UNIVERSIDADE FEDERAL DA PARAÍBA

IZABEL CRISTINA ALCANTARA DE SOUZA

ANÁLISE ESPACIAL DA RELAÇÃO ENTRE PLUVIOMETRIA E INCIDÊNCIA DOS CASOS NOTIFICADOS DO DENGUE NA PARAÍBA EM 2004

João Pessoa
2004

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

IZABEL CRISTINA ALCANTARA DE SOUZA

ANÁLISE ESPACIAL DA RELAÇÃO ENTRE PLUVIOMETRIA E INCIDÊNCIA DOS CASOS NOTIFICADOS DO DENGUE NA PARAÍBA EM 2004

Monografia apresentada ao curso de Bacharelado em Estatística da Universidade Federal da Paraíba, em cumprimento às exigências para obtenção do grau de Bacharel.

Tutor: *Ronei Marcos de Moraes*
Depto. Estatística - UFPB

Pedro Monteiro de Almeida
Secretaria de Estado da Saúde - PB

João Pessoa - PB
2004

IZABEL CRISTINA ALCANTARA DE SOUZA
Bacharel em Estatística

ANÁLISE ESPACIAL DA RELAÇÃO ENTRE PLUVIOMETRIA E INCIDÊNCIA DOS CASOS NOTIFICADOS DO DENGUE NA PARAÍBA EM 2004

Monografia apresentada ao curso de Bacharelado em Estatística da Universidade Federal da Paraíba, em cumprimento às exigências para obtenção do grau de Bacharel.

Aprovada em: *22/11/2004,* pela banca examinadora:

Tutor: *Prof. Dr. Ronei Marcos de Moraes* - UFPB

Prof. Dr. João Agnaldo do Nascimento - UFPB

Prof. Ms. Sydney Gomes da Silva - UFPB

Ao prof. Ronei Marcos de Moraes
que me orientou nos caminhos
da graduação e iniciação científica.
OFEREÇO

À minha mãe Cristina,
meu irmão Victor
e Ceiça (em memória).
DEDICO

AGRADECIMENTOS

Agradeço a Deus,
por me oferecer saúde, paz, sabedoria e apoio familiar.

Aos meus familiares,
pela compreenção, paciência e carinhos inesquecíveis.

Aos professores do Departamento de Estatística da UFPB,
pelos ensinamentos, dedicação e incentivo;

Aos amigos,
que sempre acreditaram no meu trabalho
e incentivam no meu crescimento profissional.

A Cleane e Pedro,
pela orientação e apoio, no meu estágio
na Secretaria de Estado da Saúde da Paraíba.

Aos membros da Banca Examinadora,
pela atenção e dedicação no exame desta monografia.

RESUMO

O dengue é uma virose cujo principal vetor é o mosquito *Aedes aegypti*. As fêmeas deste mosquito alimentam-se de sangue (transmitindo a doença quando infectadas pelo vírus do dengue) e depositam seus ovos nas paredes internas de objetos que acumulem água (na superfície). Segundo o Ministério da Saúde, o período de maior incidência do dengue é durante o verão, devido ao clima quente e chuvoso desta estação. Este estudo apresenta os resultados obtidos no estágio realizado na Secretaria de Estado da Saúde da Paraíba, onde foram coletados dados referentes ao número de notificações do dengue por município, no período de janeiro a maio de 2004. A partir dos dados de notificação do dengue, calculou-se os coeficientes de incidência do dengue por município paraibano neste período, a fim de verificar a relação entre a incidência do dengue e a pluviometria no estado. Os dados pluviométricos referem-se as amostras de precipitação (total do mês) das Estações Meteorológicas do estado da Paraíba, deste período, e foram cedidos pelo Departamento de Ciências Atmosféricas da Universidade Federal de Campina Grande. Utilizaram-se as técnicas de Krigeagem por Indicação para modelar as amostras pluviométricas e obter estimativas da precipitação média de cada município paraibano. Realizou-se comparação entre a pluviometria e a incidência do dengue através de gráficos, mapas e o coeficiente de correlação por postos de Kendall. Observou-se que na Paraíba, no período de janeiro a maio de 2004, a distribuição da incidência do dengue apresentou, como principal característica espacial, a ausência de agrupamentos de municípios paraibanos com incidência similar da doença. O mesmo não foi observado na distribuição pluviométrica, que apresentou aspectos de continuidade espacial, isto é, crescimento (ou decrescimento) gradativo dos níveis de precipitação numa determinada direção. Além disto, pode-se notar que no período de maior incidência do dengue (meses de fevereiro e março para mesorregiões do Agreste, Borborema e Sertão e meses de janeiro e fevereiro para Mata Paraibana), os níveis de precipitação estavam decrescendo. A partir deste estudo, verificaram-se que os coeficientes de correlação entre a pluviometria e a incidência do dengue estavam muito próximos de zero, entre -0,0656 e 0,0918. As hipóteses destes coeficientes de correlação serem iguais a zero não foram rejeitadas, com *valores p* superiores a 0,08. Apesar de ter verificado a ausência de correlação entre a pluviometria e a incidência do dengue, neste período, a estação do verão foi a que apresentou maior incidência do dengue, o que está de acordo com os alertas do Ministério da Saúde.

Palavras-chave: Estatística Aplicada. Dengue. Análise Espacial. Pluviometria.

LISTA DE FIGURAS

Figura 1	Representação da distância entre duas amostras	21
Figura 2	Semivariograma Indicador experimental e o estimado	28
Figura 3	Mesorregiões Geográficas da Paraíba	41
Figura 4	Municípios paraibanos segundo a presença de casos do dengue no período de janeiro a maio de 2004	43
Figura 5	Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de janeiro de 2004	44
Figura 6	Ajuste do Semivariograma Esférico para o valor de corte igual a 372,6 das amostras de precipitação do mês de janeiro de 2004	48
Figura 7	Diagramas de dispersão e correlações de Kendall entre as variáveis precipitação média e coeficiente de incidência do dengue por município paraibano segundo os meses	50
Figura 8	Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de janeiro de 2004	51
Figura 9	Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de fevereiro	52
Figura 10	Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de março de 2004	53
Figura 11	Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de abril de 2004	54
Figura 12	Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de maio de 2004	55
Figura 13	Série Temporal da precipitação média segundo as mesorregiões paraibanas	57
Figura 14	Série Temporal da incidência do dengue segundo as mesorregiões paraibanas	57

Figura A1 Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de fevereiro de 2004 63

Figura A2 Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de março de 2004 64

Figura A3 Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de abril de 2004 65

Figura A4 Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de maio de 2004 66

LISTA DE TABELAS

Tabela 1	Medidas Descritivas dos coeficientes de incidência do dengue (por 10.000 hab) por mês	42
Tabela 2	Medidas Descritivas das amostras de precipitação acumulada (em mm) por mês	43
Tabela 3	Testes de Kolmogorov - Smirnov para o ajuste a distribuição normal da precipitação mensal e seus respectivos *valores p*	45
Tabela B1	Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de janeiro de 2004	67
Tabela B2	Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de fevereiro de 2004	67
Tabela B3	Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de março de 2004	67
Tabela B4	Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de abril de 2004	68
Tabela B5	Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de maio de 2004	68

LISTA DE SÍMBOLOS

$\overline{x}$	- média aritmética da variável aleatória X
s^2	- variância amostral da variável aleatória X
p	- proporção, talque $0 \leq p \leq 1$
$q_{(p)}$	- quantil de ordem p
g	- número de grupos dos métodos de agrupamento espacial
$u = (x, y)$	- posição do espaço com coordenadas: latitude x e longitude y
G	- Região Geográfica de Estudo
h_{jk}	- vetor de distância entre as amostras u_j e u_k
h_j	- vetor de distância entre a amostra u_j e uma posição qualquer u
$\parallel h_{jk} \parallel$	- comprimento (norma) do vetor de distância h_{jk}
$Z(u)$	- variável aleatória na posição u
$z(u_j)$	- amostra da variável aleatória $Z(u_j)$
$v.a.$	- variável aleatória
fda	- função de distribuição acumulada
$fdac$	- função de distribuição acumulada condicionada
$F(u, z)$	- função de distribuição acumulada
$F(u, z \mid n)$	- função de distribuição acumulada condicionada as n amostras
fa	- Função Aleatória (ou Processo Estocástico)
$F(u_1, ..., u_k;$	
$z_1, ..., z_k)$	- função de distribuição acumulada multivariada
z_c	- valor de corte pertencente ao espaço de estados $\{Z(u) : u \in G\}$
$I(u, z_c)$	- variável por indicação
$P[\quad]$	- probabilidade
$E[\quad]$	- valor esperado
$V[\quad]$	- variância
$2\gamma(h, z_c)$	- variograma indicador
$\gamma(h, z_c)$	- semivariograma indicador
$\hat{\gamma}(h, z_c)$	- semivariograma indicador estimado
$\tilde{\gamma}(h, z_c)$	- semivariograma indicador experimental
$N(h)$	- número de amostras separadas por um vetor de distância h
α	- ângulo do vetor de distância
d	- parâmetro de tolerância de distância
t	- parâmetro de tolerância angular
$C[\quad]$	- covariância
a	- Alcance
C	- Patamar
C_0	- Efeito Pepita
C_1	- Contribuição
$M(\parallel h \parallel)$	- modelo de semivariograma
$m_z(u)$	- valor médio verdadeiro da variável $Z(u)$
$\hat{m}_z(u)$	- estimativa do valor médio da variável $Z(u)$
$n(u)$	- número de amostras vizinhas a posição u
$w_j(u)$	- peso de ponderação j a ser estimado na posição u
$m_z(u_j)$	- valor médio verdadeiro da variável $Z(u_j)$
$w_j(u, z_c)$	- peso de ponderação j a ser estimado na posição u e valor de corte z_c
σ_R^2	- variância do erro de estimação do modelo de Krigeagem por Indicação

$\phi(u, z_c)$	- parâmetro de Lagrange para uma posição u e valor de corte z_c
$\hat{\mu}(u)$	- estimativa do valor médio de $Z(u)$
$\hat{\sigma}^2(u)$	- estimativa da variância de $Z(u)$
$q_{(R)}$	- intervalo interquartil
$\Pr(t)$	- tamanho da população no tempo t
Δt	- intervalo de tempo da projeção da população
r	- taxa de crescimento da população
τ	- coeficiente de correlação não paramétrico por postos de Kendall
l_j	- j-ésimo posto da variável X
v_j	- j-ésimo posto da variável Y
s_{ij}	- escores do método do coeficiente de correlação de Kendall
b_k	- número de observações empatadas em cada grupo k de empates na variável X
f_r	- número de observações empatadas em cada grupo r de empates na variável Y
μ_τ	- média da distribuição da variável τ
σ_τ^2	- variância da distribuição da variável τ

SUMÁRIO

1	INTRODUÇÃO	12
2	OBJETIVOS	14
3	METODOLOGIA	15
3.1	*SISTEMAS DE INFORMAÇÃO GEOGRÁFICA*	15
3.2	*ANÁLISE EXPLORATÓRIA DOS DADOS*	16
3.3	*MÉTODOS DE AGRUPAMENTOS ESPACIAIS*	19
3.4	*KRIGEAGEM POR INDICAÇÃO*	20
3.4.1	*DISTÂNCIA ENTRE AS AMOSTRAS*	20
3.4.2	*KRIGEAGEM*	21
3.4.3	*FUNÇÃO ALEATÓRIA*	22
3.4.4	*CODIFICAÇÃO POR INDICAÇÃO*	24
3.4.5	*SEMIVARIOGRAMA INDICADOR*	25
3.4.6	*KRIGEAGEM POR INDICAÇÃO*	30
3.4.7	*CORREÇÃO DOS DESVIOS DE ORDEM*	34
3.4.8	*ESTIMAÇÃO DE VALORES*	35
3.5	*COEFICIENTE DE INCIDÊNCIA DO DENGUE*	37
3.6	*ANÁLISE DE CORRELAÇÃO POR POSTOS DE KENDALL*	38
3.7	*MESORREGIÕES GEOGRÁFICAS*	40
4	ANÁLISE DOS RESULTADOS	42
4.1	*ANÁLISE EXPLORATÓRIA DOS DADOS*	42
4.2	*MODELAGEM DOS DADOS PLUVIOMÉTRICOS*	46
4.3	*CORRELAÇÃO ENTRE PLUVIOMETRIA E INCIDÊNCIA DO DENGUE*	48
4.4	*RESULTADOS POR MESORREGIÕES GEOGRÁFICAS*	55
5	CONCLUSÕES	58
	REFERÊNCIAS	59
	APÊNDICES	62

1 INTRODUÇÃO

O dengue é uma doença infecciosa que, atualmente no Brasil, pode ser causada por três sorotipos diferentes de virus do gênero *Flavivirus* (denominados de DEN 1, DEN 2 e DEN 3). Esta doença é transmitida ao ser humano através da picada do mosquito *Aedes aegypti* ou do *Aedes albopictus*, contaminado por um dos virus do dengue. O *Aedes albopictus* é um mosquito de ambiente silvestre, encontrado em algumas regiões do Brasil e o *Aedes aegypti* é um mosquito de ambiente urbano, considerado o maior transmissor do dengue e que está presente em quase todo o território brasileiro. O *Aedes aegypti* alimenta-se exclusivamente de frutas, no entanto, as fêmeas desta espécie alimentam-se de sangue, necessário para o amadurecimento dos ovos, ou seja, a fêmea é o principal vetor do dengue. Os ovos são depositados nas paredes internas de objetos que acumulem água (na superfície) e desta forma, todo objeto que contenha água parada, torna-se um ambiente que oferece condições de reprodução favoráveis para estes mosquitos (COSTA; FERREIRA, 2002).

Segundo o Ministério da Saúde (FARHAT, 2003), o período de maior incidência do dengue é durante o verão, devido ao clima quente e chuvoso comum nesta estação. Neste período do ano, o governo passa a investir mais no combate da doença através da contratação de agentes comunitários de saúde e de campanhas de mobilização comunitária.

No entanto, estudos realizados sobre o dengue no estado da Paraíba mostram que o período de maior ocorrência de internações hospitalares do dengue no Sistema Único de Saúde - SUS - é durante a estação do outono, com picos em abril nos anos pares e junho nos anos ímpares (FURTADO; SOUZA; MORAES, 2004). A fonte de dados destes estudos foi o *site* do Departamento de Informática do SUS, que informa apenas o número de internações hospitalares do dengue por município paraibano.

Considerado um dos principais problemas de saúde pública do mundo, o dengue é uma doença de notificação compulsória, isto é, por lei, todos os casos conhecidos desta doença devem ser notificados ao Ministério da Saúde. Uma das ações do Plano Nacional de Controle do Dengue, instituído em 24 de julho de 2002, foi manter o Sistema Nacional de Agravos de Notificação - SINAN - como única fonte de dados de notificação do dengue (BRASIL, 2002), sendo desta forma, a fonte oficial de informações sobre o dengue.

Esta monografia apresenta os resultados obtidos no estágio realizado na Secretaria de Estado da Saúde da Paraíba (SES/PB), onde foram coletados dados referentes ao número de notificações do dengue por município, no período de janeiro a maio de 2004 (dados da base de informação do SINAN da SES/PB). A partir dos dados de notificação do dengue, calculou-se os coeficientes de incidência do dengue por município paraibano neste período, a fim de verificar a relação entre a incidência do dengue e a pluviometria no estado. Os dados pluviométricos referem-se as amostras de precipitação (total do mês) das Estações Meteorológicas do estado da Paraíba (período de janeiro a maio de 2004) e foram cedidos pelo Departamento de Ciências Atmosféricas da Universidade Federal de Campina Grande.

2 OBJETIVOS

O objetivo principal deste trabalho é realizar um estudo exploratório entre as variáveis incidência do dengue e pluviometria, no estado da Paraíba, no período de janeiro a maio de 2004, através de técnicas Estatísticas e Geoestatísticas. Com este estudo, pretende-se resolver as seguintes questões específicas:

1) Identificar padrões na distribuição espacial da precipitação e do coeficiente de incidência do dengue na Paraíba.

2) Determinar a Correlação estatística entre precipitação e coeficiente de incidência do dengue no período considerado.

3) Detectar padrões entre estas duas variáveis em cada mesorregião geográfica do estado da Paraíba.

3 METODOLOGIA

3.1 SISTEMAS DE INFORMAÇÃO GEOGRÁFICA

Sistemas de Informação Geográfica - SIG - são tecnologias do Geoprocessamento que permitem representar em ambientes computacionais fenômenos geográficos, por exemplo, a distribuição espacial de dados meteorológicos e da ocorrência do dengue (CARVALHO; PINA; SANTOS, 2000).

Os dados da ocorrência de um fenômeno são denominados de **dados de área** quando se referem a indivíduos localizados em pontos específicos do espaço que, por motivo de confidencialidade ou de tratamento estatístico, são agregados em unidades de análise, usualmente delimitadas por polígonos fechados (setores censitários, municípios,...) (MORAES; SOUZA; TELES, 2003). Dados sobre saúde, demográficos e sócio-econômico são exemplos típicos.

A representação computacional, num ambiente SIG, dos dados de área é denominada de **Mapa Cadastral**. Este mapa permite associar duas componentes dos dados de área: a parte gráfica, que representa simbolicamente uma área no mapa (ou imagem) e a parte não-gráfica, que armazena as características qualitativas e quantitativas desta área (SOUZA; MORAES, 2003).

Quando os dados são obtidos de levantamentos de recursos naturais, nos quais, a sua natureza aleatória é a própria superfície do fenômeno estudado, estes são denominados de **dados de superfície**. Estão disponíveis usualmente como um conjunto de valores (amostras), que podem estar regularmente ou irregularmente distribuídos no espaço e são modelados como uma amostra de uma superfície contínua. Por exemplo: a distribuição da precipitação em um estado ou a distribuição do ph no solo. No SIG, a representação

matemática computacional dos dados de superfície é denominada de **Modelo Numérico de Terreno - MNT** (SOUZA; MORAES, 2004).

O programa de SIG utilizado neste estudo é o Sistema de Processamento de Informações Georeferenciadas - SPRING. Neste, as amostras de MNT da precipitação e os dados cadastrais do dengue são organizados em bancos de dados geográficos, segundo uma arquitetura hierárquica denominada de modelagem orientada a objetos. Segundo esta arquitetura, a composição desta estrutura de dados tem início com a criação de um Banco de Dados. Dentro deste Banco de Dados pode ser definido um ou mais projetos, onde cada um deles define uma área física de trabalho e é composto por Planos de Informação - PIs. A área física de trabalho é denominada de retângulo envolvente e representa a região geográfica delimitada por duas coordenadas geográficas. No caso das amostras de MNT, a cada Plano de Informação é associado uma grade regular (ou irregular) de amostras da mesma área do Projeto, com dados espaciais diferentes. Estes PI's são agrupados por Categorias, facilitando a organização das informações no MNT. Já no caso dos dados cadastrais do dengue, a parte não-gráfica (os dados) são armazenados numa tabela, onde cada coluna corresponde a uma informação diferente, denominada de atributo e cada linha corresponde a um município (objeto geográfico). As linhas são associadas a parte gráfica (no caso o mapa político da Paraíba, obtido no *site* do DATASUS (DATASUS, 2003)) através de um identificador denominado de *GEOID* (CÂMARA et al., 1996). A descrição completa da construção destes tipos de banco de dados pode ser encontrada em Souza e Moraes (2003, 2004).

3.2 ANÁLISE EXPLORATÓRIA DE DADOS

Análise Exploratória de Dados é um conjunto de procedimentos estatísticos que visam descrever e identificar o tipo de distribuição das variáveis aleatórias. Apesar de não ser

um método de Análise Espacial, a Análise Exploratória de Dados é uma etapa primordial para realização de análises mais complexas, por que vários procedimentos Estatísticos e Geoestatísticos são baseados na suposição de que os dados provêm de uma distribuição Normal ou supõem conhecido o tipo de distribuição das variáveis aleatórias.

A descrição dos dados é realizada através da Estatística Descritiva, que se utiliza, de métodos numéricos para resumir as informações contidas nos dados, como por exemplo: Medidas de Tendência Central, Medidas de Dispersão e os Quantis. Utiliza-se também de métodos gráficos, que permitem visualizar padrões de comportamento nos dados (TUKEY, 1977).

As Medidas de Tendência Central permitem descrevem a tendência que os dados têm de se agrupar em torno de certos valores. Dessas medidas as mais usuais são: a média aritmética, a mediana e a moda.

- **Média aritmética**, denotada por $\bar{x}$, representa o centro de gravidade de uma amostra aleatória com n valores, $x_1, x_2, ..., x_n$ e é definido por (1):

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

- A **mediana** de uma variável aleatória X ou de sua correspondente distribuição é denotada por ξ e é definida como o menor valor ξ que satisfaz $F_x(\xi) \geq 0,5$, onde F_x é a função densidade acumulada de X (MOOD; GRAYBILL; BOES, 1974).

- Os valores de uma amostra da variável aleatória, que ocorrem com maior freqüência são denominados de **moda**. Se a moda for um único valor, a distribuição é denominada de unimodal. Caso exista mais de um valor, as denominações são: bimodal para dois valores, trimodal para três valores e etc.

As informações fornecidas pelas medidas de Tendência Central, em geral, necessitam ser complementadas pelas medidas de Dispersão. Estas servem para indicar, o quanto os dados apresentam-se dispersos, em torno da região central (TUKEY, 1977). Dentre essas medidas destacam-se a variância e o desvio-padrão. Numa amostra com n observações $x_1, x_2, ..., x_n$, a **variância amostral** é definida por (2):

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} \tag{2}$$

A variância é uma medida expressa numa unidade igual ao quadrado da unidade dos dados, o que pode causar problemas de interpretação. Costuma-se usar o **desvio padrão**, denotado por s, definido como a raiz quadrada positiva da variância. O desvio padrão indica qual será o "erro" (desvio) cometido, ao tentar substituir cada observação pela média aritmética.

Apenas com as medidas de Tendência Central e Dispersão não dá pra se ter idéia da simetria (ou assimetria) da distribuição de uma variável aleatória X. Para contornar esse fato, pode-se utilizar as estatísticas de ordem de uma amostra aleatória de X (denominadas de quantis). Em uma amostra de valores ordenados $x_1, x_2, ..., x_n$, o quantil de ordem p é uma medida, indicada por $q_{(p)}$ e definida por (BUSSAB; MORETTIN, 2002):

$$q_{(p)} = \begin{cases} x_1, & se \quad p < p_1 \\ x_i, & se \quad p = p_i \\ (1-f_i)q_{(p_i)} + f_i q_{(p_{i+1})}, & se \quad p_i < p < p_{i+1} \\ x_n, & se \quad p > p_n \end{cases} \qquad com \quad i = 1, 2, ..., n$$

onde:

x_i é a observação de ordem i;

p é uma proporção qualquer, talque $0 \leq p \leq 1$;

$p_i = \dfrac{i - 0,5}{n}$, com $i = 1, 2, ..., n$;

$f_i = \dfrac{p - p_i}{p_{i+1} - p_i}$, com $i = 1, 2, ..., n$.

Quando a distribuição de uma variável aleatória apresenta uma forma aproximadamente simétrica, pode-se verificar as seguintes relações, onde os valores $q_{(0,25)}$, $q_{(0,5)}$ e $q_{(0,75)}$ são denominados de quartis e $q_{(0,5)}$ é a mediana (BUSSAB; MORETTIN, 2002):

(a) $q_{(0,5)} - x_1 \simeq x_n - q_{(0,5)}$;

(b) $q_{(0,5)} - q_{(0,25)} \simeq q_{(0,75)} - q_{(0,5)}$;

(c) $q_{(0,25)} - x_1 \simeq x_n - q_{(0,75)}$.

Quanto aos métodos gráficos, pode-se destacar o histograma e o qqplot. O histograma que permite visualizar a simetria (ou assimetria) da distribuição é um gráfico de barras contíguas, onde as bases são proporcionais aos intervalos de classes e a área de cada retângulo proporcional à respectiva freqüência (BUSSAB; MORETTIN, 2002).

O qqplot é um gráfico de comparação entre a distribuição dos valores observados da variável X e a distribuição dos quantis esperados, $q_{(p)}$, se a variável X tivesse distribuição Normal. Esse gráfico pode ser usado para avaliar a hipótese de normalidade dos dados. Se a variável tem distribuição Normal, os pontos $(x_i, q(p_i))$ desenhados num gráfico cartesiano devem se localizar sobre a reta $X = q_{(p)}$ (JOHNSON; WICHERN, 1992).

3.3 MÉTODOS DE AGRUPAMENTO ESPACIAL

O agrupamento consiste num processo de agrupar áreas, com base nos valores de suas variáveis (CÂMARA et al., 1996). No SPRING encontram-se três métodos de agrupamentos: Passo Igual, Quantis e Estatístico.

O método **Passo Igual** permite agrupar áreas em até 15 grupos, que são formados dividindo os valores de uma determinada variável X, em intervalos com amplitudes iguais (intervalos de classe). Cada intervalo de classe corresponde a um grupo e as áreas que

fazem parte deste grupo são coloridos no mapa, com uma cor específica, obtendo-se então um mapa coroplético (nome dado pelos geógrafos aos mapas coloridos).

O método **Quantil** também permite agrupar áreas em até 15 grupos, sendo que cada grupo tem aproximadamente a mesma proporção $(1/g)$ de áreas, onde g é o número de grupos. No entanto, os intervalos de classe podem ter amplitudes desiguais e são dados em termos dos quantis da variável aleatória X. Estes intervalos são denominados de intervalos inter-quantis. As áreas que fazem parte de um intervalo inter-quantil são agrupados e coloridos no mapa com uma determinada cor, compondo um mapa coroplético.

O método **Estatístico**, por sua vez, permite agrupar áreas com base na média aritmética $\overline{x}$ e no desvio padrão s de uma variável aleatória X, isto é, a distribuição da variável X pode ser dividida em intervalos com amplitudes iguais a: $\pm s$, $\pm(1/2)s$ ou $\pm(1/4)s$ em relação à média. Determinada a amplitude, as áreas, que fazem parte de cada um dos intervalos formados, são agrupados e coloridos com uma cor específica, obtendo-se também um mapa coroplético.

A grande dificuldade dos três métodos de agrupamento, consiste em que, para uma mesma variável, eles podem fornecer resultados diferentes. Deve-se ter sempre em mente que, cada método, tem um propósito particular que deve ser respeitado na interpretação do resultado. Uma análise exploratória da variável em estudo pode ajudar na escolha do método apropiado ou na elaboração de critérios de agrupamentos.

3.4 KRIGEAGEM POR INDICAÇÃO

3.4.1 DISTÂNCIA ENTRE AMOSTRAS

Considere uma região geográfica G, representada em um plano segundo uma projeção cartográfica, onde cada posição u é definida pelas cordenadas (x, y), ou seja,

$u = (x, y) \in G$. A distância entre duas posições de amostras, por exemplo, $u_j = (x_j, y_j)$ e $u_k = (x_k, y_k)$, pode ser determinada pelo comprimento do vetor de distância h_{jk} que as separa (ISAAKS; SRIVASTAVA, 1989). Considerando como referência um ponto qualquer $u \in G$, o vetor de distância entre o ponto de referência u e a posição da amostra u_j será denotado por h_j, como está ilustrado na Figura 1. Da mesma maneira, o vetor de distância entre o ponto de referência u e a posição da amostra u_k, será denotado por h_k. Portanto, tendo como referência um ponto u, o vetor de distância entre duas posições de amostras quaisquer, u_j e u_k, pode ser determinada por: $h_{jk} = h_k - h_j$, como mostra a Figura 1. Com isto, o comprimento do vetor de distância que separa estas amostras será denotado por $\|h_{jk}\|$ ($\|h\|$ lê-se norma do vetor h). Está será a notação para as amostras e distância utilizada neste texto.

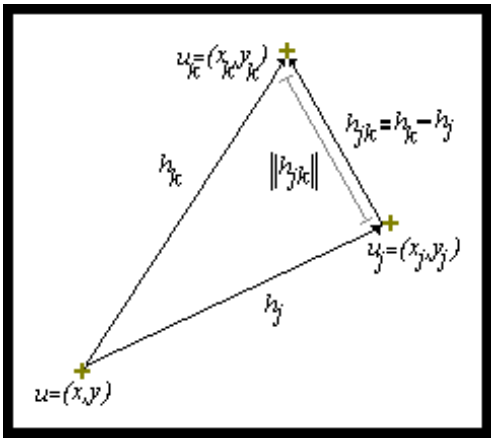


Figura 1. Representação da distância entre duas amostras

3.4.2 KRIGEAGEM

Em 1951 Daniel G. Krige, desenvolveu um método para solucionar problemas de mapeamentos geológicos, levando em consideração a localização geográfica, continuidade espacial, a anisotropia e dependência espacial das observações (CÂMARA et al., 1996;

DRUCK et al., 2002; MORAES; SOUZA; TELES, 2003). A anisotropia é uma qualidade que certos fenômenos têm de se distribuirem no espaço de forma diferenciada segundo sua direção de propagação. Caso não haja anisotropia, o fenômeno é considerado isotrópico, ou seja, independente da sua direção de propagação. O método desenvolvido por Krige foi denominado de Krigeagem (ou Krigagem) e rapidamente se expandiu para outros campos da ciência. A Krigeagem é portanto, um processo de interpolação Geoestatístico, no qual estimam-se valores da distribuição de um fenômeno a partir dos dados amostrados, levando em consideração na análise a localização, continuidade, anisotropia e dependência espacial do fenômeno. Os estimadores de Krigeagem podem ser visto de duas maneiras (FELGUEIRAS, 1999):

- **Um estimador de Krigeagem linear**, que é definido como uma "coleção de técnicas de regressão linear generalizada para minimizar a variância do erro de estimação a partir de um modelo de covariância definido a priori" (DEUTSCH; JOURNEL, 1998 apud FELGUEIRAS, 1999).

- **Um estimador de Krigeagem não-linear**, que é um estimador de krigeagem linear aplicado aos dados transformados através de uma função não-linear, do tipo gaussiana, lognormal, codificação por indicação, etc. (FELGUEIRAS, 1999).

3.4.3 FUNÇÃO ALEATÓRIA

Considere uma região geográfica G, na qual, cada posição $u \in G$, seja associada uma variável $Z(u)$. O objetivo da Krigeagem é estimar valores da variável $Z(u)$, a partir de um conjunto de n amostras $z(u_1), z(u_2), ..., z(u_n)$, obtidas nas posições $u_j \in G$, $\forall j = 1, 2, ..., n$. Podemos considerar a distribuição espacial de $z(u_j)$, nos pontos amostrados $u_j, \forall j = 1, 2, ..., n$, como sendo uma variável aleatória *v.a.*, cujo valor medido é

admitido sem erros. Nas posições não amostrados, a distribuição espacial de $Z(u)$ pode ser vista como Função Aleatória (ou Processo Estocástico), isto é, uma família $\{Z(u) : u \in G\}$, tal que, para cada posição $u \in G$, $Z(u)$ é uma variável aleatória $v.a.$, com uma probabilidade de ocorrência associada (DIGGLE; RIBEIRO JR, 2000; FELGUEIRAS, 1999; CLARKE; DISNEY, 1979). O conjunto de valores $\{Z(u) : u \in G\}$ é denominado de espaço de estados do processo estocástico e os valores $Z(u)$ são chamados de estados (CLARKE; DISNEY, 1979).

A incerteza da variável aleatória $Z(u)$, na posição u (incerteza local), é caracterizada pela sua função de distribuição acumulada, fda, dada por (3). A distribuição espacial $F(u, z)$, na região geográfica G, pode determinar áreas, onde a probabilidade de ocorrência de valores menores ou iguais a um certo valor crítico z é alta, isto é, áreas de risco (DRUCK et al., 2002).

$$F(u, z) = P[Z(u) \leq z] \tag{3}$$

Para estimar a incerteza local da variável aleatória $Z(u)$, a partir do conhecimento das n amostras, pode-se utilizar a função de distribuição acumulada condicionada, $fdac$, da variável $Z(u)$, dada por (4):

$$F(u; z \mid n) = P[Z(u) \leq z \mid n] \tag{4}$$

Dada qualquer coleção finita de posições não amostradas $u_1, u_2, ..., u_k$, com $u_i \in G, \forall i = 1, 2, ..., k$, então $Z(u_1), Z(u_2), ..., Z(u_k)$ é um conjunto de k $v.a.$. A Função Aleatória, fa, que modela a incerteza global destas k $v.a.$ é caracterizada pela função de distribuição acumulada multivariada, definida em (5):

$$F(u_1, u_2, ..., u_k; z_1, z_2, ..., z_k) = P[Z(u_1) \leq z_1, Z(u_2) \leq z_2, ..., Z(u_k) \leq z_k] \tag{5}$$

3.4.4 CODIFICAÇÃO POR INDICAÇÃO

Como foi mencionado anteriormente, a Krigeagem por Indicação é um método de Krigeagem não-linear, no qual os dados são codificados por indicação e utilizados para estimar a *v.a.* $Z(u)$, através do estimador de Krigeagem linear. A codificação por Indicação consiste em transformar os dados da *v.a.* $Z(u)$, em uma variável por indicação $I(u, z_c)$, através da função não-linear definida por (6). Nesta expressão z_c representa um valor qualquer, pertencente ao espaço de estados $\{Z(u) : u \in G\}$, denominado de valor de corte (ISAAKS; SRIVASTAVA, 1989).

$$I(u, z_c) = \begin{cases} 1 & , se \quad Z(u) \leq z_c \\ 0 & , se \quad Z(u) > z_c \end{cases} \tag{6}$$

A distribuição espacial da variável por indicação $I\,(u, z_c)$, para um determinado valor de corte z_c, também pode ser vista como um processo estocástico $\{I\,(u, z_c) : u \in G\}$, tal que, para cada posição $u \in G$, $I\,(u, z_c)$ é uma variável aleatória com distribuição de Bernoulli (DRUCK et al., 2002). O valor esperado da *v.a.* $I\,(u, z_c)$ é a função de distribuição acumulada da *v.a.* $Z\,(u)$, isto é, $F(u, z_c)$, como mostra a expressão (7).

$$\begin{aligned} E\,[I(u, z_c)] &= 0 \cdot P[I(u, z_c) = 0] + 1 \cdot P[I(u, z_c) = 1] \\ &= 0 + P[Z(u) \leq z_c] \\ &= F(u, z_c) \end{aligned} \tag{7}$$

O valor esperado $E\left[I\left(u,z_c\right)\right]$ é uma função Estatística, que expressa o comportamento central da distribuição da *v.a.* $I(u,z_c)$. A variância da *v.a.* $I\left(u,z_c\right)$, denotada por $V\left[I(u,z_c)\right]$, também é uma função Estatística, que determina a média quadrática dos afastamentos contados em relação a média. Para um valor de corte z_c, na posição u, a variância da *v.a.* $I\left(u,z_c\right)$ é dada por (8). Neste caso, a variância da *v.a.* $I\left(u,z_c\right)$, varia entre 0 e $0,25$, atingindo o valor máximo $(0,25)$, quando $F(u,z_c)=0,5$ (DRUCK et al., 2002).

$$V\left[I(u,z_c)\right]=E\left[I(u,z_c)-E\left[I(u,z_c)\right]\right]^2=F(u,z_c)\left\{1-F(u,z_c)\right\} \tag{8}$$

3.4.5 SEMIVARIOGRAMA INDICADOR

O principal padrão espacial de um fenômeno é a sua dependência espacial, que é a característica que estes fenômenos têm de ser mais semelhante quanto menor for à distância entre as suas observações. O variograma é uma ferramenta geoestatística utilizada para analisar quantitativamente a variação de um fenômeno que possua dependência espacial (MORAES; SOUZA; TELES, 2003). É a partir do modelo de variografia que se obtêm as autocovariâncias necessárias para estimar os pesos de ponderação do modelo de Krigeagem por Indicação. Os resultados apresentados nesta seção podem ser encontrados em Isaaks e Srivastava (1989), Camargo (1997), Nunes (1998) e Deutsch e Schnetzler (2000).

O variograma indicador é definido como a variância da diferença entre as duas variáveis por indicação $I\left(u,z_c\right)$ e $I\left(u+h,z_c\right)$, separadas pelo vetor de distância h, para um valor de corte z_c e será denotado por $2\gamma\left(h,z_c\right)$.

$$2\gamma\left(h,z_c\right)=V\left[I\left(u,z_c\right)-I\left(u+h,z_c\right)\right] \tag{9}$$

Em algumas casos pode-se encontrar referências a uma outra entidade denominada semivariograma indicador, que nada mais é do que a metade do variograma, denotada por $\gamma(h, z_c) = \frac{1}{2}[2\gamma(h, z_c)]$. Em geral, costuma-se trabalhar com o semivariograma indicador.

Dado o conjunto de amostras por indicação, $i(u_1, z_c), i(u_2, z_c), ...,$ $i(u_n, z_c)$, para um determinado valor de corte z_c, o semivariograma indicador experimental pode ser obtido pela equação (10), onde $N(h)$ é o número de amostras que estão separadas por um vetor de distância h. A expressão $\{(j, q) \mid h_{jq} = h\}$ significa o conjunto de todos os pares (j, q) tais que o vetor de distância h_{jq}, entre as amostras $i(u_j, z_c)$ e $i(u_q, z_c)$, é igual ao vetor de distância h.

$$\tilde{\gamma}(h, z_c) = \frac{\sum\limits_{\{(j,q)|h_{jq}=h\}} [i(u_j, z_c) - i(u_q, z_c)]^2}{2N(h)} \tag{10}$$

O cálculo do semivariograma indicador experimental para um vetor de distância h e valor de corte z_c depende do tipo de grade das amostras. Se a grade de amostras é regular, então deve-se considerar apenas os parâmetros a respeito do vetor de distância, que são: o comprimento do vetor h e sua direção (dada em ângulo (α)). Desta forma, quando a grade de amostras é regular, serão pareadas todas as amostras que estiverem a uma distância exatamente h, na direção α. Se a grade de amostras é irregular, então deve-se estabelecer uma tolerância para os parâmetros do vetor h, isto é, serão pareadas todas as amostras que estiverem a uma distância $h \pm d$, onde d é um parâmetro de tolerância de distância, e na direção $\alpha \pm t$, onde t é um parâmetro de tolerância angular.

Quando as amostras estão posicionadas de forma irregular é coveniente fazer uma adaptação do semivariograma, para levar em consideração que o vetor de distância não é exatamente h, mais sim, aproximadamente h, isto é, $(h_{jq} \approx h)$. Sendo assim, o semivariograma indicador experimental passa a ser definido por (11) (ISAAKS; SRIVASTAVA, 1989).

$$\tilde{\gamma}(h, z_c) = \frac{\sum\limits_{\{(j,q)|h_{jq}\approx h\}} [i(u_j, z_c) - i(u_q, z_c)]^2}{2N(h)} \tag{11}$$

Quando há uma dependência espacial do fenômeno em estudo, espera-se que quanto maior for o comprimento do vetor de distância (h_{jq}) que separa duas amostras quaisquer, $i(u_j, z_c)$ e $i(u_q, z_c)$, menor seja a covariância entre estas amostras $C[i(u_j, z_c), i(u_q, z_c)] = C(h_{jq}, z_c)$. Neste mesmo contexto, espera-se que as diferenças $[i(u_j, z_c) - i(u_q, z_c)]$ cresçam à medida que o comprimento do vetor de distância que separa as amostras (h_{jq}) aumenta. Consequentemente, espera-se que o semivariograma indicador experimental $\tilde{\gamma}(h, z_c)$ aumente com o aumento da distância entre os pares de amostras (h_{jq}). Porém, é importante destacar que o semivariograma indicador pode aumentar com a distância até se estabilizar em um platô (Patamar), quando desaparece a dependência espacial (MORAES; SOUZA; TELES, 2003).

Ao calcular as estimativas do semivariograma indicador experimental, $\tilde{\gamma}(h, z_c)$, para uma determinado valor de corte z_c, pode-se contruir um gráfico cartesiano da função $\tilde{\gamma}(h, z_c)$ pela distância h. Este gráfico revela características importantes a respeito da variografia da variável por indicação $I(u, z_c)$, que são determinadas por quatro parâmetros: Alcance (a), Patamar (C), Efeito pepita (C_0) e Contribuição (C_1). Na Figura 2, tem-se uma ilustração de um gráfico do semivariograma indicador.

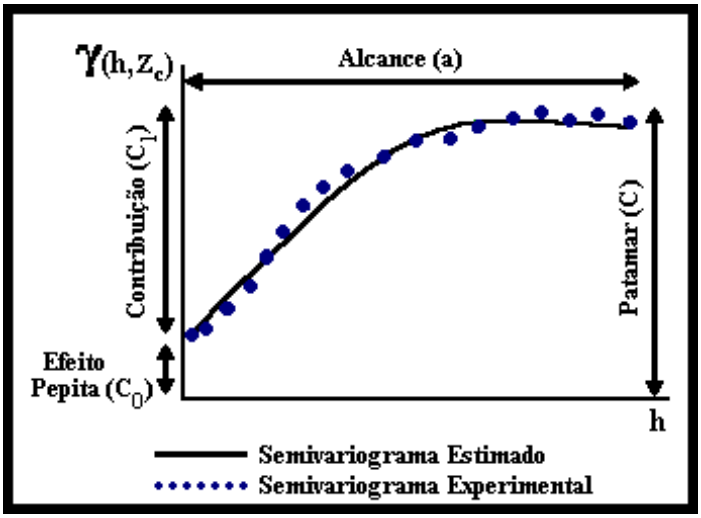
- **Alcance (a):** é a distância dentro da qual as amostras apresentam-se correlacionadas espacialmente;

- **Patamar (C):** é o valor do semivariograma correspondente a seu Alcance (a), a partir do qual não se espera que exista dependência espacial entre as amostras;

- **Efeito pepita (C_0):** teoricamente, espera-se que quando h for igual a 0, $C(0, z_k) = C[F(u, z_c), F(u + 0, z_c)] = V[F(u, z_c)] = 0$, mas na prática à me-

dida que h tende a zero, $C(0, z_k)$ se aproxima de um valor positivo chamado de Efeito pepita (C_0), que revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras.

- **Contribuição (C_1):** é a diferença entre o Patamar (C) e o Efeito pepita (C_0).



Fonte: Adaptado de Felgueiras (1999).

Figura 2. Semivariograma Indicador experimental e o estimado

Segundo Deutsch e Schnetzler (2000), a relação entre as autocovariâncias e o semivariograma indicador é dada pela equação $\hat{\gamma}(h, z_k) = C(0, z_k) - C(h, z_k)$. Para estimar as autocovariâncias não se utiliza o semivariograma indicador experimental diretamente. Ao invés disso, ajusta-se um modelo teórico a esse semivariograma e utiliza-o para calcular as autocovariâncias. O objetivo é que o modelo ajustado represente a tendência de $\gamma(h, z_k)$ em relação a h, obtendo-se assim, melhores estimativas para $C(h, z_k)$.

Os modelos teóricos mais usuais na análise do semivariograma indicador são: o esférico e o exponencial. A seguir, têm-se a descrição de cada um destes modelos de semivariograma, onde a é o Alcance prático, definido como a distância na qual o valor do modelo $M(\|h\|)$ é 95% do Patamar C (ISAAKS; SRIVASTAVA, 1989).

- **Modelo Esférico de Semivariograma:** utilizado quando a variação da função $\gamma(h, z_k)$, antes de atingir o Patamar, apresenta-se com um comportamento semelhante a uma semi-esfera. Este modelo de semivariograma é dado pela equação (12).

$$\hat{\gamma}(h, z_c) = \begin{cases} 0 & , \|h\| = 0 \\ C_0 + C_1 \left\{ \frac{3}{2}\left(\frac{\|h\|}{a}\right) - \frac{1}{2}\left(\frac{\|h\|}{a}\right)^3 \right\} & , 0 < \|h\| \leq a \\ 1 & , \|h\| > a \end{cases} \tag{12}$$

- **Modelo Exponencial de Semivariograma:** indicado para modelar semivariogramas, nos quais, a função $\gamma(h, z_k)$ cresce mais rápido com o aumento da distância, que o modelo esférico. O modelo exponencial de semivariograma é dado pela equação (13).

$$\hat{\gamma}(h, z_c) = \begin{cases} 0 & , \|h\| = 0 \\ C_0 + C_1 \left\{ 1 - \exp\left(-\frac{\|h\|}{a}\right) \right\} & , \|h\| \neq 0 \end{cases} \tag{13}$$

Para modelar o semivariograma utilizou-se o método automático do SPRING, que obtém os parâmetros do modelo de semivariograma através do algoritmo de ajuste pelos Mínimos Quadrados Ponderados (JIAN; OLEA; YU, 1996). Este algoritmo fornece também uma medida quantitativa, denominada Critério de Informação de Akaike, AIC, dada por (14), que reporta para qual modelo o ajuste é o mais preciso. Pelo Critério de Informação de Akaike, quanto menor o seu valor, melhor é o ajuste.

$$AIC = n \ln \left(\frac{\left[\tilde{\gamma}(h, z_c) - \hat{\gamma}(h, z_c)\right]^T \sum^{-1} \left[\tilde{\gamma}(h, z_c) - \hat{\gamma}(h, z_c)\right]}{n} \right) + 2o \tag{14}$$

onde $\tilde{\gamma}(h, z_c)$ é o vetor que contém os semivariogramas experimentais $\forall h$, $\hat{\gamma}(h, z_c)$ representa o vetor que contém os semivariogramas estimados $\forall h$, o é o número de parâmetros

do modelo de semivariograma teórico a ser estimado e $\sum$ é uma matriz. No caso mais simples, onde o ajuste é realizado pelo método de Mínimos Quadrados, $\sum$ é a matriz identidade. Porém, isso requer a hipótese que as diferenças sigam distribuição normal, sejam independentes e possuam as mesmas variâncias, o que não pode ser sustentado em uma situação prática. A matriz $\sum$ também pode ser a matriz de covariâncias do semivariograma experimental, quando o ajuste é realizado por Mínimos Quadrados Generalizados. Quando o ajuste é feito pelos Mínimos Quadrados Ponderados, $\sum$ assume a forma de uma matriz diagonal, com apenas variâncias na diagonal principal.

3.4.6 KRIGEAGEM POR INDICAÇÃO

O resultado obtido em (7), determina uma relação entre a *v.a.* $I(u, z_c)$ e a fda de $Z(u)$, isto é, estimar o valor médio da *v.a.* $I(u, z_c)$, equivale a estimar a fda da *v.a.* $Z(u)$. Sabe-se que para medir a incerteza local da *v.a.* $Z(u)$, a partir do conhecimento do conjunto de amostras, é necessário conhecer o comportamento da sua $fdac$, isto é, o comportamento de $F(u, z_c \mid n)$. Este valor é obtido ao estimar o valor esperado da *v.a.* $I(u, z_c)$, condicionado as n amostras, como mostra a expressão (15). Os resultados abordados nesta subseção podem ser vistos em Isaaks e Srivastava (1989), Felgueiras (1999), Deutsch e Schnetzler (2000) e Felgueiras et al. (1999).

$$\begin{aligned} E\left[I(u, z_c) \mid n\right] &= 0 \cdot P[I(u, z_c) = 0 \mid n] + 1 \cdot P[I(u, z_c) = 1 \mid n] \\ &= 0 + P[Z(u) \leq z_c \mid n] \\ &= F(u, z_c \mid n) \end{aligned} \tag{15}$$

Na Krigeagem por Indicação as amostras $z(u_1), z(u_2), ..., z(u_n)$ são transformados nas $i(u_1, z_c), i(u_2, z_c), ..., i(u_n, z_c)$ e utilizada para estimar $F(u, z_c \mid n)$. Desta forma, pode-

se definir o processo de estimação da *v.a.* $Z(u)$, pelo método de Krigeagem por Indicação, através dos seguintes passos:

1. Obtenção do conjunto de amostras $z(u_1), z(u_2), ..., z(u_n)$, com $u_j \in G, \forall j = 1, 2, ..., n$;

2. Transformação não-linear das amostras da *v.a.* $Z(u)$, para obter as amostras por indicação $i(u_1, z_c), i(u_2, z_c), ..., i(u_n, z_c)$;

3. Obtenção da $F(u, z_c \mid n)$ (que representa o valor esperado da *v.a.* $I(u, z_c)$, condicionado as n amostras), através do método de estimação por regressão linear generalizada;

4. Estimação da *v.a.* $Z(u)$, a partir das estimativas da sua $fdac$, dada por $F(u, z_c \mid n)$, para vários valores de corte z_c;

O modelo de regressão linear generalizada é dado pela fórmula (16), onde: $\hat{m}_z(u)$ é uma estimativa do valor médio da variável $Z(u)$, na posição $u \in G$; $m_z(u)$ é o valor médio verdadeiro da variável $Z(u)$; $n(u)$ é o número de amostras vizinhas a u; $w_j(u)$, $\forall j = 1, 2, ..., n(u)$ são os pesos de ponderação a serem estimados na posição u; $z(u_j)$, $\forall j = 1, 2, ..., n(u)$ são os valores amostrais vizinhos a posição u e $m_z(u_j), \forall j = 1, 2, ..., n(u)$ são os valores médios verdadeiros nas posições u_j.

$$\hat{m}_z(u) - m_z(u) = \sum_{j=1}^{n(u)} w_j(u) \{z(u_j) - m_z(u_j)\} \tag{16}$$

Portanto, o modelo de Krigeagem por Indicação, para estimar o valor médio da variável por indicação $I(u, z_c)$, isto é, $F(u, z_c|n)$, em uma posição espacial u, a partir de uma quantidade $n(u)$ de amostras vizinhas a u, $i(u_1, z_c), i(u_2, z_c), ..., i(u_{n(u)}, z_c)$, para um valor de corte z_c, é difinido por (17):

$$F(u, z_c \mid n) - F(u, z_c) = \sum_{j=1}^{n(u)} w_j(u, z_c)\{i(u_j, z_c) - F(u_j, z_c)\} \tag{17}$$

onde $w_j(u, z_c)$ é o peso de ponderação atribuído a amostra j, em uma posição u e para um valor de corte z_c.

O Modelo de Krigeagem por Indicação propôsto na equação (17), apresenta algumas dificuldades para a sua estimação, pois necessita do conhecimento a priori das medidas $F(u, z_c), \forall u \in G$, para um valor de corte z_c. Desta forma, para aplicação desta técnica, algumas hipóteses devem ser consideradas a priori. Segundo Felgueiras (1999), para cada valor de corte z_c, deve-se considerar a hipótese de estacionaridade de 2ª ordem do processo estocástico $\{I\,(u, z_c) : u \in G\}$, isto é, considerar que:

1. O valor médio da *v.a.* $I(u, z_c)$ é constante em toda posição u pertencente a região geográfica G, ou seja, $E\,[I(u, z_c)] = F\,(u, z_c) = F(z_c), \forall u \in G$ (hipótese de estacionaridade de 1ª ordem);

2. A covariância entre $I(u, z_c)$ e $I(u + h, z_c)$ existe e depende somente de h (vetor de distância), ou seja, $C\,[I(u, z_c), I(u + h, z_c)] = C\,(h, z_c)\,, \forall h \in \Re^2$ (hipótese de estacionaridade de 2ª ordem).

A hipótese de estacionariedade de 2ª ordem do processo estocástico $\{I\,(u, z_c) : u \in G\}$, implica também na estacionariedade do semivariograma. Porém, nas aplicações práticas usa-se a estacionariedade intrínseca (LANDIM, 1998; NUNES, 1998), na qual:

1. $E\,[I(u, z_c)] = F\,(z_c)\,, \forall u \in G$ e

2. A variância das diferenças entre a realização do processo nas posições u e $u+h$, para qualquer u pode ser modelado pelo semivariograma, dado pelas equações (10 ou 11).

Sob a hipótese de estacionaridade de 1ª ordem, o modelo (17) pode ser simplificado, como mostra expressão (18), obtendo-se assim o modelo de Krigeagem por Indicação Simples. Para aplicação deste modelo é necessário o conhecimento a priori apenas do valor médio estacionário da *v.a.* $I(u, z_c)$, $F(z_c)$.

$$F(u, z_c \mid n) = \sum_{j=1}^{n(u)} w_j(u, z_c) i(u_j, z_c) + \{1 - \sum_{j=1}^{n(u)} w_j\,(u, z_c)\} F(z_c) \tag{18}$$

Os pesos de ponderação $w_j(u, z_c), \forall j = i, 2, ..., n\,(u)$ são obtidos de forma a minimizar variância do erro de estimação, isto é, minimizar $\sigma_R^2 = V\,[F(u, z_c \mid n) - F(u, z_c)]$. Sendo assim, na Krigeagem por Indicação Simples os pesos são obtidos através da solução do sistema de $n\,(u)$ equações, dada por (19):

$$\sum_{j=1}^{n(u)} w_j(u, z_c) C\,(h_{jq}, z_c) = C\,(h_q, z_c) \qquad \forall q = 1, 2, ..., n(u) \tag{19}$$

onde $C(h_{jq}, z_c) = C\,[i(u_j, z_c), i\,(u_q, z_c)]\,, \forall j, q = 1, 2, ..., n(u)$ são as autocovariâncias definidas pelo vetor de distância h_{jq} e $C(h_q, z_c) = C\,[i(u, z_c), i\,(u_q, z_c)]\,, \forall q = 1, 2, ..., n(u)$ são as autocovariâncias definidas pelo vetor de distância h_q, com referência a posição u. Estas autocovariâncias são obtidas através do modelo de semivariograma indicador teórico.

Se não for do conhecimento do pesquisador o valor médio estacionário da *v.a.* $I(u, z_c)$, $F(z_c)$, então pode-se considerar a soma dos pesos atribuidos as amostras igual a 1, isto é, $\sum_{j=1}^{n(u)} w_j(u, z_c) = 1$. Com isto, obtém-se o modelo de Krigeagem por Indicação Ordinária, dado pela equação (20):

$$F(u, z_c \mid n) = \sum_{j=1}^{n(u)} w_j(u, z_c) i(u_j, z_c) \tag{20}$$

No Modelo de Krigeagem por Indicação Ordinário, para estimar $F(u, z_c \mid n)$ a partir das $n(u)$ amostras vizinhas, basta determinar os pesos de ponderação

$w_j(u, z_c), \forall j = 1, 2, ..., n(u)$, na posição u, para cada valor de corte z_c. Estes pesos são obtidos com o intuito de minimizar a variância do erro de estimação, σ_R^2, sujeito a restrição da soma dos pesos ser igual a 1. Com isto, os pesos de ponderação da Krigeagem por Indicação Ordinária são obtidos através da solução do sistema de $n(u)+1$ equações dadas em (21). A variável $\phi(u, z_c)$ representa o parâmetro de Langrange, para uma posição u e valor de corte z_c. As autocovariâncias dadas neste sistema de equações são obtidas através do modelo de semivariograma indicador teórico.

$$\begin{cases} \sum_{j=1}^{n(u)} w_j(u, z_c) C\left(h_{jq}, z_c\right) + \phi(u, z_c) = C\left(h_q, z_c\right) \qquad \forall q = 1, 2, ..., n(u) \\ \sum_{j=1}^{n(u)} w_j = 1 \end{cases} \tag{21}$$

3.4.7 CORREÇÃO DOS DESVIOS DE ORDEM

A Krigeagem por Indicação tem como principal característica não ser necessário considerar nenhum tipo de distribuição a priori para variável aleatória de estudo. Por esta razão, pode-se defini-lá como um método geoestatístico de interpolação espacial não-paramétrico. Segundo Felgueiras (1999), a Krigeagem por Indicação "possibilita a construção de uma aproximação discretizada da fda de $Z(u)$", que são utilizados diretamente para se estimar valores estatísticos característicos da distribuição de $Z(u)$, tais como: valor médio, variância, mediana e quantis. Além destes valores característicos, a Krigeagem por Indicação permite estimar a incerteza de valores da distribuição de $Z\left(u\right)$.

No entanto, as funções de distribuição acumulada condicionada, para cada valor de corte, são estimadas separadamente provocando alguns erros de aproximação, conhecidos como desvios de relação de ordem. Para que as funções de distribuição acumulada consti-

tuam uma distribuição legítima (em conjunto), elas devem ser corrigidas para verificarem as seguintes relações de ordem (FELGUEIRA; DRUCK; MONTEIRO, 2002):

(a) O valor inferido de $F(u, z_c \mid n)$ deve satisfazer a seguite relação: $0 \leq F(u, z_c \mid n) \leq 1$, para todo valor de corte z_c;

(b) O valor estimado de $F(u, z_c \mid n)$ não deve ser maior do que o da $F(u, z_{c+1} \mid n)$, quando $z_c \leq z_{c+1}$.

O estimador de Krigeagem por Indicação não garante que a relação (a) seja verificada. Para corrigir este desvio pode-se ajustar os valores estimados para as bordas, ou seja, sempre que $F(u, z_c \mid n) \leq 0$, deve-se fazer $F(u, z_c \mid n) = 0$ e se $F(u, z_c \mid n) \geq 1$, então deve-se fazer $F(u, z_c \mid n) = 1$.

A relação (b) também pode não ser verificada, e portanto, deve-se corrigida da seguinte forma: verificar as estimativas de todos os pares de $fdac$ consecutivos $(F(u, z_c \mid n), F(u, z_{c+1} \mid n))$ e sempre que a (b) relação não for satisfeita, deve-se fazer (22):

$$F(u, z_c \mid n) = F(u, z_{c+1} \mid n) = \frac{1}{2}\{F(u, z_c \mid n) + F(u, z_{c+1} \mid n)\} \tag{22}$$

3.4.8 ESTIMAÇÃO DE VALORES

Determinada as estimativas das distribuições acumuladas da $v.a.$ $Z(u)$, condicionadas as n amostras, para um conjunto de valores de corte ordenados $\{z_0, z_1, z_2, ..., z_K, z_{K+1}\}$, isto é, determinada os valores $F(u, z_0 \mid n), F(u, z_1 \mid n), ..., F(u, z_{K+1} \mid n)$. As estimativas dos valores estatísticos característicos da distribuição de $Z(u)$ são dados a seguir (FELGUEIRAS, 1999):

- Uma estimativa do valor médio da *v.a.* $Z(u)$, denotado por $\hat{\mu}(u)$, pode ser aproximada por(23). O valor médio expressa a tendência central da distribuição da *v.a.* $Z(u)$.

$$\begin{aligned}\hat{\mu}(u) &= E[Z(u)] \\ &= \int_{-\infty}^{\infty} z \cdot f(u, z \mid n)\, \partial z \\ &= \int_{-\infty}^{\infty} z \cdot \partial F(u, z \mid n) \\ &\approx \sum_{k=1}^{K+1} z_c \left[F(u, z_k \mid n) - F(u, z_{k-1} \mid n)\right]\end{aligned} \tag{23}$$

- Uma estimativa da variância de $Z(u)$, denotada por $\hat{\sigma}^2(u)$, pode ser aproximada por (24). Este valor indica qual será o "erro quadrático" (quadrado do desvio) cometido ao substituir $Z(u)$ pela média $\hat{\mu}(u)$.

$$\begin{aligned}\hat{\sigma}^2(u) &= V[Z(u)] \\ &= \int_{-\infty}^{\infty} [z - \mu(u)]^2 \cdot f(u, z \mid n)\, \partial z \\ &= \int_{-\infty}^{\infty} [z - \mu(u)]^2 \cdot \partial F(u, z \mid n) \\ &\approx \sum_{k=1}^{K+1} [z_k - \mu(u)]^2 \left[F(u, z_k \mid n) - F(u, z_{k-1} \mid n)\right]\end{aligned} \tag{24}$$

- Além destas estatísticas existem as estatísticas de ordem, denominadas de p-quantil. Esta medida determina um valor $q_{(p)}$, pertencente ao espaço de estados do processo estocástico $\{Z(u) \mid u \in G\}$, tal que, $F\left(u, q_{(p)} \mid n\right) = p$.

- Uma estatística de ordem bastante utilizada é a mediana, que é o p-quantil $q_{(0,5)}$. Quando a distribuição da *v.a.* $Z(u)$ é assimétrica, a mediana é um estimador de tendência central mais robusto que a média.

- O intervalo interquartil $q_{(R)}$ ($q_{(R)} = q_{(0,75)} - q_{(0,25)}$), pode ser utilizado como medida de dispersão, quando a distribuição da *v.a.* $Z(u)$ é assimétrica.

Em Geoprocessamento, os procedimentos de análise espacial permitem a criação de modelos ambientais, que representem adequadamente o fenômeno natural em estudo. A modelagem ambiental consiste na criação de modelos matemáticos (determinísticos) ou estocásticos, que relacione variáveis ambientais, na tentativa de representar o comportamento de um processo ocorrendo na natureza. Os modelos ambientais são teoricamente contínuos, mas no entanto, os computadores digitais trabalham com um número fixo de campos ou bytes, isto é, são discretos e finitos (GOODCHILD; GUOGING, 1992 apud FELGUEIRAS; FELGUEIRAS, 1999). Desta maneira, ao transformar o modelo ambiental em modelo computacional, estima-se um número finito de valores do fenômeno considerado, em pontos posicionados de forma regular, formando uma grade regular de valores estimados (mesmo que as amostras sejam de uma grade irregular). Os espaçamentos entre os pontos desta grade são denominados de resolução. Quanto menor a resolução, mais pontos serão estimados para formar esta grade, obtendo-se mais informação do fenômeno, porém necessitando-se de maior tempo para sua geração.

3.5 COEFICIENTE DE INCIDÊNCIA DO DENGUE

O Coeficiente de incidência do dengue consiste numa taxa utilizada para medir a morbidade do dengue. Pode ser definido como o quociente entre o número de casos iniciados num determinado período (mês) e área (município) sobre a população desta área e na

metade deste período (LAURENTI et al., 1987). Neste estudo, o coeficiente de incidência do dengue foi obtido com base nas seguintes variáveis: número mensal de casos notificados do dengue por local de residência (município) e população residente mensal estimada por município. Os dados referentes a população residente estimada são de projeções feitas pelo Instituto Brasileiro de Geografia e Estatística, coletadas no site do DATASUS (DATASUS, 2003).

Os dados da população residente estimada por município foram coletados em períodos anuais, mas a construção do coeficiente de incidência do dengue requer a população referente à metade do período (mês). Sendo assim, estimou-se a população residente de cada município para o dia 15 de cada mês do período de janeiro a maio de 2004. O método de projeção adotado foi o Exponencial (Geométrico), que supõe que a população apresenta um crescimento geométrico (LAURENTI et al., 1987), e que também é adotado pelas fontes dos dados utilizados. A equação geral do método de projeção Geométrico é $\mathrm{Pr}(t+\Delta t) = \mathrm{Pr}(t)r^t$, onde t representa a data base, Δt é intervalo de tempo entre a data base e a data a ser estimada e r é a taxa de crescimento da população dado por (25):

$$r = \exp\{\frac{1}{t}\ln\left(\frac{\mathrm{Pr}(t+\Delta t)}{\mathrm{Pr}(t)}\right)\} \tag{25}$$

3.6 ANÁLISE DE CORRELAÇÃO POR POSTOS DE KENDALL

O coeficiente de correlação por postos de Kendall, denotado por τ, é uma medida de correlação não-paramétrica, que determina o grau de concordância entre duas variáveis X e Y, com níveis de mensuração no mínimo ordinal. Sejam $x_1, x_2, ..., x_n$ as observações de uma amostra aleatória de X e $y_1, y_2, ..., y_n$ as observações de uma amostra aleatória de Y, o roteiro para obter esse coeficiente é (SIEGEL, 1975):

1. Atribuir postos as observações das variáveis X e Y, e em caso de empates, atribuir as observações empatadas, a média dos postos que lhes caberiam se não houvesse empate;

2. Ordenar os postos de uma das variáveis, por exemplo X;

3. Sejam $l_1, l_2, ..., l_n$ os postos ordenados da variável X e $v_1, v_2, ..., v_n$ os postos correspondentes da variável Y, obtêm-se os escores $s_{ij}, \forall i < j, com \quad i, j = 1, 2, ...n$, definidos por (26):

$$s_{ij} = \begin{cases} -1 & , se \quad v_i > v_j \quad \text{quando} \quad i < j \\ 0 & , se \quad v_i = v_j \quad \text{quando} \quad i < j \\ 1 & , se \quad v_i < v_j \quad \text{quando} \quad i < j \end{cases} \tag{26}$$

4. Quando não há observações empatadas, o coeficiente de correlação por postos de Kendall, τ, é dado por (27):

$$\tau = \frac{\sum\limits_{\substack{i=1}}^{n}\sum\limits_{\substack{j=2 \\ i<j}}^{n} s_{ij}}{\frac{1}{2}n(n-1)} \tag{27}$$

5. Quando há empates, calculam-se as constantes $T_X = \frac{1}{2}\sum\limits_{k} b_k(b_k - 1)$, onde b_k é o número de observações empatadas em cada grupo k de empates na variável X e $T_Y = \frac{1}{2}\sum\limits_{r} f_r(f_r - 1)$, onde f_r é o número de observações empatadas em cada grupo r de empates na variável Y.

6. O coeficiente de correlação por postos de Kendall, τ, para o caso de haver observações empatadas, é dado por (28):

$$\tau = \frac{\sum\limits_{\substack{i=1}}^{n}\sum\limits_{\substack{j=2 \\ i<j}}^{n} s_{ij}}{\sqrt{\left(\frac{1}{2}n(n-1) - T_X\right)\left(\frac{1}{2}n(n-1) - T_Y\right)}} \tag{28}$$

Segundo Kendall, "para $n \geq 8$, a distribuição amostral de τ praticamente se confunde com a distribuição normal" (SIEGEL, 1975, p. 249). Sendo assim, para n grande ($n \geq 8$), sob a hipótese H_0: não há correlação entre as variáveis X e Y ($\tau = 0$), τ tem distribuição normal com média $\mu_\tau = 0$ e variância σ_τ^2 dada por (29), podendo então, realizar teste de significância para τ.

$$\sigma_\tau^2 = \frac{2(2n+5)}{9n(n-1)} \tag{29}$$

3.7 MESORREGIÕES GEOGRÁFICAS

Na Paraíba, a principal característica do regime climático é a presença de apenas duas estações: a chuvosa e a seca. As duas estão fortemente relacionadas com a localização geográfica, ou seja, quanto mais próximo da Mata Paraibana, mais úmido será o clima e quanto mais longe, mais seco (o mapa com as mesorregiões paraibanas pode ser visualizado na Figura 3). Em particular, a distribuição espacial da precipitação no estado da Paraíba apresenta uma alta heterogeneidade. A altura média da precipitação anual se distribui decrescendo bruscamente da Mata Paraibana até a Borborema e desta até o extremo oeste, os valores voltam a crescer, porém lentamente. Nas áreas mais secas, que abrange grande parte do estado, além de poucas chuvas, estas ocorrem de forma irregular, provocando longos períodos de seca. Observa-se também no estado que, as médias térmicas anuais são relativamente altas, apresentando como principal característica do regime térmico a sua homogeneidade no tempo e no espaço. Portanto, na Paraíba, o conhecimento acurado da distribuição espacial e temporal das precipitações (fenômeno climático que apresenta maior variação no estado) é fundamental para a obtenção de padrões definidores de regiões

climáticas e ecológicas, para a climatologia agrícola, engenharia, transporte, turismo, etc (NIMER, 1979).

Fonte: Adaptado de Rodriguez (2000).

Figura 3. Mesorregiões Geográficas da Paraíba

4 ANÁLISE DOS RESULTADOS

4.1 ANÁLISE EXPLORATÓRIA DOS DADOS

Para cada mês do período de janeiro a maio de 2004, obteve-se, a partir das variáveis número de notificações do dengue por município e da população estimada por município, os coeficientes de incidência do dengue (por 10.000 hab) por município paraibano. As medidas descritivas destes coeficientes em cada mês pode ser observado na Tabela 1. No mês de março, por exemplo, observou-se a maior média do coeficiente de incidência do dengue neste período, 2,149 casos por 10.000 hab (com desvio padrão cinco vezes maior, 10,157 casos por 10.000 hab). A assimetria destas distribuições pode ser verificada pelo fato do valor mínimo ser igual a mediana e a moda (nenhum caso), estando consideravelmente distante do valor máximo (acima de 9 casos por 10.000 hab, observado em Serra da Raiz no mês de maio). Dos 223 municípios paraibanos, 116 (ou 52,02%) não tiveram casos do dengue no período de janeiro a maio de 2004, como pode ser observado no mapa da Figura 4. A maior incidência do dengue ocorreu em Mulungu no mês de março (140,880 casos por 10.000 hab).

Tabela 1 – Medidas Descritivas dos coeficientes de incidência do dengue (por 10.000 hab) por mês

	Janeiro	Fevereiro	Março	Abril	Maio
Número de Municípios	223	223	223	223	223
Média	0,661	1,297	2,149	0,683	0,415
Mediana	0	0	0	0	0
Moda	0	0	0	0	0
Desvio Padrão	3,014	3,942	10,157	3,525	1,283
Mínimo	0	0	0	0	0
Máximo	35,800	33,840	140,880	48,440	9,010

Fonte: dados da Secretaria de Estado da Saúde - PB.

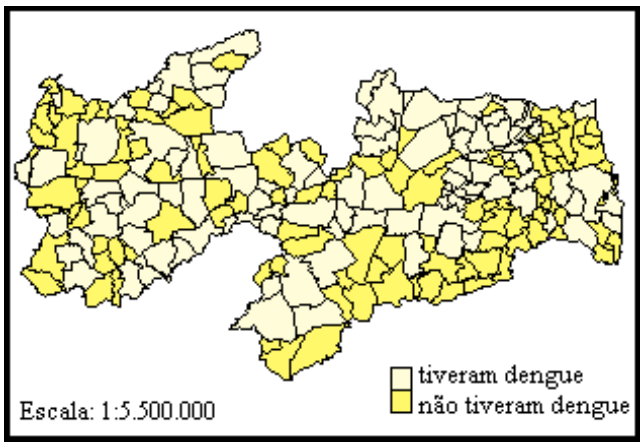


Figura 4. Municípios paraibanos segundo a presença de casos do dengue no período de janeiro a maio de 2004

Na Tabela 2, têm-se as medidas descritivas das amostras de precipitação referentes aos meses de janeiro a maio de 2004. Pode-se verificar que, neste período (com exceção do mês de abril), a média é maior que a mediana que, por sua vez, é maior que a moda, indicando a assimetria positiva desta distribuição. Os meses com maiores níveis de precipitação foram janeiro (com precipitaçao máxima de 633,40 mm) e fevereiro (com precipitação máxima de 437,90 mm).

Tabela 2 – Medidas Descritivas das amostras de precipitação acumulada (em mm) por mês

	Janeiro	Fevereiro	Março	Abril	Maio
Número de Estações	159	148	160	154	157
Média	372,69	208,52	93,13	72,74	115,27
Mediana	372,60	193,90	74,85	72,95	94,20
Moda	315,40	117,80	43,20	0,00	38,00
Desvio padrão	95,66	80,31	63,56	44,51	81,12
Mínimo	123,10	53,90	0,00	0,00	7,20
Máximo	633,40	437,90	292,20	292,20	394,10

Fonte: dados do Depto. de Ciências Atmosférica da UFCG.

Na Figura 5, pode-se observar a distribuição da precipitação (em mm) e da incidência do dengue (por 10.000 hab), no mês de janeiro de 2004. Pode-se visualizar nos gráficos

histograma e qqplot da variável incidência do dengue, que essa apresenta distribuição assimétrica positiva e não normal. Já a distribuição da precipitação apresentou um bom ajuste a distribuição normal (ver histograma e qqplot da precipitação na Figura 5).

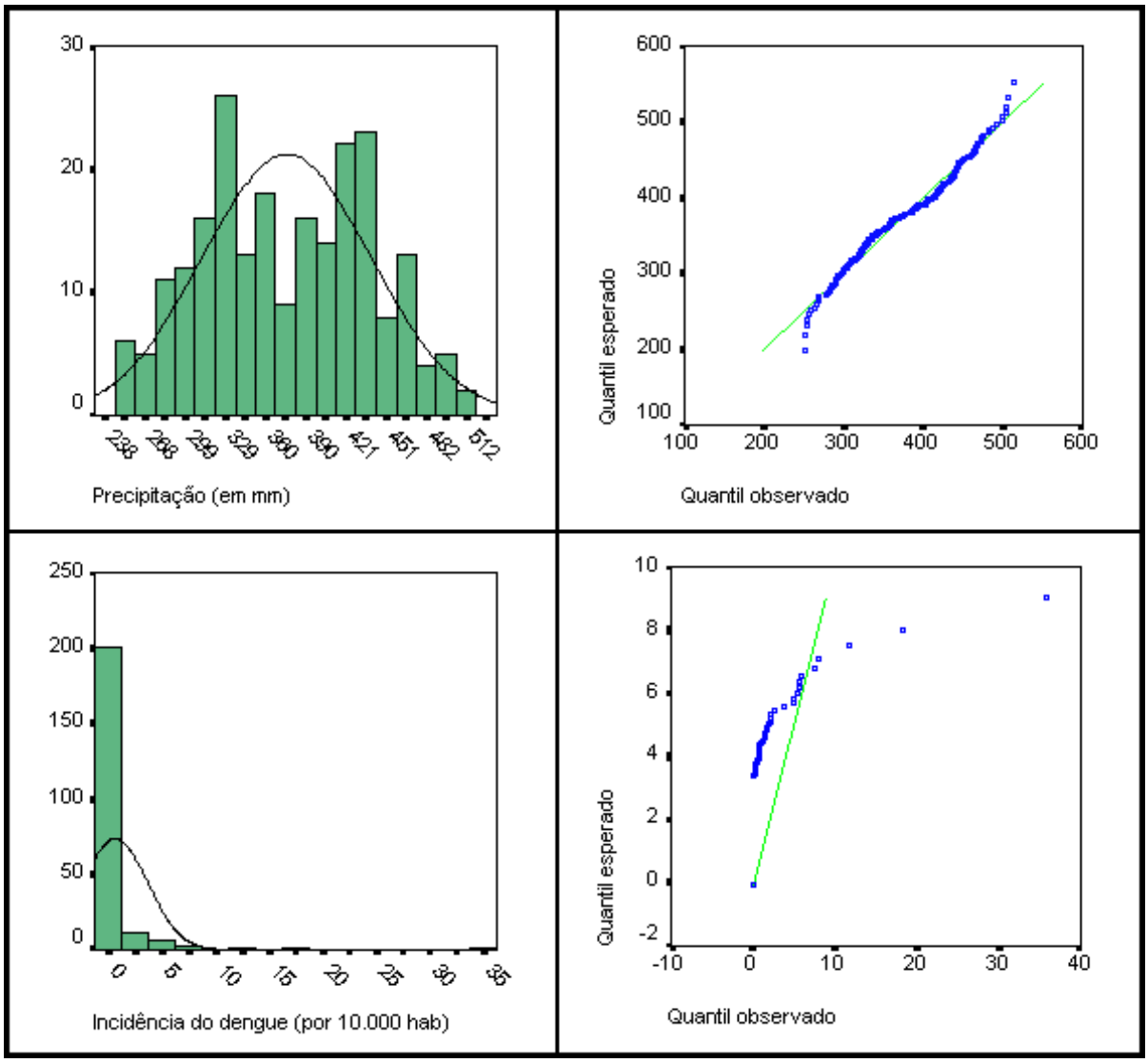


Figura 5. Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de janeiro de 2004

Os gráficos, histograma e qqplot, destas variáveis, referentes aos meses de fevereiro a maio podem ser observados nas Figuras A1 a A4, localizadas no Apêndice A. A partir dessas figuras, pode-se verificar que a distribuição da incidência do dengue é assimétrica positiva, com muitos valores iguais a zero. Para verificar a hipótese de normalidade da

distribuição da inciência do dengue realizou-se o teste não-paramétrico de Kolmogorov - Smirnov (SIEGEL, 1975). Ao nível de 1% de significância, rejeitou-se a hipótese de normalidade das distribuições mensais de incidência do dengue. Por esta razão, aplicou-se o coeficiente de correlação não paramétrico por postos de Kendall para medir a associação entre a pluviometria e a incidência do dengue.

Realizou-se também, o teste Kolmogorov - Smirnov, de ajuste a distribuição normal, para as distribuições mensais da precipitação. As estatísticas dos testes e os seus respectivos *valores p* podem ser observados na Tabela 3. Pode-se concluir, a partir desses testes que, a hipótese de normalidade da distribuição da precipitação, nos meses de janeiro, fevereiro e abril, não é rejeitada, com *valor p* superior a 0,26. Nos meses de março e maio, rejeitou-se a hipótese de normalidade, com *valores p* iguais a 0,02. Utilizando o recurso de transformação de variáveis, verificou-se que a raiz quadrada da precipitação no mês de março tem distribuição normal (com *valor p* igual a 0,74) e que o logaritmo neperiano da precipitação no mês de maio, também tem distribuição normal (com *valor p* igual a 0,16).

Tabela 3 - Testes de Kolmogorov - Smirnov para o ajuste a distribuição normal da precipitação mensal e seus respectivos *valores p*

	τ	*valor p*
Janeiro	0,62	0,84
Fevereiro	1,01	0,26
Março*	1,51	0,02
Abril	0,75	0,62
Maio**	1,55	0,02

Fonte: Dados do Depto. de Ciências Atmosférica da UFCG.
*Em março, a raiz quadrada da precipitação apresentou distribuição normal, com *valor p* igual a 0,74.
**Em maio, o logaritmo neperiano da precipitação apresentou distribuição normal, com *valor p* igual a 0,16.

Para aplicar a Krigeagem Linear (Paramétrica) sobre os dados amostrais originais dos meses de março e maio, seria necessário utilizar as amostras transformadas, isto é, a

raiz quadrada da precipitação no mês de março e o logaritmo neperiano da precipitação no mês de maio. Transformando as amostras antes de georreferenciá-las no sistema SPRING é possível de realizar a Krigeagem Linear. No entanto, como o resultado da Krigeagem é dado por uma grade com as estimativas dos valores, não seria possível reverter as transformações no sistema SPRING, para apresentar os resultados finais, ou seja, os mapas de precipitação gerados. Com a multiplicidade de transformações e as restrições impostas pelo sistema SPRING na manipulação dos dados amostrais transformados, a análise dos resultados da Krigeagem seria inviabilizada. Por estes motivos, optou-se pela aplicação de uma versão não-paramétrica da Krigeagem, denominada de Krigeagem por Indicação.

4.2 MODELAGEM DOS DADOS PLUVIOMÉTRICOS

As amostras de precipitação acumulada dos meses de janeiro a maio de 2004 foram codificadas por indicação para cinco valores de corte (ou cotas). Para cada mês, escolheu-se os cinco valores de corte com base nos sextis da distribuição da variável precipitação acumulada, ou seja, a 1ª cota (z_1) foi o 1ª sextil, a 2ª cota (z_2) foi o 2ª sextil, ..., a 5ª cota (z_5) foi o 5ª sextil. A partir destes valores de corte foram geradas, para cada mês do período considerado, cinco variáveis por indicação, $i(u, z_1), i(u, z_2), ..., i(u, z_5)$. Estas variáveis por indicação foram então utilizadas para estimar os semivariogramas indicadores, com base nos seguintes parâmetros:

- $N^o deLag$ (número de distâncias utilizadas);
- Incremento (denotado por: $\|h\|$, distância de cada Lag.);
- Tol. do Lag. (denotado por: d, tolerância do Lag.);

- Dir. (denotado por: (α), direção (ou ângulo) de busca);

- Tol. angular (denotada por: t, tolerância angular).

Utilizou-se como padrão a direção de ($\alpha = 0°$), com uma tolerância angular de ($t = 90°$), isto equivale a escolher todas as direções (caso isotrópico). A partir desse parâmetro de direção, foram pareadas todas as amostras que estavam a um intervalo de distância pré definido ($\| h \| \pm d$), sem levar em consideração a direção. Para estabelecer os parâmetros de distância, baseou-se nos trabalhos de Camargo (1997) e Felgueiras et al. (1999), fixando o número de distâncias (utilizadas em todos os semivariogramas) igual a seis ($N^o deLag. = 6$). A partir desse parâmetro, calculou-se o incremento, $\| h \|$, que foi igual a $\frac{1}{6}$ do alcance e a tolerância do Lag, d, que foi igual a metade do incremento $\left(d = \frac{\|h\|}{2}\right)$. Os semivariogramas indicadores gerados foram então ajustados, utilizando o modelo Esférico. Os parâmetros obtidos após a modelagem destes semivariogramas, estão descritos nas Tabelas B1 a B5, que se encontram no Apêndice B. Um exemplo do gráfico do semivariograma pode ser observado na Figura 6, que refere-se a modelagem do semivariograma indicador esférico das amostras de precipitação acumulada do mês janeiro de 2004, para o valor de corte igual a 372,6 (372,6 corresponde ao 3ª sextil). Os demais gráficos de semivariograma foram feitos de forma análoga.

Por fim, utilizou-se os modelos de semivariogramas estimados, para realizar a Krigeagem por Indicação e obter as estimativas da média da precipitação acumulada, para cada um dos meses do período de janeiro a maio de 2004. O resultado da Krigeagem por Indicação é uma grade com os valores estimados, que pode ser mais refinada que a grade das amostras, dependendo do parâmetro resolução adotado. Adotou-se para todas as grades geradas a resolução 1.000 x 1.000 m (Res. X = 1.000 e Res. Y = 1.000). Além disto, os parâmetros de raio e orientação do elipsóide de busca da Krigeagem por Indicação, também foram estabelecidos a priori, para o caso isotrópico, com os seguintes

valores: raio mínimo e raio máximo igual, em metros, a diagonal do retângulo envolvente do Projeto (R.max = 111.080,14 e R.min = 111.080,14) e ângulo igual a 0°.

Após a modelagem dos dados de chuva, estes foram apresentados na forma de mapa, isto é, obteve-se estimativas da precipitação média em toda a Paraíba, através da Krigeagem por Indicação, e em seguida gerou-se mapas coropléticos para cada mês do período. Além disto, obteve-se, a partir dos modelos de Krigeagem por Indicação, estimativas da precipitação média por município paraibano em cada mês do período de janiero a maio de 2004.

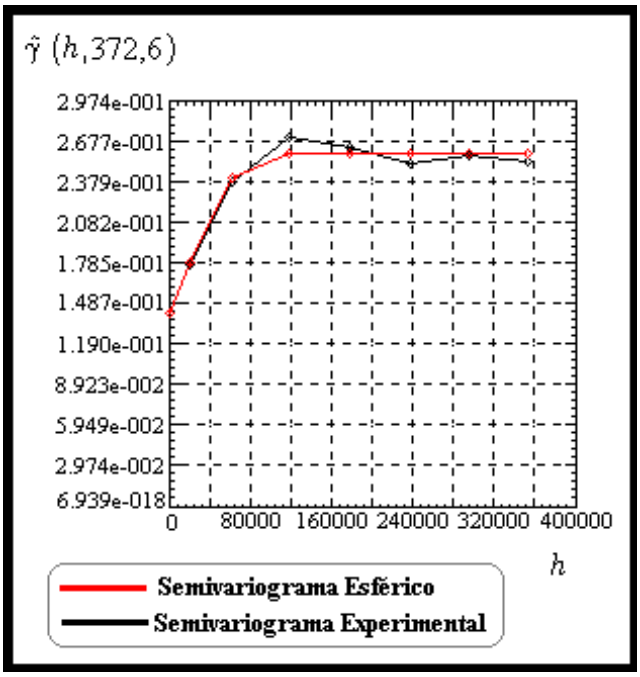
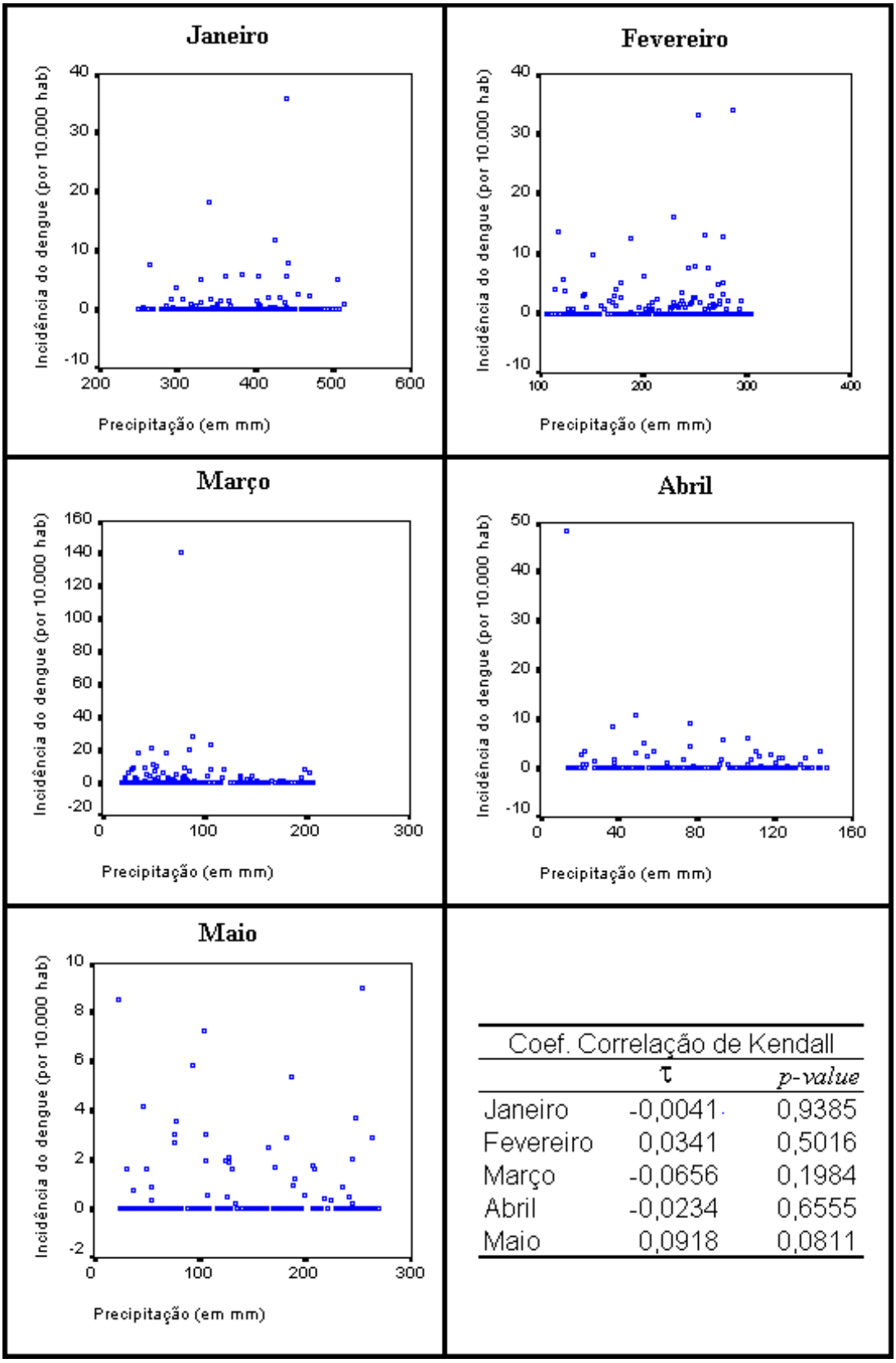


Figura 6 - Ajuste do Semivariograma Esférico para o valor de corte igual 372,6 das amostras de precipitação do mês de janeiro de 2004

4.3 CORRELAÇÃO ENTRE PLUVIOMETRIA E INCIDÊNCIA DO DENGUE

A partir dos coeficientes de incidência do dengue e da média da precipitação por município paraibano para cada mês de estudo, realizou-se uma análise de correlação entre a plu-

viometria e a incidência do dengue, através do coeficiente de correlação, não-paramétrico, por postos de Kendall. Na Figura 7, pode-se visualizar os diagramas de dispersão entre estas duas variáveis segundo os meses do período de janeiro a maio de 2004. Observaram-se coeficientes de correlação entre -0,0656 e 0,0918 e a hipótese de que o coeficiente de correlação em cada mês é igual a zero não foi rejeitada, com *valores p* superiores a 0,08 (ver tabela da Figura 7). Note que o aumento ou diminuição da pluviosidade não acarreta aumento ou diminuição da incidência do dengue.



Coef. Correlação de Kendall		
	τ	*p-value*
Janeiro	-0,0041	0,9385
Fevereiro	0,0341	0,5016
Março	-0,0656	0,1984
Abril	-0,0234	0,6555
Maio	0,0918	0,0811

Figura 7. Diagramas de dispersão e correlações de Kendall entre as variáveis precipitação média e coeficiente de incidência do dengue por município paraibano segundo os meses

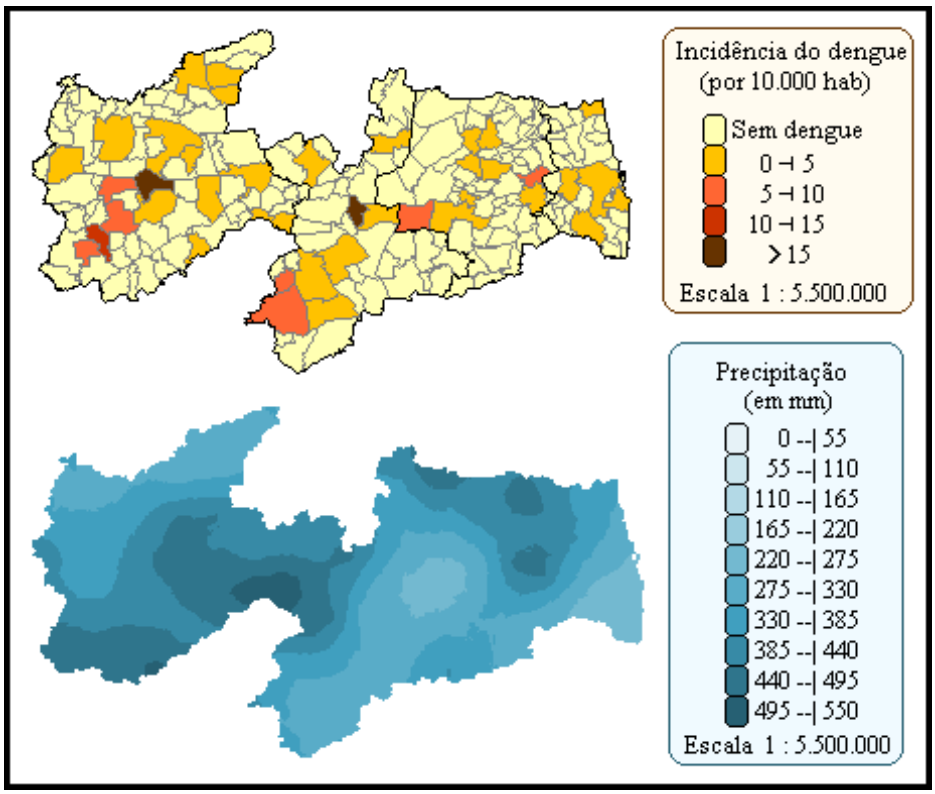


Figura 8. Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de janeiro de 2004

Nos mapas da Figura 8, pode ser visualizada a distribuição espacial da precipitação e da incidência do dengue no mês de janeiro de 2004. Verifica-se que as áreas onde ocorreram mais casos do dengue não são as mesmas que apresentaram maiores níveis de chuva. Os níveis de precipitação nas áreas que tiveram mais de 15 incidência por 10.000 hab variaram entre 385 e 495 mm. Já nas áreas onde foram observados os maiores níveis de precipitação, acima de 495 mm, a incidência do dengue foi de menos de 5 casos por 10.000 hab.

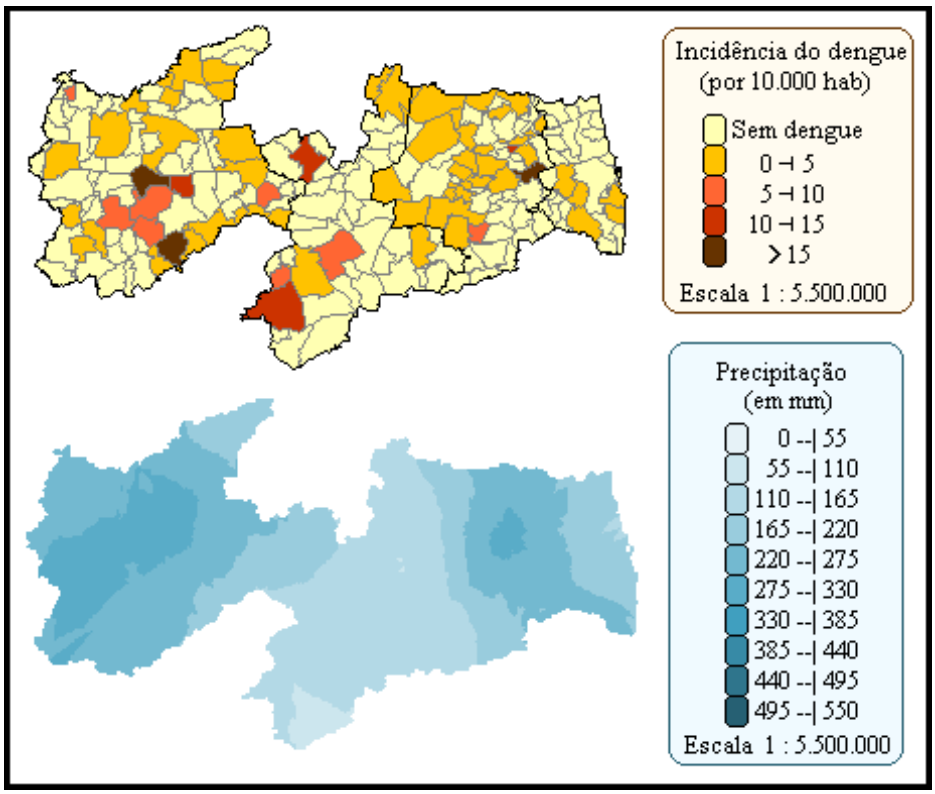


Figura 9. Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de fevereiro de 2004

Na Figura 9, têm-se os mapas da distribuição espacial da incidência do dengue e da precipitação para o mês de fevereiro de 2004. Pode-se notar que os maiores níveis de chuva ocorreram nas mesorregiões do Sertão e Agreste. A incidência do dengue foi maior em áreas específicas, como a parte sul do Sertão, nas vizinhança da cidade de Monteiro, onde foram observados níveis de precipitação máxima de 330 mm. As maiores incidências do dengue foram observadas nas cidades de Coremas (33,84 casos por 10.000 hab), Mulungu (33,22 casos por 10.000 hab) e Juru (16,21 casos por 10.000 hab).

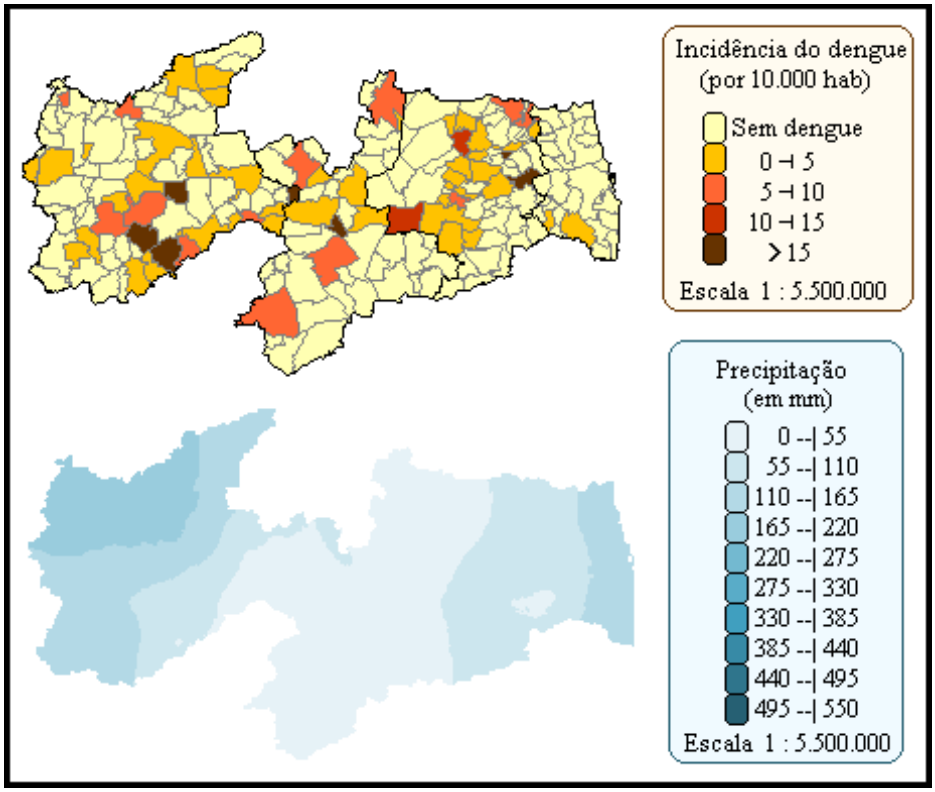


Figura 10. Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de março de 2004

No mapa da distribuição espacial da precipitação para o mês de março de 2004 (Figura 10), pode-se observar claramente o padrão espacial da chuva no estado da Paraíba no período do verão-outono, que caracteriza-se pelas ocorrências de maiores intensidade de chuvas no Sertão e Mata Paraibana. Pode-se verificar ao visualizar o mapa de incidência do dengue (Figura 10), que este fenômeno não apresenta nenhuma semelhança com a distribuição da chuva no Estado. Padrões espaciais semelhantes foram observados nas distribuições da chuva e incidência do dengue na Paraíba, no mês de abril (Figura 11) e no mês de maio (Figura 12).

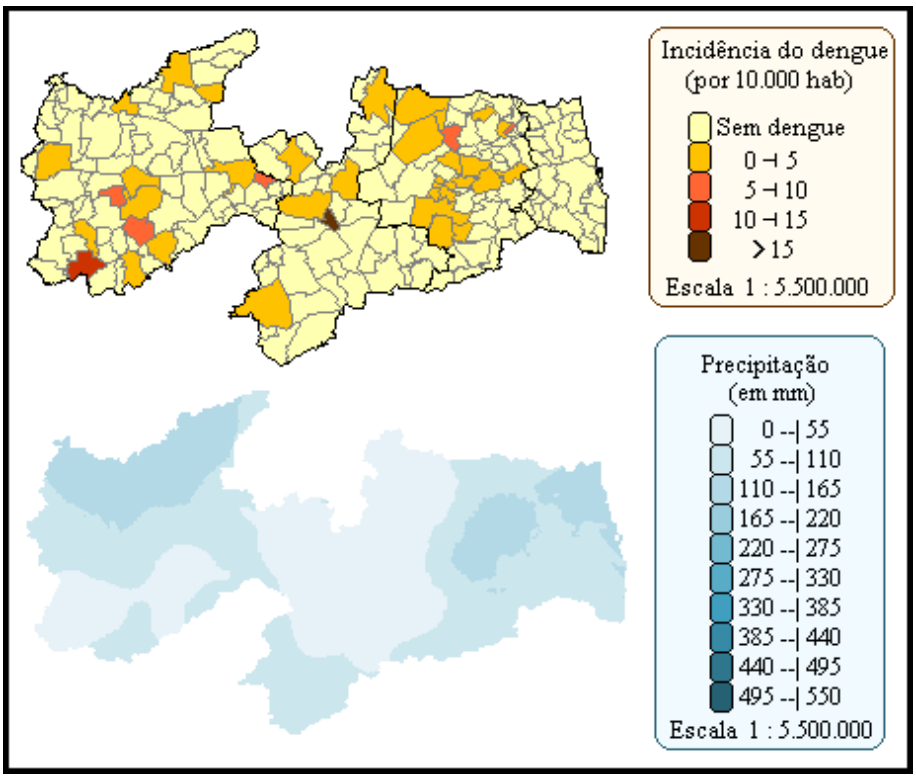


Figura 11. Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de abril de 2004

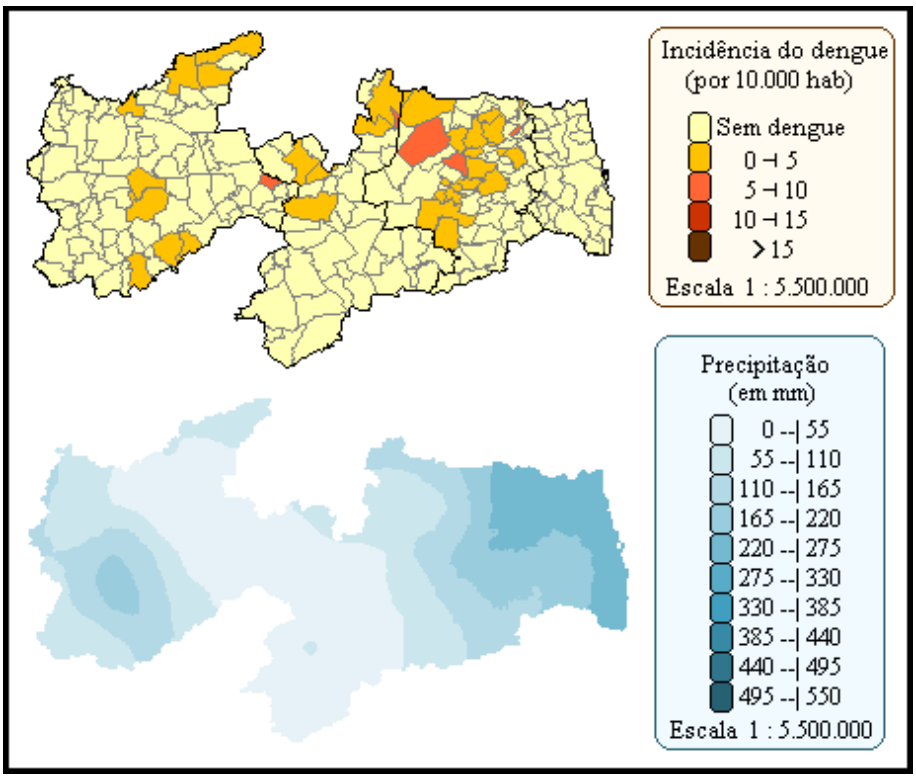


Figura 12. Distribuição espacial das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de maio de 2004

4.4 RESULTADOS POR MESORREGIÕES GEOGRÁFICAS

Pode-se verificar, nos mapas das Figuras 10 a 15, que foram observadas poucas variações da precipitação em cada mesorregião paraibana. Para responder a questão "quais padrões são observados entre a pluviometria e a incidência do dengue em cada mesorregião da Paraíba?", pode-se observar as séries temporais da precipitação (Figura 17) e da incidência do dengue (Figura 17) por mesorregião da Paraíba. Ao comparar estas duas figuras, pode-se notar que:

- Na Mata Paraibana, a precipitação média, que era de 311,44 mm (em janeiro), diminuiu até 108,19 mm (mês de abril), voltando a crescer até 232,86 mm (mês de maio). A incidência do dengue, no entanto, foi a mais baixa, dentre todas as

mesorregiões, sempre inferior a 0,151 casos por 10.000 hab, não havendo nenhum registro do dengue nos meses de abril e maio.

- No Agreste, nos meses de janeiro a março, observou-se um aumento na incidência do dengue (de 0,3 para 2,024 casos por 10.000 hab) e uma diminuição nos níveis de precipitação (de 359,22 para 59,05 mm). Já no período de março a maio, a incidência do dengue diminuiu (de 2,024 para 0,557 casos por 10.000 hab) e a precipitação aumentou (de 59,05 para 160,99 mm). Portanto, nesta mesorregião, observou-se que quando a incidência do dengue estava crescendo a pluviometria estava decrescendo, e vice-versa.

- Na Borborema, os níveis de precipitação, com exceção do mês de janeiro, estiveram sempre inferiores aos das outras mesorregiões, mas no entanto, foi a mesorregião que teve os maiores picos de incidência do dengue (em fevereiro, com 2,711 casos por 10.000 hab e em março, com 2,603 casos por 10.000 hab).

- No Sertão, observou-se o maior nível de precipitação no Estado neste período (404,52 mm em janeiro). A incidência do dengue aumentou (de 1,309 para 1,903 casos por 10.000 hab) de janeiro a fevereiro e decresceu a partir deste mês (chegando a ser 0,254 casos por 10.000 hab em maio).

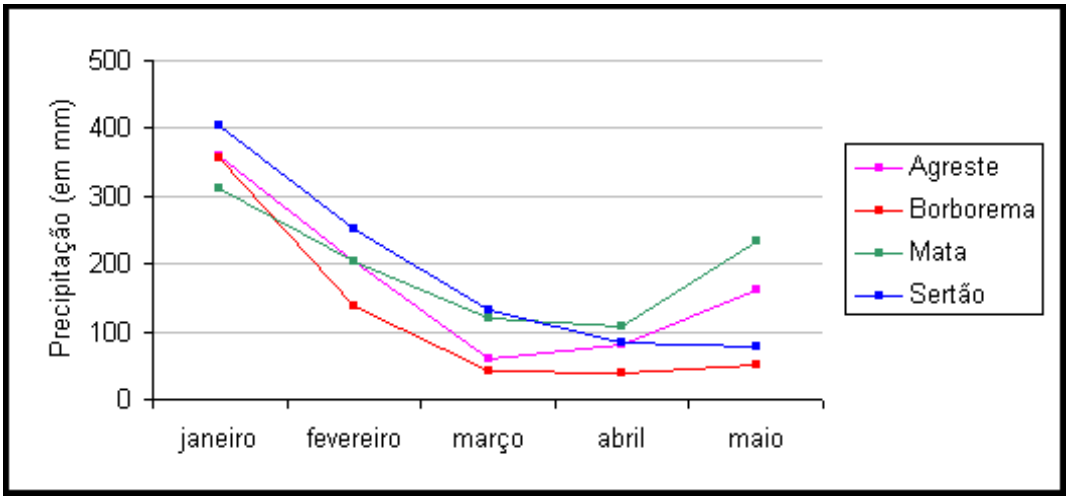


Figura 17. Série Temporal da precipitação média segundo as mesorregiões paraibanas

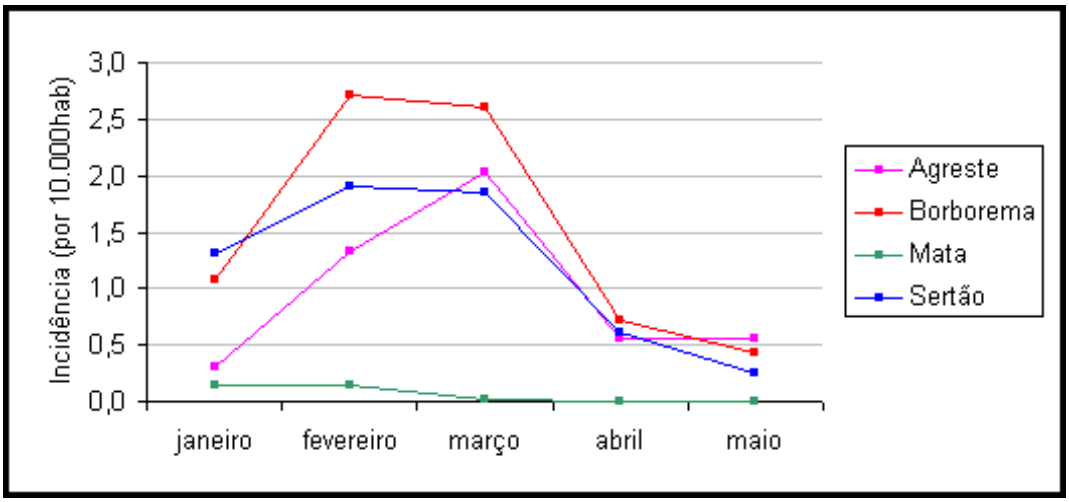


Figura 14. Série Temporal da incidência do dengue segundo as mesorregiões paraibanas

5 CONCLUSÕES

Neste trabalho fez-se um estudo exploratório sobre a incidência do dengue e a sua relação com a pluviometria no estado da Paraíba, no período de janeiro a maio de 2004. Observou-se que mais da metade dos 223 municípios paraibanos (cerca de 52,02%) não notificaram casos do dengue. A distribuição da incidência do dengue apresentou, como principal característica espacial, a ausência de aglomerados de municípios paraibanos com incidência similar da doença. O mesmo não foi observado na distribuição pluviométrica que apresentou aspectos de continuidade espacial, isto é, crescimento (ou decrescimento) gradativo dos níveis de precipitação numa determinada direção. A partir deste estudo, pode-se verificar que, no estado da Paraíba, entre os meses de janeiro e maio de 2004, a incidência do dengue não variou com o aumento da pluviosidade. Observou-se também, que no período de maior incidência do dengue (meses de fevereiro e março para mesorregiões do Agreste, Borborema e Sertão e meses de janeiro e fevereiro para Mata Paraibana), os níveis de precipitação estavam decrescendo. Apesar de não haver relação entre chuva e incidência do dengue, neste primeiro semestre do ano de 2004, a estação do verão foi a que apresentou maior incidência do dengue, o que está de acordo com os alertas do Ministério da Saúde.

REFERÊNCIAS

BRASIL. Fundação Nacional de Saúde. **Programa Nacional de Controle do Dengue - PNCD.** Brasília: Fundação Nacional de Saúde, 2002. 51 p.

BUSSAB, W. O.; MORETTIN, P. A. **Estatistica Básica.** 5. ed. São Paulo: Saraiva, 2002.

CÂMARA, G. et al. SPRING: Integrating remote sensing and GIS by object oriented data modelling. Sistema de Processamento de Informações Georeferenciadas (SPRING). **Computers & Graphics**. Great Britain, 1996. v. 20, n. 3, p. 395-403. Disponível em: <http://ww.dpi.inpe.br/spring/>. Acesso em: 3 ago. 2003.

CAMARGO, E. C. G. **Desenvolvimento, implementação e teste de procedimentos Geoestatísticos (KRIGEAGEM) no Sistema de Processamento de Informações Georreferenciadas (SPRING)**. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos: INPE, jun. 1997. Disponível em: <http://www.dpi.inpe.br/teses/eduardo/>. Acesso em: 12 set. 2003.

CARVALHO, M. S.; PINA, M. F.; SANTOS, S. M. **Conceitos básicos de Sistemas e Informação Geográfica e Cartografia aplicados à saúde.** Brasília: Organização Panamericana de Saúde / Ministério da Saúde, 2000.

CLARKE, A. B.; DISNEY, R. L. **Probabilidade e Processos Estocásticos.** Rio de Janeiro: Livros Técnicos Científicos, 1979.

COSTA, A. E. A.; FERREIRA, L. G. Consideração sobre o dengue clássico e o hemorrágico. **Pharmacia Brasileira**. Rio de Janeiro, 2002.

DATASUS. Departamento de Informática do SUS. **Informações de Saúde.** Disponível em: <http://www.datasus.gov.br/>. Acesso em: 18 nov. 2003.

DEUTSCH, C. V.; JOURNEL, A. G. **GSLIB Geostatistical Software Library and User's Guide.** New York: Oxford University Press, 1998. 369 p.

DEUTSCH, C. V.; SCHNETZLER, E. T. **Fundamentals of Geostatistics:** principles and hands-on practice. A Short Course. Edmonton: Statios, 2000. Disponível em: <http://www.statios.com/Training/index.html>. Acesso em: 19 set 2003.

DIGGLE, P. J.; RIBEIRO JR., P. J. Model - based Geostatistics. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14, 2000. **Anais**... Caxambu: ABE, 2000. 137f.

DRUCK, S. et al. (Ed.). **Análise espacial de dados geográficos.** 2. ed. São José dos Campos: INPE, jul. 2002. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/>. Acesso em: 12 set. 2003.

FARHAT, R. Dengue: Brasil prepara-se para a batalha. **Informe Saúde:** publicação semanal da Divisão de Jornalismo da Assessoria de Comunicação Social do Ministério da Saúde. Brasília, DF, ano 7, n. 196, 1 jan. 2003.

FELGUEIRAS, C. A. **Modelagem ambiental com tratamento de incertezas em Sistemas de Informação Geográfica: o paradigma Geoestatístico por indicação.** Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos: INPE, 1999. Disponível em: <http://www.dpi.inpe.br/teses/carlos/>. Acesso em: 10 ago. 2003.

FELGUEIRAS, C. A. et al. Inferências e estimativas de incertezas utilizando técnicas de krigeagem não linear. In: CONGRESSO E FEIRA PARA USUÁRIOS DE GEOPROCESSAMENTO DA AMÉRICA LATINA, 5, 1999. **Anais...** Salvador: Gisbrasil'99, 1999. 1 CD-ROM.

FELGUEIRAS, C. A.; DRUCK, S.; MONTEIRO, A. M. V. Análise espacial de superfícies: o enfoque da Geoestatística por indicação. In: DRUCK, S. et al. (Ed.). **Análise espacial de dados geográficos.** 2. ed. São José dos Campos: INPE, jul. 2002. cap. 4. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/>. Acesso em: 12 set. 2003.

FURTADO, P. C. H.; SOUZA, I. C. A.; MORAES, R. M. Influência das estações do ano na incidência de casos de dengue no estado da Paraíba no período 1998-2003. In: CONGRESSO BRASILEIRO DE PESQUISAS AMBIENTAIS E SAÚDE, 4., 2004, Santos. **Anais...** Santos: COPEC, 2004. p. 139-142. 1 CD-ROM.

GOODCHILD, M. F.; GUOGING, S. Development and test of an error model for categorical data. **International Journal of Geographical Information Systems**, [S.l.], v. 6, n. 2, p. 87-104, 1992.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An Introduction to Applied Geostatistics.** New York: Oxford University Press, 1989.

JIAN, X; OLEA, R. A.; YU, Y-S. Semivariogram modeling by weighted least squares. **Computers & Geosciences**. Great Britain, 1996. v. 22, n. 4, p. 387-397.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis.** New Jersey: Prentice Hall, 1992.

LANDIM, P. M. B. **Análise Estatística de Dados Geológicos.** São Paulo: UNESP, 1998.

LAURENTI, R. et al. **Estatísticas de Saúde.** 2. ed. São Paulo: EPU, 1987.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of Statistics**. Tokyo: Mc Graw-Hill, 1974.

MORAES, R. M.; SOUZA, I. C. A.; TELES, M. M. F. Análise Espacial de Dados e Aplicações. In: ENCONTRO REGIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 3., 2003, João Pessoa. Notas de Minicurso. **Anais...** João Pessoa: UFPB, 2003. 1 CD-ROM.

NIMER, E. **Pluviometria e Recursos Hídricos dos Estados de Pernambuco e Paraíba**. Rio de Janeiro: IBGE, 1979. 128 p.

NUNES, L. M. **ESTIMAÇÃO:** Inferência estatística variografia e krigagem. Universidade do ALGARVE - Unidade de Ciências e Tecnologias dos Recursos Aquáticos. Faro, dez. 1998. 20f.

RODRIGUEZ, J. L. **Atlas escolar da Paraíba**. 2. ed. João Pessoa: GRAFSET, 2000.

SIEGEL, S. **Estatística Não-Paramétrica para as Ciências do Comportamento.** São Paulo: McGraw-Hill do Brasil, 1975.

SOUZA, I. C. A.; MORAES, R. M. **Utilização de Sistemas de Informação Geográfica na Análise Espacial de Dados de Saúde Pública do Estado na Paraíba entre os anos de 1998 e 2001.** Relatório de Iniciação Científica PIBIC/CNPq. João Pessoa: UFPB, 2003. Disponível em: <http://www.de.ufpb.br/~ronei/monica/index.html>. Acesso em: 3 jul. 2003.

SOUZA, I. C. A.; MORAES, R. M. **Utilização de Métodos Geoestatísticos na Análise Espacial de Dados dos Municípios Paraibanos.** Relatório de Iniciação Científica PIBIC/CNPq. João Pessoa: UFPB, 2004. Disponível em: <http://www.de.ufpb.br/~ronei/monica/index.html>. Acesso em: 10 out. 2004.

TUKEY, J. W. **Exploratory data analysis.** Massachusetts: Addison-Wesley Publishing Company, 1977.

APÊNDICES

APÊNDICE A - Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano para os meses de fevereiro a maio de 2004.

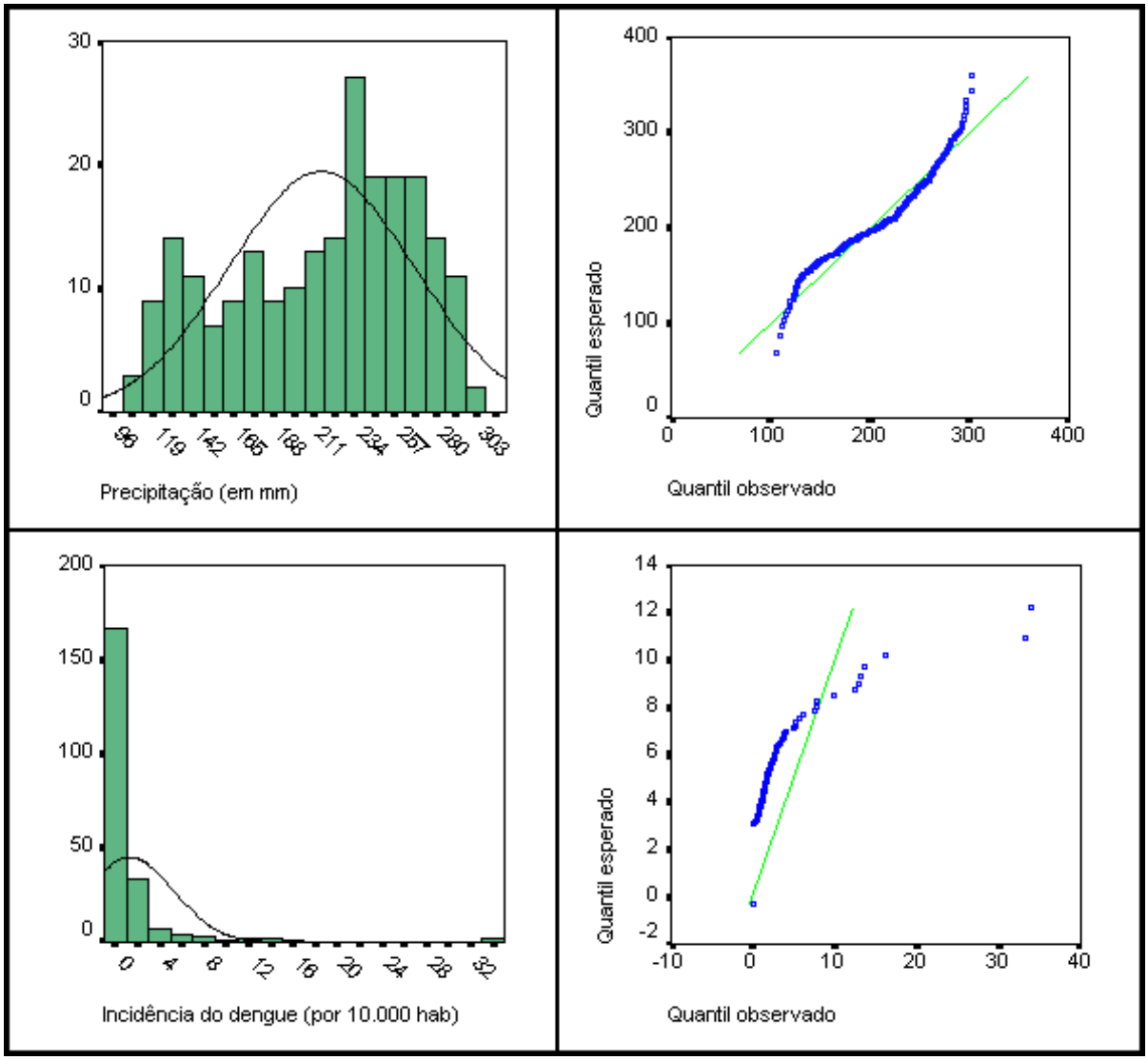


Figura A1. Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de fevereiro de 2004

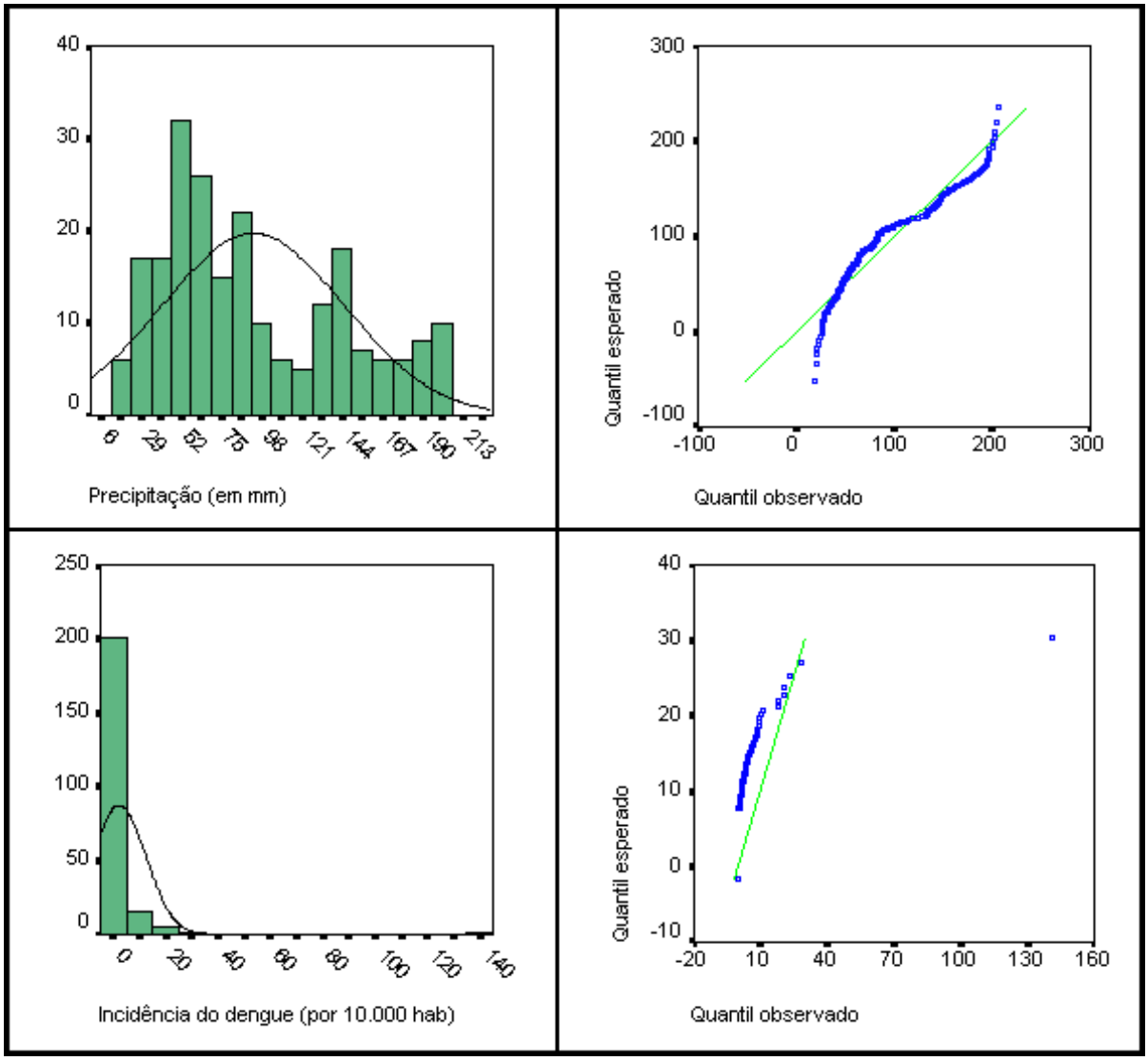


Figura A2. Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de março de 2004

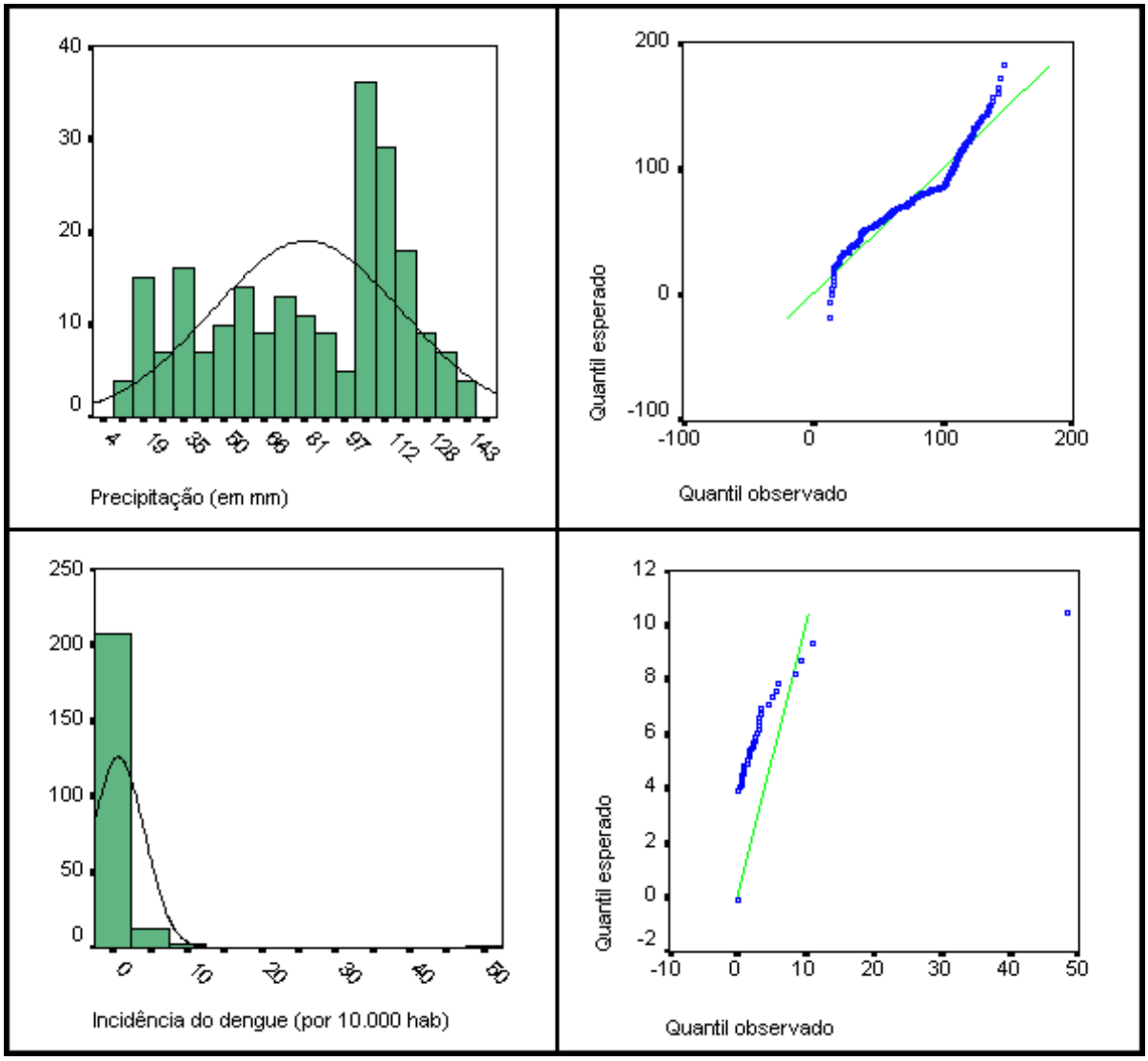


Figura A3. Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de abril de 2004

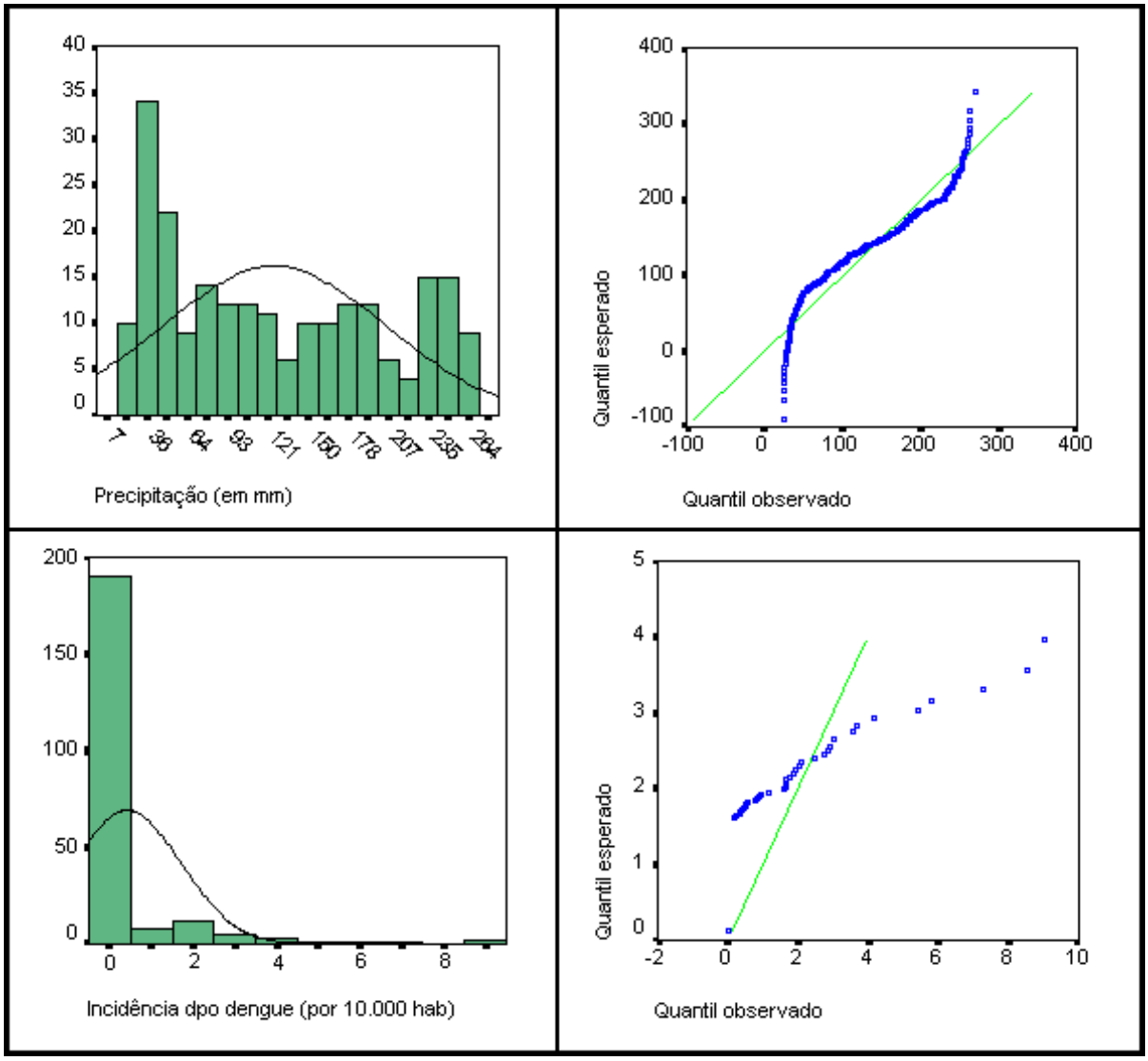


Figura A4. Histograma e qqplot das variáveis precipitação média e coeficiente de incidência do dengue por município paraibano no mês de maio de 2004

APÊNDICE B - Parâmetros dos modelos de semivariograma indicador da variável precipitação acumulada (em mm) segundo os meses de janeiro a maio de 2004.

Tabela B1– Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de janeiro de 2004

z_c	h	C_0	C_1	a	Akaike
279,3	60.000	0,109	0,045	234.367,223	-21,342
334,9	60.000	0,114	0,122	100.170,802	-26,783
372,6	60.000	0,154	0,098	63.571,692	-39,557
408,0	50.000	0,134	0,094	83.958,067	-52,973
460,3	30.000	0,059	0,093	72.191,362	-32,584

Fonte: dados do Depto. de Ciências Atmosféricas da UFCG.

Tabela B2– Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de fevereiro de 2004

z_c	h	C_0	C_1	a	Akaike
122,1	30.000	0,088	0,120	257.566,852	-34,766
165,4	30.000	0,111	0,183	203.559,145	-59,102
193,9	30.000	0,104	0,184	138.135,280	-35,54
243,7	60.000	0,169	0,108	471.203,733	-43,852
294,9	60.000	0,129	0,349	13.563.456,657	-20,266

Fonte: dados do Depto. de Ciências Atmosféricas da UFCG.

Tabela B3– Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de março de 2004

z_c	h	C_0	C_1	a	Akaike
35,0	28.000	0,025	0,173	139.625,360	-27,680
55,8	28.000	0,092	0,241	237.178,814	-35,936
74,9	36.000	0,122	0,315	436.535,397	-38,423
107,2	36.000	0,072	18,358	31.688.042,412	-28,062
152,4	50.000	0,063	10,210	36.118.521,754	-29,000

Fonte: dados do Depto. de Ciências Atmosféricas da UFCG.

Tabela B4 – Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de abril de 2004

z_c	h	C_0	C_1	a	Akaike
23,0	24.000	0,000	0,236	154.664,62	-30,747
51,9	26.000	0,014	0,284	129.611,33	-36,950
73,0	26.000	0,089	0,210	147.177,13	-50,782
90,2	60.000	0,177	0,140	708.705,16	-40,176
118,3	60.000	0,120	0,310	3.868.860,32	-34,187

Fonte: dados do Depto. de Ciências Atmosféricas da UFCG.

Tabela B5 – Parâmetros do modelo de semivariograma indicador da variável precipitação acumulada (em mm) referente ao mês de maio de 2004

z_c	h	C_0	C_1	a	Akaike
34,9	30.000	0,021	0,188	215.494,462	-40,100
63,7	30.000	0,013	0,793	657.425,784	-27,667
89,0	30.000	0,014	0,380	266.769,880	-37,381
136,9	60.000	0,096	0,451	832.483,852	-39,307
196,3	60.000	0,117	7,342	55.804.987,842	-15,434

Fonte: dados do Depto. de Ciências Atmosféricas da UFCG.