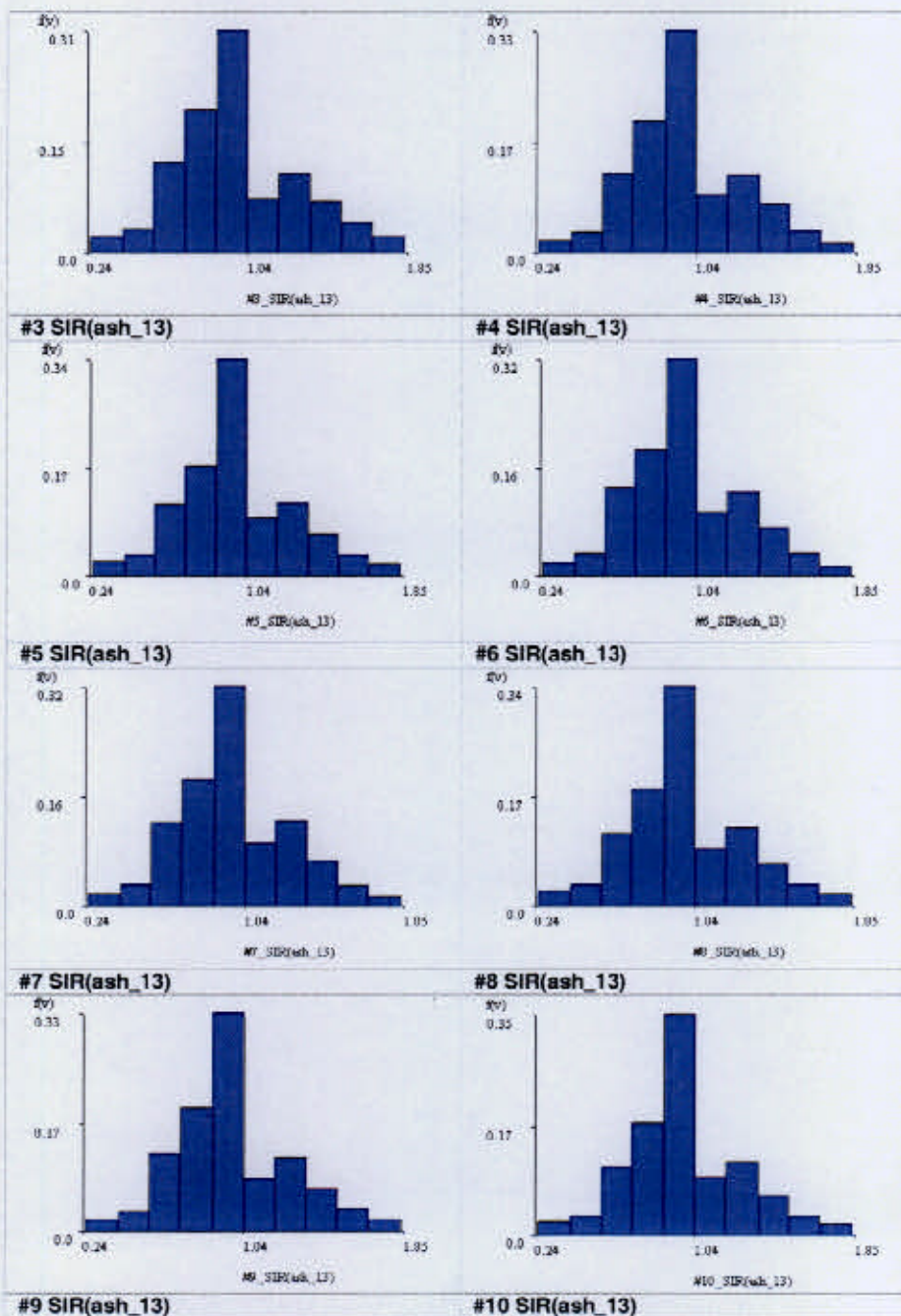
### **Histogramas das simulações das variáveis da saúde**

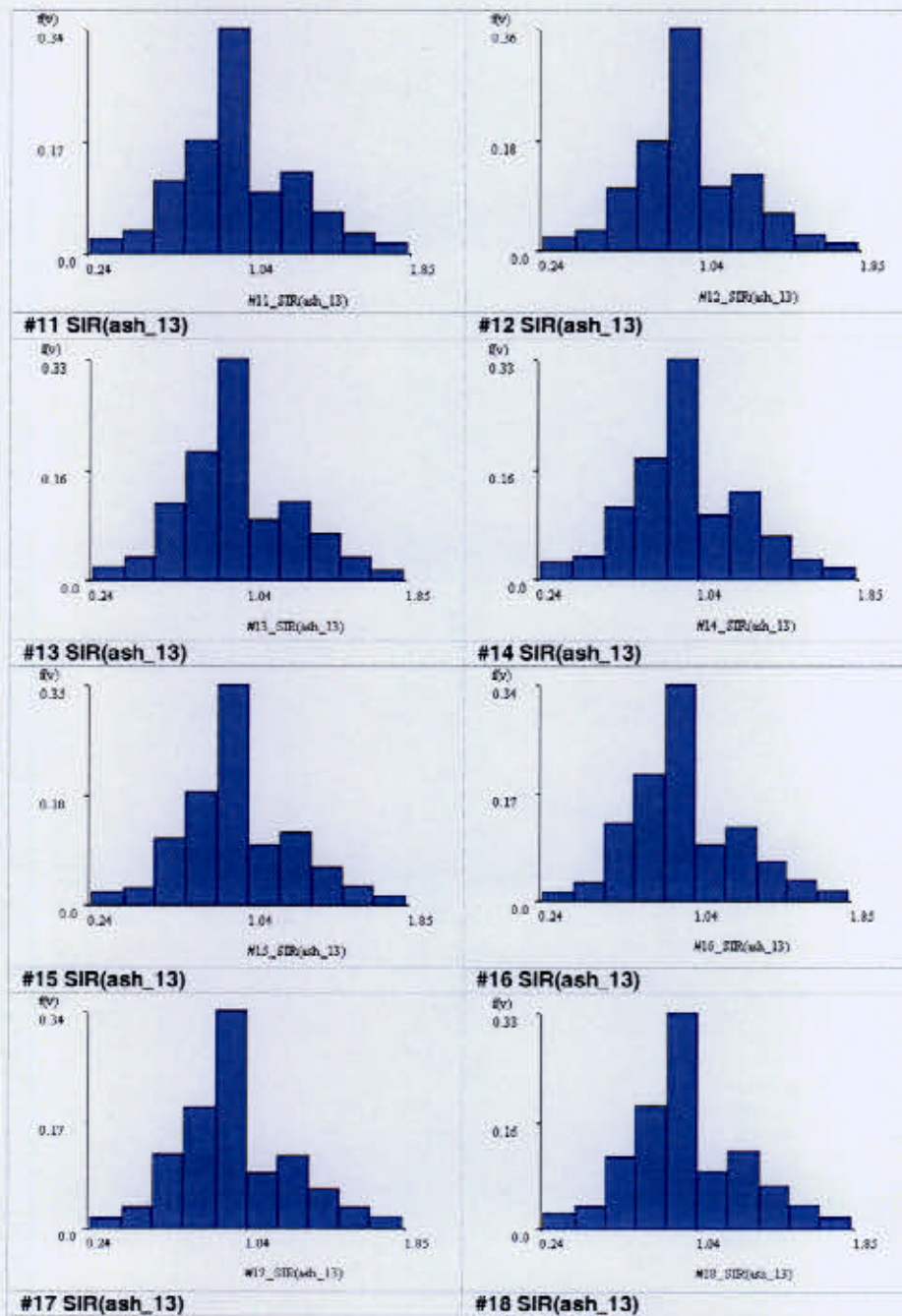


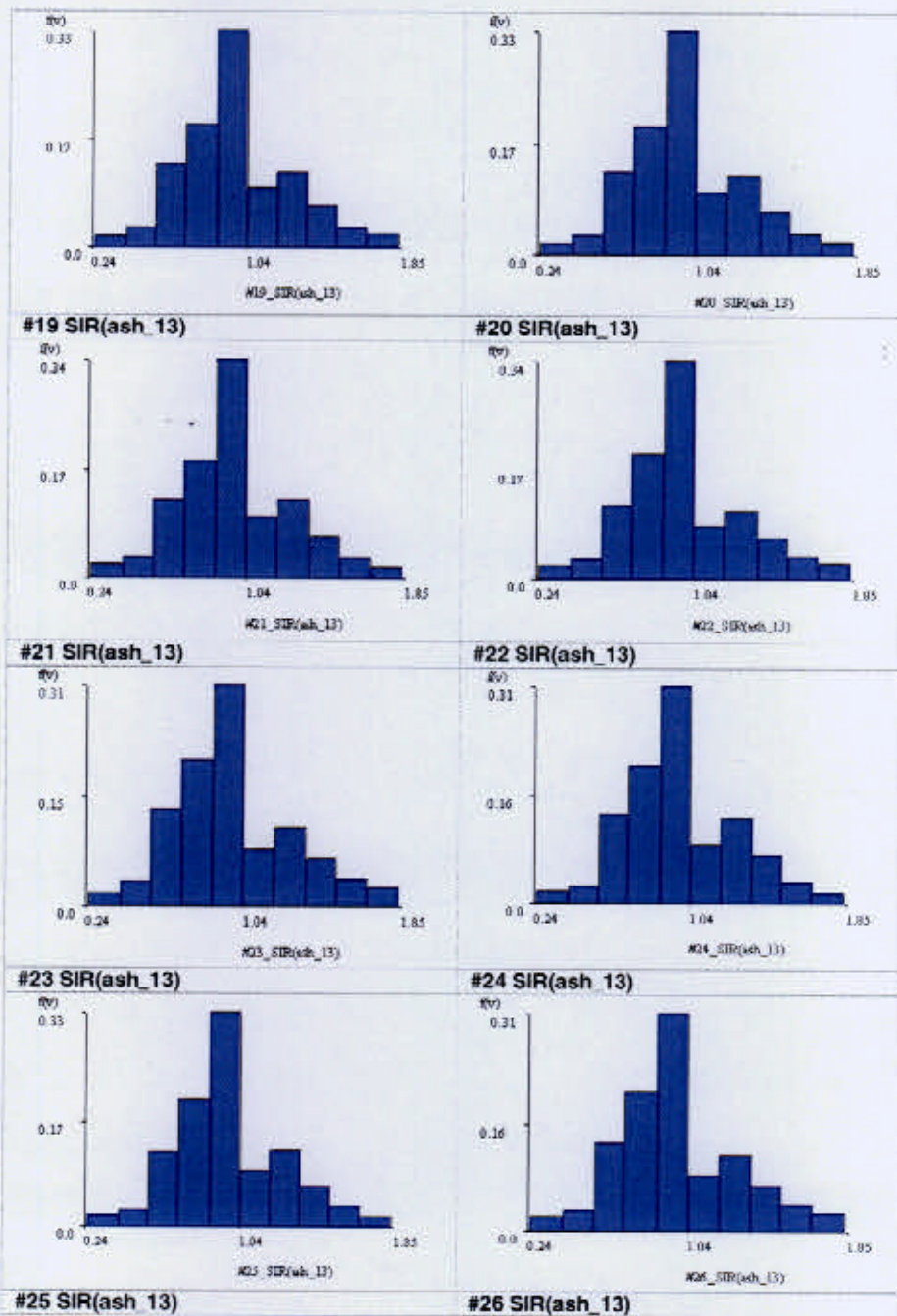


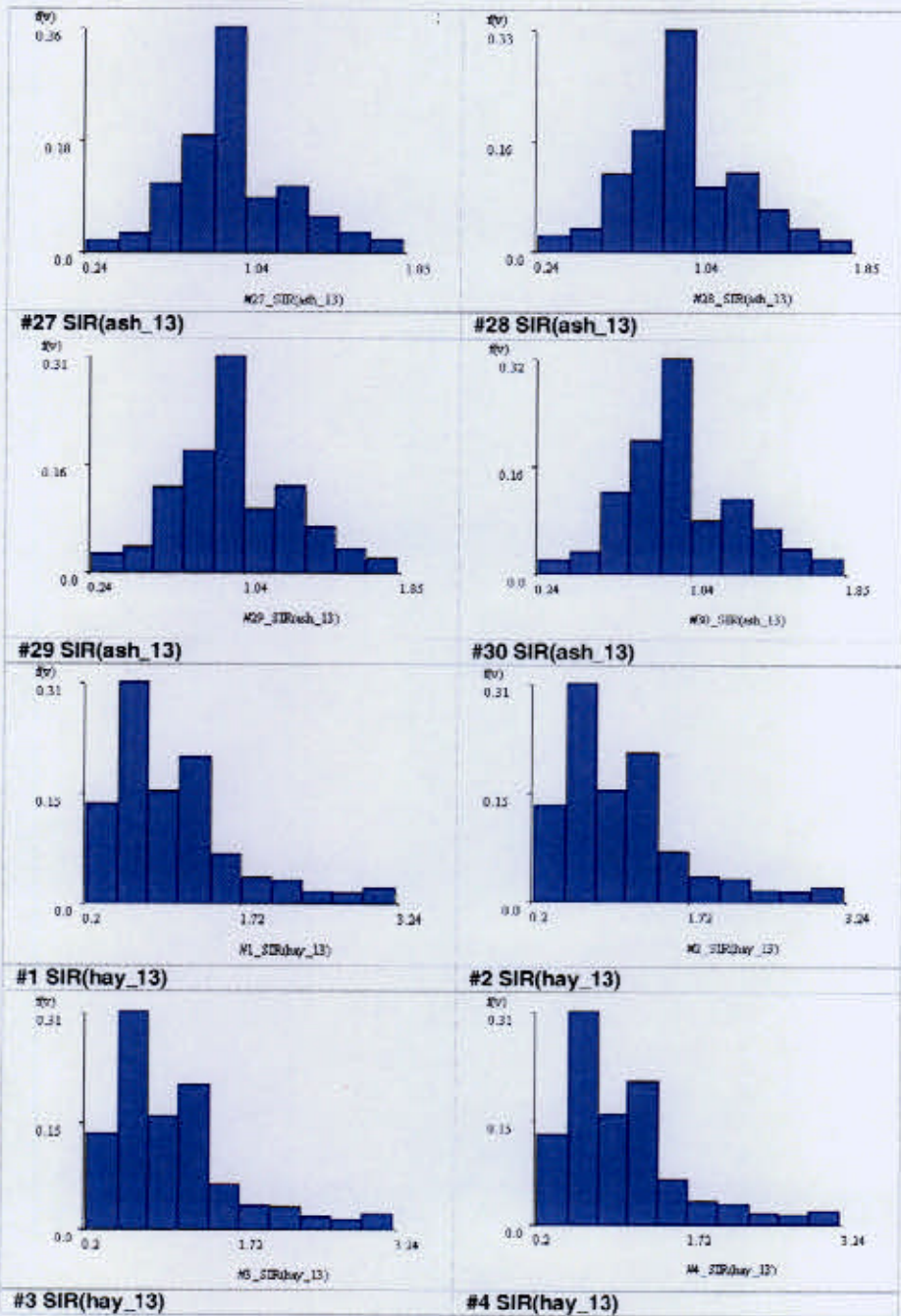


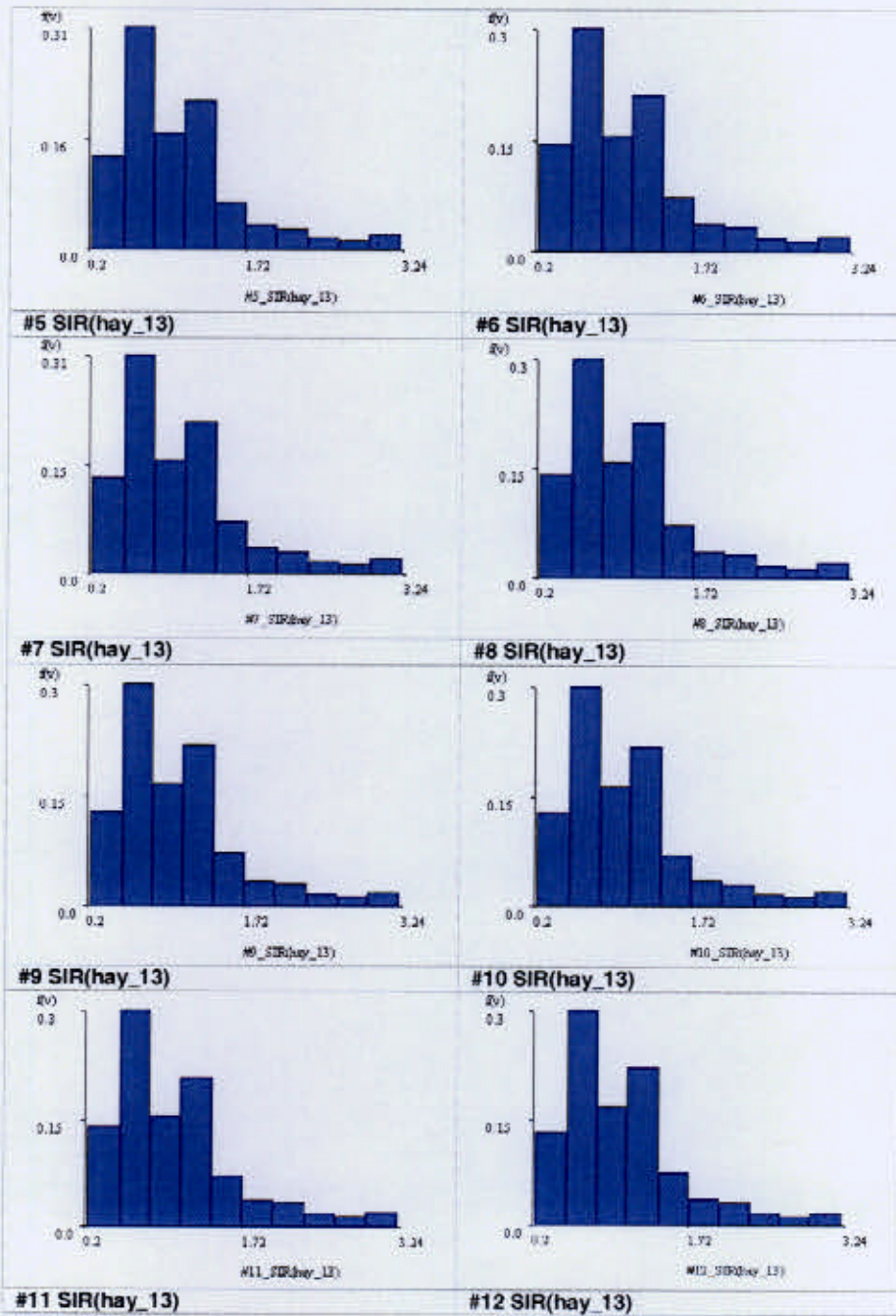


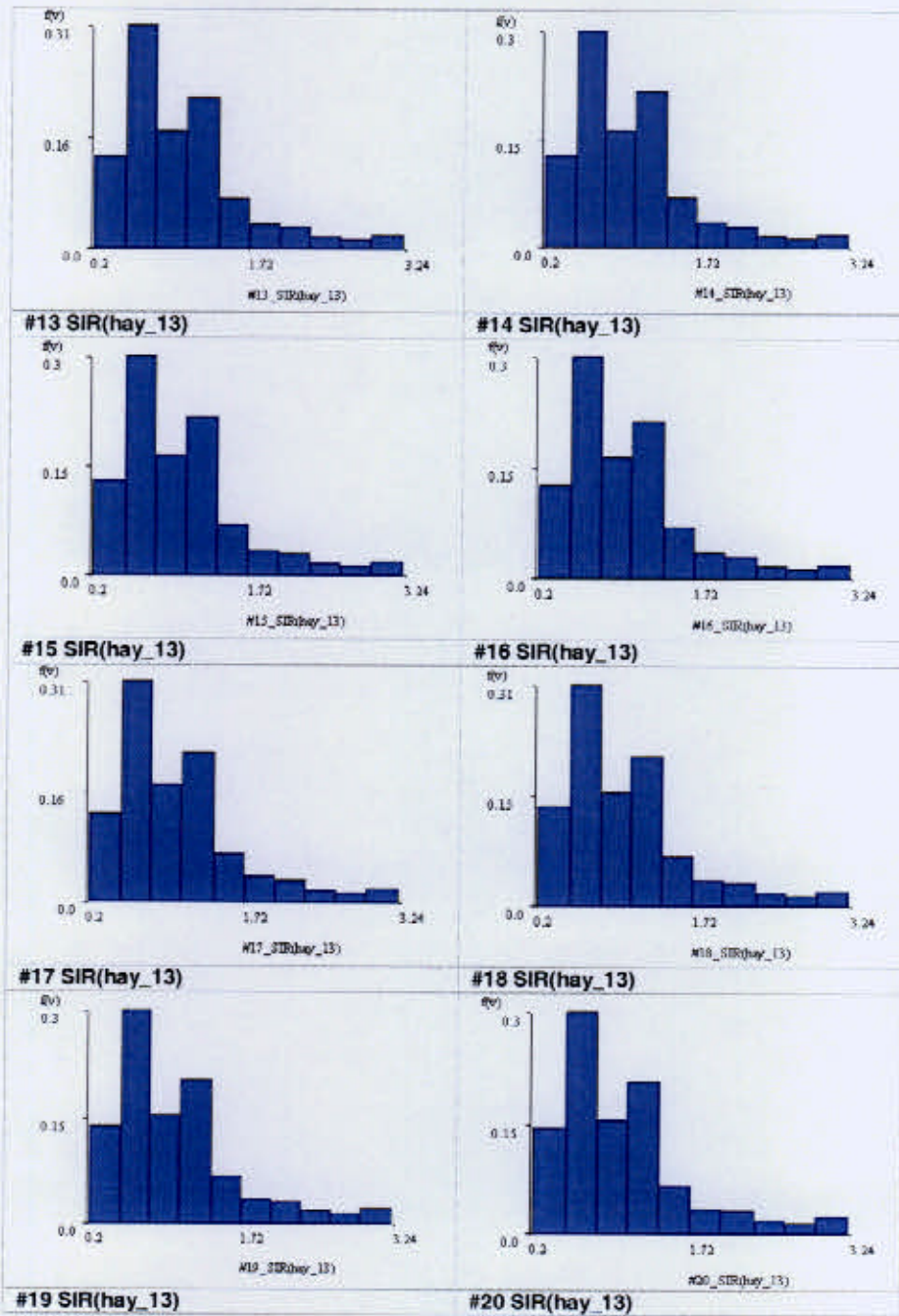


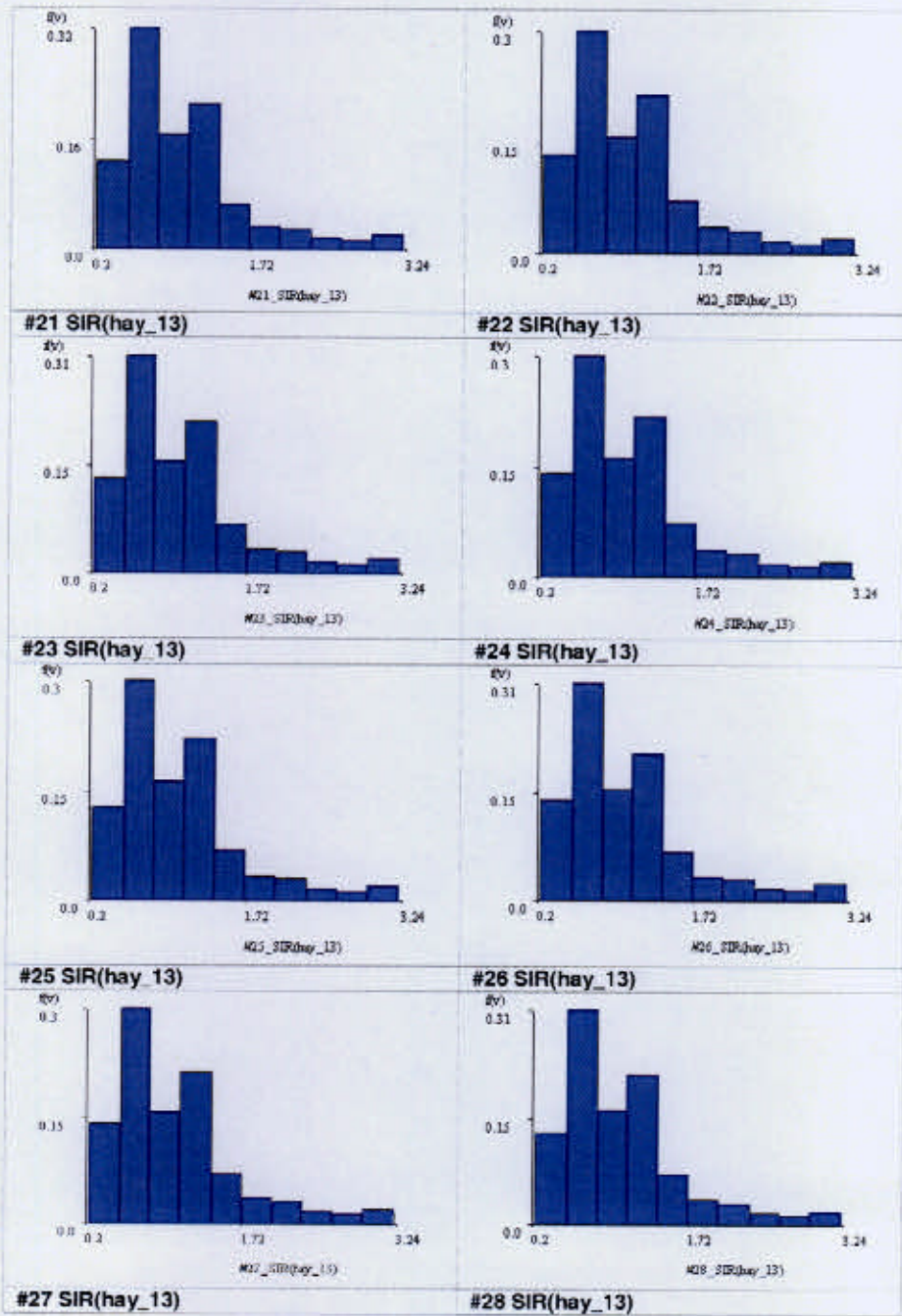


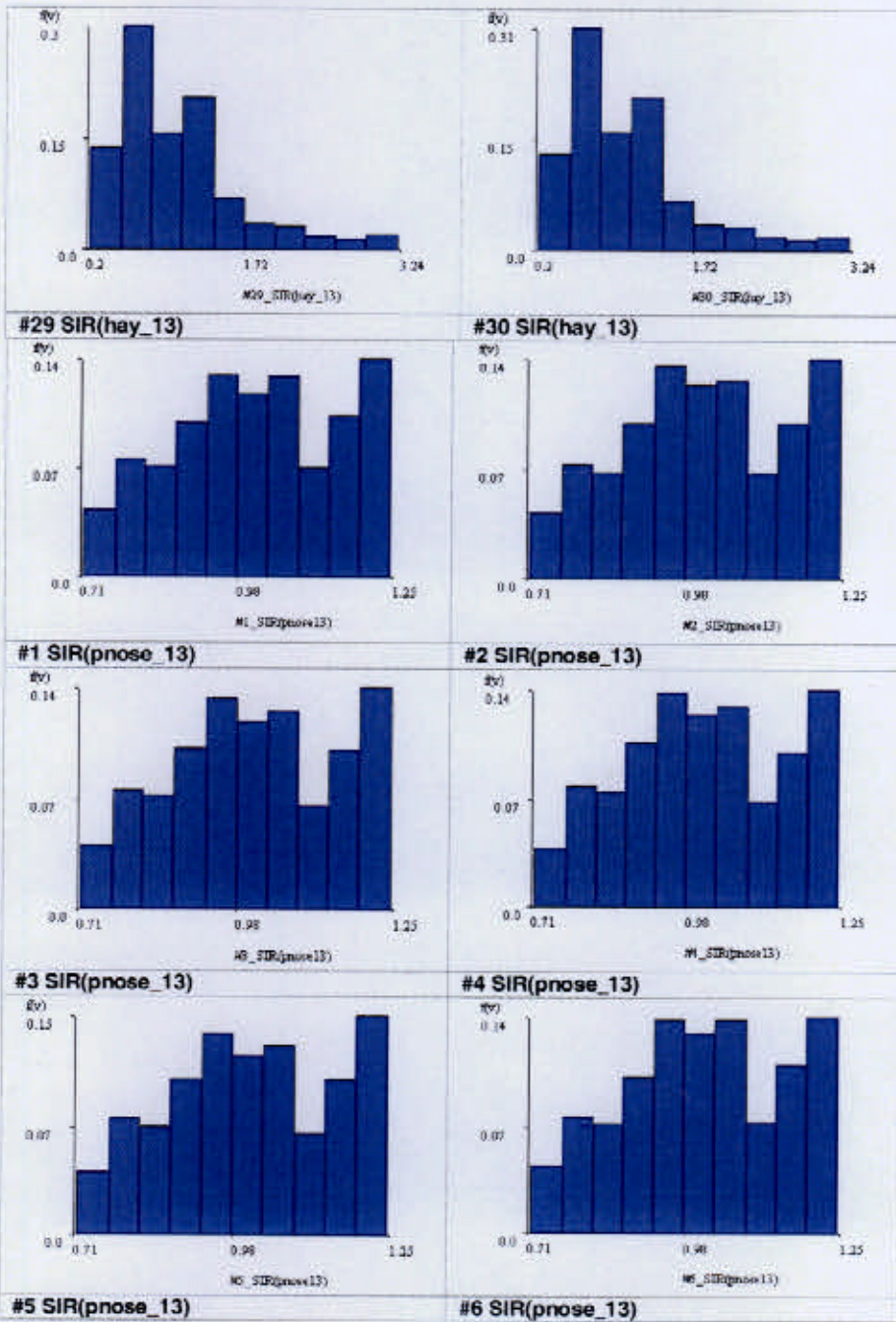


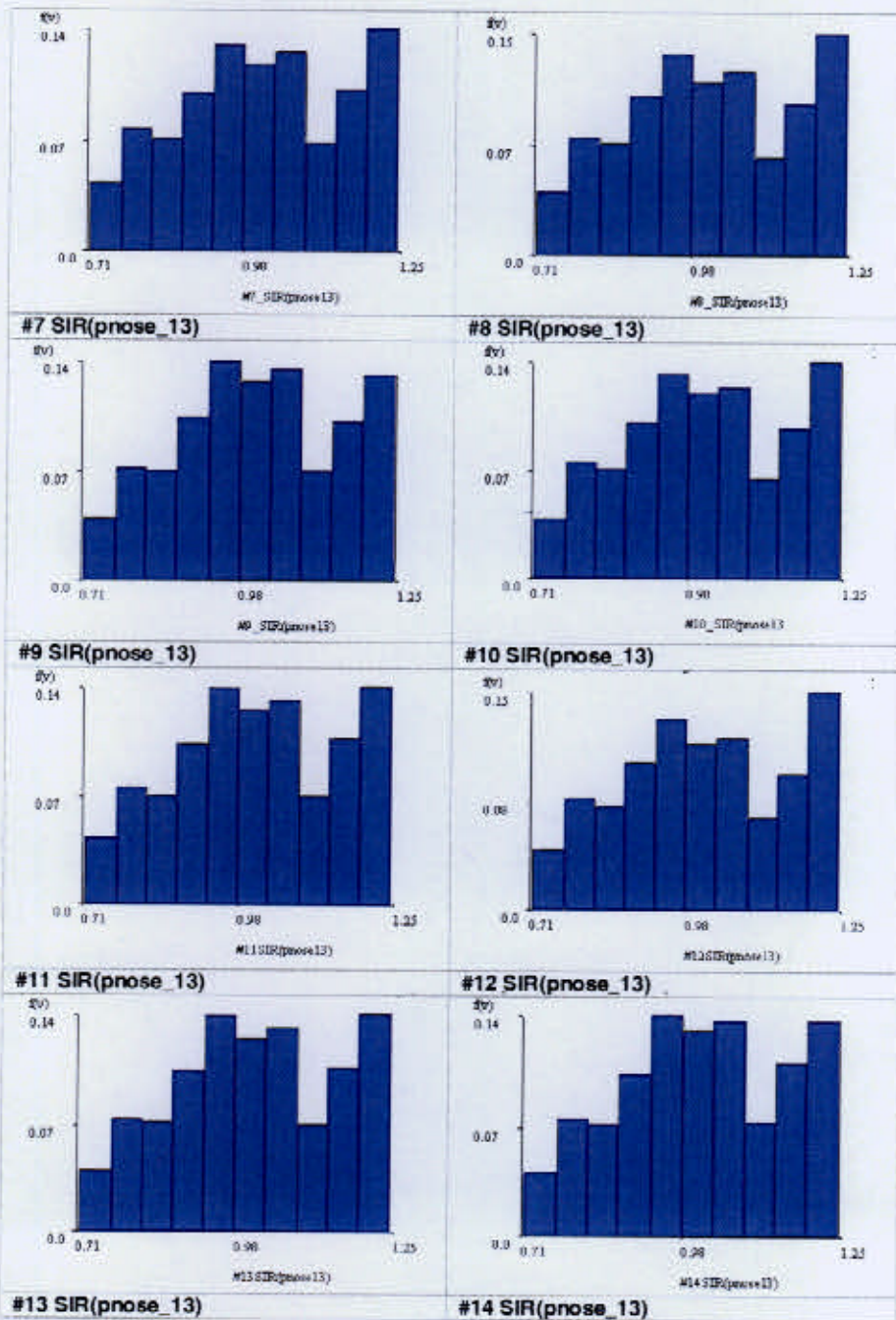


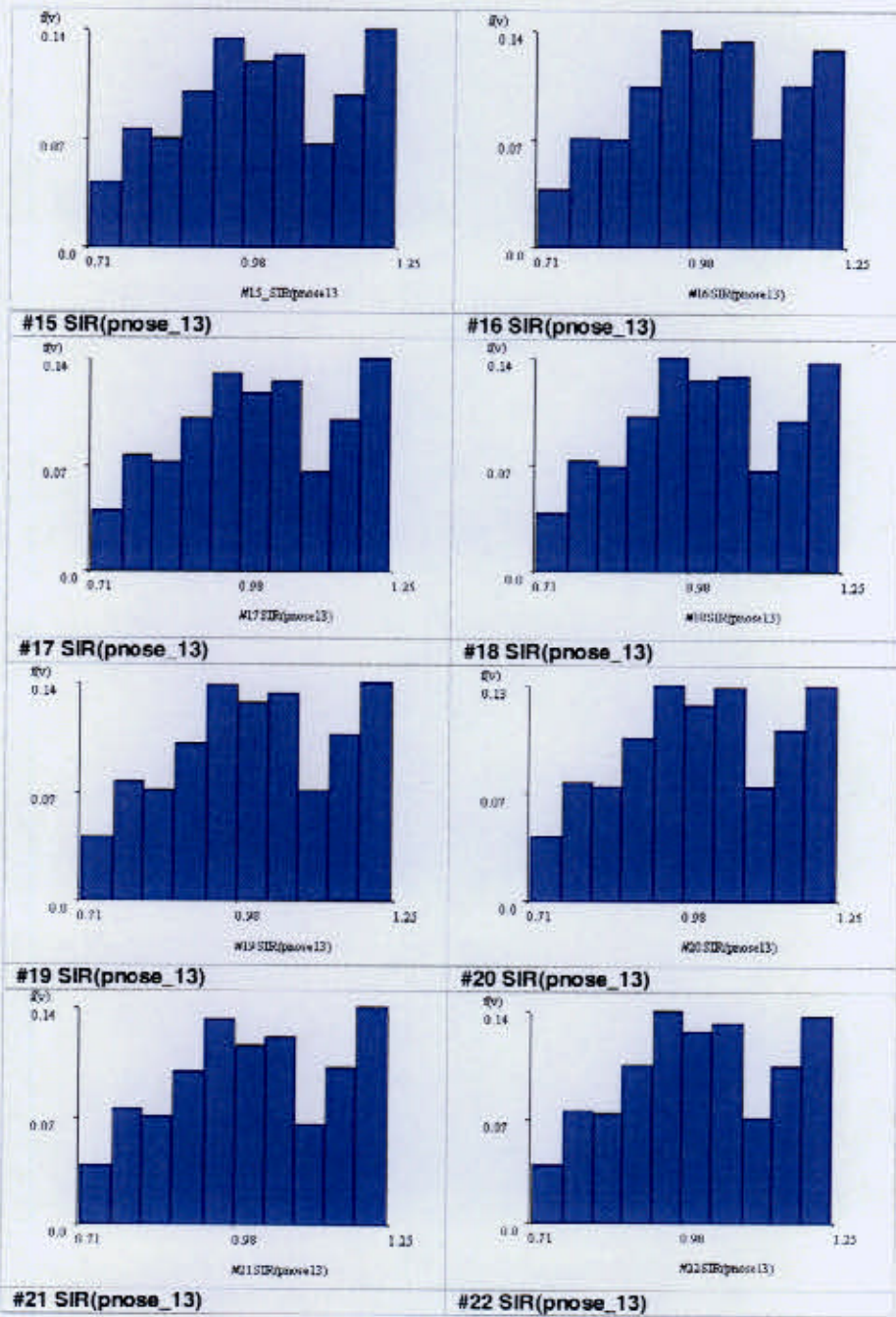


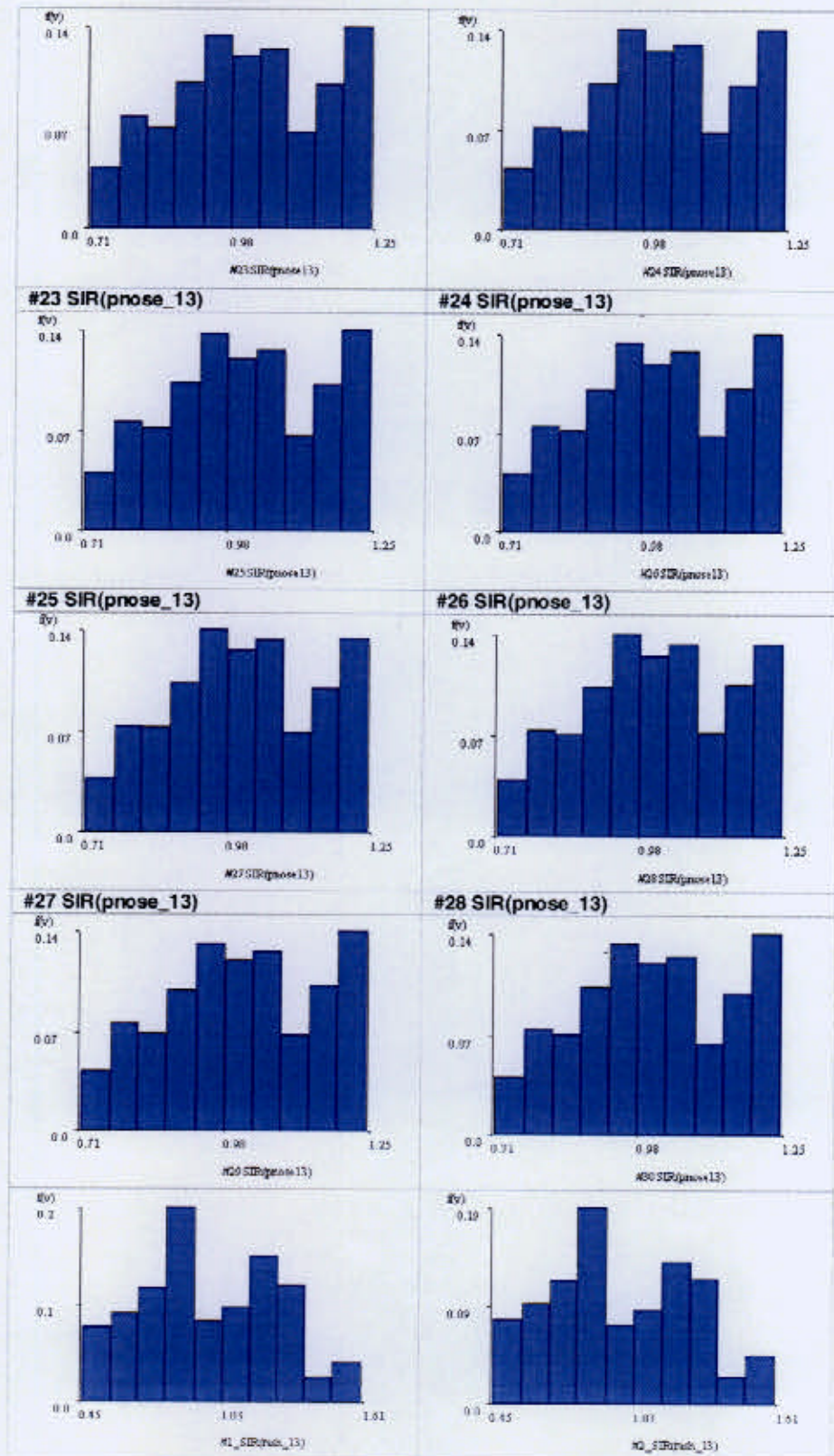


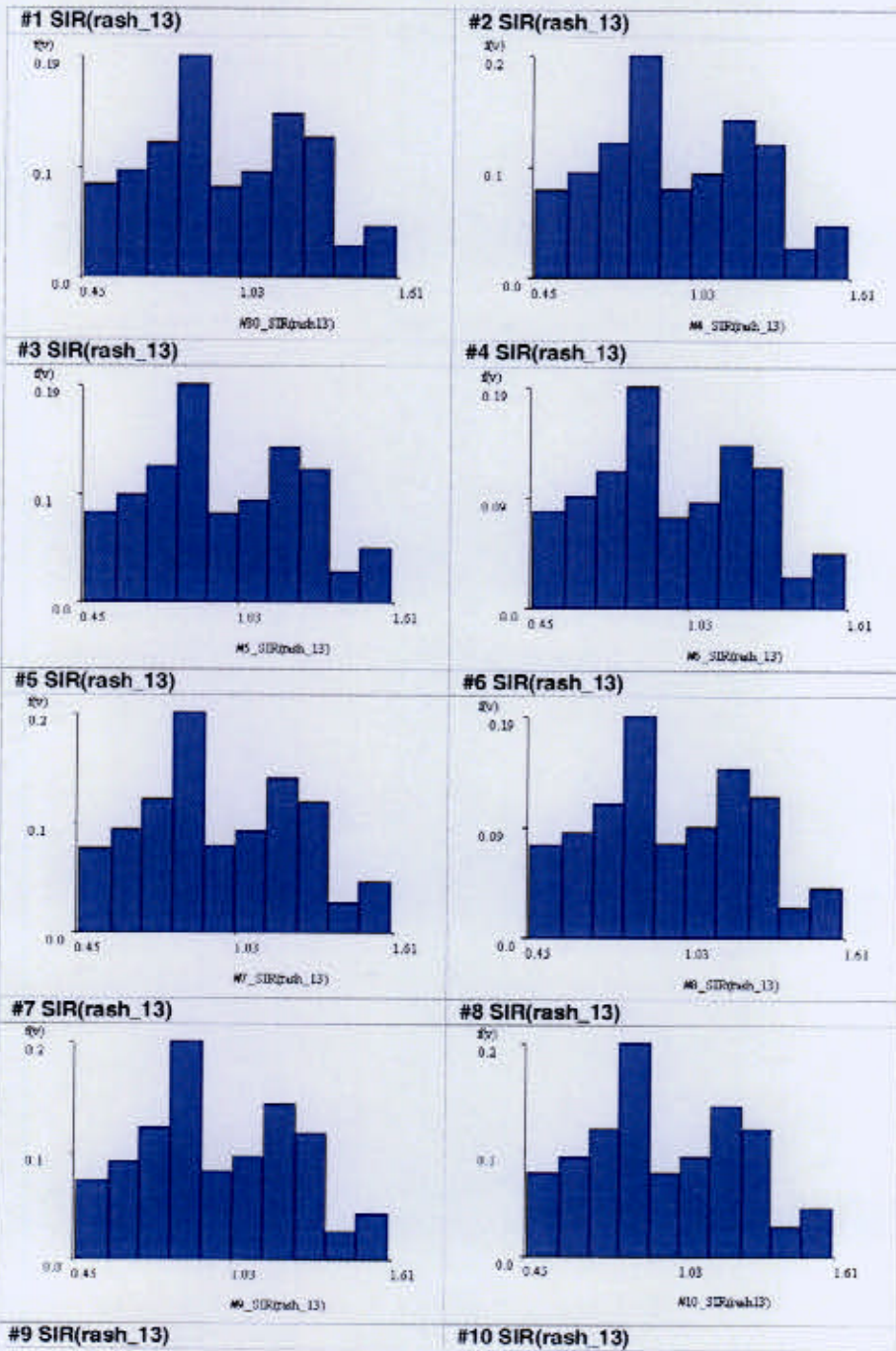


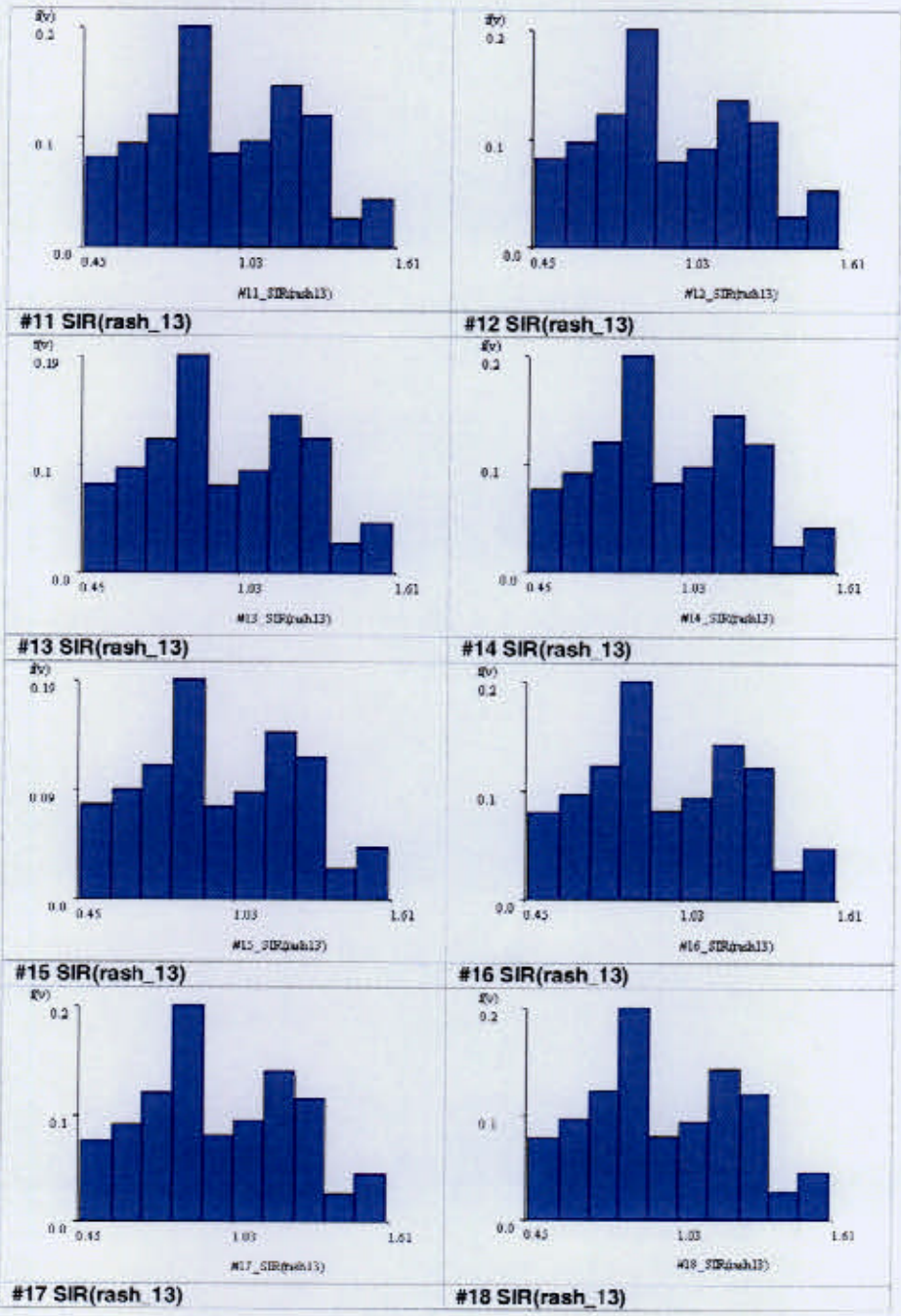


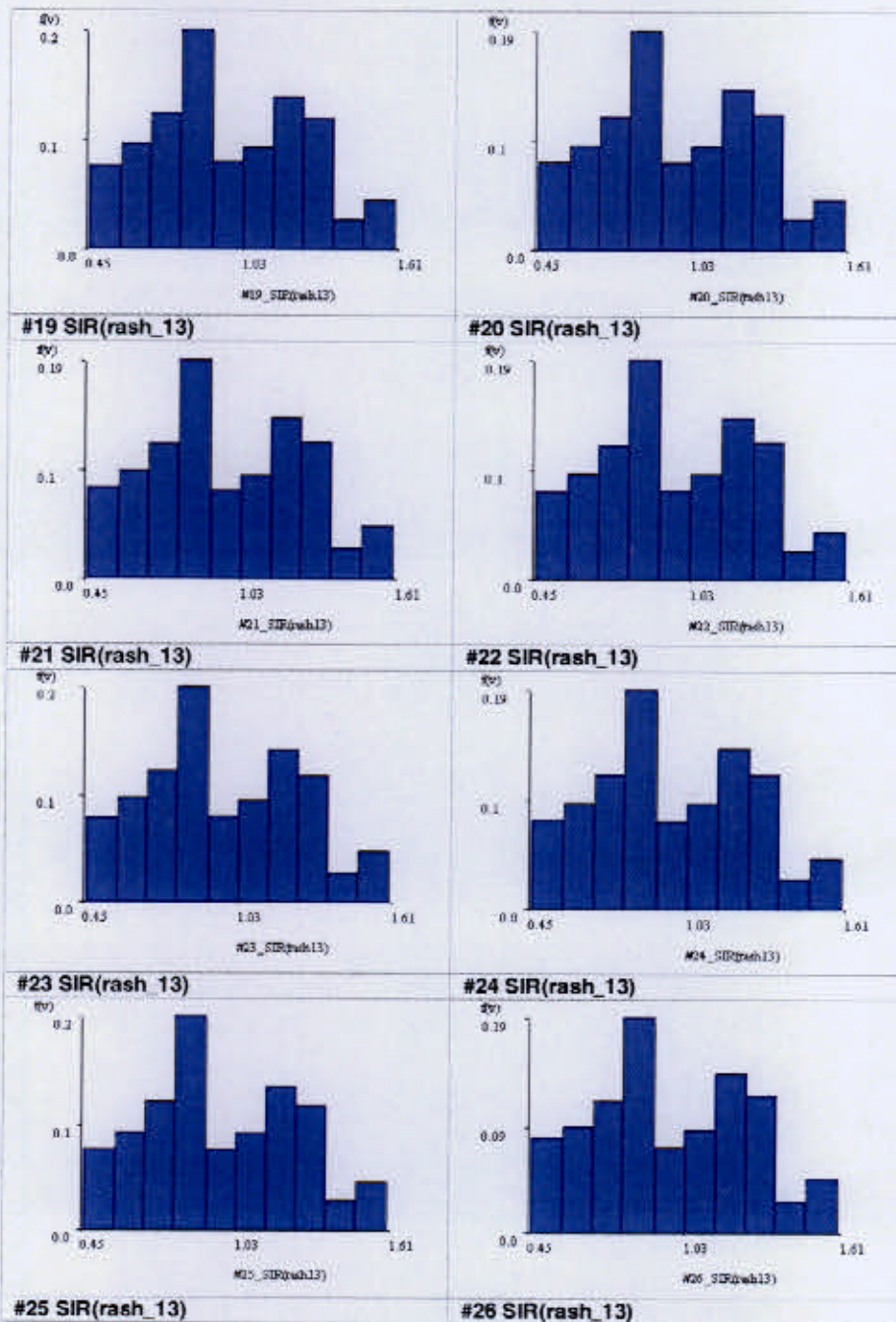


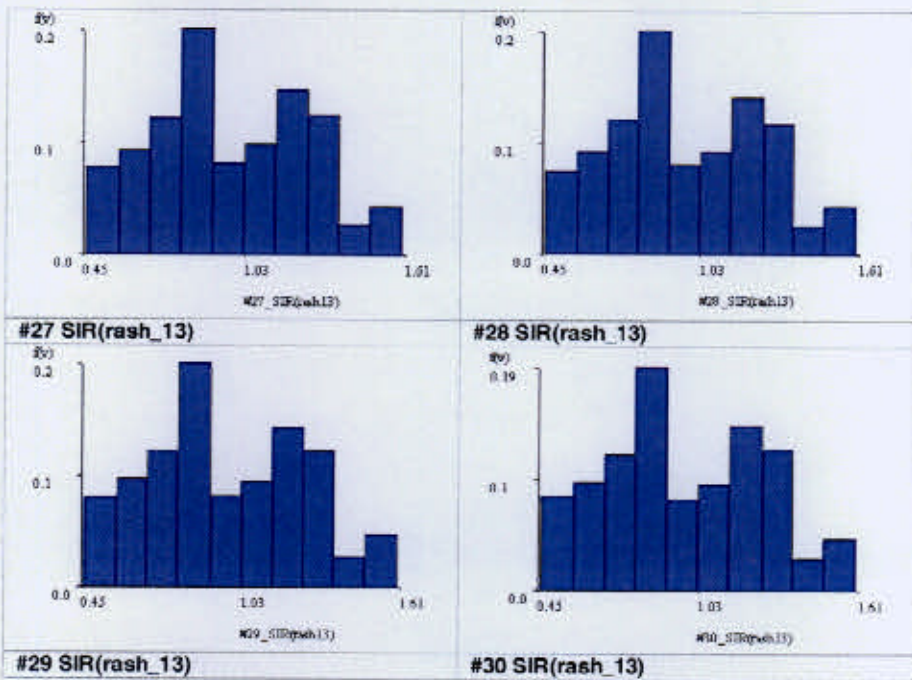






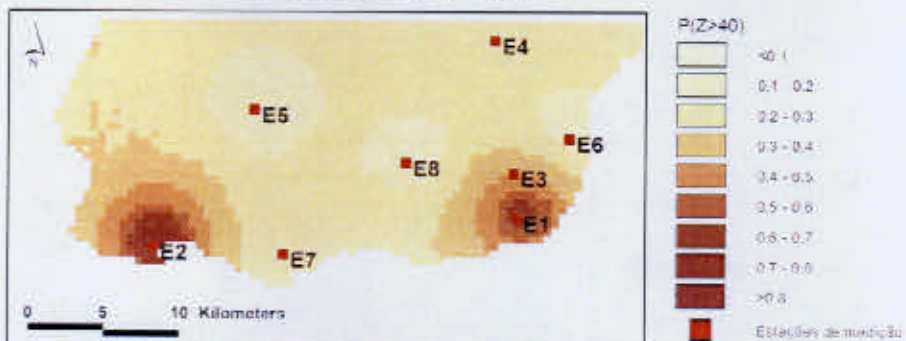
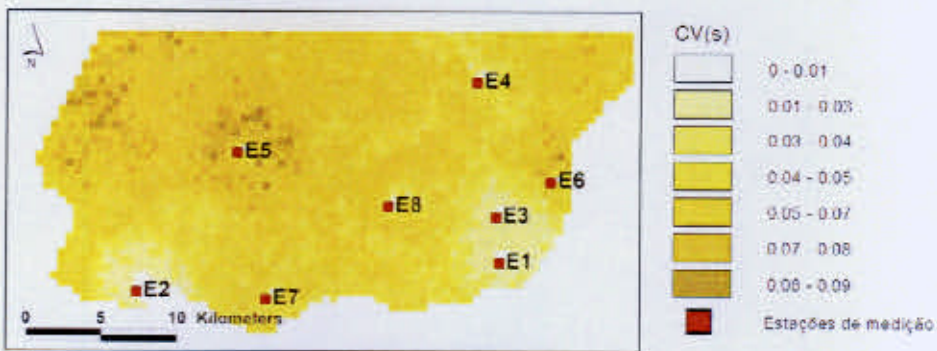
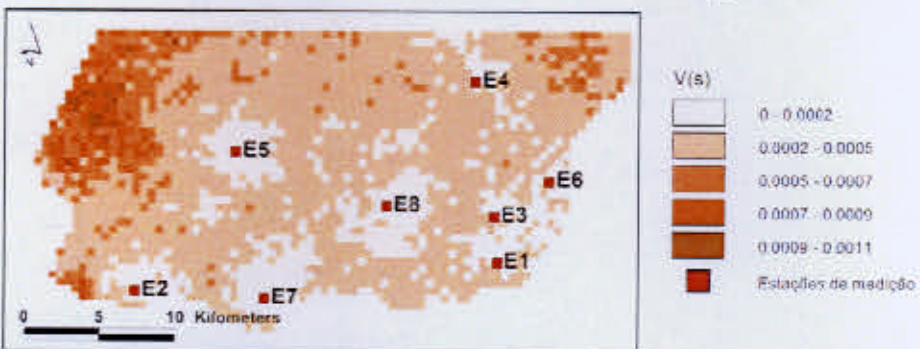
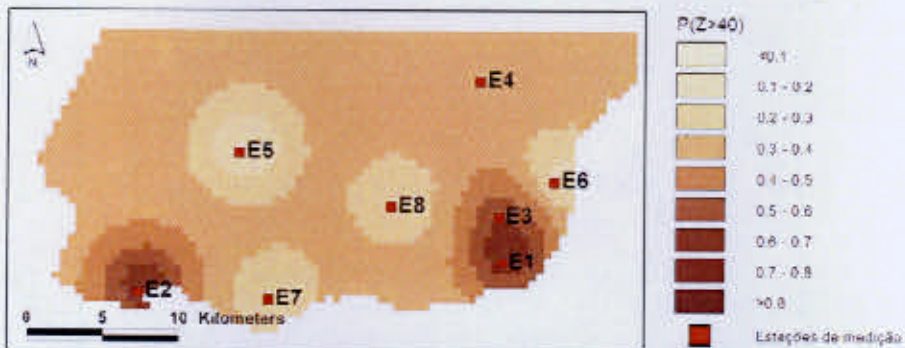




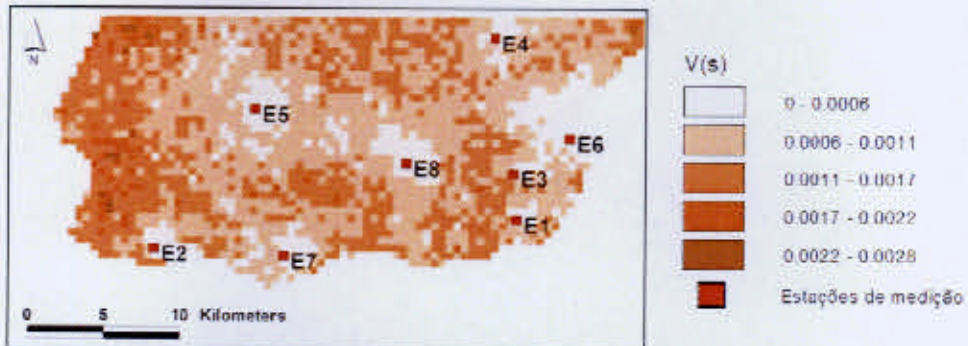


## ANEXO F

**Mapas de probabilidade, variância e coeficientes de variação das PM<sub>10</sub> por período  
(valor de corte=40 µg/m<sup>3</sup>)**



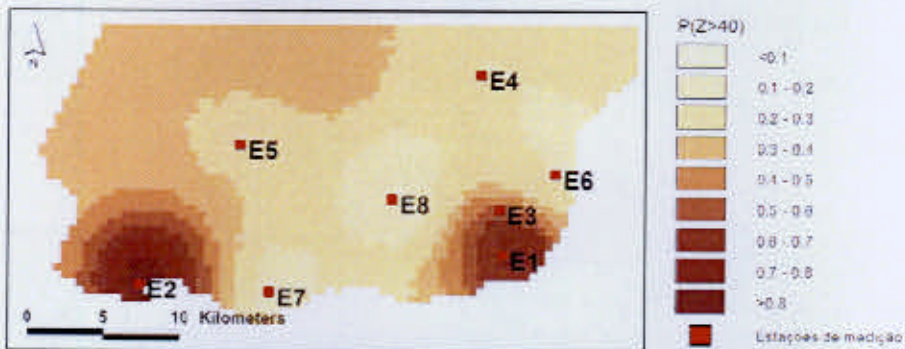
2.a. Período C- Mapa de probabilidades para um valor de corte=40  $\mu\text{g}/\text{m}^3/\text{dia}$



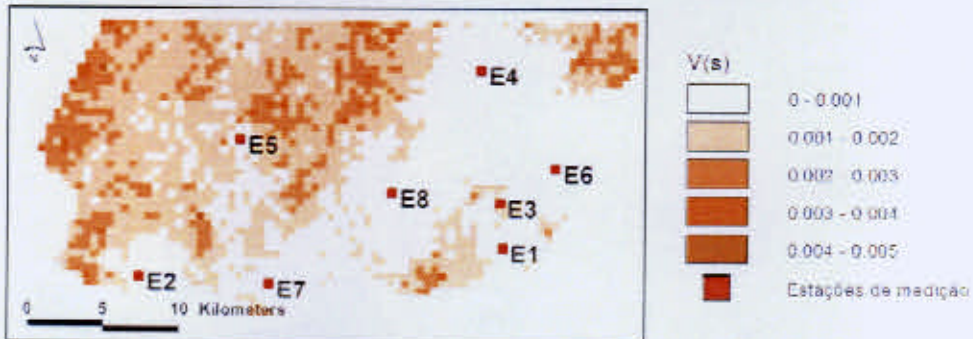
2.b. Período C- Mapa de variâncias



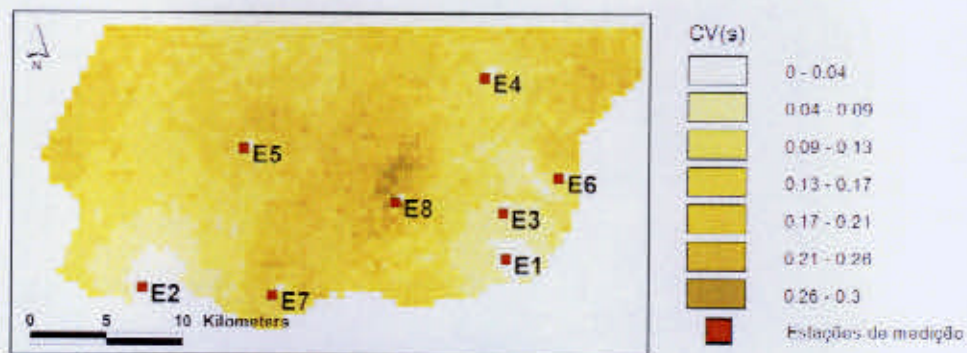
2.c. Período C- Mapa de coeficientes de variação



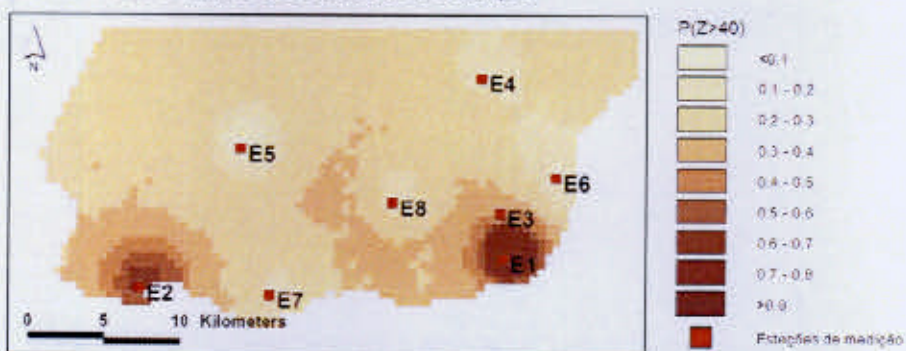
3.a. Período D- Mapa de probabilidades para um valor de corte=40  $\mu\text{g}/\text{m}^3/\text{dia}$



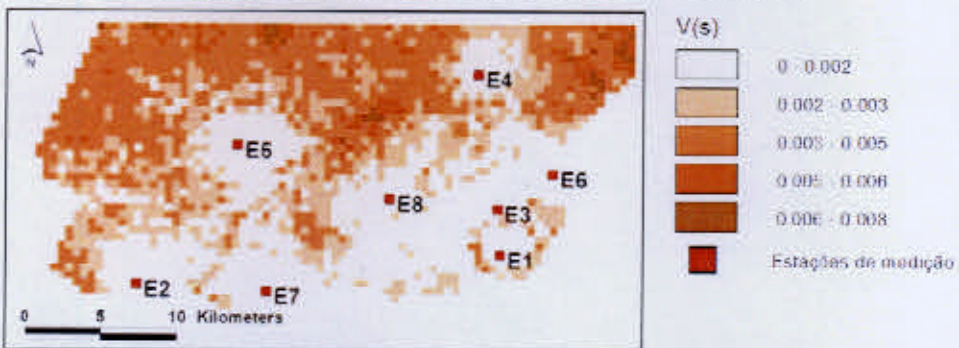
3.b. Período D- Mapa de variâncias



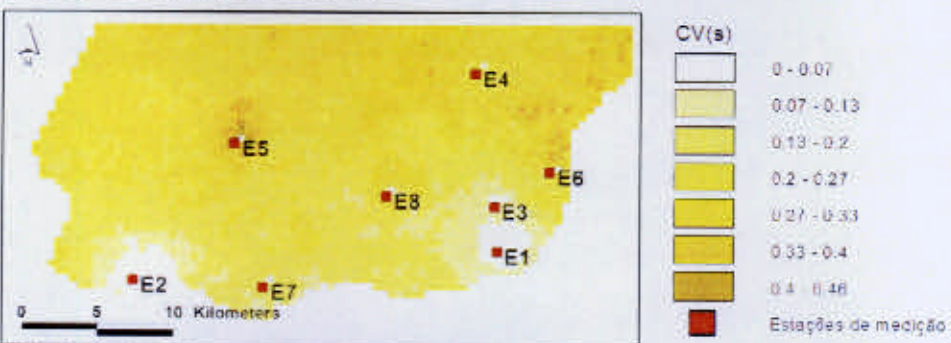
3.c. Período D- Mapa de coeficientes de variação



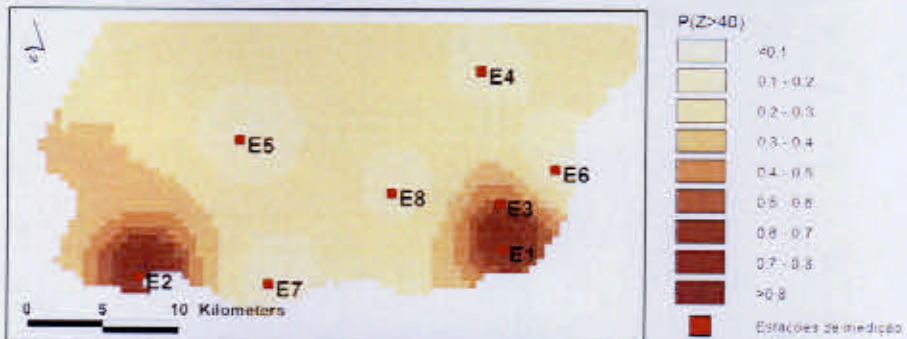
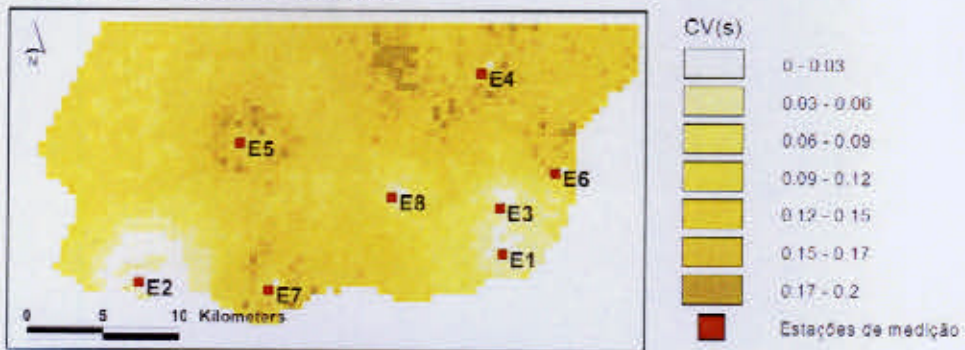
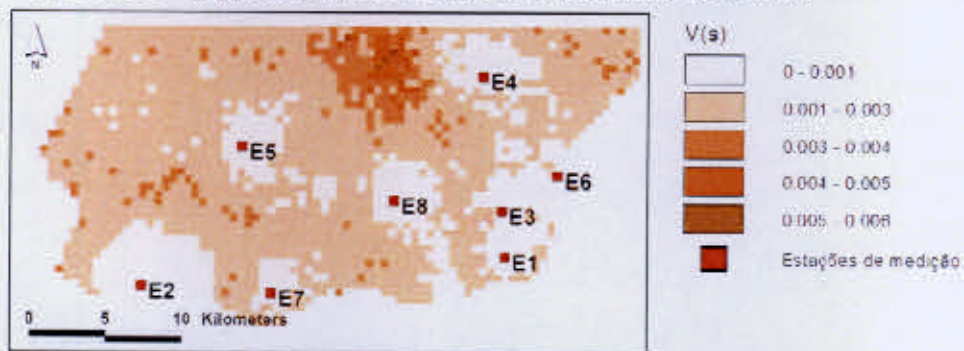
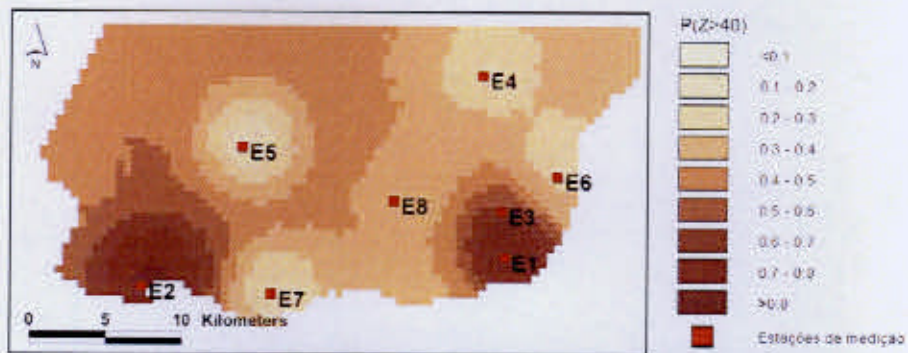
4.a. Período E- Mapa de probabilidades para um valor de corte= $40 \mu\text{g}/\text{m}^3/\text{dia}$

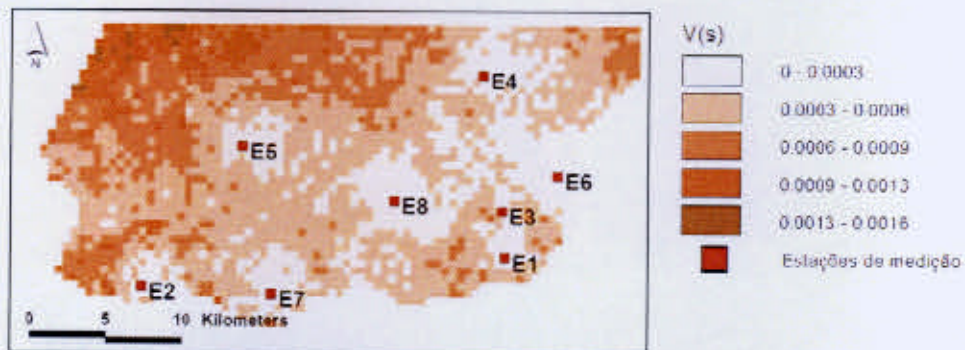


4.b. Período E- Mapa de variâncias

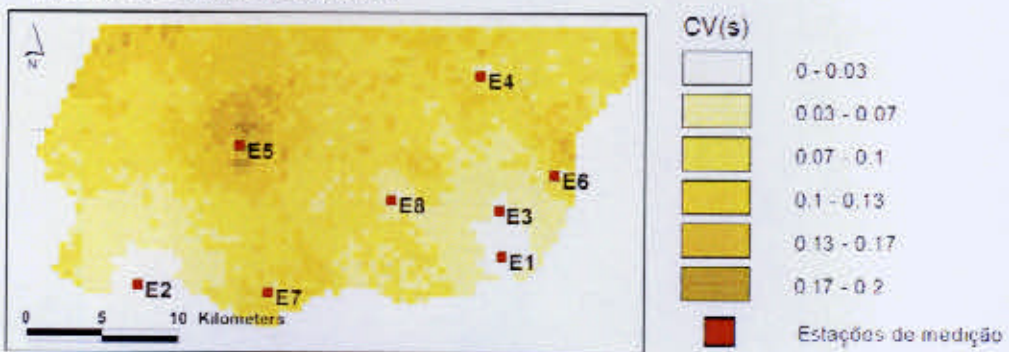


4.c. Período E- Mapa de coeficientes de variação





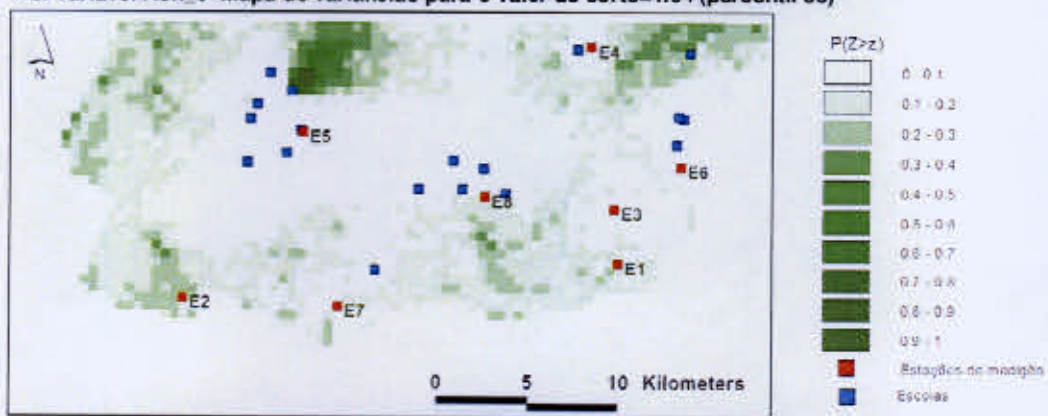
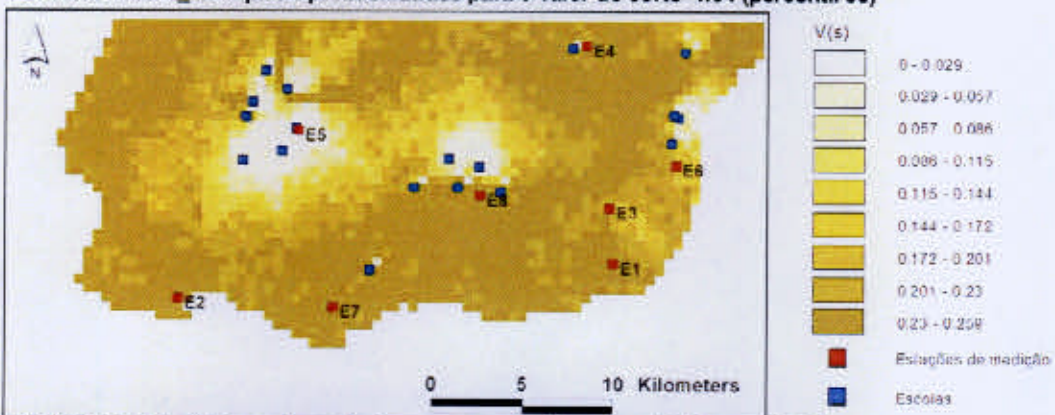
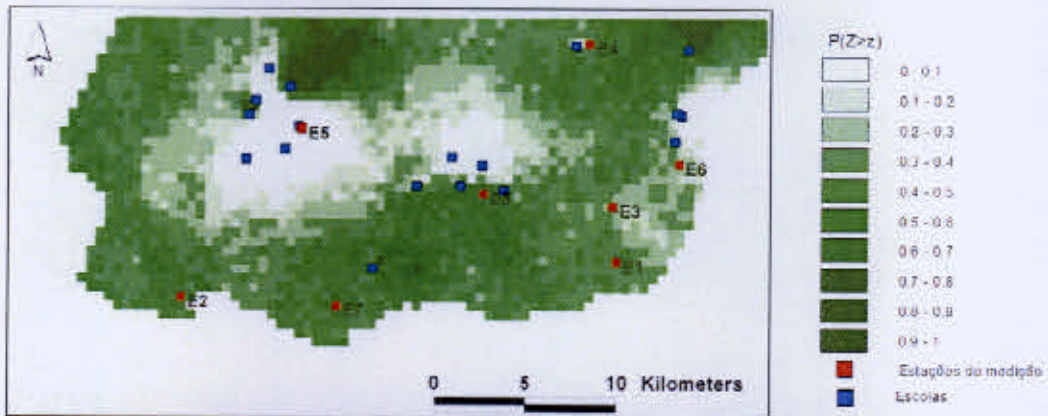
**6.b. Período G- Mapa de variâncias**

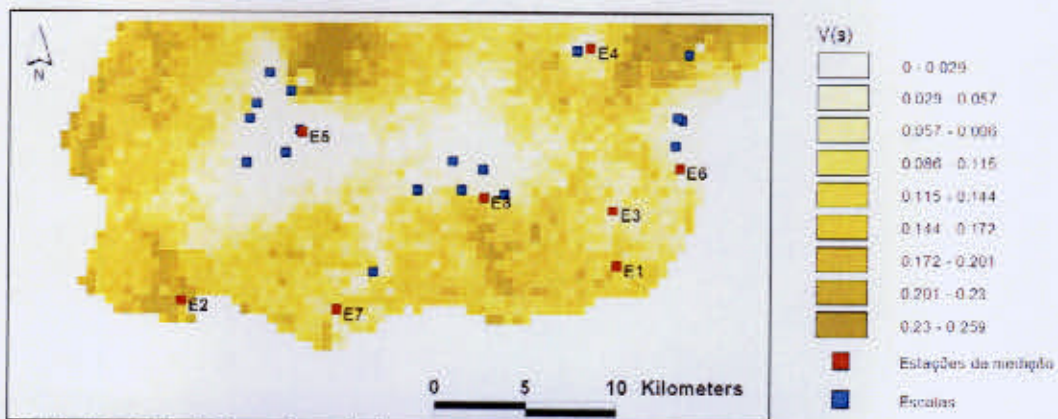


**6.c. Período G- Mapa de coeficientes de variação**

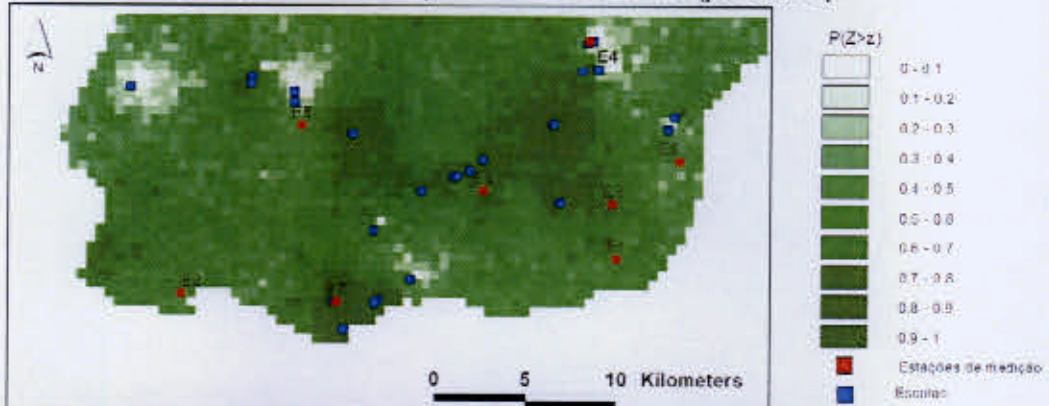
## **ANEXO G**

**Mapas de probabilidades e de variâncias das variáveis da saúde  
(valores de corte= P.50 e P.75)**

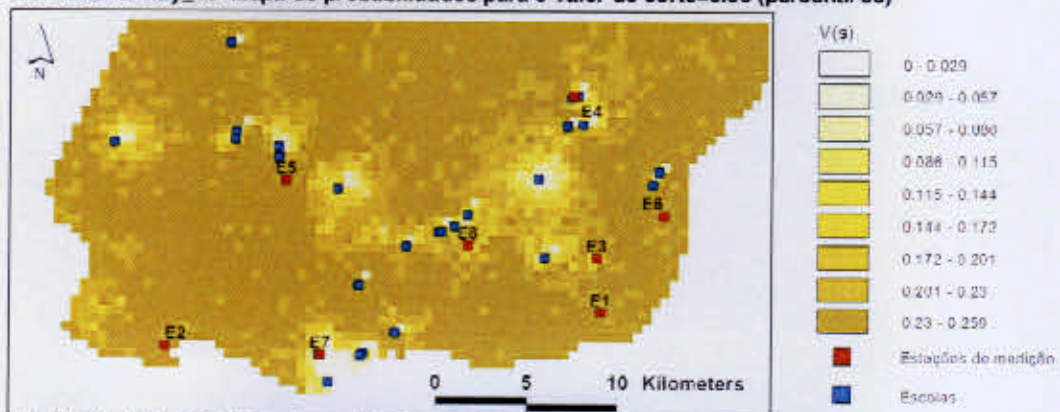




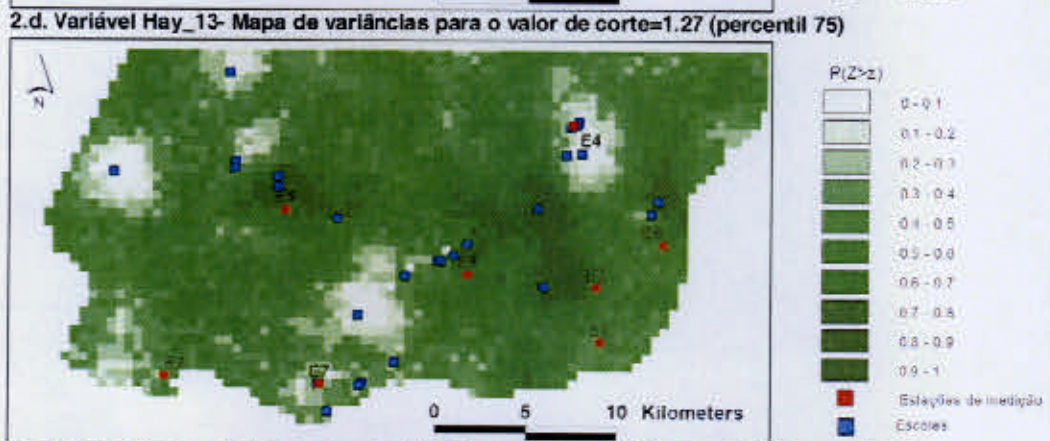
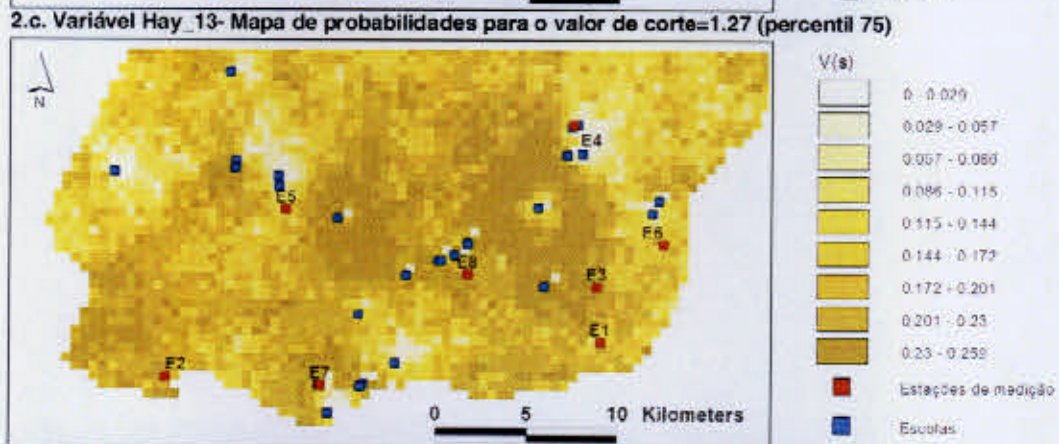
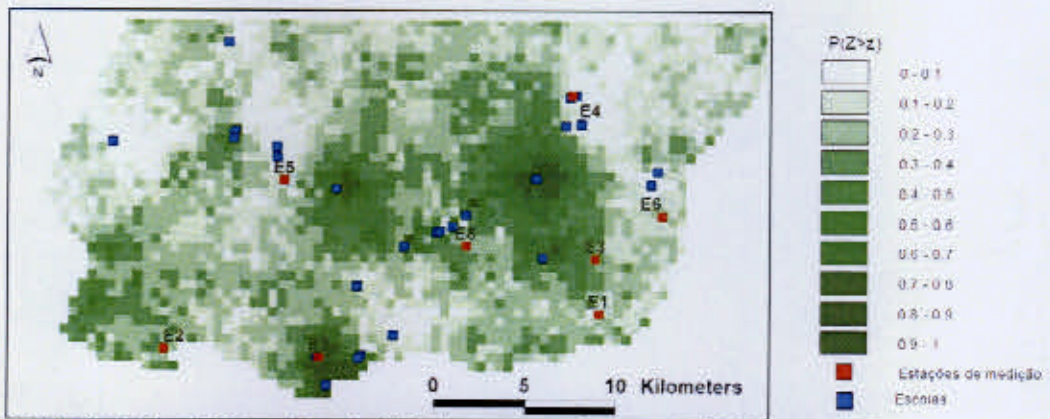
1.d. Variável Ash\_6- Mapa de variâncias para o valor de corte=1.31 (percentil 75)

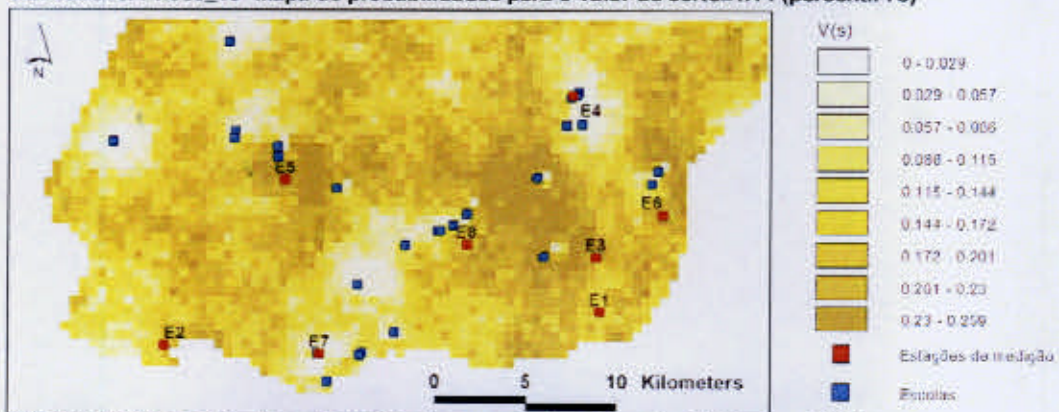
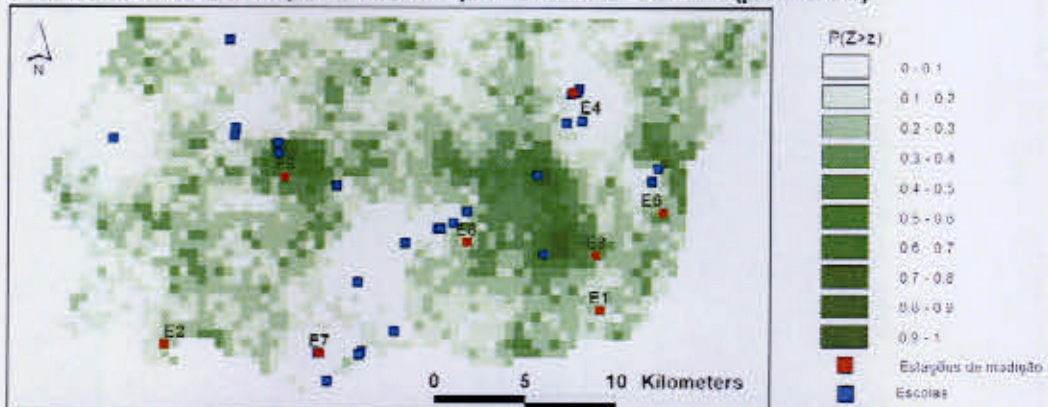
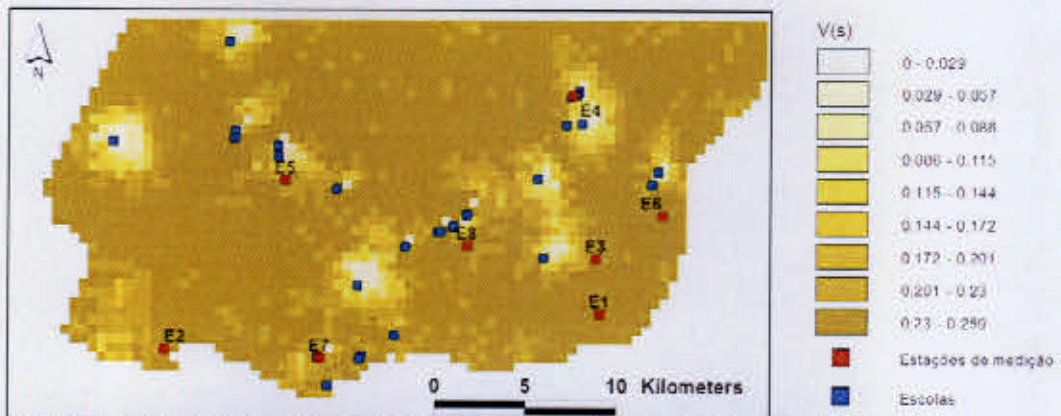


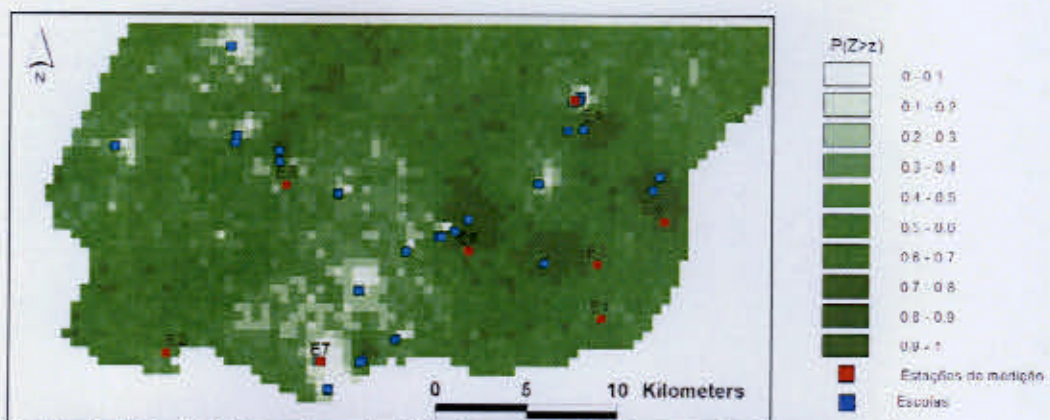
2.a. Variável Hay\_13- Mapa de probabilidades para o valor de corte=0.90 (percentil 50)



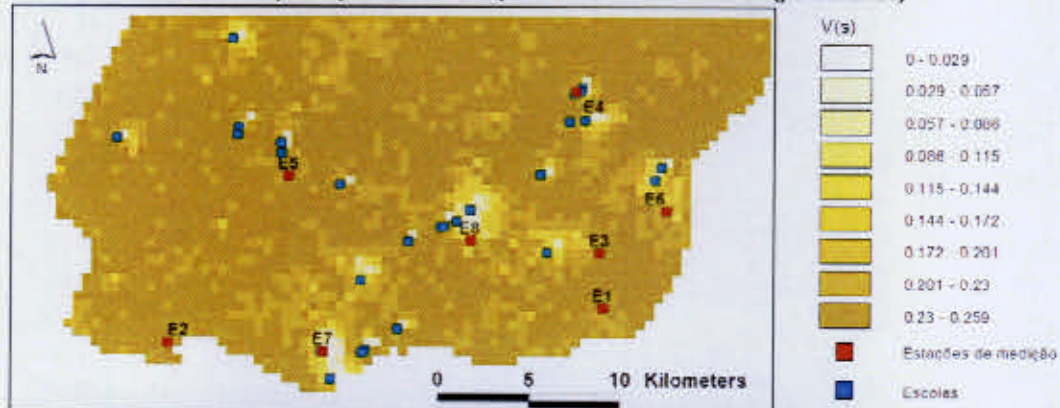
2.b. Variável Hay\_13- Mapa de variâncias para o valor de corte=0.90 (percentil 50)



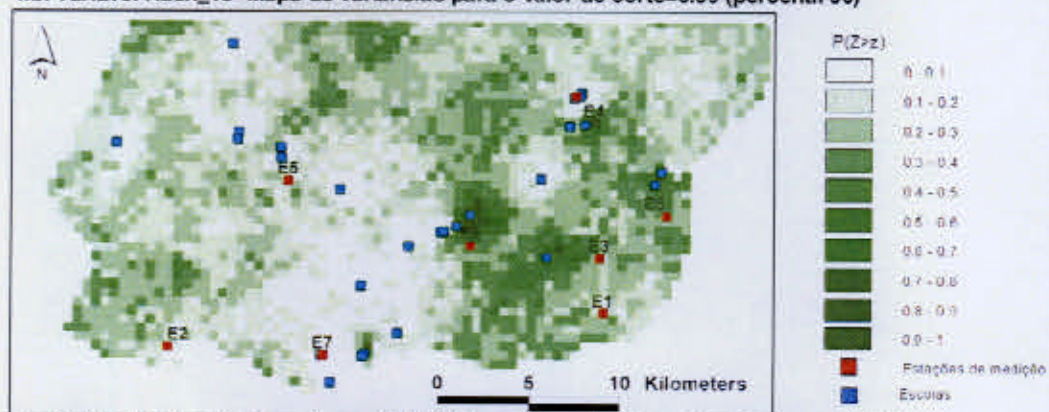




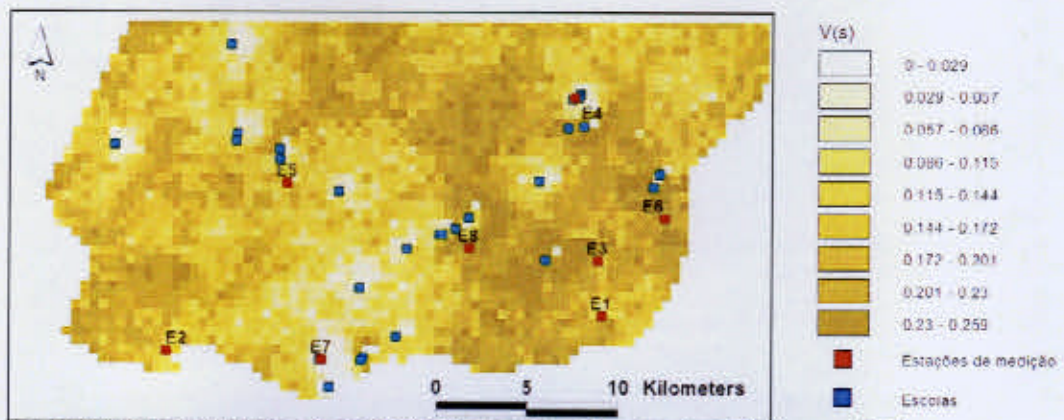
4.a. Variável Rash\_13- Mapa de probabilidades para o valor de corte=0.89 (percentil 50)



4.b. Variável Rash\_13- Mapa de variâncias para o valor de corte=0.89 (percentil 50)



4.c. Variável Rash\_13- Mapa de probabilidades para o valor de corte=1.20 (percentil 75)

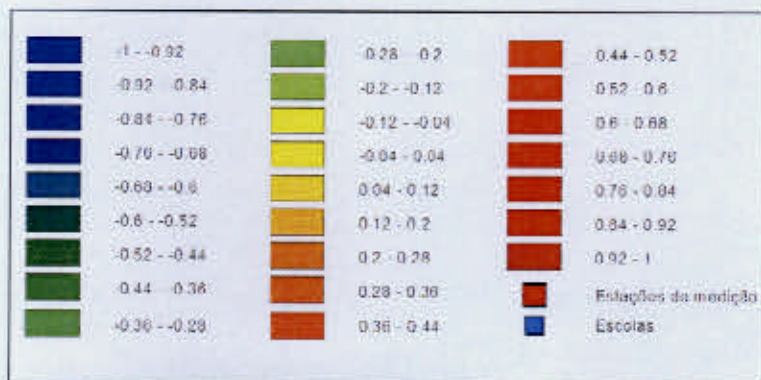


4.d. Variável Rash\_13- Mapa de probabilidades para o valor de corte=1.20 (percentil 75)

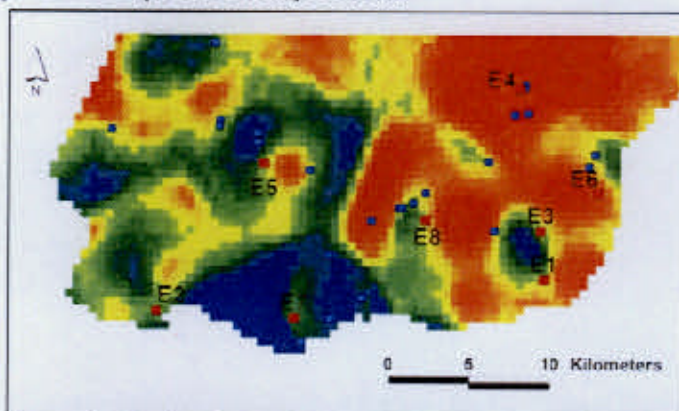
## **ANEXO H**

**Mapas de correlações locais com buffers de 2500, 5000 e 10000 metros**

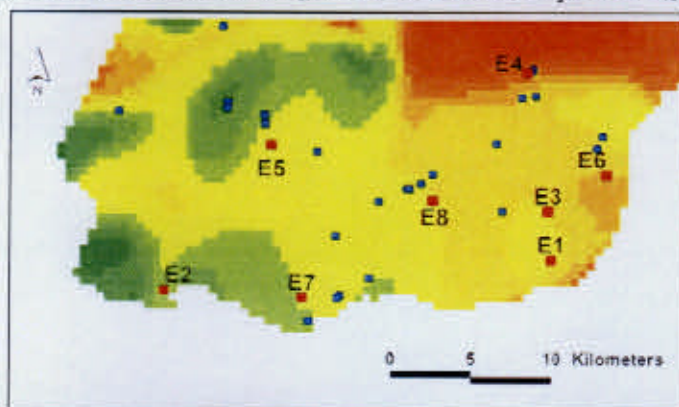
A definição do *buffer* baseou-se numa análise qualitativa de resultados observados para *buffers* de 2500, 5000 e 10000 metros. Calcularam-se os mapas de correlações locais para os três casos e a escolha foi feita à luz de critérios de coerência e relevância dos resultados obtidos. Apresenta-se a seguir um exemplo ilustrativo dos critérios utilizados.



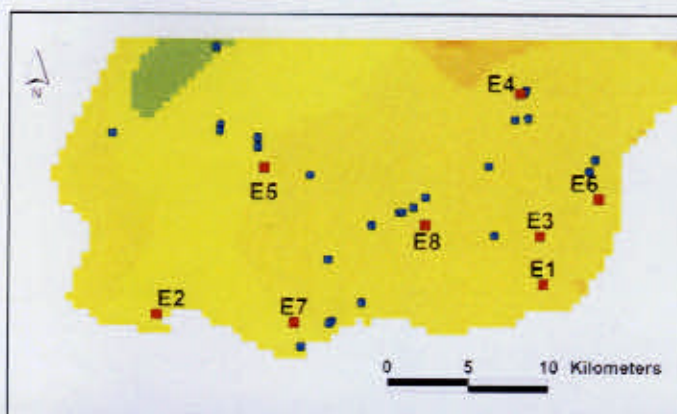
Legenda dos mapas de correlações locais



a. Mapa de correlações locais para *buffer* de 2500 metros, entre a excedência da razão padronizada de incidência de ash 13 = 1.07 e da concentração de  $PM_{10} = 20 \mu g/m^3$



b. Mapa de correlações locais para *buffer* de 5000 metros, entre a excedência da razão padronizada de incidência de ash 13 = 1.07 e da concentração de  $PM_{10} = 20 \mu g/m^3$



c. Mapa de correlações locais para *buffer* de 10000 metros, entre a excedência da razão padronizada de incidência de  $ash_{13} = 1.07$  e da concentração de  $PM_{10} = 20 \mu g/m^3$

Neste exemplo, para o *buffer* de 2500 metros os resultados evidenciaram correlações positivas bastante elevadas na zona oriental da área de estudo (que abrange as áreas de Lisboa, Loures e Mem Martins). No entanto, na região entre as estações E1 e E3, as correlações foram negativas e muito altas, ou seja, as variações nas probabilidades de se exceder o valor de corte da razão padronizada de incidência da asma (que neste caso é 1.07) são opostas às variações nas probabilidades de se excederem concentrações de partículas, o que é pouco plausível à luz dos conhecimentos actuais sobre esta matéria.

No *buffer* de 5000 metros, o alisamento das correlações provocou uma melhoria significativa na interpretabilidade do mapa, caracterizado por uma variação mais suave e mais coerente com os conhecimentos actuais sobre esta matéria.

No *buffer* de 10000 metros verificou-se um alisamento dos coeficientes obtidos de tal ordem que os resultados obtidos deixaram de ser relevantes do ponto de vista estatístico. Neste exemplo os coeficientes de correlação situaram-se entre -0.17 e 0.14.