

GABRIELLE GOMES DOS SANTOS RIBEIRO

**ANÁLISE GEOESTATÍSTICA PARA GERAÇÃO DE SUPERFÍCIES A
PARTIR DE DADOS DE CLOROFILA-A ADQUIRIDOS EM TRANSECTOS**



GABRIELLE GOMES DOS SANTOS RIBEIRO

**ANÁLISE GEOESTATÍSTICA PARA GERAÇÃO DE SUPERFÍCIES A
PARTIR DE DADOS DE CLOROFILA-A ADQUIRIDOS EM TRANSECTOS**

Dissertação apresentada ao Programa de Pós
Graduação em Ciências Cartográficas da
FCT – Faculdade de Ciências e Tecnologia
da UNESP – Universidade Estadual Paulista,
campus de Presidente Prudente.

Orientadora: Prof^ª Dr^ª Vilma Mayumi
Tachibana.

Coorientadora: Prof^ª Dr^ª Maria de Lourdes B.
T. Galo.

Presidente Prudente
2015

FICHA CATALOGRÁFICA

R369a Ribeiro, Gabrielle Gomes dos Santos.
Análise geoestatística para geração de superfícies a partir de dados de
clorofila-*a* adquiridos em transectos / Gabrielle Gomes dos Santos Ribeiro. -
Presidente Prudente : [s.n], 2015
95 f.

Orientador: Vilma Mayumi Tachibana
Coorientador: Maria de Lourdes Bueno Trindade Galo
Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de
Ciências e Tecnologia
Inclui bibliografia

1. Delineamento Amostral. 2. Transectos. 3. Krigagem. I. Tachibana,
Vilma Mayumi. II. Galo, Maria de Lourdes Bueno Trindade. III.
Universidade Estadual Paulista. Faculdade de Ciências e Tecnologia. IV.
Título.



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Presidente Prudente

BANCA EXAMINADORA

Prof. Dra. **VILMA MAYUMI TACHIBANA**
(ORIENTADOR)

Prof. Dr. **JOSÉ SILVIO GOVONE**
(UNESP/IGCE)

Prof. Dr. **MILTON NOBUHIRO IMAI**
(UNESP/FCT)

GABRIELLE GOMES DOS SANTOS RIBEIRO

Presidente Prudente (SP), 26 de fevereiro de 2015.

Resultado: **APROVADO**

*Aos meus amados pais, Marta e Edivaldo,
Por incentivar todas as minhas conquistas.
Devo tudo a vocês, vocês são a minha vida.*

Ao meu esposo Danilo, pela paciência, apoio e carinho.

E principalmente ao meu amado filho Lorenzo.

AGRADECIMENTOS

Muitas pessoas foram importantes para mim durante esses três anos que fiquei envolvida nessa pesquisa, pessoas indispensáveis para a realização deste trabalho e que me ajudaram, de alguma forma, a conseguir mais essa vitória na minha vida. Mas eu não posso deixar de agradecer primeiramente a Deus, Ele foi meu amigo fiel e meu ajudador, me acompanhou durante toda esta jornada. Nos momentos de mais aflição, em que eu não tinha a quem recorrer, Ele que foi o meu guia e meu socorro.

Com todo amor e carinho, aos meus pais e meu irmão, que sempre me apoiaram, acreditaram no meu potencial e vibraram por minhas vitórias, dizendo “sentimos muito orgulho de você”. Vocês são a razão de todo o meu esforço e dedicação, se estou aqui é por causa de vocês, e nada que eu faça será suficiente para recompensar tudo o que vocês fizeram por mim. Amo muito vocês.

Em especial, ao amor da minha vida, meu esposo Danilo, sem ele ao meu lado, eu não teria conseguido chegar até aqui. Ele foi o meu “porto seguro”, meu amigo e companheiro. Dividiu comigo as madrugadas de estudo, ouviu os meus choros de desespero e me deu apoio nos dias difíceis em que eu achava que não conseguiria, dizendo “calma querida, um dia você será recompensada por todo seu esforço”.

Não posso deixar de agradecer a minha orientadora, Prof.^a Dra. Vilma Mayumi Tachibana, pela sua paciência e total dedicação. Além de orientadora, ela foi uma amiga e um exemplo de profissionalismo, competência e eficiência. Sem o seu apoio, sei que não teria conseguido chegar até aqui, sinto-me privilegiada por ter sido um de seus orientandos.

A minha coorientadora Maria de Lourdes pelos ensinamentos e imensa contribuição no desenvolvimento da pesquisa.

Também agradeço aos amigos do PPGCC, em especial aqueles que estiveram ao meu lado, me ajudando e me ensinando, quando me inseri em um mundo totalmente novo para mim, o mundo da Cartografia: Samara Calçado, Fernanda Watanabe, Érika Saito, Rejane, Ricardo Javier, Mariana Chaves, Hérica, Ariane e Juliane.

Aos professores do departamento de Cartografia, por compartilharem seus conhecimentos e experiências.

A CAPES pela bolsa de estudos concedida e ao Programa de Pós Graduação em Ciências Cartográficas da UNESP (PPGCC/UNESP) pela infraestrutura para o desenvolvimento do trabalho.

À Faculdade de Ciências e Tecnologia da UNESP (FCT/UNESP) e ao Departamento de Cartografia da FCT/UNESP por todo apoio e auxílio concedido.

Muito obrigado!

“Entrega o teu caminho ao Senhor; confia nele, e ele tudo fará”
(Salmos, 37.5)

RESUMO

A disposição dos elementos amostrais na área de estudo e sua influência nos resultados de análises espaciais é algo que vem sendo discutido frequentemente, já que a qualidade de uma inferência espacial vai depender do tamanho da amostra e da distribuição espacial dos pontos amostrais. Nesse sentido, este trabalho tem o objetivo de analisar o impacto que diferentes delineamentos amostrais podem causar nos resultados da inferência espacial por Krigagem Ordinária. Para isso, primeiramente utilizou-se um conjunto de dados coletado em forma de transectos em uma parte do Reservatório de Nova Avanhandava, composto por 978 observações. Esse conjunto sofreu reduções sistemáticas, com o intuito de analisar o que essas reduções causariam nos resultados das inferências espaciais. Com o objetivo de analisar diferentes delineamentos amostrais, simulou-se uma quantidade densa de dados e aplicou-se as técnicas de Amostragem Simples, Amostragem Sistemática e Amostragem Estratificada. Para complementar, utilizou-se um conjunto de dados de tamanho reduzido (70 observações), coletado de forma aleatória, a fim de analisar os resultados obtidos pela Krigagem Ordinária ao utilizar um conjunto considerado “pequeno”, do ponto de vista estatístico. Então, foi possível realizar o processo da Krigagem Ordinária e obter mapeamentos da variável clorofila-*a* na região de interesse para os diferentes tipos e tamanhos de amostras. A validação dos processos de inferência foi realizada a partir de dois métodos, o Erro Médio Quadrático e o Índice Kappa. Os resultados mostraram que ao diminuir gradualmente o tamanho da amostra dos dados em transectos, não houve grandes alterações nos resultados das inferências. No caso dos dados simulados, ao comparar os delineamentos amostrais, verificou-se que as amostras obtidas pela técnica de Amostragem Sistemática foram as mais eficazes ao mapear a variável clorofila-*a* por Krigagem Ordinária. Por fim, no conjunto menor, os resultados mostraram que um pequeno aumento na quantidade de dados na amostra melhorou os resultados da inferência espacial.

Palavras-chave: Delineamento amostral, dados espaciais, transectos, krigagem.

ABSTRACT

The arrangement of sampling units in the study area and its influence on the results of spatial analysis is something that has been frequently discussed by researchers of the area, since the quality of a spatial inference will depend on sample size and spatial distribution of sample points. In this sense, this work aims to analyze the impact that different sampling designs may cause in the results of spatial inference by ordinary kriging. For this, first we used a dataset collected in the form of transects in a part of New Avandava Reservoir, consisting of 978 observations. This set suffered systematic reductions, with the aim to analyze what these reductions would cause in the results of spatial inferences. In order to analyze different sampling designs, simulated up a dense amount of data and it was applied the Simple Sampling, Systematic Sampling and Stratified Sampling techniques. To complement, it was used a data set with a small size (70 observations) and collected randomly, in order to analyze the results obtained by ordinary kriging when using a set considered “small”, from a statistical point of view. Then, it was possible to perform the process of Ordinary Kriging and obtain mappings of the variable chlorophyll-a in the region of interest, for different types and sizes of samples. The validation of inference processes was carried out from two methods, the Mean Squared Error and the Kappa Index. The results showed that by gradually decreasing the size of the sample of data on transects, there were no major changes in the results of inferences. In the case of simulated data by comparing the sampling designs, it was found that the samples obtained by systematic sampling technique were the most effective in mapping the variable chlorophyll-a by ordinary kriging. Finally, in the lowest set, the results showed that a small increase in the amount of data in the sample considerably improved the results of spatial inference.

Keywords: Sampling designs, spatial data, transects, kriging.

LISTA DE FIGURAS

Figura 1(a) e 1(b) - Principais componentes da variação espacial.	25
Figura 2 - Semivariograma Experimental.	26
Figura 3 - Anisotropia Geométrica.	27
Figura 4- Anisotropia Zonal.	27
Figura 5- Modelos teóricos de semivariogramas.	28
Figura 6 - Krigagem pontual para interpolação do ponto (X=28; Y=21,25).	33
Figura 7 - Relação entre o tamanho da amostra e o erro amostral.	37
Figura 8- Matriz de confusão.	44
Figura 9 - Ilustração do Reservatório de Nova Avanhandava.	46
Figura 10 - Ilustração da subárea de estudo.	47
Figura 11 - Área de estudo: trecho do reservatório de Nova Avanhandava.	47
Figura 12 – Fluxograma do desenvolvimento metodológico.	50
Figura 13 – Amostragem em forma de transecto.	51
Figura 14 - Distribuição espacial do conjunto de dados.	52
Figura 15 – Estratos da amostragem estratificada.	54
Figura 16 – Distribuição das concentrações de clorofila- <i>a</i> ao longo do transecto.	56
Figura 17 - Formas da curva de distribuição em relação à curva normal.	57
Figura 18 - Tipos de distribuição em relação à assimetria.	58
Figura 19 – Histograma da concentração de clorofila- <i>a</i>	58
Figura 20 - Gráfico Normal Q-Q Plot.	59
Figura 21 - Gráfico Box-Plot.	59
Figura 22 - Distribuição espacial da amostra sistemática 400.	62
Figura 23 - Distribuição espacial da amostra sistemática 300.	62
Figura 24 - Distribuição espacial da amostra sistemática 200.	63
Figura 25 - Distribuição espacial da amostra sistemática 100.	63
Figura 26- Mapas variográficos para as amostras de tamanhos 400 (1), 300 (2), 200 (3) e 100 (4).	64
Figura 27 - Semivariogramas omnidirecional e para a direção 0° da amostra 400 dos dados em transecto.	64
Figura 28 – Semivariogramas para as direções 45° e 90° da amostra 400 dos dados em transecto.	65
Figura 29 – Semivariograma para a direção 135° da amostra 400 dos dados em transecto.	65
Figura 30 – Semivariogramas omnidirecional e para a direção 0° da amostra 300 dos dados em transecto.	65
Figura 31 – Semivariogramas para as direções 45° e 90° da amostra 300 dos dados em transecto.	65
Figura 32 – Semivariograma para a direção 135° da amostra 300 dos dados em transecto.	66
Figura 33 – Semivariogramas omnidirecional e para a direção 0° da amostra 200 dos dados em transecto.	66
Figura 34 – Semivariogramas para as direções 45° e 90° da amostra 200 dos dados em transecto.	66
Figura 35 – Semivariograma para a direção 135° da amostra 200 dos dados em transecto.	66
Figura 36 - Semivariograma omnidirecional e para a direção 0° da amostra 100 dos dados em transecto.	67

Figura 37 – Semivariogramas para as direções 45° e 90° da amostra 100 dos dados em transecto.....	67
Figura 38 - Semivariograma para a direção 135° da amostra 100 dos dados em transecto.	67
Figura 39 – Modelagem da anisotropia para as amostras tamanho 400 (1), 300 (2), 200 (3) e 100 (1).	68
Figura 40 – Krigagem Ordinária da clorofila- <i>a</i> para as amostras de tamanhos 400 (1), 300 (2), 200 (3) e 100 (1).	69
Figura 41- Krigagem Ordinária utilizando todos os 978 dados.....	69
Figura 42 – Histograma dos valores preditos de clorofila- <i>a</i>	70
Figura 43 – Mapas de classes da Krigagem Ordinária das amostras de tamanhos 400 (1), 300 (2), 200 (3) e 100 (4).	71
Figura 44 - Mapas de variância do erro de estimação para as amostras de tamanhos 400 (1), 300 (2), 200 (3) e 100 (4).	71
Figura 45 - Mapas variográficos das amostras simples de tamanhos 1000, 400, 300, 200 e 100.	73
Figura 46 - Krigagem Ordinária da clorofila- <i>a</i> utilizando as amostras simples de tamanhos 1000, 400, 300, 200 e 100.	74
Figura 47 - Mapa de classes da Krigagem Ordinária das amostras de tamanhos 1000, 400, 300, 200 e 100.	75
Figura 48 - Mapas variográficos das amostras estratificadas de tamanhos 1000, 400, 300, 200 e 100.	76
Figura 49 - Krigagem Ordinária da clorofila- <i>a</i> utilizando as amostras estratificadas de tamanhos 1000, 400, 300, 200 e 100.	77
Figura 50 - Mapa de classes da Krigagem Ordinária das amostras estratificadas de tamanhos 1000, 400, 300, 200 e 100.	78
Figura 51 - Mapas variográficos para as amostras sistemáticas de tamanhos 1000, 400, 300, 200 e 100.	80
Figura 52 - Krigagem Ordinária da clorofila- <i>a</i> utilizando as amostras sistemáticas de tamanhos 1000, 400, 300, 200 e 100.	81
Figura 53 - Mapa de classes da Krigagem Ordinária das amostras sistemáticas de tamanhos 1000, 400, 300, 200 e 100.	82
Figura 54 - Semivariograma omnidirecional.	84
Figura 55 - Mapa variográfico dos dados pontuais.	84
Figura 56 - Semivariogramas das direções 0° e 45° para a amostra de dados pontuais.	85
Figura 57 - Semivariogramas das direções 90° e 135° para a amostra de dados pontuais.	85
Figura 58 - Modelagem da anisotropia para a amostra dos dados pontuais.	86
Figura 59 - Interpolação espacial por Krigagem Ordinária da amostra de dados pontuais de clorofila- <i>a</i>	87
Figura 60 - Mapa de variância do erro para a amostra dos dados pontuais.....	87
Figura 61 - Mapa de classes da Krigagem Ordinária dos dados pontuais.....	88

LISTA DE TABELAS

Tabela 1 - Pontos de dados vizinhos para estimativa da localização ($X=28,75$; $Y=21,25$) por meio da krigagem ordinária pontual.	33
Tabela 2– Interpretação dos valores de kappa.	44
Tabela 3 - Resultados numéricos para a população de clorofila- <i>a</i>	56
Tabela 4 - Resultado do teste Shapiro-Wilk.	60
Tabela 5 - Resultados das estimativas através da AS para as amostras de clorofila- <i>a</i> ...	61
Tabela 6 - Valores de patamar, alcances e razão de anisotropia para cada amostra.	68
Tabela 7 - Resultados do EMQ e Índice Kappa para a inferência espacial utilizando as amostras estratificadas.	72
Tabela 8 - Resultados do EMQ e Índice Kappa para a inferência espacial utilizando as amostras simples.	75
Tabela 9 - Resultados do EMQ e Índice Kappa para a inferência espacial utilizando as amostras estratificadas.	78
Tabela 10 - Resultados das estimativas através da AS para as amostras de clorofila- <i>a</i> .	79
Tabela 11 - Resultados do EMQ e Índice Kappa para a inferência espacial utilizando as amostras sistemáticas.	82
Tabela 12 - Resultados do EMQ e Índice Kappa para todos os casos.	83

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Apresentação do Problema	15
1.2 Objetivos	19
1.3 Justificativa	19
1.4 Estrutura do trabalho	21
2 FUNDAMENTAÇÃO TEÓRICA	23
2.1 Geoestatística	23
2.1.1 Variáveis Regionalizadas	23
2.1.2 Semivariogramas	25
2.1.3 Modelos Teóricos	28
2.1.4 Inferência por Krigagem	29
2.2 Amostragem	34
2.2.1 Tamanho da Amostra	37
2.2.2 Tipos de Amostragem	38
2.3 Variável de estudo: clorofila- <i>a</i>	41
2.4 Validação da Inferência Espacial	42
2.4.1 Erro Quadrático Médio (EQM)	43
2.4.2 Índice de Concordância Kappa	43
3 MATERIAIS E MÉTODO	46
3.1 Área de estudo	46
3.2 Materiais	48
3.4 Desenvolvimento Metodológico	50
3.4.1 Dados adquiridos em Transecto	50
3.4.1.1 Análise do Delineamento Amostral	52
3.4.1.2 Experimento dos Dados Simulados	53
3.4.2 Dados Pontuais	55
4 RESULTADOS E DISCUSSÕES	56
4.1 Análise descritiva dos dados	56
4.2 Em forma de transecto	60
4.2.1 Dados Reais	61
4.2.2 Dados Simulados	72
4.3 Amostragem pontual	83

5 CONCLUSÕES E RECOMENDAÇÕES	89
REFERÊNCIAS	91

1 INTRODUÇÃO

1.1 Apresentação do Problema

A Estatística Espacial, segundo Assunção (2001), é um ramo da Estatística que estuda métodos científicos para coleta, descrição, visualização e análise de dados georreferenciados para caracterizar os fenômenos e ocorrências de interesse em diversas áreas do conhecimento, entre elas a Cartografia. Com os avanços de Sistemas de Informações Geográficas, ela vem se apresentando como uma importante ferramenta no mapeamento de regiões, em diferentes escalas de grandeza.

As técnicas de análise estatística espacial têm servido a um grande leque de aplicações, mas principalmente à epidemiologia (no mapeamento de doenças) e meio ambiente (na identificação de aglomerados espaciais – *clusters* – e no controle e monitoramento de problemas ambientais) (SILVA, 2011). Nessas áreas é frequente querer prever ou estimar a quantidade total ou concentração de uma ou mais variáveis de interesse em uma região de estudo. As medidas dessa variável observadas em uma quantidade pequena de localidades podem ser usadas para prever, por exemplo, a concentração de poluente, em uma nova localidade ou para estimar a concentração média na região toda.

Nessas situações, para Thompson (2002) é útil considerar o valor da observação em uma localidade como variável aleatória. A partir de vários valores observados em diferentes localidades é possível estimar ou prever o valor de uma nova variável aleatória, que pode ser um novo local ou a quantidade total em uma grande região geográfica.

Há diferentes métodos propostos na literatura para obtenção dessas estimativas. Uma divisão simples entre os métodos pode ser em métodos determinísticos e métodos estocásticos. Os métodos determinísticos têm por base critérios puramente geométricos em que as distâncias são euclidianas e não fornecem medidas de incerteza como, por exemplo, os métodos: inverso do quadrado da distância (IQD), vizinho mais próximo, média simples, triangulação, média ponderada e mínima curvatura. Nos modelos

estocásticos os valores coletados são interpretados como provenientes de processos aleatórios e são capazes de quantificar a incerteza associada ao estimador. Os métodos geoestatísticos pertencem a essa categoria, entre eles está a Krigagem (YAMAMOTO e LANDIM, 2013).

Neste trabalho foi utilizado o método de krigagem, que segundo Wang et al. (2012) é o principal método de estimação em geoestatística, usado primeiro por D. J. Krige e H. S. Sichel no estudo de localização de depósitos minerais nos anos 1950s e 1960s e posteriormente desenvolvido e aplicado por Matheron como realmente uma teoria científica. Esse método dá estimação não tendenciosa de valores de uma variável em locais onde a coleta não foi realizada com o menor erro quadrático médio de estimação.

Em muitas abordagens de análise espacial é necessário coletar uma quantidade considerável de dados georreferenciados para produzir um mapeamento da região de estudo e, dependendo do tamanho e da localização da região em questão, a aquisição das informações pode demandar muito tempo e ainda ter um custo muito elevado. Por isso, realizar uma pesquisa usando qualquer tipo de análise estatística, em primeiro lugar, depende da seleção cuidadosa e do planejamento de um método de amostragem. Segundo Yamamoto e Landim (2013), a qualidade de uma inferência espacial vai depender do tamanho da amostra e da distribuição espacial dos pontos amostrais.

Por isso, um dos problemas enfrentados é a obtenção do layout do padrão de amostragem. Supondo que se tenha uma área definida em que o número de pontos amostrais foi fixado com base em orçamento restrito, em que locais deve-se coletar as informações? Isso será definido de acordo com as definições do método de amostragem que o pesquisador adotar, os métodos mais comuns são: Amostragem Simples, Amostragem Sistemática e Amostragem Estratificada.

Muitos trabalhos foram e estão sendo publicados no sentido de estudar planejamentos amostrais que sejam mais adequados para a coleta de dados espaciais. Isso porque autores como Haining (1990), consideram que é necessário certo cuidado ao trabalhar com esse tipo de dados já que são correlacionados espacialmente e não são provenientes de populações independentes e identicamente distribuídas (i.i.d.), ou seja, violam a suposição de independência.

Dentre os trabalhos que estudam delineamentos amostrais para coleta *in situ* de parâmetros de qualidade da água está o trabalho de Samizava et al. (2008), que propôs duas metodologias diferentes para delineamento amostral em planícies de inundação (lagoas e rios). A primeira proposta foi baseada em análise de agrupamentos em dados limnológicos para definir as lagoas com características limnológicas semelhantes, a partir disso verificaram-se lagoas com características bastante homogêneas, que permitiu selecionar para a amostragem localizações com maior variabilidade. A segunda proposta baseou-se em imagens multiespectrais do sensor TM/Landsat, que possibilitou gerar regiões espectralmente homogêneas (estratos) para a aplicação de uma amostragem aleatória estratificada. Ambas as abordagens utilizadas mostraram-se adequadas no reconhecimento das observações que apresentem padrões de homogeneidade no espaço, facilitando no momento da definição das posições em que serão coletadas medidas limnológicas e espectrais da água.

Outro fator importante de ser estudado é o tamanho da amostra, sabe-se que amostras maiores geram estimativas mais precisas, porém, erros menores têm seu custo: maiores complexidades, mais equipes, mais equipamentos, mais tempo de trabalho em campo, etc., o que implica custos financeiros mais elevados. Por outro lado, deve-se levar em conta que os resultados extraídos de amostras menores estão sujeitos a grandes variabilidades, transmitem pouca confiança e, portanto, não são consideradas adequadas para tomadas de decisões estratégicas. Portanto, o ideal seria encontrar um ponto de equilíbrio entre o erro permitido pelo pesquisador e a precisão requerida nos resultados.

Wang et al. (2012) definem duas teorias para amostragem e inferência espacial: amostragem baseada em delineamento e amostragem baseada em modelo. Na amostragem baseada em delineamento as características de um campo aleatório são obtidas depois que a área inteira é exaustivamente amostrada. Os valores na região de interesse são desconhecidos, mas fixos. A incerteza da estimação amostral surge da aleatoriedade dos locais de amostragem, que pode ser reduzida aumentando-se o tamanho da amostra até que a área seja completamente coberta. Ela pode ser implementada por amostragem tradicional. A abordagem em que a amostragem é baseada em modelo, o valor observado é uma realização de um processo estocástico subjacente e considerado não fixo. Para calcular um estimador ou um preditor, os pesos dos dados amostrais são determinados pelas covariâncias entre as observações.

Na abordagem baseada em modelo, para alcançar os propósitos de minimizar a estimação da variância do erro, uma exigência frequente é que a variação espacial, como dado pelo variograma, seja conhecida a priori, o que não é comum.

O plano de amostragem adotado, além de gerar uma amostra representativa que forneça boas estimativas, deve ser operacionalmente viável, ou seja, sua aplicação em campo deve ser efetiva. Para isso, o delineamento escolhido deve fornecer uma disposição dos elementos amostrais na área de estudo que respeite alguns fatores importantes no momento da coleta, como: o difícil acesso a algumas áreas da região, acidentes geográficos, o tamanho da amostra relacionado ao tempo disponível para a coleta, a distância entre os pontos, etc. Mas mesmo tomando todos os cuidados possíveis, sabe-se que o pesquisador das Ciências Naturais sempre encontrará algumas adversidades em campo, e por isso adaptações no esquema amostral quase sempre são necessárias.

Uma das formas usadas para aquisição de dados na água é a coletada de dados em fluxo contínuo em forma de transectos, deslocando o barco de modo a alcançar a maior densidade de elementos amostrais. Um dos problemas desse método são os espaços em branco deixados entre as curvas dos transectos, em que se podem perder informações relevantes sobre o comportamento da variável dentro da área de estudo. Para contornar esse problema é necessário coletar uma grande quantidade de dados, para que os espaços entre as amostras e entre as linhas dos transectos sejam pequenos e a perda de informações seja mínima, mas isso demanda tempo e pode onerar a coleta de dados.

Devido à complexidade da definição do esquema de amostragem, da coleta de dados, neste trabalho considerou-se uma amostra densa de dados, com mais de 900 observações adquiridas em transecto e coletadas por Cicerelli (2013), e a variável a ser analisada foi a clorofila – *a*. Algumas vezes, observações tomadas a uma distância muito pequena podem dificultar a análise, principalmente na análise espacial em que o efeito “pepita” pode indicar variação nos dados, que na realidade são aleatórias não controláveis. Por isso, resolveu-se realizar reduções no tamanho desse conjunto de dados, obtendo várias amostras através da amostragem sistemática, a fim de responder as seguintes: É possível utilizar apenas parte dessa amostra densa sem prejudicar a análise? E quanto é possível reduzir?

Para fazer afirmações sobre um universo, a partir de uma amostra, a inferência espacial foi aplicada através da técnica de Krigagem Ordinária, com a criação de uma função matemática que forneça informações sobre pontos não coletados referentes à concentração de clorofila-*a* em uma pequena área do Reservatório de Nova Avanhandava, possibilitando gerar uma representação da distribuição espacial dessa variável.

Com o objetivo de estudar os diferentes delineamentos amostrais propostos neste estudo com diferentes tamanhos de amostra, e com isso analisar qual a melhor disposição espacial das amostras para realizar a inferência por krigagem, a área original de estudo de Cicerelli (2013) foi reduzida e simulou-se através da krigagem ordinária a distribuição da concentração da clorofila-*a* nesta área, formada por 7384 células de 30 metros.

Da mesma forma em que é possível obter uma amostra densa da variável de interesse, como fez a Cicerelli (2013), em outras situações só é possível coletar dados em poucas localidades. Nessa última situação, tem-se o trabalho de Utsumi (2013) que coletou de forma aleatória cerca de 70 observações referentes à concentração de clorofila-*a* em um pequeno trecho do Reservatório de Nova Avanhandava. Com esses dados fez-se então uma análise variando-se o tamanho da amostra utilizada por Utsumi (2012) para obtenção da inferência e comparação dos resultados.

1.2 Objetivos

Resumindo o que já foi dito no tópico anterior, o objetivo geral desta pesquisa é avaliar a influência do delineamento amostral no mapeamento por geoestatística da variável clorofila-*a* em um trecho do Reservatório de Nova Avanhandava, a partir de dados fluorimétricos adquiridos em transecto.

Especificamente, pretende-se verificar a influência da redução do tamanho da amostra nas situações com dados reais e dados simulados, utilizando a krigagem ordinária.

1.3 Justificativa

A escolha de um delineamento amostral adequado é uma das etapas mais importantes da pesquisa, pois está cada vez mais comprovado que a metodologia adotada pelo pesquisador afeta diretamente os resultados a serem alcançados, por exemplo, uma amostra pouco representativa pode gerar um mapeamento limitado de uma variável, provendo conclusões equivocadas sobre a mesma.

Nos casos de dados espaciais, deve-se atentar que diferentemente de dados provenientes de populações independentes e identicamente distribuídas (i.i.d.), eles são correlacionados espacialmente, ou seja, violam a suposição de independência. Alguns cuidados devem ser tomados ao definir um método de coleta para esse tipo de dados, como: o comportamento (a variação) da variável na área de estudo, a época da coleta, os custos envolvidos (equipamentos, equipe, deslocamento e logística) e o difícil acesso ao ponto a ser amostrado.

Para propósitos gerais, Wang et al. (2012) comentam que têm sido usados com frequência, em aplicações práticas, os métodos de amostragem por gride, amostragem por transecto, amostragem sequencial e amostragem aninhada. Por outro lado, eles também comentam que os métodos de maior custo-benefício são: amostragem estratificada, amostragem por conglomerado, amostragem em dois estágios e amostragem sequencial.

Para Englund (1988), quando trabalhava no U. S. Environmental Protection Agency (EPA), a maioria dos documentos e guia, como SW-846¹ recomendava amostragem aleatória simples como uma solução geral. Mas, vários autores, entre eles Olea² (1984) apud Englund (1988) e Yamamoto e Landim (2013), mostraram que na presença de autocorrelação espacial, a amostragem sistemática sobre um gride produz resultados mais eficientes. Se não existir autocorrelação espacial, pode-se usar qualquer um dos métodos. O autor só alerta o uso de amostragem sistemática se o fenômeno a ser estudado apresentar periodicidade espacial na escala próxima ao gride processado. Englund (1988) observou que essa situação de periodicidade não é muito comum e o gride regular pode ser usado na maioria dos programas de amostragem espacial, por sua simplicidade e eficiência. Além dessas duas técnicas de amostragem aleatória, há também amostragem estratificada que considera o custo de obtenção dos dados.

¹ Publicação do EPA é um compêndio oficial com métodos analíticos e de amostragem para usar na regularização da Lei de Recuperação e Conservação de Recursos.

²OLEA, R. A. *Systematic Sampling of Spatial Functions*. Kansas Geological Survey, Lawrence, Kansas.

Analisar diferentes métodos de amostragem, assim como a influência das amostras nos resultados das análises espaciais, são temas relevantes. Assim, espera-se mapear a distribuição de clorofila-*a* na área de estudo através de amostras de diferentes tipos e tamanhos, e dessa forma avaliar qual a melhor disposição espacial das amostras para realizar a inferência por krigagem ordinária.

Segundo Thompson (1992), a krigagem tem capacidade de produzir melhores estimativas em termos de interpolação, porque está embasada em duas premissas: não-tendenciosidade do estimador e variância mínima das estimativas.

A avaliação da disposição dos elementos amostrais na área de estudo é muito importante quando se pretender inferir por krigagem ordinária, uma vez que a amostra deve ser capaz de captar a dependência espacial da variável.

Por fim, este trabalho pretende contribuir na discussão de: quanto é possível diminuir uma amostra em transecto sem afetar a qualidade dos resultados, qual o melhor delineamento amostral para aplicação da krigagem ordinária e como o tamanho da amostra pode influenciar nos resultados da inferência.

1.4 Estrutura do trabalho

No primeiro capítulo desta dissertação apresenta-se, de modo geral, uma introdução sobre o tema proposto seguida dos objetivos, da justificativa e da estrutura do trabalho.

O segundo capítulo apresenta a fundamentação teórica na qual são revisados alguns conceitos considerados fundamentais para o desenvolvimento do trabalho. Apresenta-se primeiramente um estudo sobre a geoestatística e sobre a técnica utilizada no trabalho, a krigagem ordinária. São apresentados também os planos de amostragem mais utilizados na estatística espacial, explicando em seguida sobre a definição do tamanho da amostra e descrevendo as três técnicas escolhidas para serem utilizadas neste trabalho: Amostragem Simples, Amostragem Estratificada e Amostragem Sistemática. E por último, as definições das técnicas que foram utilizadas para a validação dos resultados: Erro Médio Quadrático (EMQ) e Índice Kappa.

No terceiro capítulo é exposto o desenvolvimento metodológico do trabalho, iniciando com a descrição da variável estudada (clorofila-*a*), caracterização da área de

estudo e os modos de aquisição de dados espaciais. Em seguida são apresentados os materiais utilizados no desenvolvimento da pesquisa e um breve relato sobre utilização do *software* R para análises espaciais. O capítulo se encerra com a descrição da abordagem metodológica utilizada para atingir os objetivos da pesquisa.

O quarto capítulo se inicia com uma análise descritiva dos dados. Logo após são expostos os resultados obtidos com os dados em transecto, os dados simulados e os dados obtidos por amostragem pontual. São apresentados os processos de amostragem, os resultados dos processos de inferência espacial e as respectivas validações. Por fim, apresenta-se a discussão e análise dos resultados.

No capítulo 5 apresenta-se as conclusões referentes aos diversos resultados obtidos e são sugeridas propostas para futuros trabalhos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Geoestatística

2.1.1 Variáveis Regionalizadas

Os métodos geoestatísticos, ou simplesmente geoestatística, tiveram origem a partir de estudos desenvolvidos na França no final da década de 50 e início da década de 60, baseados em dados referentes às atividades mineradoras na África do Sul. As técnicas geoestatísticas podem ser usadas para descrever e modelar padrões espaciais (variografia), para estimar valores em locais não amostrados (krigagem), para obter a incerteza associada a um valor estimado em locais não amostrados e também para otimizar malhas de amostragem (ANDRIOTTI, 2003).

A geoestatística tem por objetivo a caracterização espacial de uma variável de interesse por meio do estudo da sua variabilidade espacial, permitindo o mapeamento, a quantificação e a modelagem dessa variável, através da interpolação dos pontos amostrados no espaço.

As técnicas da estatística clássica assumem que todas as amostras são aleatórias e independentes de uma distribuição de probabilidade simples, esta suposição é chamada estacionaridade. Sua aplicação não envolve qualquer conhecimento da posição atual das amostras ou do relacionamento entre amostras. Já a geoestatística assume que a distribuição das diferenças de variáveis entre dois pontos amostrados é a mesma para toda a área, e que isto depende somente da distância entre eles e da orientação dos pontos (CLARK, 1979 apud GENÚ, 2004, p.2).

Segundo Webster e Oliver (2007), a geoestatística permite a estimativa de valores em locais não amostrados, de modo que não haja tendência e com um erro mínimo. Assim, pode-se lidar com propriedades que variam de modo não sistemático e em diferentes escalas.

As vantagens da geoestatística sobre outras técnicas convencionais de predição são: o estudo da variabilidade espacial, a suavização, o desagrupamento (ou efeito de anular concentrações localizadas de observações), a determinação da anisotropia, a precisão e a estimativa do erro (ANDRIOTTI, 2003).

Segundo Landim (1998), a geoestatística é um ramo da estatística que trata de problemas referentes às variáveis regionalizadas (variáveis distribuídas no espaço ou tempo), que possuem características tanto de variáveis verdadeiramente casuais quanto totalmente determinísticas. A variável possui uma tendência de apresentar valores muito similares em dois pontos muito próximos e à medida que os pontos começam a se distanciar, os valores estimados se tornam cada vez mais distintos.

A *Teoria das Variáveis Regionalizadas*, desenvolvida por Matheron (1963, 1971), serviu de base para as suposições das técnicas geoestatísticas, já que a teoria trata de variáveis com condicionamento espacial. A teoria diz que uma medida pode ser vista como uma realização única de uma função aleatória. Esta teoria consolida o alicerce da geoestatística (CAMARGO et al., 2000).

De acordo com Olea (1975, apud Camargo, 1997), as principais características de uma variável regionalizada são: localização, anisotropia (caso onde as variáveis apresentam variações graduais em uma determinada direção e variações irregulares em outras) e continuidade.

Segundo Burrough (1986), a variação espacial de uma variável regionalizada pode ser expressa pela soma de três componentes: a) uma componente estrutural, associada a um valor médio constante ou a uma tendência constante; b) uma componente aleatória, espacialmente correlacionada; e c) um ruído aleatório ou erro residual.

Se x representa uma posição em uma, duas ou três dimensões, então o valor da variável Z em x é dada por:

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon''$$

em que:

- $m(x)$ é uma função determinística que descreve a componente estrutural de Z em x ;
- $\varepsilon'(x)$ é um termo estocástico, que varia localmente e depende espacialmente de $m(x)$;
- ε'' é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância σ^2 .

As Figura 1(a) e 1(b) ilustram as três componentes principais da variação espacial. A Figura 1(a) apresenta uma componente determinística que varia abruptamente, enquanto na Figura 1(b) a componente apresenta uma tendência constante.

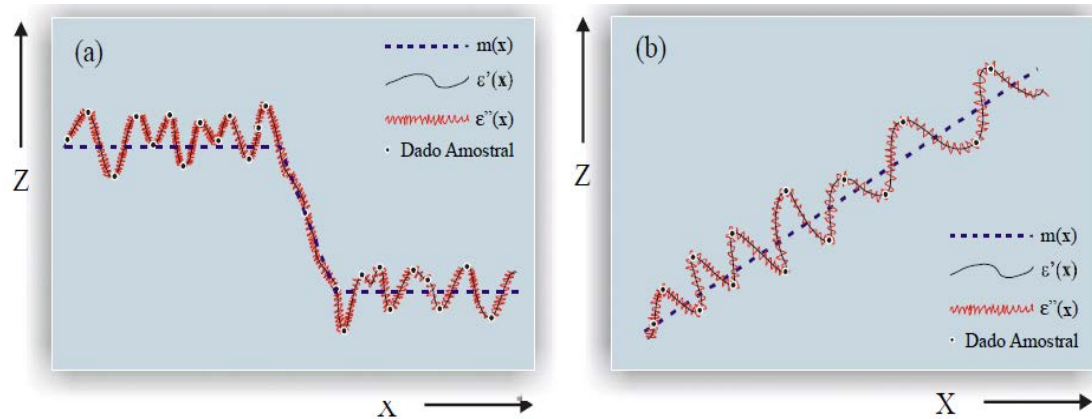


Figura 1(a) e 1(b) - Principais componentes da variação espacial.

FONTE: BURROUGH, 1987 apud CAMARGO et al., 2000, p.3.

2.1.2 Semivariogramas

O semivariograma é uma ferramenta gráfica utilizada para medir o grau de dependência espacial entre os elementos da amostra de uma variável, possibilitando medir a magnitude e a forma de tal dependência espacial.

O gráfico da covariância ($\gamma(h)$) versus a distância(h), ilustrado na Figura 2, é conhecido como semivariograma experimental. O seu padrão, como mostra a figura, representa o que se espera que aconteça quando se utiliza dados de campo, isto é, que as diferenças dos valores obtidos pela variável Z nos pontos x_i e x_{i+h} ($Z(x_i) - Z(x_{i+h})$) decresçam à medida que a distância h entre eles diminui. Por outro lado, espera-se que observações mais próximas geograficamente apresentem mais semelhanças entre si do que aquelas separadas por maiores distâncias. Portanto, é esperado que $\gamma(h)$ aumente com a distância h .

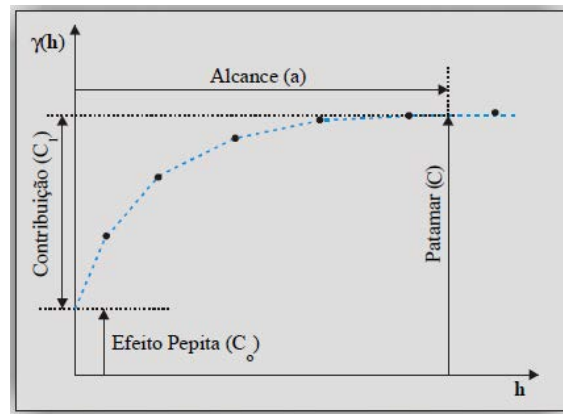


Figura 2 - Semivariograma Experimental.

Fonte: CAMARGO et al., 2000, p.3.

Os parâmetros do semivariograma são:

- Alcance (a): distância dentro da qual as amostras apresentam-se correlacionadas espacialmente;
- Patamar (C): valor do semivariograma correspondente ao seu alcance (a). Considera-se que a partir desse ponto a função se estabiliza e não existe mais dependência espacial entre as amostras;
- Efeito Pepita (C_0): este parâmetro refere-se à descontinuidade na origem do semivariograma. O efeito pepita é o valor da semivariância para a distância zero e representa a componente da variabilidade espacial que não pode ser correlacionada com uma causa específica;
- Contribuição (C_1): diferença entre o patamar (C) e o efeito pepita (C_0).

Em Geoestatística, diferente da Estatística, a covariância mede a relação entre valores da mesma variável, obtidos em pontos separados por uma distância h , conforme uma determinada direção. Isso significa que, ao alterar a direção, a covariância também pode se alterar e, nesse caso, há indicação de presença de fenômeno espacial anisotrópico. Existem casos em que a covariância é a mesma em qualquer direção e, por isso, o fenômeno espacial é isotrópico. Assim, para detectar se o fenômeno espacial apresenta anisotropia ou não, a covariância é calculada para várias direções, geralmente para 0° , 45° , 90° e 135° . (YAMAMOTO & LANDIM, 2013).

Após a construção dos semivariogramas experimentais em diferentes direções e do mapa variográfico é possível verificar se o fenômeno espacial é isotrópico ou não.

Quando os semivariogramas tiverem o mesmo comportamento, diz-se haver isotropia da variável, caso contrário tem-se anisotropia.

Existem dois tipos de anisotropia, a geométrica (Figura 3) e a zonal (Figura 4), ou quando ocorrem as duas juntas tem-se a anisotropia combinada. Na anisotropia geométrica os semivariogramas direcionais possuem a mesma forma e patamar, porém diferentes valores de alcance. Já na anisotropia zonal os semivariogramas apresentam patamares diferentes, mas sob um mesmo alcance.

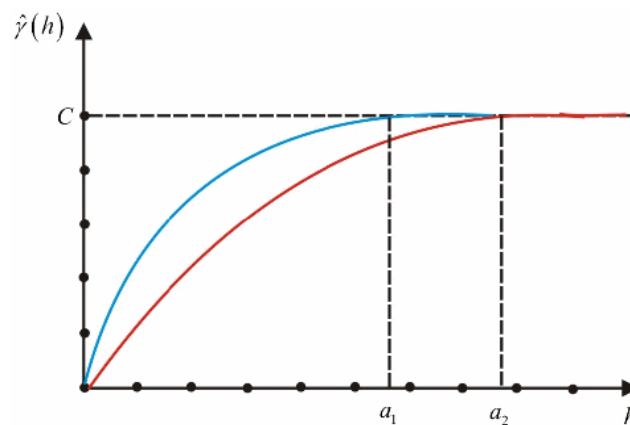


Figura 3 - Anisotropia Geométrica.
Fonte: ROSSONI (2010).

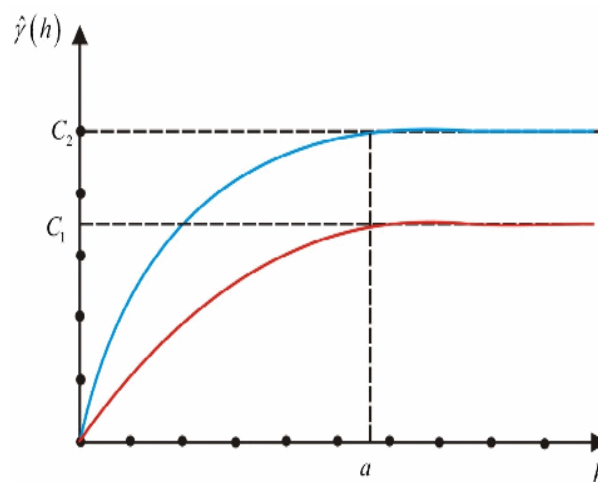


Figura 4- Anisotropia Zonal.
Fonte: ROSSONI (2010).

Se constatada a anisotropia deve-se selecionar quais são as direções de maior e menor alcance para que sejam modelados os semivariogramas relativos a essas duas direções por meio de ajuste de um modelo teórico nos mesmos.

2.1.3 Modelos Teóricos

Como foi dito anteriormente, se constatada a anisotropia deve-se ajustar o semivariograma experimental a um modelo teórico. Segundo Camargo (1996), o procedimento de ajuste não é direto e automático, como no caso de uma regressão, por exemplo, mas sim interativo, pois nesse processo o intérprete faz um primeiro ajuste e verifica a adequação do modelo teórico. Dependendo do ajuste obtido, pode ou não redefinir o modelo, até obter um que seja considerado satisfatório.

Embora existam diversos modelos de variogramas teóricos com patamar, alguns são mais comuns e podem explicar a maioria dos fenômenos espaciais. São eles: o exponencial, o esférico e o gaussiano. A Figura 5 apresenta os três modelos construídos com o mesmo alcance:

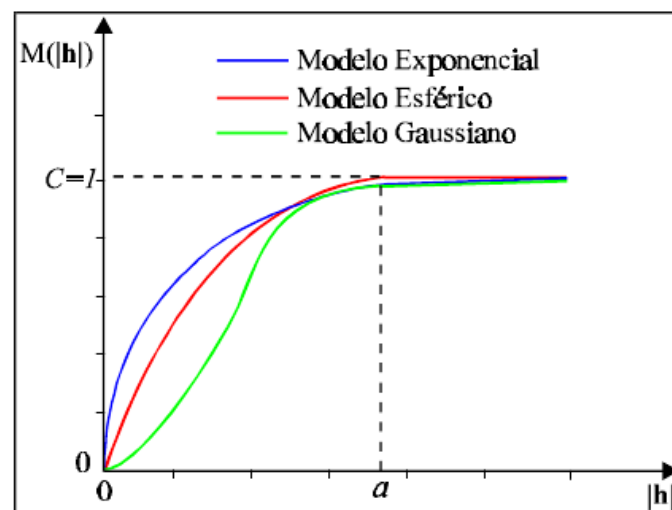


Figura 5- Modelos teóricos de semivariogramas.
Fonte: Adaptado de Isaaks e Srivastava (1989).

- **Modelo Esférico**

É um dos modelos mais utilizados, sua equação normalizada é:

$$sph(|h|) = \begin{cases} 0, & |h| = 0 \\ 1,5 \left(\frac{|h|}{a}\right) - 0,5 \left(\frac{|h|}{a}\right)^3, & 0 < |h| < a \\ 1, & |h| > a \end{cases} \quad (1)$$

- **Modelo Exponencial**

Também bastante utilizado, o modelo exponencial atinge o patamar assintoticamente, com o alcance prático definido como a distância na qual o valor do modelo é 95% do patamar (ISAAKS e SRIVASTAVA, 1989). A equação normalizada desse modelo é:

$$Exp(|h|) = \begin{cases} 0 & , |h| = 0 \\ 1 - \exp\left(-\frac{|h|}{a}\right) & , |h| \neq 0 \end{cases} \quad (2)$$

- **Modelo Gaussiano**

Usado principalmente para modelar fenômenos extremamente contínuos, o modelo gaussiano apresenta a seguinte fórmula:

$$Gau(|h|) = \begin{cases} 0 & , |h| = 0 \\ 1 - \exp\left(-\frac{|h|}{a}\right)^2 & , |h| \neq 0 \end{cases} \quad (3)$$

Semelhante no modelo exponencial, o modelo gaussiano atinge o patamar assintoticamente e o parâmetro a é definido como o alcance prático ou distância na qual o valor do modelo é 95% do patamar. O que caracteriza este modelo é seu comportamento parabólico próximo à origem (CAMARGO, 1996).

2.1.4 Inferência por Krigagem

As propriedades naturais da superfície terrestre são espacialmente contínuas, por isso encontram-se algumas restrições ao tentar descrevê-las por meio de simples funções matemáticas, são necessárias funções numéricas ordinárias que assumem um valor definido a cada ponto no espaço. A Krigagem é um dos modelos inferenciais que vem sendo propostos para esse objetivo, e sua base conceitual está fundamentada na teoria das variáveis regionalizadas, já citada anteriormente (CAMARGO, 1997).

Diferente de outras técnicas de inferência, a Krigagem pressupõe que o fenômeno estudado atende a hipótese de estacionaridade. Segundo Camargo (1997),

uma função aleatória $Z(x)$ é dita estacionária quando se admite que a componente determinística $m(x)$ é constante (não há tendências na região), consequentemente as variáveis aleatórias terão a mesma média m , independente da localização (estacionaridade de 1ª ordem). Assim, a diferença média entre os valores observados $Z(x)$ e $Z(x+h)$, separados por um vetor de distância h é nula.

$$E[Z(x)] = E[Z(x+h)] = m(x) = m \text{ e } E[Z(x)-Z(x+h)] = 0. \quad (4)$$

O fato de não rejeitar essa hipótese significa supor que a média das amostras possa representar a área estudada, isto é, que os valores são homogêneos. Essa condição raramente ocorre em um experimento, pois se deve considerar a variabilidade espacial da função aleatória.

Outra pressuposição da Krigagem Ordinária é a hipótese de estacionaridade de 2ª ordem, que consiste na estacionaridade da covariância, isto é, a covariância entre dois pares quaisquer $Z(x)$ e $Z(x+h)$ existe e depende somente de h (vetor de distância). Então:

$$C(h) = \text{Cov}[Z(x), Z(x+h)] = E[Z(x) \cdot Z(x+h)] - m^2, \quad \forall x; \quad (5)$$

Portanto, assumir a estacionaridade da covariância implica na estacionariedade da variância e também do variograma, definido por:

$$2\gamma(h) = E\{[Z(x)-Z(x+h)]^2\} \quad (6)$$

Porém, na geoestatística adota-se uma hipótese de estacionariedade menos restritiva denominada *Intrínseca*, que atende o critério de estacionaridade de 1ª ordem, em que $E[Z(x)] = m(x) = m$, $\forall x$, e determina a existência de uma variância finita das diferenças que não depende de x , mas somente do vetor distância h , isto é:

$$\text{Var}[Z(x) - Z(x+h)] = E\{[Z(x)-Z(x+h)]^2\} = 2\gamma(h), \quad (7)$$

sendo $2\gamma(h)$ o variograma, conforme definido anteriormente.

Deste modo, esta hipótese requer somente a estacionariedade do semivariograma, sem nenhuma restrição quanto à variância (GENÚ, 2004).

Se as condições especificadas pela hipótese intrínseca forem seguidas, pode-se obter a estimativa do semivariograma (semivariograma amostral) utilizando dados amostrais na função definida em (8).

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (8)$$

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

Com relação à Krigagem, a grande diferença entre ela e outros métodos de inferência é a maneira como os pesos são atribuídos às diferentes amostras. No caso da interpolação por média simples, por exemplo, os pesos são todos iguais a $1/n$ (n = número de amostras); na interpolação baseada no inverso do quadrado das distâncias, os pesos são definidos como o inverso do quadrado da distância que separa o valor interpolado dos valores observados. Na Krigagem, o procedimento é semelhante ao de interpolação por média móvel ponderada, exceto que aqui as informações são extraídas a partir do semivariograma para encontrar pesos ótimos a serem associados às amostras para a estimativa do valor em um determinado ponto (CÂMARA et al., 2004).

Neste trabalho será utilizada a Krigagem Ordinária, que ao contrário da Krigagem Simples, não exige o conhecimento da média. A Krigagem Ordinária é um método local de estimativa e, dessa forma, a estimativa em um ponto não amostrado resulta da combinação linear dos valores encontrados na vizinhança próxima (YAMAMOTO e LANDIM, 2013). O estimador da Krigagem Ordinária é:

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (9)$$

em que os λ_i são os pesos e $Z(x_i)$ são os valores de Z para $i = 1, 2, 3, \dots, n$.

Os pesos ótimos são calculados de modo que: 1ª) o estimador não seja tendencioso; 2ª) a variância de estimativa seja mínima. Para que a 1ª restrição seja atendida, a diferença entre o valor real e o valor estimado tem que ser igual a zero:

$$E[\hat{Z}(x_0) - Z(x_0)] = 0 \quad (10)$$

Desenvolvendo a equação (10), chega-se a condição de não viés fazendo

$$\sum_{i=1}^n \lambda_i = 1$$

A variância do erro de estimativa é dada por:

$$var [\hat{Z}(x_0)] = 2 \sum_{i=1}^n \lambda_i \gamma(x_i, x_0) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i, x_j) \quad (11)$$

em que $\gamma(x_i, x_j)$ é a semivariância de Z entre os dados x_i e x_j e $\gamma(x_i, x_0)$ é a semivariância entre o i -ésimo ponto e o ponto x_0 .

Minimizando a variância do erro, sob a condição de que a somatória dos pesos é 1, os pesos λ_i são obtidos a partir de um sistema de equações denominado de sistema de Krigagem Ordinária:

$$\sum_{j=1}^n \lambda_j C(x_i, x_j) - \alpha = C(x_i, x_0) \quad (12)$$

para $i = 1, 2, 3, \dots, n$. Sendo α o multiplicador de Lagrange necessário para a minimização da variância do erro.

O sistema de Krigagem Ordinária também pode ser escrito em forma de notação matricial:

$$K \cdot \lambda = k \rightarrow \lambda = K^{-1} \cdot k$$

ou,

$$\begin{pmatrix} C(x_1 - x_1) & C(x_1 - x_2) & \dots & C(x_1 - x_n) \\ C(x_2 - x_1) & C(x_2 - x_2) & \dots & C(x_2 - x_n) \\ \vdots & \vdots & & \vdots \\ C(x_n - x_1) & C(x_n - x_2) & \dots & C(x_n - x_n) \end{pmatrix} \cdot \begin{pmatrix} \lambda \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} C(x_1 - x_0) \\ C(x_2 - x_0) \\ \vdots \\ C(x_n - x_0) \end{pmatrix}$$

A seguir apresenta-se um exemplo de aplicação da krigagem ordinária (YAMAMOTO e LANDIM, 2013). A Figura 6 mostra, esquematicamente, a krigagem pontual para interpolação do teor na localização (X=28,75; Y=21,25). O modelo de semivariograma válido para esse conjunto é descrito pela Equação 13.

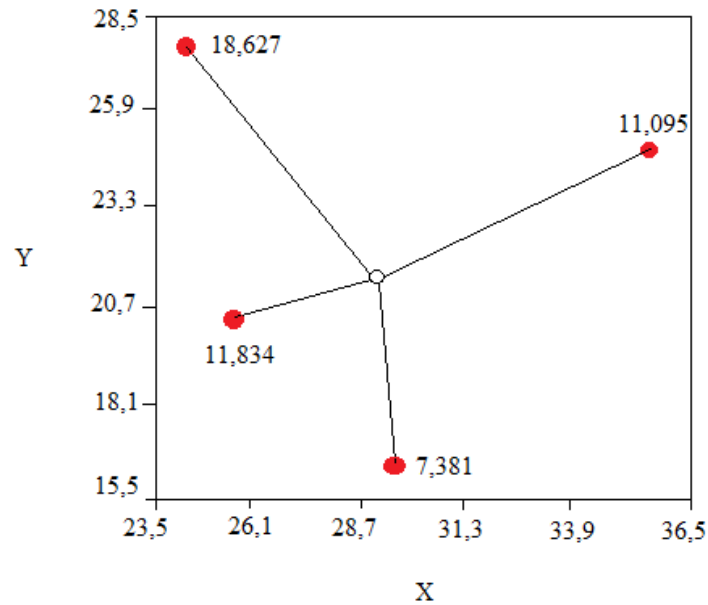


Figura 6 - Krigagem pontual para interpolação do ponto (X=28,7; Y=21,25).
Fonte: Adaptado de Yamamoto e Landim (2013).

$$\begin{cases} \gamma(h) = 19,8 \left(1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right) & \text{para } h < 14,16 \\ \gamma(h) = 19,8 & \text{para } h \geq 14,16 \end{cases} \quad (13)$$

Os vizinhos mais próximos ao ponto a ser interpolado encontram-se na Tabela 1.

Tabela 1 - Pontos de dados vizinhos para estimativa da localização (X=28,75; Y=21,25) por meio da krigagem ordinária pontual.

Ponto	X	Y	Valor
1	35,50	24,50	11,095
2	24,50	27,50	18,627
3	25,50	20,50	11,834
4	29,50	16,50	7,381

Para efetuar o cálculo da krigagem ordinária, monta-se o sistema de equação de krigagem, conforme:

$$\begin{pmatrix} 0 & 18,746 & 18,234 & 17,488 & 1 \\ 18,746 & 0 & 13,598 & 19,192 & 1 \\ 18,234 & 13,598 & 0 & 11,234 & 1 \\ 17,488 & 19,192 & 11,234 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \mu \end{pmatrix} = \begin{pmatrix} 14,247 \\ 14,347 \\ 6,867 \\ 9,699 \\ 1 \end{pmatrix}$$

Resolvendo o sistema, os ponderadores da krigagem ordinária são obtidos:

$$\lambda_1 = 0,18386; \lambda_2 = 0,08361; \lambda_3 = 0,47761; \lambda_4 = 0,25492; \mu = -0,48669$$

Aplicando-se os ponderadores obtidos na Equação 9, obtém-se a estimativa no ponto de coordenadas (X=28,75; Y=21,25):

$$\begin{aligned} Z(x_0) &= (0,18386 \cdot 11,095) + (0,08361 \cdot 18,627) + (0,47761 \cdot 11,834) \\ &+ (0,25492 \cdot 7,381) = 11,131 \end{aligned}$$

A variância da krigagem é igual a:

$$\begin{aligned} \sigma^2 &= \sum_{i=1}^n \lambda_i \gamma(x_0 - x_i) + \mu \\ &= (0,18386 \cdot 14,247) + (0,08361 \cdot 14,347) + (0,47761 \cdot 6,847) \\ &+ (0,25492 \cdot 9,699) + 0,48669 = 9,084 \end{aligned}$$

2.2 Amostragem

As vantagens dos métodos por amostragem em relação ao de contagem integral (censos) são, segundo Cochran (1977), custo reduzido, maior rapidez, maior amplitude e maior exatidão.

O primeiro a tentar fazer afirmações sobre uma população, usando informações sobre apenas uma parte dela, foi o comerciante inglês John Graunt (1620-1674). Ele propôs um método para estimar a população total de Londres com base em informações parciais. Desde então, a teoria da amostragem vem se desenvolvendo em ritmo acelerado (WANG et al., 2012).

Entre os vários potenciais usuários desses métodos, estão os pesquisadores que realizam experimentos que necessitam da localização geográfica na sua análise.

Um dos planos de amostragem muito utilizado pelos pesquisadores da análise espacial é a Amostragem Aleatória Simples, como cita Englund(1988). No Brasil, os pesquisadores Luiz e Epiphanyo (2001) utilizaram essa técnica para estimar área plantada por município por meio de amostragem por pontos em imagens de sensoriamento remoto e chegaram à conclusão de que o percentual da área de um município ocupado por uma determinada cultura pode ser estimado de maneira rápida pelo método de amostragem aleatória simples, ou seja, a técnica se mostrou uma ferramenta eficiente. Já no trabalho de Santos e Epiphanyo (2007), a proposta foi aprimorar o método de amostragem simples utilizado pelo Projeto Geosafras para estimativa municipal de área plantada com soja. O Projeto Geosafras, coordenado pela CONAB (Companhia Nacional de Abastecimento), vem desenvolvendo para alguns Estados brasileiros, com base em amostragem por município, um método de amostragem simples para estimativa de área agrícola de alguns cultivos. Os resultados do estudo foram considerados satisfatórios quando comparados às estimativas oficiais do IBGE. Porém, em algumas situações, verifica-se que a esse tipo de amostragem é a que oferece o pior resultado, podendo ter áreas não amostradas e áreas com pontos agrupados.

Encontram-se também em alguns artigos, com menos frequência, as técnicas de Amostragem Estratificada e Amostragem Sistemática. Em sua tese de doutorado, Luiz (2003) propôs definir um método simples e confiável, baseado no uso de imagens de sensoriamento e na amostragem estratificada, que permita a localização, a identificação e a quantificação das áreas ocupadas por uma determinada cultura agrícola em escala municipal. Como resultado, o uso da técnica, em conjunto com as imagens, os sistemas de informação geográfica (SIGs) e os aparelhos de posicionamento global (GPS), permitiu estimar a área plantada com as principais culturas de um Estado ou do País, de maneira rápida, relativamente econômica, extremamente simples, e com confiabilidade predeterminada. Já Pereira et al. (2007) aplicaram a técnica de Amostragem Sistemática para a realização de amostragem em corpos d'água. O objetivo do trabalho era definir um esquema de amostragem que considerasse um número suficiente de pontos, permitindo a otimização do levantamento de campo e garantindo, ao mesmo tempo, a representatividade espacial dos parâmetros coletados para a análise da qualidade da água nesses ambientes.

A aplicação da Amostragem Estratificada no espaço consiste em dividir a área de estudo em subáreas, de modo que essas subáreas sejam as mais homogêneas possíveis dentro delas e heterogêneas entre si. Diferente de um esquema aleatório simples, esse esquema de amostragem assegura que todas as subáreas que compõe o local em estudo sejam amostradas. Para que seja feita a subdivisão na área é importante levar em consideração informações prévias da região.

As duas formas de amostragem aleatória, simples ou estratificada, não levam em consideração a continuidade e correlação espacial entre os pontos. Por isso, quando se realiza uma distribuição aleatória dos pontos de amostragem, dois pontos podem ser localizados muito próximos gerando informação redundante e desperdiçando os recursos empregados.

Para contornar os problemas de uma amostragem aleatória é recomendável o emprego de um esquema com distribuição sistemática dos pontos de amostragem (malha de amostragem) que, além de evitar a coleta de amostras em pontos muito próximos, apresenta as mesmas vantagens da subdivisão da área. A Amostragem Sistemática é, sem dúvida, a que oferece o melhor resultado, porém nem sempre ela é possível, pois depende de uma série de fatores, tais como: acesso, acidentes geográficos (rios, lagos, topografia), vegetação, etc. (YAMAMOTO & LANDIM, 2013).

O sensoriamento remoto oferece um conjunto muito particular de dados, com características que precisam ser levadas em consideração na escolha dos métodos estatísticos a serem utilizados na sua análise (LUIZ, 2003).

Antes de iniciar-se qualquer programa de amostragem o pesquisador deve definir quais objetivos deverão ser alcançados, pois são os fatores determinantes no momento da escolha do método a ser executado (definição da posição dos pontos de amostragem, dos procedimentos de campo, dos métodos de conservação das amostras e das necessidades analíticas). De acordo com os objetivos estabelecidos, serão determinados o grau de precisão e o erro aceitável a ser adotado no programa de amostragem estabelecido para a área.

O objetivo primário ao estabelecer um plano de amostragem é promover um levantamento de dados o mais representativo possível da área avaliada, considerando-se um custo de investigação já fixado, ou se possível que seja minimizado. O segundo

objetivo é a adoção de um esquema de amostragem simples e eficiente, que facilite a análise dos dados e a sua implantação em campo.

Áreas contaminadas que possuem sérios impactos econômicos ou à saúde da população exposta requerem a seleção e aplicação de malhas de amostragem mais detalhadas e precisas. A escolha do esquema de amostragem dependerá das informações conseguidas durante a investigação preliminar.

2.2.1 Tamanho da Amostra

Quando o pesquisador procura tirar conclusões a respeito de um grande número de elementos, ele se depara com diversos contratempos como: tempo, equipe e recursos financeiros limitados. Portanto, o pesquisador trabalha somente com um pequeno grupo retirado da população, chamado amostra, e a partir dele faz afirmações sobre a amostra.

Sendo assim, depois de definidos os objetivos e a metodologia da pesquisa, algumas das perguntas feitas são: quantos elementos ou quantas observações da variável de interesse devem-se coletar da população? Qual é o número ideal? Este número chama-se tamanho da amostra e é representado por **n**.

Não há dúvida de que uma amostra não representa perfeitamente uma população. Ou seja, a utilização de uma amostra implica na aceitação de uma margem de erro que denominaremos erro amostral, que é a diferença entre um resultado amostral e o verdadeiro resultado populacional (NETO, 2004).

Não se pode evitar a ocorrência do erro amostral, porém pode-se limitar seu valor através da escolha de uma amostra de tamanho adequado. Obviamente, o erro amostral e o tamanho da amostra seguem sentidos contrários (Figura 7). Quanto maior o tamanho da amostra, menor o erro cometido e vice-versa (NETO, 2004).

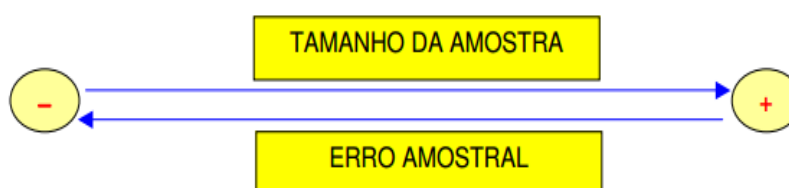


Figura 7 - Relação entre o tamanho da amostra e o erro amostral.
Fonte: NETO, 2004.

A determinação do tamanho de uma amostra é problema de grande importância, por que:

- amostras desnecessariamente grandes acarretam desperdício de tempo e de dinheiro;
- e amostras excessivamente pequenas podem levar a resultados não confiáveis.

Por isso, cabe ao pesquisador decidir qual o erro aceitável e o nível de confiança apropriado para sua pesquisa.

Principalmente em pesquisas de sensoriamento remoto na água, a escolha do tamanho da amostra é ainda mais complicada, pois outros fatores podem interferir na coleta: o difícil acesso a alguns pontos, as alterações climáticas, as dificuldades com o manuseio dos equipamentos, o tempo limitado, o meio de transporte para a coleta, etc. Muitas vezes, planeja-se um esquema amostral baseado em imagens da região, mas nem sempre se consegue cumpri-lo, pois a realidade encontrada pelo pesquisador é extremamente diferente. Nesses casos, o pesquisador tem que adaptar a amostragem às condições encontradas. Por isso, pode-se dizer que o melhor esquema de amostragem e o tamanho de amostra ideal será aquele que for viável ao pesquisador no momento da coleta.

Na próxima sessão, serão apresentados, de forma sucinta, os principais métodos de amostragem.

2.2.2 Tipos de Amostragem

Como citado anteriormente, os métodos de amostragem mais utilizados em análises estatísticas, com dados espaciais ou não, são: Amostragem Aleatória Simples (AAS), Amostragem Aleatória Estratificada (AE) e Amostragem Sistemática (AS).

- Amostragem Aleatória Simples

A amostragem aleatória simples (AAS) é o método mais simples e mais importante para a seleção de uma amostra. Além de servir como um plano próprio, o seu procedimento é usado de modo repetido em procedimentos de múltiplos estágios. Ele pode ser caracterizado através da definição operacional: “De uma lista com N

unidades elementares, sorteiam-se com igual probabilidade n unidades” (BOLFARINE e BUSSAB, 2005).

É um dos tipos de amostragem que utiliza técnica probabilística. Na AAS, uma amostra é escolhida de tal forma que cada item ou pessoa na população tem a mesma probabilidade de ser incluída, ou seja, se a população tem um tamanho N , cada pessoa desta população tem a mesma probabilidade igual a $1/N$ de entrar na amostra.

Segundo Bolfarine e Bussab (2005), a caracterização para o uso do plano AAS é a existência de um sistema de referências completo, descrevendo cada uma das unidades elementares. Deste modo, tem-se bem listrado o universo:

$$U = \{1, 2, \dots, N\},$$

em que N é o tamanho fixo e algumas vezes desconhecido da população.

O plano é descrito do seguinte modo:

- i) Utilizando-se um procedimento aleatório (tabela de números ou uma função randômica através de um software estatístico, como por exemplo, o R), sorteia-se com igual probabilidade um elemento da população U .
- ii) Repete-se o processo anterior até que sejam sorteadas n unidades, tendo sido este número prefixado anteriormente;
- iii) Caso seja permitido o sorteio de uma unidade mais de uma vez, tem-se o processo AAS com reposição. Quando o elemento sorteado é removido de U antes do sorteio do próximo, tem-se o plano AAS sem reposição. O primeiro procedimento, também conhecido como AAS irrestrito, será indicado por AASc, enquanto que o segundo, conhecido como AAS restrito, será designado por AASs.

Do ponto de vista prático o plano ASSs é muito mais interessante, pois satisfaz o princípio intuitivo de que “não se ganha mais informação se uma mesma unidade aparece mais de uma vez na amostra”. Por outro lado, o plano ASSc introduz vantagens matemáticas e estatísticas, como a independência entre as unidades sorteadas, que facilita em muito a determinação das propriedades dos estimadores das quantidades populacionais de interesse.

- Amostragem Aleatória Estratificada

Uma das técnicas mais utilizadas é a Amostragem Aleatória Estratificada, que usa informação existente sobre a população para dividi-la em grupos bem definidos, chamados estratos. De cada um desses estratos, é selecionada uma amostra mediante um processo aleatório simples. A lógica que assiste à estratificação de uma população é a de identificação de grupos que variam muito entre si no que diz respeito ao parâmetro em estudo, mas muito pouco dentro de si, ou seja, cada um é homogêneo e com pouca variabilidade.

Este método de amostragem estratificada tem a vantagem de fornecer resultados com menor probabilidade de erro associada. Porém, ele apresenta problemas na distribuição espacial dos pontos de amostragem, pois nem sempre se consegue subdividir a área em estratos bem distintos, é comum que os estratos fiquem sobrepostos entre si.

Segundo Bolfarine e Bussab (2005), a execução de um plano de amostragem estratificada (AE) exige os seguintes passos:

- i) Divisão da população em subpopulações bem definidas (estratos);
- ii) De cada estrato retira-se uma amostra, usualmente independente;
- iii) Em cada amostra, usam-se estimadores convenientes para os parâmetros do estrato;
- iv) Monta-se para a população um estimador combinando os estimadores de cada estrato, e determinam-se suas propriedades.

- Amostragem Sistemática

Por último tem-se a Amostragem Sistemática, bastante utilizada em aplicações no meio aquático. É uma variação da amostragem aleatória, sua aplicação requer que a população seja ordenada de modo tal que cada um de seus elementos possa ser identificado pela sua posição. Ou seja, os elementos da população estão apresentados em uma lista, dispostos em alguma ordem (cronológica, alfabética, etc.).

Considere uma população com N elementos, sendo $N=k.n$ e k um número inteiro, tal que $k= N/n$. Seleciona-se então aleatoriamente um número, simbolizado por r , entre

1 e k . A unidade da população referente a esse número será a primeira unidade da amostra, os outros elementos seguintes serão obtidos a partir dessa primeira unidade, selecionadas em intervalos de comprimento k . Portanto, a amostra ficará da seguinte forma:

$$(r, r+k, r+2k, \dots, r+(i-1)k), i=1, \dots, n.$$

Por exemplo, se o tamanho da população for $N=1600$ e o tamanho da amostra for $n=100$, tem-se que $k=16$. Sorteia-se então um número entre 1 e 16 (de forma aleatória), que será o primeiro número da amostra, logo as próximas unidades amostrais serão retiradas de 16 em 16, até obter as 100 unidades. Supondo que o primeiro número sorteado foi 10, a amostra ficaria da seguinte maneira:

$$10, 26, 42, 58, 74, \dots$$

Algumas de suas vantagens são:

- Facilidade na determinação dos elementos da amostra;
- Não é necessário usar números aleatórios;
- Mais rapidez para grandes populações.

Por outro lado, um problema encontrado na aplicação desse método é a periodicidade, isto é, dependendo do intervalo escolhido, os pontos podem ser distribuídos exatamente onde ocorre de forma regular alguma característica do habitat ou então deixar de fora da amostra algum padrão da variável em estudo, gerando uma amostra pouco representativa do local.

2.3 Variável de estudo: clorofila-a

Segundo Arraut et al. (2005), fitoplâncton é um grupo heterogêneo formado principalmente por algas fotossintetizantes que se distribui por grande parte dos ambientes marinhos e aquáticos continentais do mundo. Sua função nesses ambientes é captar a energia eletromagnética e transformá-la em energia química, por meio da fotossíntese.

Sabe-se que clorofila-*a* é o principal pigmento na realização da fotossíntese de organismos vivos, estando presente em todas as algas, ou seja, em todo o fitoplâncton. O seu papel é absorver radiação eletromagnética e transformá-la em energia química e por esse motivo, sua detecção é de fundamental importância para verificar a qualidade da água (HALL & RAO, 1980). De acordo com a CETESB (2004), a clorofila-*a* é a mais comum das clorofilas e representa, aproximadamente, de 1 a 2% do peso seco do material orgânico em todas as algas planctônicas, por esse motivo ela também é considerada uma indicadora da biomassa fitoplanctônica e do estado trófico dos ambientes aquáticos.

De acordo com a Cetesb (2013), o valor de concentração de clorofila-*a* considerado alarmante é de 50 µg/L, indicando floração de cianobactérias nocivas. Florações de cianobactérias comprometem os usos múltiplos dos ambientes aquáticos, implicando em riscos à saúde humana e ao ecossistema aquático.

2.4 Validação da Inferência Espacial

No estudo das técnicas de previsão as medidas de precisão são uma aplicação de extrema importância. Os valores futuros das variáveis tornam-se bastante difíceis de prever e quando previstos raramente são iguais aos valores reais, dada a complexidade da grande maioria dessas variáveis na vida real e às variações aleatórias que caracterizam as variáveis. Por isso, é fundamental obter informações acerca da medida em que a previsão pode desviar-se do valor real da variável. Este conhecimento adicional fornece uma melhor percepção sobre precisão da previsão.

Existem diversas técnicas disponíveis, umas com maior precisão do que outras, por isso, fica a critério do responsável escolher a medida de precisão que seja mais adequada ao seu estudo.

Neste trabalho, para validar os resultados obtidos através da inferência espacial da clorofila-*a* utilizou-se dois índices: o Erro Quadrático Médio (EQM) e o Índice de Concordância Kappa.

2.4.1 Erro Quadrático Médio (EQM)

O erro quadrático médio (EQM) é uma das medidas mais comuns de erro de previsão, erro que consiste na diferença entre o valor real e a previsão do valor:

$$e_i = Z_i - \hat{Z}_i$$

sendo:

e_i = Erro da i-ésima observação.

Z_i = Valor real da i-ésima observação.

$\hat{Z}_i$ = Previsão para a i-ésima observação.

As decisões podem ser influenciadas de duas formas distintas pelos erros de previsão. A primeira consiste na escolha entre alternativas de previsão.

O EQM é determinado somando os erros de previsão ao quadrado e dividindo pelo número de erros usados no cálculo. O erro quadrático médio pode ser expresso pela seguinte equação:

$$EQM = \frac{\sum_{t=1}^n e_t^2}{n}$$

2.4.2 Índice de Concordância Kappa

Segundo Kotz et al., (1983), o Coeficiente Kappa pode ser definido como uma medida de associação usada para descrever e testar o grau de concordância (confiabilidade e precisão) entre dois ou mais métodos de classificação. Ela é baseada no número de respostas concordantes, ou seja, no número de casos cujo resultado é o mesmo entre os métodos. Para isso, é necessário que elas sejam classificadas por grupos, definidos pelo próprio pesquisador de acordo com as especificações do seu conjunto de dados.

É uma importante medida para determinar o quão bem funciona uma aplicação de alguma medição ou de algum método de predição.

Esta medida de concordância varia de -1 a 1, em que 1 representa total concordância, ou seja, os métodos classificam exatamente da mesma forma as observações e zero indica nenhuma concordância, ou melhor, que não existe nenhuma relação entre as classificações dos dois métodos. E um eventual valor de Kappa menor que zero, negativo, sugere que os métodos classificaram exatamente o oposto. Sugere, portanto, discordância, mas seu valor não tem interpretação como intensidade de discordância. Landis e Koch (1977) sugeriram a seguinte análise:

Tabela 2– Interpretação dos valores de kappa.

Valores de Kappa	Interpretação
<0	Sem concordância
0 – 0,19	Concordância pobre
0,20 – 0,39	Concordância considerável
0,40 – 0,59	Concordância Moderada
0,60 – 0,79	Concordância Satisfatória
0,80 - 1	Concordância Excelente

O cálculo do índice Kappa se inicia a partir de uma matriz de confusão. Para ilustrar, considere-se um exemplo simples no qual uma população de “n” indivíduos é classificada por dois médicos como “doentes” ou “não doentes”:

		Médico 1		
		“Não doente”	“Doente”	Total
Médico 2	“Não doente”	a	b	a + b
	“Doente”	c	d	c + d
	Total	a + c	b + d	n

Figura 8- Matriz de confusão.

Fonte: PINTO et al., 2014.

Os passos a serem seguidos são:

- Primeiro, calcula-se a concordância observada, indicada por θ_1 :

$$\theta_1 = \frac{a + d}{n}$$

- Em segundo lugar, calcula-se um índice que represente a porcentagem de concordância esperada pelo acaso, simbolizada por θ_2 :

$$\theta_2 = \frac{((a + c) * (a + b)) + ((b + d) * (c + d))}{n^2}$$

- Obtidos estes dois índices, a estatística k (resultado do índice kappa) será calculada através da seguinte fórmula:

$$\hat{k} = \frac{\theta_1 - \theta_2}{1 - \theta_2}$$

3 MATERIAIS E MÉTODO

3.1 Área de estudo

A área de estudo contempla um trecho do reservatório da usina hidrelétrica de Nova Avanhandava, localizado no Baixo Tiête, mais precisamente no município de Buritama-SP. A usina entrou em operação em 1982, com área de cobertura do reservatório de 210 km² e volume de 2.830 x 10⁶ m³ (AES TIETÊ, 2013). A seleção do ambiente aquático se deu através de estudos preliminares realizados por Cicerelli (2013), onde foram detectados indicadores da ocorrência de cianobactérias nesta região. A Figura 10 apresenta uma imagem LANDSAT da localização do reservatório.

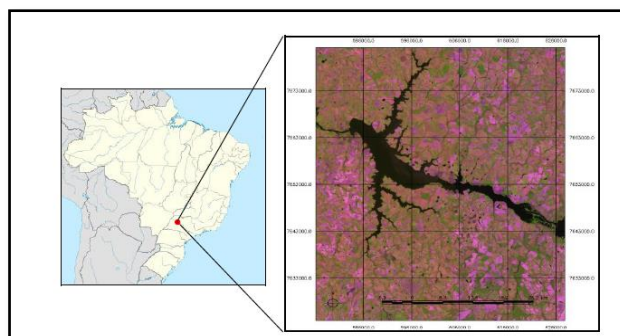


Figura 9 - Ilustração do Reservatório de Nova Avanhandava.
Fonte: Cicerelli et al., 2013.

Como o reservatório possui uma área muito extensa, foi necessário delimitar uma subárea para realizar os estudos. Sendo assim, Cicerelli realizou uma visita preliminar ao reservatório a fim de observar áreas com maiores concentrações de fitoplâncton. Após as observações, Cicerelli (2013) selecionou uma área de estudo irregular de aproximadamente 8,4 km², extensão de 4,5 km e largura de 2,2 km, como mostra a Figura 10. Porém, como a amostragem realizada se concentrou apenas no centro dessa região devido à forte chuva, achou-se necessário realizar uma nova delimitação nessa subárea para diminuir a área não amostrada, a fim de garantir melhores resultados na inferência espacial da clorofila-a e diminuir a variância dos resultados. Com isso, na área final as coordenadas no eixo x (E) variam

aproximadamente entre 603.450 m e 606.558 e no eixo y entre 7.651.090 m e 7653197 m, na projeção UTM de acordo com o sistema de referência SIRGAS 2000.

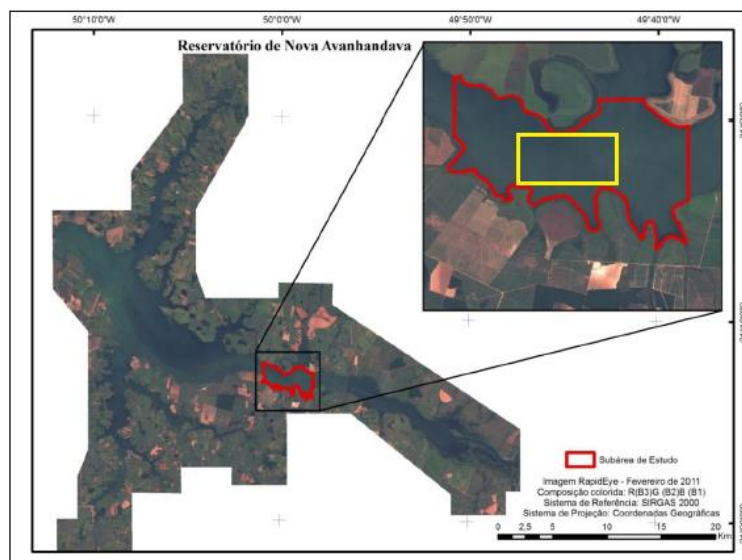


Figura 10 - Ilustração da subárea de estudo.
Fonte: Cicerelli et al., 2013.

Além disso, para análises complementares utilizou-se também um conjunto de dados pontuais de concentração de clorofila-*a* coletados por Utsumi (2012) em 59 pontos em um trecho do reservatório de Nova Avanhandava (Figura 11).

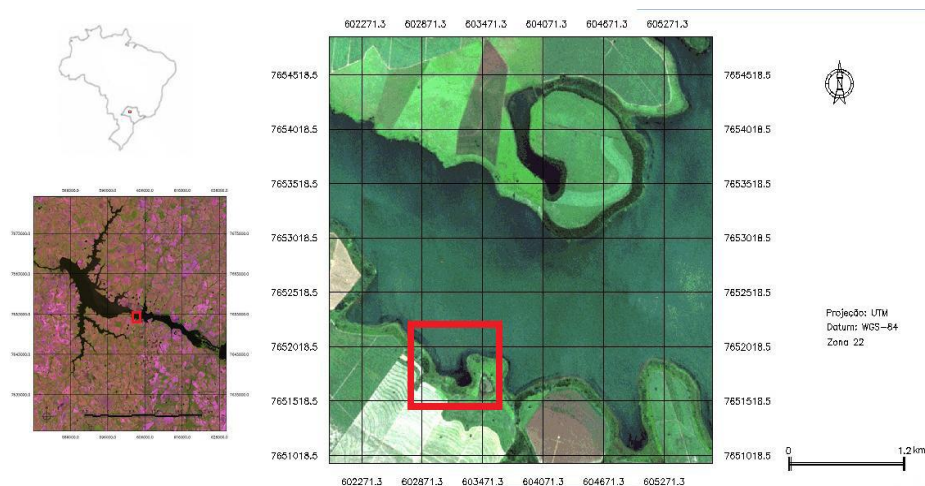


Figura 11- Área de estudo: trecho do reservatório de Nova Avanhandava.
Fonte: UTSUMI, 2012.

3.2 Materiais

Os materiais utilizados nesse estudo foram:

- Notebook DELL Intel Core i3;
- *Software* Excel;
- Aplicativo VARIOWIN;
- *Software* R.

O VARIOWIN® é um conjunto de programas desenvolvido por Yvan Pannatier para análise e modelamento de variogramas em 2 D, em ambiente Windows. Ele adota o mesmo formato de entrada de dados do Geo-EAS®.

O VARIOWIN® é composto por quatro programas (arquivos ***.exe**) executados isoladamente. São eles: **Prevar2D**, utilitário que constrói uma matriz de distâncias para todos os possíveis pares de dados existentes no arquivo ***.dat**; **Vario2D**, utilitário que usa o arquivo ***.pcf**, originado pelo Prevar2D, para fazer a variografia exploratória em 2D; **Model**, utilitário que executa, de modo interativo, o ajuste a um modelo teórico do variograma experimental obtido pelo Vario2D e faz a modelagem interativa para anisotropia geométrica e zonal e **Gdisplay**, utilitário que exibe um arquivo, com dados dispostos num arranjo reticulado (***.grd**), na forma de um mapa em pixels.

Atualmente existe uma seleção extensa de *softwares* capazes de realizarem análise geoestatística, alguns deles comerciais e outros livres. O *R* é um programa de linguagem aberta e gratuita conhecido no meio estatístico por sua capacidade de analisar e modelar conjuntos de dados, além de fornecer recursos para geração de gráficos de alto nível.

O Projeto R começou em 1995 com os trabalhos de Robert Gentleman e Ross Ihaka, do Departamento de Estatística da Universidade de Auckland. Atualmente o programa é mantido pela R Core-development Team, que é um time internacional de desenvolvedores voluntários que continuamente desenvolvem tutoriais e pacotes funcionais que acrescentam muitas potencialidades à versão base do R.

O *R* usa uma interface de texto, ou seja, o pesquisador necessita conhecer pelo menos os princípios básicos da técnica de programação para realizar os trabalhos. Mas a

boa notícia é que facilmente são encontrados na internet documentos de ajuda contendo as bases dos principais programas.

Segundo Pova et al. (2011), a análise de dados espaciais utilizando o *R* também é muito difundida e tem ganhado um grande número de recursos nos últimos anos. Especificamente em geoestatística, rotinas até então disponíveis em softwares comerciais e de alto valor no mercado são disponibilizadas em diversas bibliotecas. Além disso, o *R* dispõe de funções especiais para exibir, armazenar e organizar dados desse tipo. Contudo, a linguagem *R* por si só não possui artifícios para distinguir coordenadas de qualquer outro tipo de dado numérico. Uma equipe de pesquisadores desenvolveu um pacote denominado *SP* com classes e métodos específicos para dados espaciais, sendo esta biblioteca a matéria-prima de ferramentas em *R* para processamento e análise de dados dessa grandeza. Alguns dos pacotes disponíveis, que dependem ou importam o pacote *SP* direta ou indiretamente, são: *maptools*, *rgdal*, *splancs*, *geoR*, *gstat*, *spsurvey*, *trip*, entre muitos outros.

Mesmo com tantos recursos disponíveis para análises espaciais, o *software R* ainda é pouco explorado por pesquisadores da área de Geoestatística, eles preferem os *softwares* desenvolvidos especificamente para esta finalidade, como o *SPRING* e o *Variowin*, que não necessitam de programação (mais comum na área da Estatística). Sendo assim, a escolha desse programa como ferramenta para este trabalho foi proposital, para mostrar que é o *R* também pode ser uma opção de *software* para análises espaciais e assim encorajar a sua utilização em pesquisas da área.

- Conjunto de dados

Para obter os resultados desejados neste trabalho, utilizou-se um conjunto de dados de concentração de clorofila-*a*, coletados no Reservatório de Nova Avanhandava durante um levantamento de campo realizado no dia 19 de Setembro de 2012, pela Rejane Ennes Cicerelli do Programa de Pós-Graduação em Ciências Cartográficas para sua tese de doutorado. O processo de coleta foi realizado de forma contínua, em transectos, tentando garantir uma boa representação da área. No total, foram obtidas as concentrações de clorofila-*a* relativa em 978 pontos. (CICERELLI, 2013).

3.4 Desenvolvimento Metodológico

Para alcançar os objetivos propostos no trabalho, fez-se o seguinte desenvolvimento metodológico apresentado no fluxograma da Figura 12; em seguida, são descritas as etapas com mais detalhes.

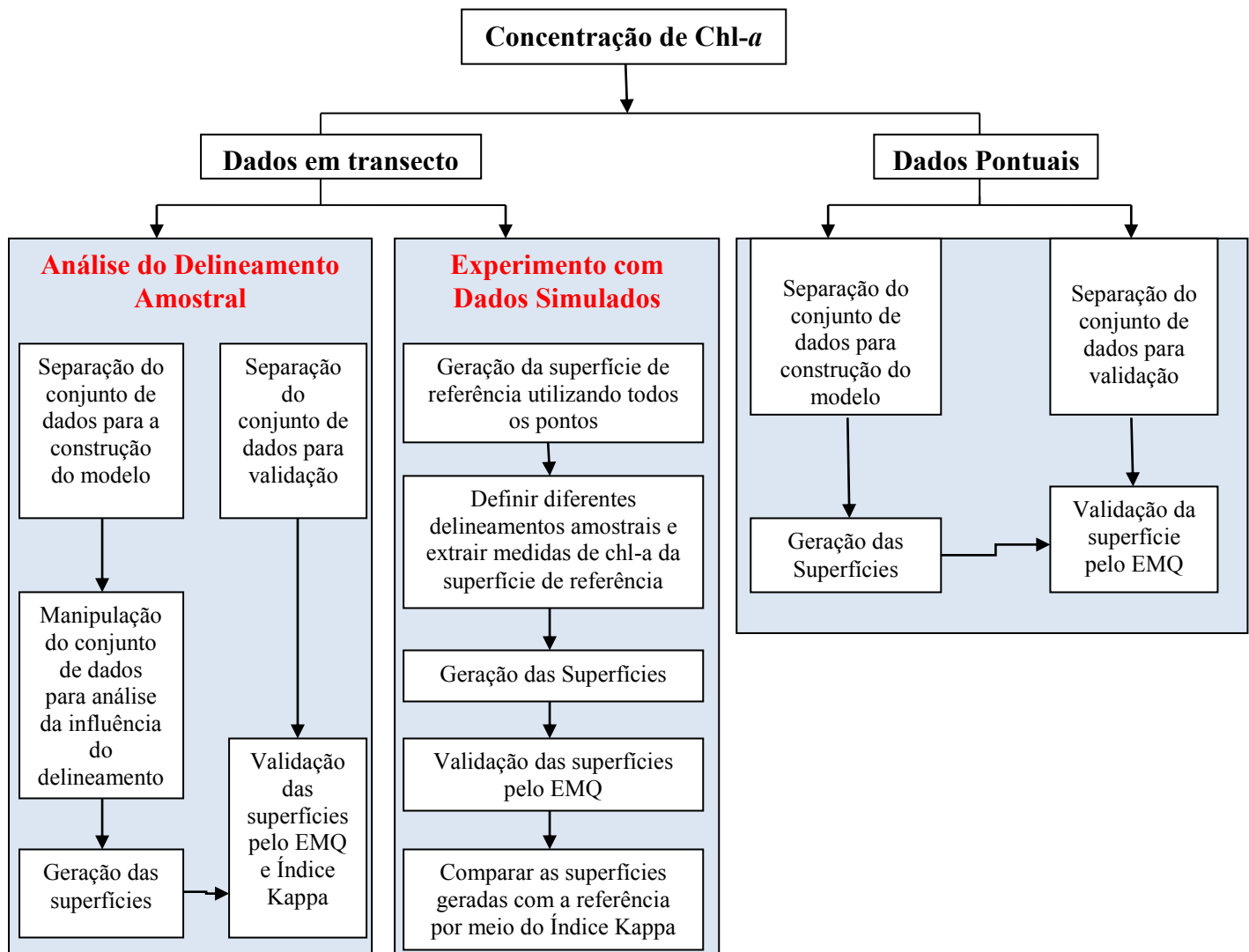


Figura 12– Fluxograma do desenvolvimento metodológico.

3.4.1 Dados adquiridos em Transecto

O método de coleta em transecto consiste em amostrar dados em fluxo contínuo. Os transectos utilizados devem ser distribuídos o melhor possível dentro da área de estudo, de forma que toda a área ou grande parte dela seja amostrada.

Transecto tanto pode ser contínuo, com todo o comprimento da linha sendo amostrado, ou as amostras podem ser tomadas em determinados pontos ao longo da linha. Por exemplo, a cada metro, ou qualquer outro medidor. No caso deste trabalho, as medidas foram pontuais, tomadas em intervalos irregulares. Para ambas as abordagens, o intervalo no qual as amostras serão tomadas dependerá do habitat em particular, bem como do esforço e do tempo que pode ser atribuído ao levantamento.

Esse método possui baixo custo operacional e permite a detecção de um grande número de medidas, porém quando essas medidas são tomadas em intervalos muito grandes pode significar que muitas informações relevantes atualmente presentes não serão notadas, assim como obscurecendo padrões de zoneamento por falta de observações. Já quando os intervalos são pequenos demais, pode consumir muito tempo com a amostragem, bem como dar origem a mais dados do que é necessário. Isso pode causar problemas com a apresentação dos dados (transecto de linha).

É importante certificar-se de que o intervalo escolhido não coincida com alguma característica que ocorre regularmente no habitat ou com algum padrão da variável em estudo, senão a amostra pode ficar viciada, como ilustra a Figura 13. Os pontos vermelhos são os pontos amostrados em transecto e a região amarela é onde ocorrem as maiores concentrações da variável em questão, portanto percebe-se que a amostragem não foi satisfatória ao tentar representar a distribuição espacial dessa variável, pois todos os pontos foram coletados onde havia pequenas concentrações da variável. E isso pode resultar conclusões equivocadas sobre a região.

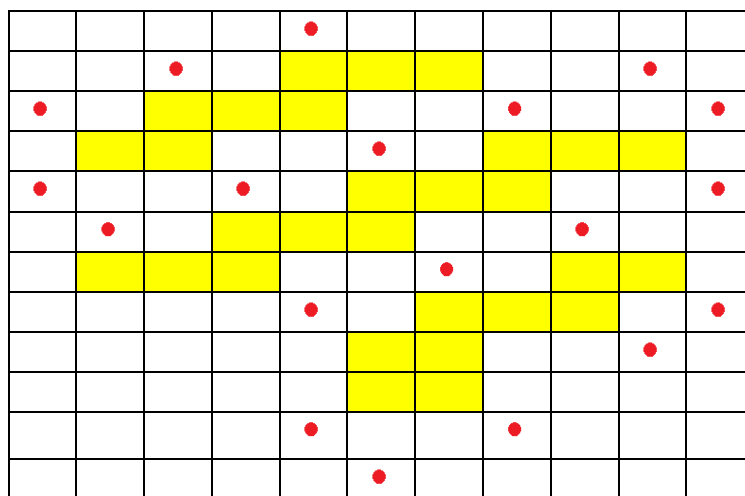


Figura 13 – Amostragem em forma de transecto.

O intervalo ideal será escolhido de forma a obter o equilíbrio entre a complexidade do habitat individual e a finalidade da pesquisa, considerando os recursos disponíveis para realizá-lo. No caso de sensoriamento remoto na água, deve-se ter cuidado também com o clima e as condições da área de estudo na época da coleta, ambos devem estar favoráveis.

3.4.1.1 Análise do Delineamento Amostral

Os dados reais, no total de 978, foram obtidos em forma de transecto, como mostra a Figura 14.

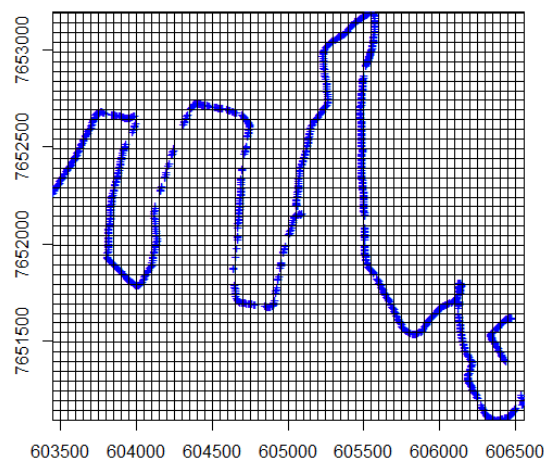


Figura 14 - Distribuição espacial do conjunto de dados.

Primeiramente separou-se aleatoriamente uma amostra, equivalente à aproximadamente 20% do conjunto total, ou seja, 200 elementos amostrais. Esse conjunto foi usado posteriormente para a validação do processo de inferência da clorofila-a.

Com o conjunto restante, de 778 elementos amostrais, iniciou-se o processo de amostragem sistemática. O intuito era reduzir ao máximo esse conjunto de dados através do processo de amostragem, sem que o conjunto final perdesse o formato característico do transecto do conjunto inicial e sem que a qualidade da inferência fosse afetada. Isso só foi possível por que os elementos amostrais foram enumerados de acordo com o desenrolar do transecto. Por isso a escolha de um método que segue uma ordem, selecionando pontos em iguais intervalos, a partir do primeiro elemento

selecionado aleatoriamente. A partir da redução máxima, seria possível comprovar que um conjunto bem menor de dados, teria sido suficiente para atingir os objetivos desejados e ainda haveria uma considerável redução nos custos, no tempo e na equipe.

Foram realizadas quatro reduções no conjunto de dados, ou seja, quatro amostragens sistemáticas, com os respectivos tamanhos: 400, 300, 200 e 100 observações. Feito isso, foi analisada a distribuição espacial de cada amostra e a partir delas realizou-se o processo de inferência através da Krigagem Ordinária. Com isso, foi possível o mapeamento da clorofila-a na área selecionada.

A validação dos processos de inferência espacial foi realizada de duas maneiras: comparando pontualmente os dados interpolados com os reais através do cálculo do Erro Médio Quadrático (EMQ) e através do cálculo do Índice Kappa, que comparou as classificações das previsões obtidas da krigagem utilizando todos os 978 dados com as classificações das previsões obtidas das krigagens com as amostras reduzidas. Além disso, verificou-se se uma amostra menor proporcionaria um melhor resultado.

3.4.1.2 Experimento dos Dados Simulados

Nesta etapa do trabalho, o intuito era observar o que aconteceria se ao invés de dados em forma de transecto, tivesse uma quantidade densa de dados com os quais fosse possível realizar diversos planejamentos amostrais, como numa situação real. Para isso, gerou-se uma superfície de referência a partir da Krigagem Ordinária utilizando todos os 978 pontos e consideraram-se os valores matriciais resultantes como se fosse uma “verdade terrestre”, ou seja, como se as 7384 células da matriz de concentração de clorofila-a fossem todos os pontos disponíveis para amostragem na área de estudo. Com isso, foi possível testar diferentes delineamentos amostrais.

Assim como se fez com os dados em transecto, também foi separada uma parte dos dados para realizar a validação do processo de inferência espacial. Neste caso, foram selecionadas aleatoriamente 1.500 células, o equivalente à aproximadamente 20% dos dados.

A partir dos 5884 dados que restaram, aplicaram-se os três tipos de amostragem estudados neste trabalho: Amostragem Aleatória Simples, Amostragem Aleatória

Estratificada e Amostragem Sistemática. Nos métodos de amostragem foram consideradas amostras de tamanhos 1000, 400, 300, 200 e 100.

No caso da Amostragem Estratificada, os estratos foram definidos a partir do mapa resultante da Krigagem Ordinária utilizando todos os pontos. Com isso, definiram-se três estratos: o primeiro com valores baixos (região verde), o segundo com valores medianos (região amarela) e o terceiro com os valores mais altos (região vermelha). Com isso, através do *software* R e com o auxílio da imagem, foi possível delimitar a área de cada estrato e assim definir em quais estratos os pontos cairiam de acordo com suas localizações geográficas. Segue abaixo na Figura 15 como ficaram definidos os estratos.

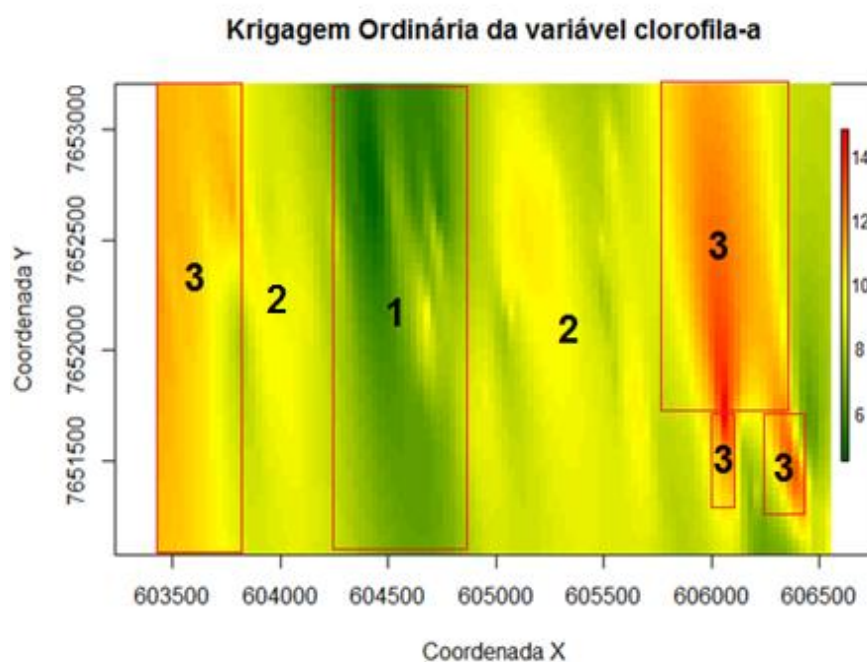


Figura 15 – Estratos da amostragem estratificada.

A partir da definição dos estratos, a amostragem foi realizada proporcionalmente à quantidade de observações dentro de cada subregião. Por exemplo, se amostra a ser coletada fosse de tamanho 100 e as subregiões 1, 2 e 3 concentrassem respectivamente 20%, 30% e 50% dos dados, então seriam coletadas 20 observações do estrato 1, 30 do estrato 2 e 50 do estrato 3.

Após a seleção dos elementos da amostra, realizou-se o processo de inferência espacial da clorofila-a a partir da Krigagem Ordinária. Os resultados foram validados novamente a partir do EMQ e do Índice Kappa, que permitiram comparar os métodos de amostragem. Também foi possível analisar o efeito causado nos resultados da inferência ao adotar amostras de tamanhos diferentes.

3.4.2 Dados Pontuais

Há situações em que o delineamento amostral é desenvolvido com auxílio de outros recursos, antes de ir propriamente ao campo, como fez Utsumi (2012). Ele definiu os pontos a serem amostrados a partir de imagens de satélites, que permitiram detectar a presença de pigmentos fitoplanctônicos. Pelas imagens Landsat, Utsumi (2012) definiu 70 pontos de coleta para aquisição dos dados fluorescência *in vivo* de clorofila-a, também em um trecho do Reservatório de Nova Avanhandava.

Embora desejasse coletar uma grande quantidade de dados, por razões de tempo, custo e logística, Utsumi (2012) realizou coleta de dados em 70 pontos distribuídos na sua área de estudo. O conjunto de dados inicial foi reduzido para 59, pois 11 pontos foram excluídos devido à instabilidade de leitura do aparelho ou proximidade com alguma superfície (fundo do reservatório ou plantas).

Com esse conjunto de dados, considerado “pequeno”, do ponto de vista estatístico, realizou-se uma análise espacial, separando apenas 9 pontos para validação dos resultados da inferência espacial. Com os 50 pontos restantes realizou-se a Krigagem Ordinária em uma área onde as coordenadas no eixo x variam aproximadamente entre 602.920 m e 603.337m e no eixo y entre 7.651.678 m e 7.652.125 m. A validação da inferência foi realizada pelo EMQ.

4 RESULTADOS E DISCUSSÕES

4.1 Análise descritiva dos dados

Os primeiros resultados foram obtidos realizando-se uma análise exploratória do conjunto de dados, por meio de um sumário de medidas-síntese, gráficos explicativos e testes estatísticos. Esse conhecimento das características particulares dos dados é importante para a próxima fase deste estudo, que consiste em aplicar diferentes planos de amostragem.

O conjunto utilizado para o estudo é composto por 978 observações. A Figura 16 ilustra a distribuição das concentrações de clorofila obtidas ao longo do transecto, mostrando a sua dispersão, porém que em muitos pontos as observações tendem a ter valores semelhantes de seus vizinhos, do que as de pontos mais distantes.

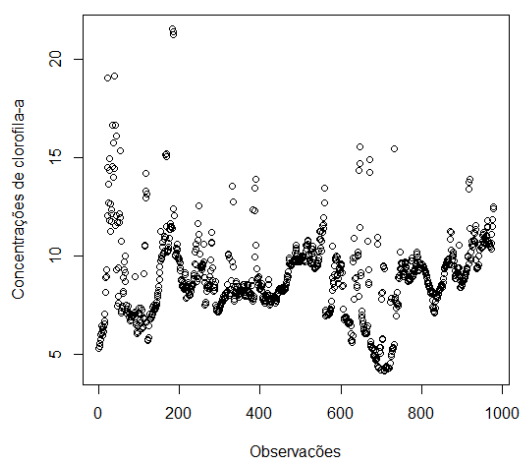


Figura 16–Distribuição das concentrações de clorofila-a ao longo do transecto.

As medidas de síntese numérica foram obtidas através de pacotes estatísticos disponíveis no *software* R e seguem na Tabela 3.

Tabela 3 - Resultados numéricos para a população de clorofila-a.

N	Mín.	Máx.	1º Quartil	3º Quartil	Somatório	C.V.
978	4,16	21,55	7,74	9,83	8.695,26	0,24
Média	Mediana	Var.	D.P.	Assimetria	Curtose	D.P. da média
8,89	8,76	4,61	2,15	1,24	5,18	0,07

Nota-se logo de início uma característica relevante do conjunto de dados, visto que 75% das observações possuem concentrações de clorofila-*a* entre 4,16 e 9,83, o que leva a concluir que poucas observações possuem valores altos, próximos do valor máximo. A mediana caracteriza muito bem essa afirmação, mostrando que os valores estão muito mais próximos do mínimo do que do máximo. A média também confirma essa característica do conjunto de dados, pois o valor resultante deu bem próximo do valor central (mediana).

Para avaliar a variabilidade desses dados o coeficiente de variação é uma ferramenta fundamental. Assim, o Coeficiente de Variação = 24% indica que não há uma grande variabilidade entre as concentrações de clorofila-*a*.

O valor da curtose indica à forma da curva da distribuição em relação ao seu achatamento ou o quanto a curva é achatada comparada a uma curva normal de referência. Como neste caso o valor resultante é maior que 0,263, a curva da distribuição recebe o nome de leptocúrtica (mais afunilada), conforme ilustrada na Figura 17.

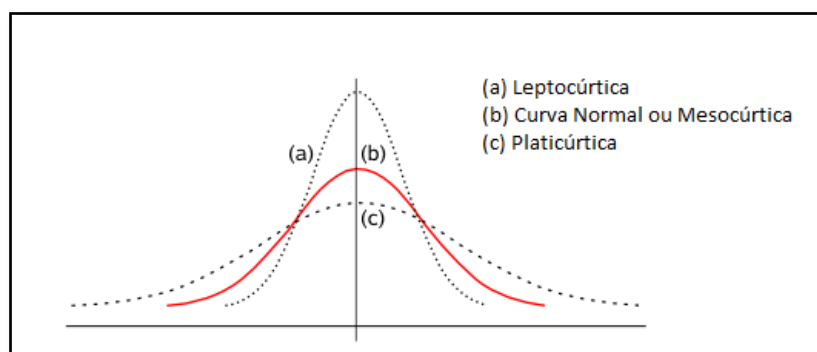


Figura 17 - Formas da curva de distribuição em relação à curva normal.
Fonte: MSPC (2008).

O valor do coeficiente de assimetria permite afirmar se a população possui uma distribuição simétrica ou assimétrica; e sendo assimétrica qual é o tipo. Como seu valor foi maior que 0, curva da distribuição de frequências é assimétrica positiva ou assimétrica à direita. A Figura 18 ilustra os tipos de distribuição em relação à simetria.

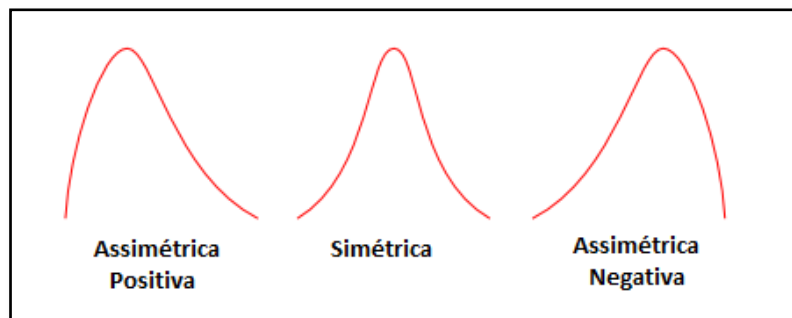


Figura 18 - Tipos de distribuição em relação à assimetria.
Fonte: MSPC (2008).

Também é possível enxergar a assimetria existente, pelo histograma, disposto na Figura 19. Essa assimetria ocorre principalmente devido aos poucos elementos amostrais com elevada concentração de clorofila-*a*, gerando barras minúsculas no final do gráfico e desequilibrando a distribuição. Percebe-se também que grande parte dos valores estão distribuídos em torno da média, nas 3 barras centrais do histograma, com valores de concentração de clorofila-*a* entre 6 e 12.

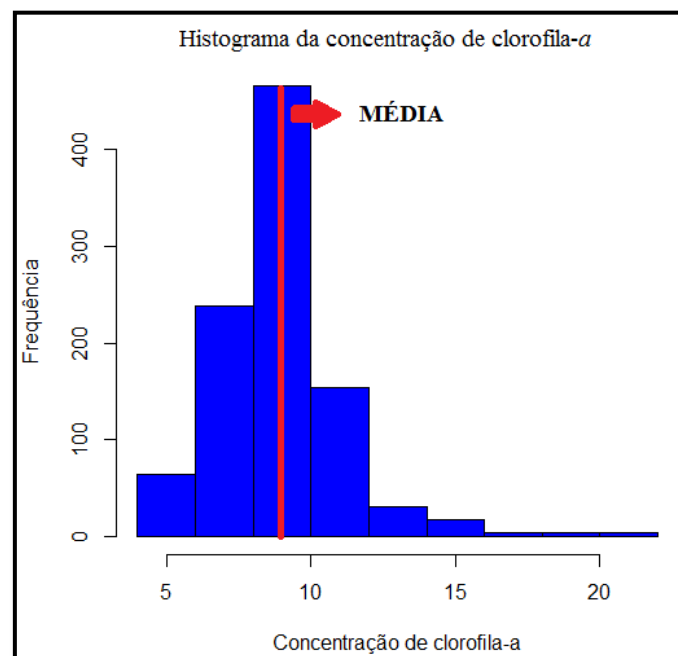


Figura 19 – Histograma da concentração de clorofila-*a*.

A fim de verificar a normalidade dos dados, inicialmente foram construídos alguns gráficos como o Q-Q Plot (Figura 20) e o BoxPlot (Figura 21) para uma inspeção apenas visual.

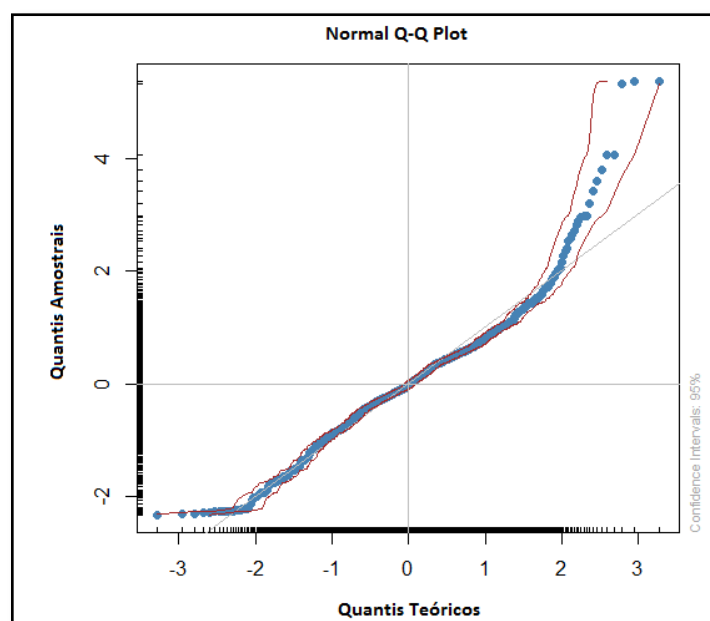


Figura 20 - Gráfico Normal Q-Q Plot.

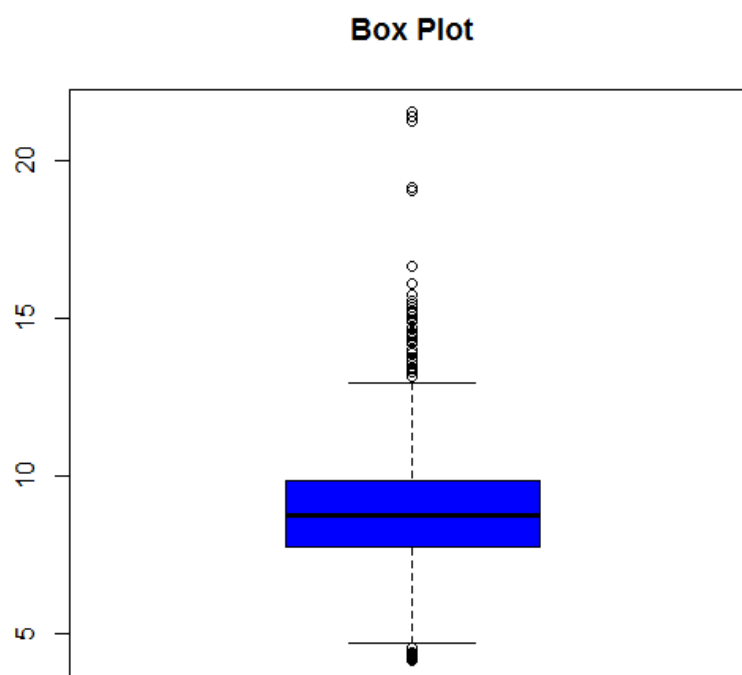


Figura 21 - Gráfico Box-Plot.

Na Figura 20 a configuração dos pontos não se aproxima exatamente de uma reta, os pontos localizados no canto direito superior do gráfico se distanciam de um comportamento linear. Portanto, há evidências de que os dados não seguem uma distribuição normal. O BoxPlot da Figura 21 também evidencia a ausência de normalidade nos dados, sendo que existem muitos pontos considerados atípicos, localizados acima do limite superior. A assimetria dos dados já é uma indicação que os dados não têm uma distribuição Normal.

Para comprovar as evidências obtidas através da análise dos gráficos, realizou-se o teste de Shapiro-Wilk a um nível de significância igual a 0,05, com o auxílio do *software* R. A hipótese testada foi:

$$\begin{cases} H_0: \text{os dados seguem uma distribuição normal.} \\ H_a: \text{os dados não seguem uma distribuição normal.} \end{cases}$$

Tabela 4 - Resultado do teste Shapiro-Wilk.
RESULTADOS DO TESTE DE SHAPIRO-WILK

ESTATÍSTICA W	0,9232
P-VALOR	2.2 e-16

Os resultados do teste, apresentados na Tabela 4, mostram que como o p-valor resultante é menor que o nível de significância adotado (0,05), conclui-se ao nível de significância de 5% que os dados não seguem uma distribuição normal. Com isso, as evidências encontradas anteriormente pelos gráficos foram confirmadas.

Felizmente, para prosseguir com este trabalho não é necessária a condição de normalidade dos dados, porém, se houvesse essa exigência, o correto seria aplicar algum tipo de transformação nos dados que os tornassem normais.

4.2 Em forma de transecto

Como foi dito anteriormente, os dados em transecto coletados por Cicerelli (2013) foram usados para duas análises neste trabalho. A primeira reduzindo ao máximo o tamanho do conjunto de dados, realizando quatro amostragens sistemáticas e depois aplicando inferência espacial em cada amostra obtida.

A segunda análise foi realizada com as previsões resultantes da inferência espacial usando todo o conjunto de dados em transecto, ou seja, com os dados simulados. Nesses dados aplicaram-se os três métodos de amostragem citados neste trabalho, sendo que em cada método foram usadas cinco amostras de tamanhos diferentes. Em seguida, aplicou-se a inferência espacial por krigagem ordinária em cada amostra. Em ambas as análises os resultados foram validados pelo EMQ e pelo Índice Kappa. Os resultados obtidos em cada análise estão dispostos a seguir.

4.2.1 Dados Reais

O primeiro passo foi aplicar no conjunto de dados em transecto o método da amostragem sistemática a fim de reduzir gradualmente o tamanho do conjunto sem que ele perca seu formato original. A população a ser amostrada é composta por 778 pontos de concentrações de clorofila-a, visto que 200 pontos foram separados para validação. A amostragem foi aplicada sobre esse universo de 778 observações com o auxílio do *software* R e do Excel, considerando-se amostras de tamanhos 400, 300, 200 e 100.

Inicialmente, foi necessário obter o valor de k (intervalo entre os elementos amostrais = $\frac{N}{n}$) para cada amostra. Assim, foi construído um programa no *software* R para selecionar aleatoriamente, entre 0 e k , o *id* (identificação) do ponto que seria a primeira unidade de cada amostra. As unidades amostrais seguintes seriam obtidas somando-se o valor de k ao valor do *id* do primeiro ponto e assim sucessivamente, até que completasse as observações da amostra. A partir da amostra foram obtidas as estimativas da média e da variância, para comparação com a média do conjunto inteiro. Os resultados para cada amostra estão apresentados na Tabela 5.

Tabela 5 - Resultados das estimativas através da AS para as amostras de clorofila-a.

N	n	Média	K	$\bar{V}(\bar{Y})$
778	400	8,895	1,945	4,509
778	300	8,919	2,593	4,418
778	200	8,870	3,890	4,025
778	100	8,846	7,789	4,296

Sabendo que a média do conjunto inteiro é igual a 8,891 e a variância igual a 4,608, comparando os resultados com esses valores, pode-se dizer que a redução do conjunto de dados em amostras menores não afetou essas características dos dados em nenhum dos casos. As distribuições espaciais de cada amostra estão apresentadas a seguir.

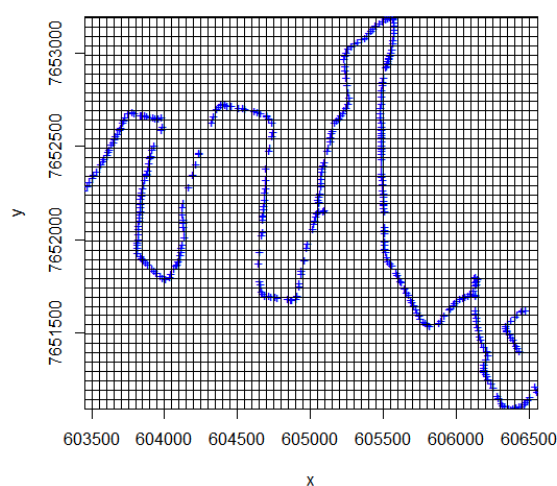


Figura 22 - Distribuição espacial da amostra sistemática 400.

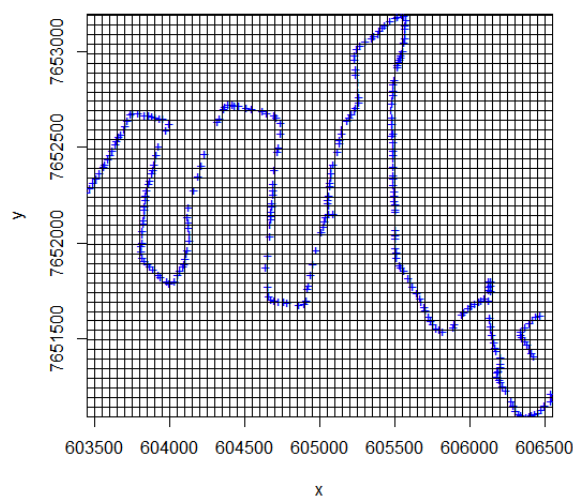


Figura 23- Distribuição espacial da amostra sistemática 300.

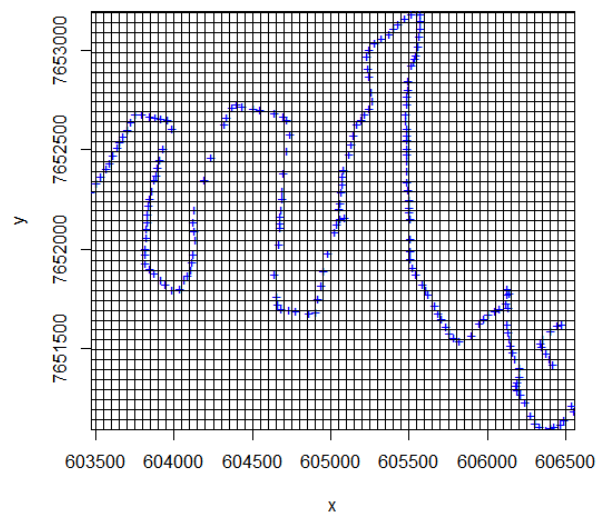


Figura 24- Distribuição espacial da amostra sistemática 200.

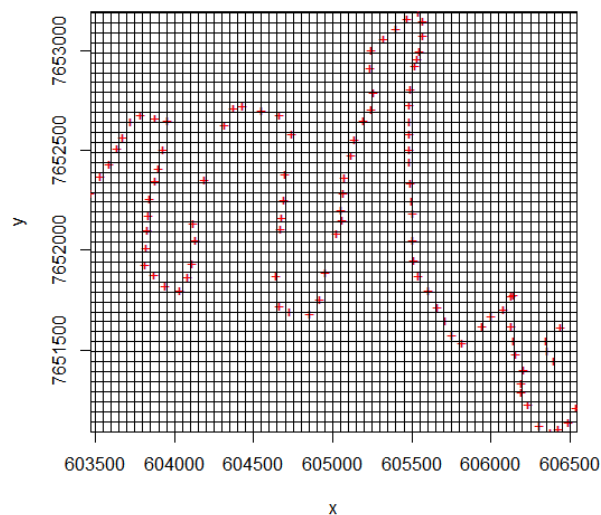


Figura 25- Distribuição espacial da amostra sistemática 100.

O próximo passo foi realizar a inferência espacial por krigagem ordinária com cada amostra. Inicialmente, para auxiliar no processo de modelagem e verificar a existência de anisotropia, construiu-se um mapa variográfico em pixels, que mostra simultaneamente os variogramas em todas as direções. Como pode-se observar na Figura 26, a variável clorofila-*a* possui um comportamento anisotrópico em todos os casos, ou seja, não varia uniformemente em todas as direções, por isso, torna-se necessário a modelagem da anisotropia.

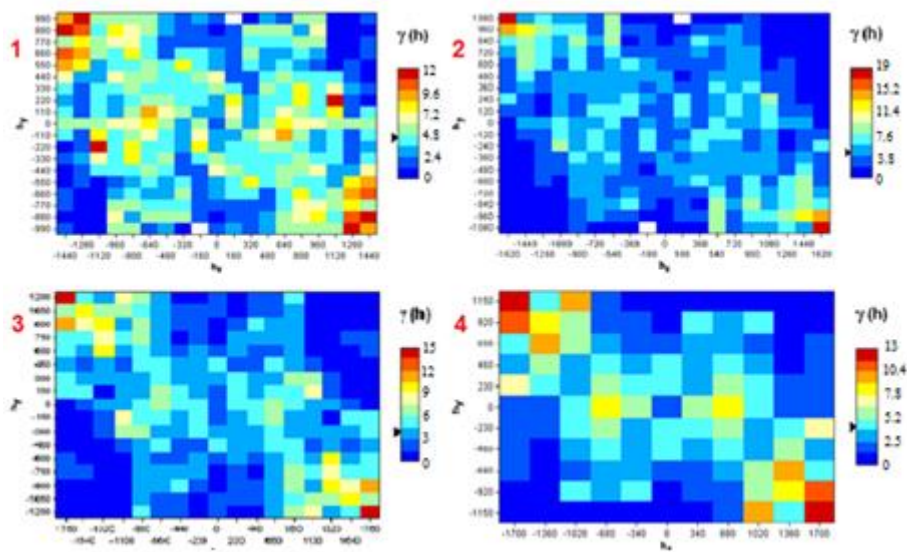


Figura 26- Mapas variográficos para as amostras de tamanhos 400 (1), 300 (2), 200 (3) e 100 (4).

Considerando como 0° o sentido Leste-Oeste e as demais direções deslocando em sentido anti-horário, nota-se que as direções anisotrópicas são 45° e 135° para todas as amostras.

Em seguida, para um estudo mais detalhado, fez-se para cada amostra o variograma omnidirecional, com direção 0° e tolerância angular de 90°, e também para as direções de 0°, 45°, 90° e 135° com tolerância angular de 45°. Como foram feitos alguns semivariogramas para direções não-ortogonais, é importante dizer que no *software* R os eixos não precisam ser obrigatoriamente ortogonais, é possível construir semivariogramas para qualquer direção. Segue abaixo, nas Figuras 27 a 38, os variogramas para cada amostra:

- Amostra 400 – variogramas com 10 lags de comprimento 130.

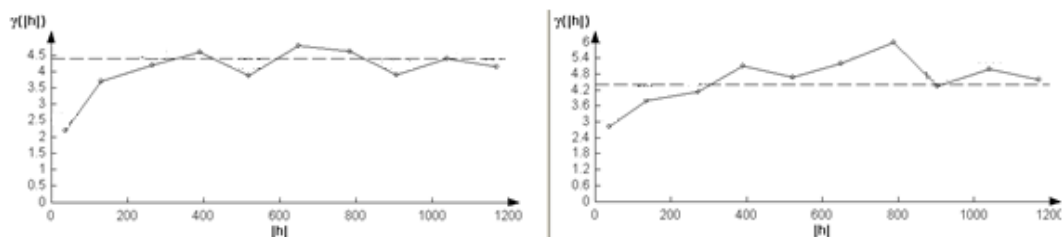


Figura 27- Semivariogramas omnidirecional e para a direção 0° da amostra 400 dos dados em transecto.

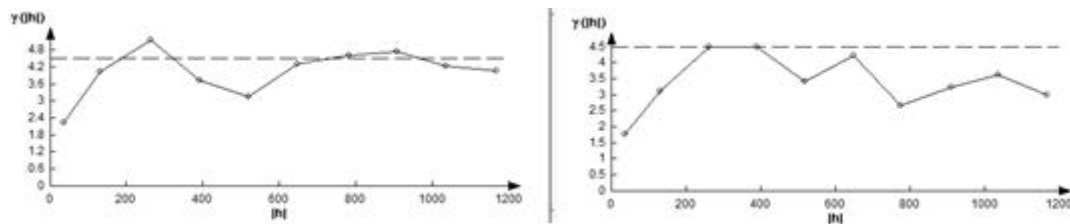


Figura 28– Semivariogramas para as direções 45° e 90° da amostra 400 dos dados em transecto.

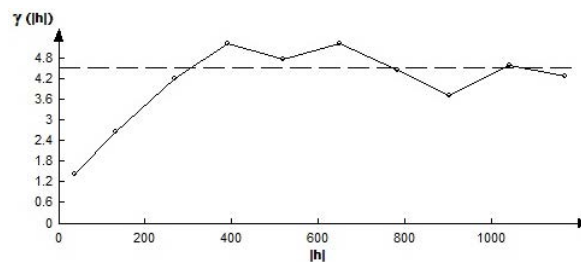


Figura 29– Semivariograma para a direção 135° da amostra 400 dos dados em transecto.

- Amostra 300 – variogramas com 8 lags de comprimento 120.

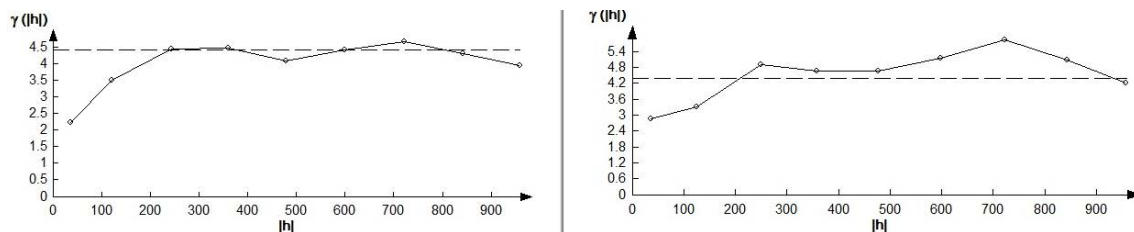


Figura 30– Semivariogramas omnidirecional e para a direção 0° da amostra 300 dos dados em transecto.

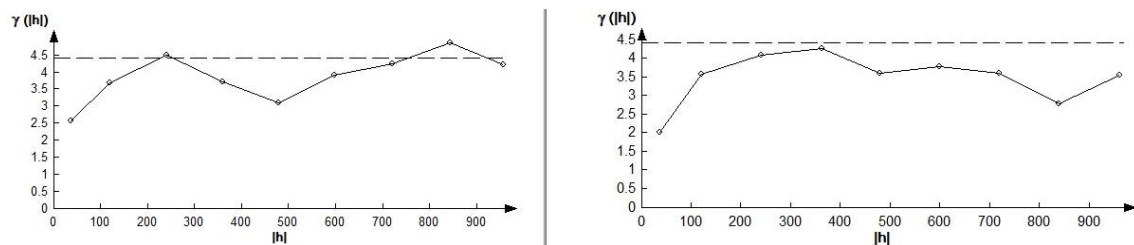


Figura 31– Semivariogramas para as direções 45° e 90° da amostra 300 dos dados em transecto.

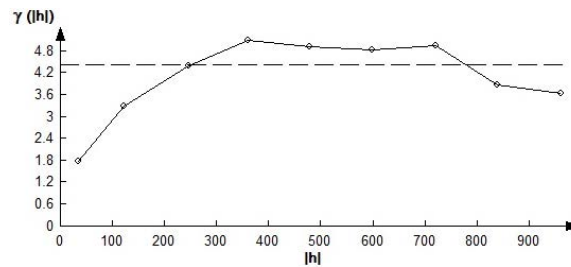


Figura 32– Semivariograma para a direção 135° da amostra 300 dos dados em transecto.

- Amostra 200 – variogramas com 7 lags de comprimento 180.

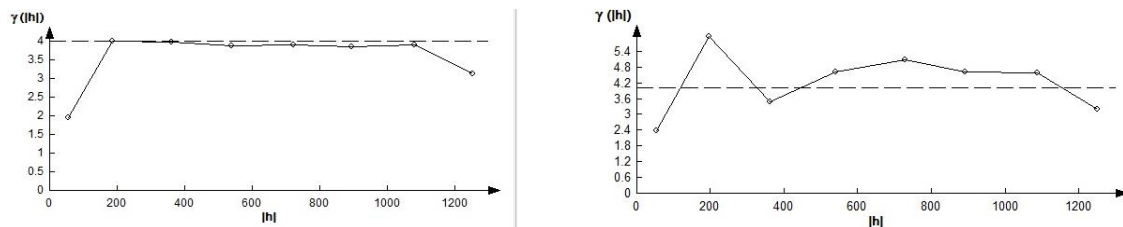


Figura 33– Semivariogramas omnidirecional e para a direção 0° da amostra 200 dos dados em transecto.

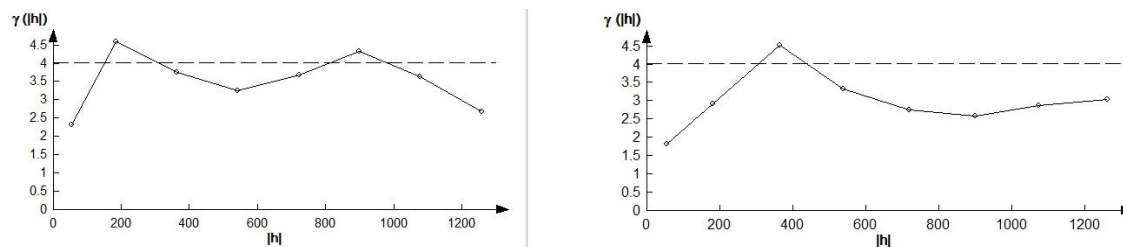


Figura 34– Semivariogramas para as direções 45° e 90° da amostra 200 dos dados em transecto.

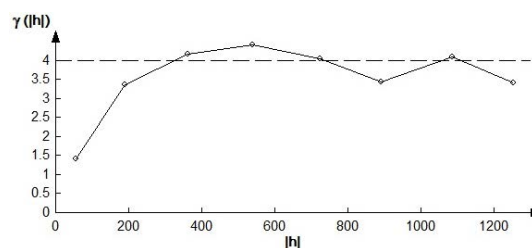


Figura 35– Semivariograma para a direção 135° da amostra 200 dos dados em transecto.

- Amostra 100 – variogramas de 7 lags de comprimento 180.

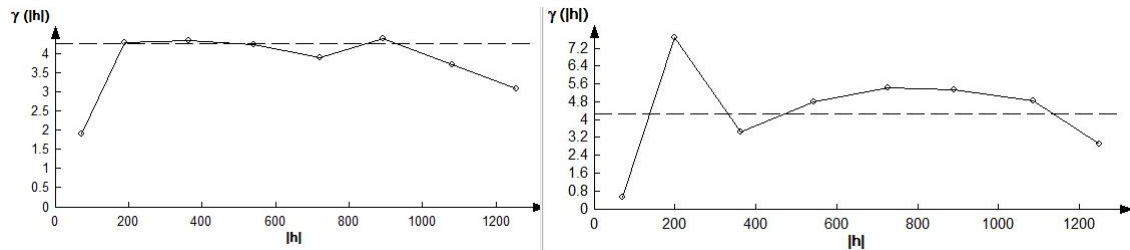


Figura 36- Semivariogramas omnidirecional e para a direção 0° da amostra 100 dos dados em transecto.

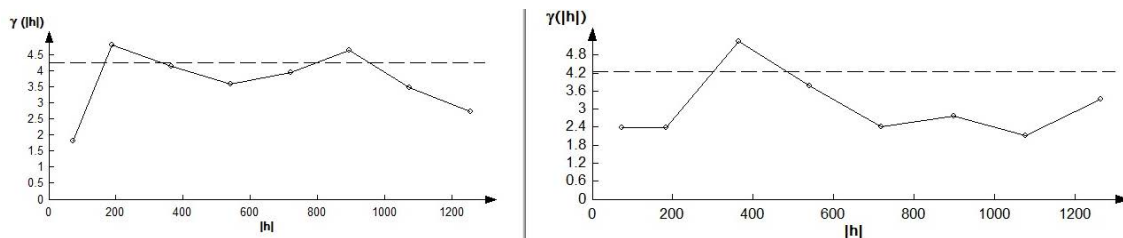


Figura 37- Semivariogramas para as direções 45° e 90° da amostra 100 dos dados em transecto.

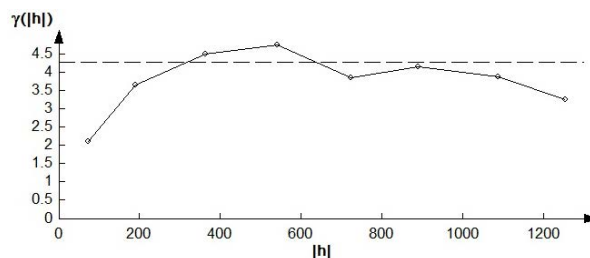


Figura 38- Semivariograma para a direção 135° da amostra 100 dos dados em transecto.

Com base nos variogramas dadas nas Figuras 27 a 38, pode-se confirmar que as direções detectadas pelo mapa variográfico são realmente as direções anisotrópicas. Portanto, as modelagens da anisotropia serão feitas baseadas nessas direções. A Figura 39 apresenta o ajuste de um modelo teórico para cada uma das direções em cada caso. Em todos os ajustes utilizou-se o modelo Exponencial e os valores para o efeito pepita foram: 0,225, 0,27, 0,369 e 0,215, respectivamente.

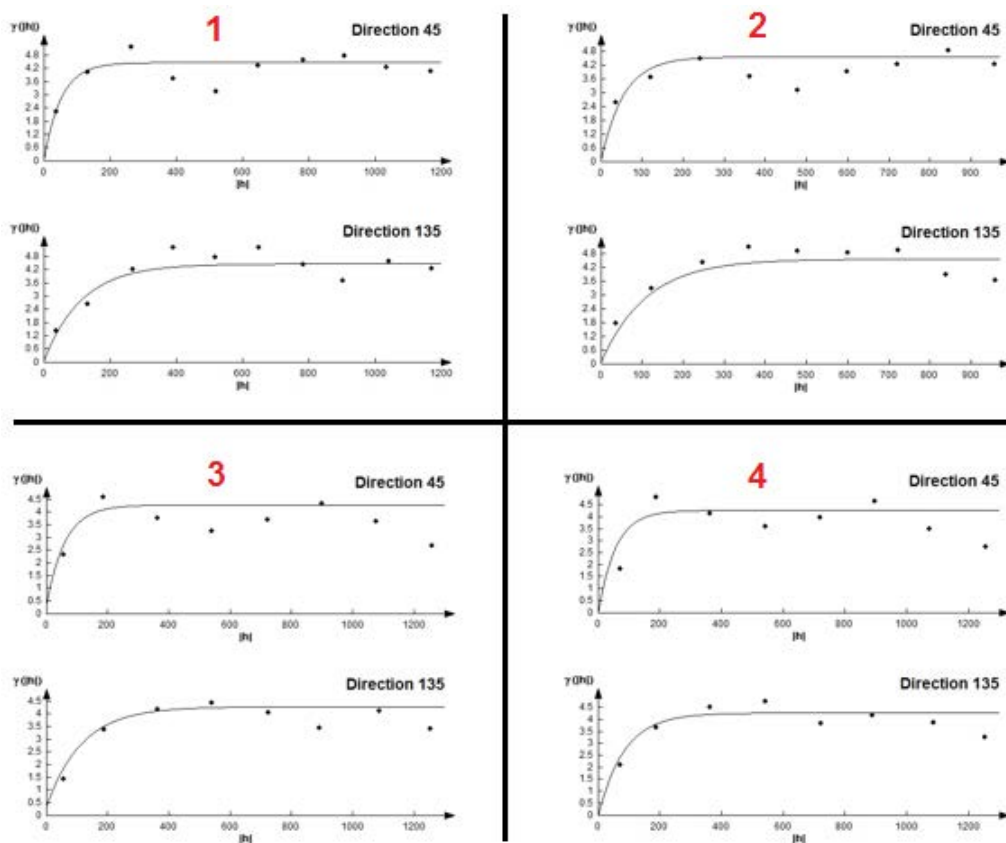


Figura 39 – Modelagem da anisotropia para as amostras tamanho 400 (1), 300 (2), 200 (3) e 100 (1).

O tipo de anisotropia encontrado em todos os casos foi a anisotropia geométrica, em que a forma e o patamar dos semivariogramas são iguais e apenas o alcance é diferente. Os valores do patamar, os alcances de cada direção e o valor da Razão de Anisotropia para cada amostra estão dispostos na Tabela 6.

Tabela 6 - Valores de patamar, alcances e razão de anisotropia para cada amostra.

	Patamar	Alcance do eixo de menor continuidade (x)	Alcance do eixo de maior continuidade (y)	Razão de Anisotropia = $\frac{x}{y}$
Amostra 400	4,5	380	700	0,54
Amostra 300	4,5	300	600	0,50
Amostra 200	3,9	280	600	0,47
Amostra 100	4,3	300	550	0,54

Após a modelagem, a inferência dos dados de clorofila-a por meio da Krigagem Ordinária foi realizada no *software* R. Para isso foi gerada uma máscara com resolução de 30 metros, que resultaram em um total de 71 linhas e 104 colunas, gerando-se 7384 estimativas. Os resultados das krigagens utilizando cada amostra estão apresentados na Figura 40.

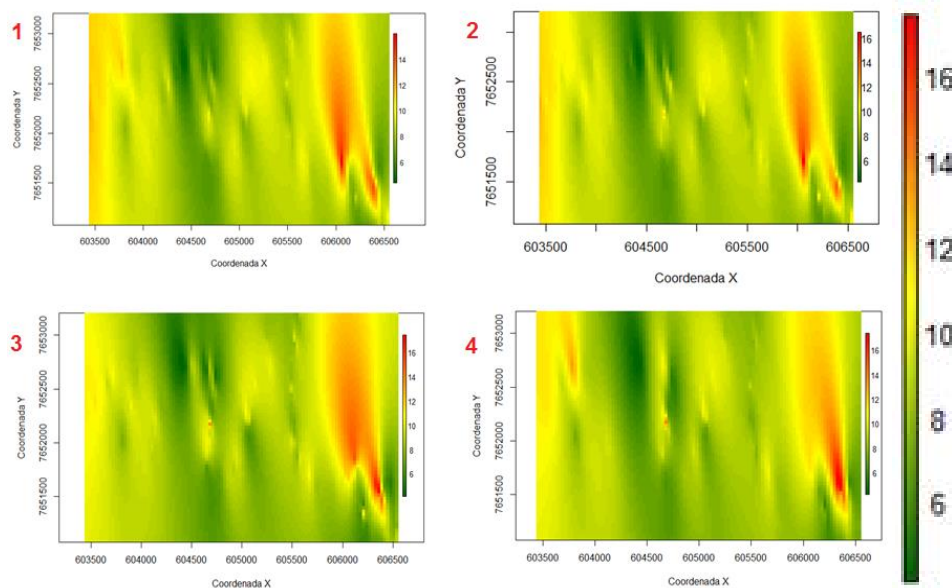


Figura 40 – Krigagem Ordinária da clorofila-a para as amostras de tamanhos 400 (1), 300 (2), 200 (3) e 100 (4).

Para efeito de comparação, realizou-se também a inferência espacial da clorofila-a utilizando todos os dados, que segue na Figura 41.

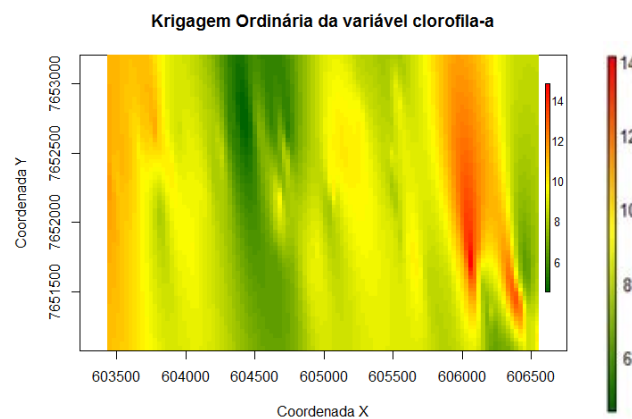


Figura 41- Krigagem Ordinária utilizando todos os 978 dados.

Ao comparar as Figuras 40 e 41, percebe-se que mesmo reduzindo o conjunto de dados em amostras menores, os resultados da krigagem ordinária de todas as amostras são bem semelhantes ao da krigagem utilizando todos os dados, ou seja, não houve grandes alterações nas características do mapeamento.

Observam-se poucas áreas em vermelho escuro, que são as regiões que predominam valores mais altos de concentração de clorofila-a. Os valores mais comuns estão entre 5 e 10, representados no mapa pelas regiões com coloração entre verde escuro e amarelo.

Feito isso, dividiu-se em classes os resultados das inferências para a construção de mapas de fatiamento a partir de programas desenvolvidos no *software* R. Foram utilizadas cinco classes, definidas visualmente a partir do histograma dos valores preditos de clorofila-a obtidos pela Krigagem Ordinária utilizando todos os dados (Figura 42).

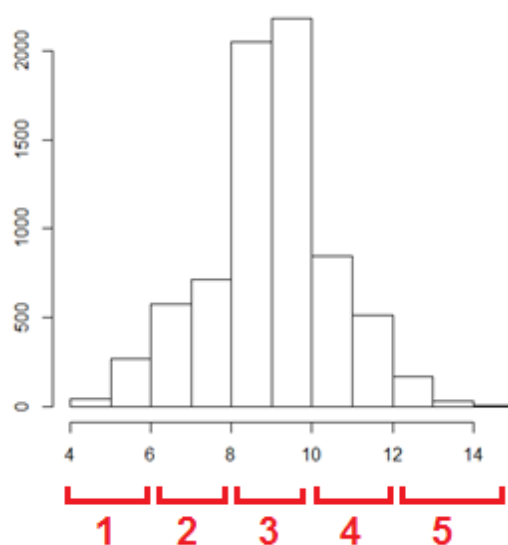


Figura 42– Histograma dos valores preditos de clorofila-a.

A Figura 43 apresenta os resultados da divisão em classes das Krigagens Ordinária para a toda área de estudo. Percebe-se que, em geral, os valores de concentração de clorofila-a são baixos, grande parte da área de estudo possui concentração de 4 a 10 ppb, e em duas regiões próximas às laterais da área os valores tendem a aumentar, apresentando valores entre 10 e 15 ppb.

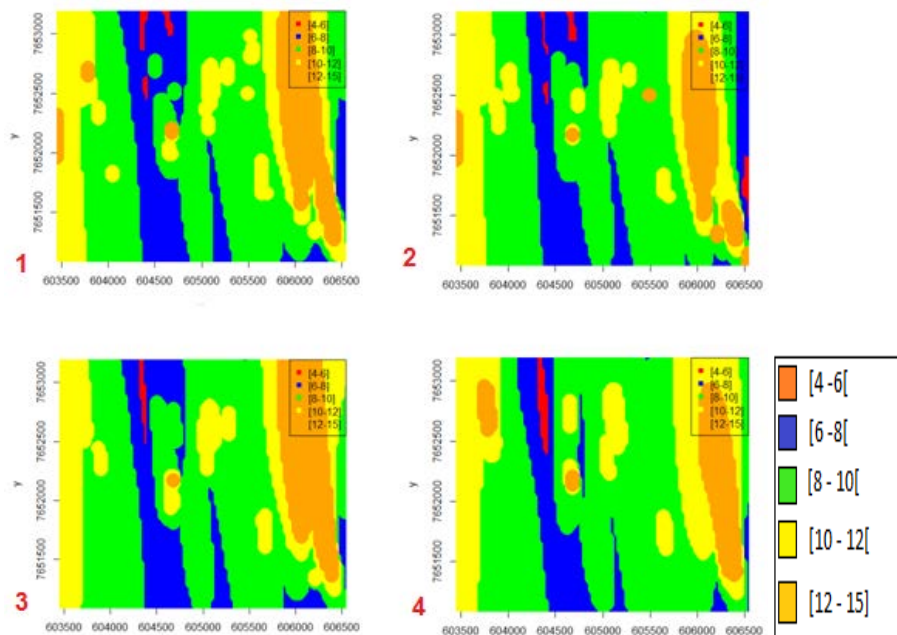


Figura 43– Mapas de classes da Krigagem Ordinária das amostras de tamanhos 400 (1), 300 (2), 200 (3) e 100 (4).

Também construiu-se os mapas de variância do erro de estimação para a variável clorofila-a (Figura 44), a fim de verificar a confiabilidade dos valores interpolados. Como a variância da krigagem depende da distribuição espacial das amostras, percebe-se que a variância é nula no local onde foi amostrado e maior nos locais onde não possuem pontos de amostragem, sendo assim, conforme se afasta da região de amostragem, a estimativa se torna menos confiável.

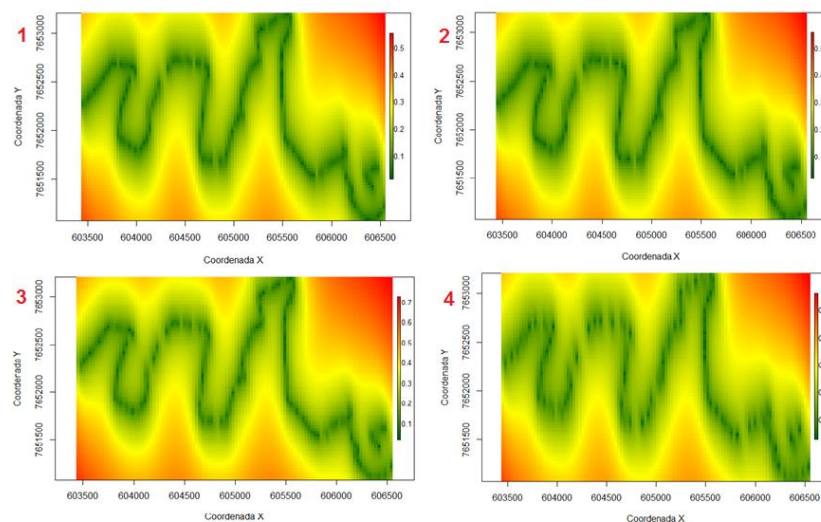


Figura 44- Mapas de variância do erro de estimação para as amostras de tamanhos 400 (1), 300 (2), 200 (3) e 100 (4).

- **VALIDAÇÃO DO PROCESSO DE INFERÊNCIA**

Para a validação dos resultados, foram utilizados duas medidas, o Erro Médio Quadrático (EMQ) e o Índice Kappa. Sabe-se que existem fórmulas na literatura estatística que auxiliam na definição do número de classes, como a Fórmula de Sturges, onde $k = 1 + 3,322(\log n)$. Porém, neste caso resolveu-se definir a quantidade de classes intuitivamente através do histograma da distribuição dos valores preditos pela krigagem ordinária utilizando todos os dados (Figura 42), definindo assim cinco classes. Isso porquê o número de classes sugerido pela Fórmula de Sturges seria muito grande (14 classes), definindo classes com amplitudes muito pequenas e deixando a análise muito sensível.

A Tabela 7 apresenta os resultados do EMQ e do Índice Kappa da inferência espacial utilizando cada amostra.

Tabela 7 - Resultados do EMQ e Índice Kappa para a inferência espacial utilizando as amostras estratificadas.

	Amostra 400	Amostra 300	Amostra 200	Amostra 100
EMQ	0,92	1,06	1,30	2,55
Índice Kappa	0,68	0,66	0,58	0,55

Como era de se esperar, nota-se que conforme diminuimos o tamanho da amostra o valor do EMQ vai aumentando e o Índice Kappa vai diminuindo. Porém, considerando que o tamanho reduziu drasticamente, pode-se dizer que não houve grande diferença entre os resultados obtidos pelas amostras. No estudo realizado, uma amostra com 200 elementos amostrais foi suficiente para atingir os objetivos propostos e assim diminuiria os custos, o tempo e a quantidade de pessoas necessárias para realizar o trabalho.

4.2.2 Dados Simulados

Para análises complementares, consideraram-se as 7384 previsões resultantes da Krigagem Ordinária utilizando todos os dados, como uma “verdade terrestre”. Primeiramente, separou-se aleatoriamente 1.500 observações para posteriormente

validar os resultados da inferência espacial. Com as 5.884 observações restantes aplicaram-se os três métodos de amostragem citados na fundamentação teórica: Amostragem Aleatória Simples, Amostragem Aleatória Estratificada e Amostragem Sistemática. Sendo que, em cada método calcularam-se cinco amostras, de tamanhos iguais a 1000, 400, 300, 200 e 100.

Em seguida, deu-se início ao processo de inferência espacial da clorofila-a utilizando cada amostra obtida. Os resultados para cada método de amostragem estão dispostos nos tópicos a seguir.

- Amostragem Aleatória Simples

Inicialmente, construiu-se o mapa variográfico para cada amostra a fim de analisar a ocorrência de anisotropia nos dados (Figura 45). Devido ao uso de dados já provenientes de um processo de inferência espacial, ou seja, já suavizados, os resultados dos mapas variográficos foram fora do comum pelo fato de não ser possível identificar os eixos de maior e menor continuidade, por isso optou-se por considerar o fenômeno isotrópico. Portanto, é importante dizer que esse conjunto foi escolhido apenas para análises complementares e não para comparações.

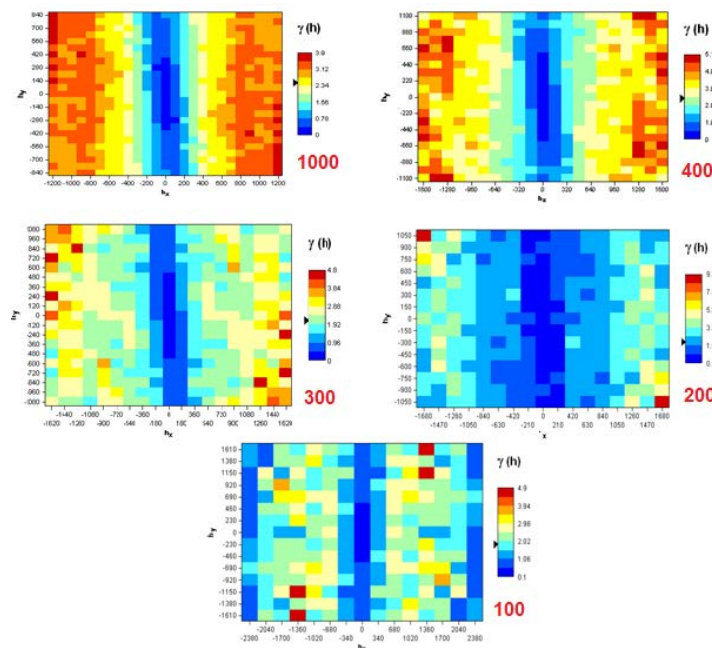


Figura 45- Mapas variográficos das amostras simples de tamanhos 1000, 400, 300, 200 e 100.

Após a construção dos mapas variográficos, realizou-se a inferência dos dados de clorofila-a por meio da Krigagem Ordinária. Para isso, novamente foi gerada no *software* R uma máscara com resolução de 30 metros, resultando em um total de 71 linhas e 104 colunas, com isso foram geradas 7384 estimativas. Os resultados das krigagens utilizando cada amostra estão na Figura 46.

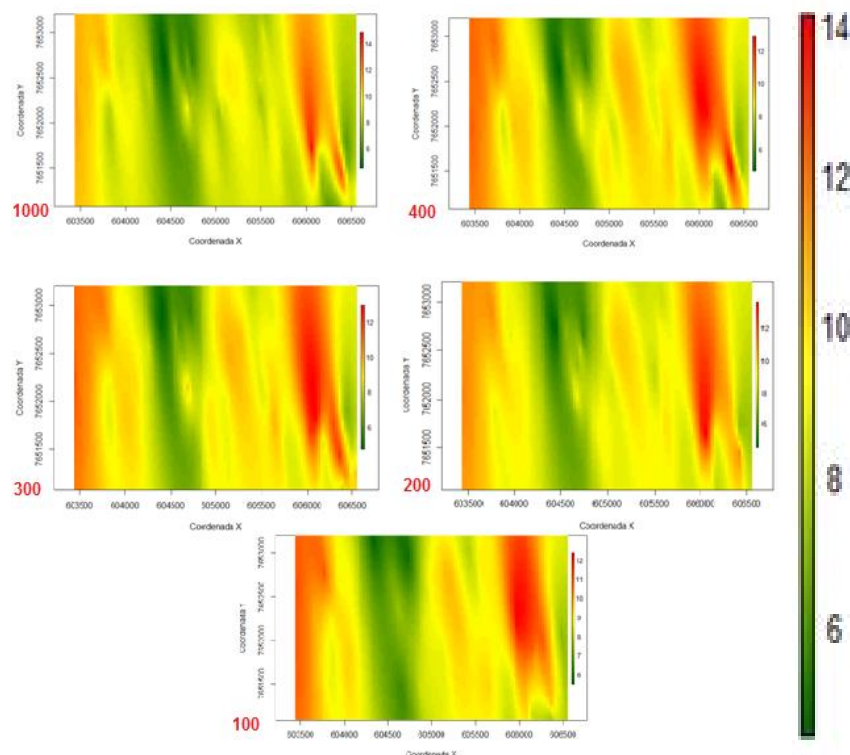


Figura 46- Krigagem Ordinária da clorofila-a utilizando as amostras simples de tamanhos 1000, 400, 300, 200 e 100.

Nota-se que os mapas seguem um mesmo padrão nas krigagens de todas as amostras, com mínimas diferenças entre eles. Porém, o mapeamento que mais se aproximou da inferência espacial que utilizou dos os 978 dados originais (Figura 41), foi o da amostra de tamanho 1000, os demais apresentaram algumas regiões com tom muito avermelhado.

A Figura 47 apresenta o resultado da divisão em classes das Krigagens Ordinárias para todas as amostras simples, em toda área de estudo. Assim como com os dados em transecto, em quase toda área de estudo concentram-se valores baixos de clorofila-a e apenas duas regiões nas laterais apresentam valores maiores.

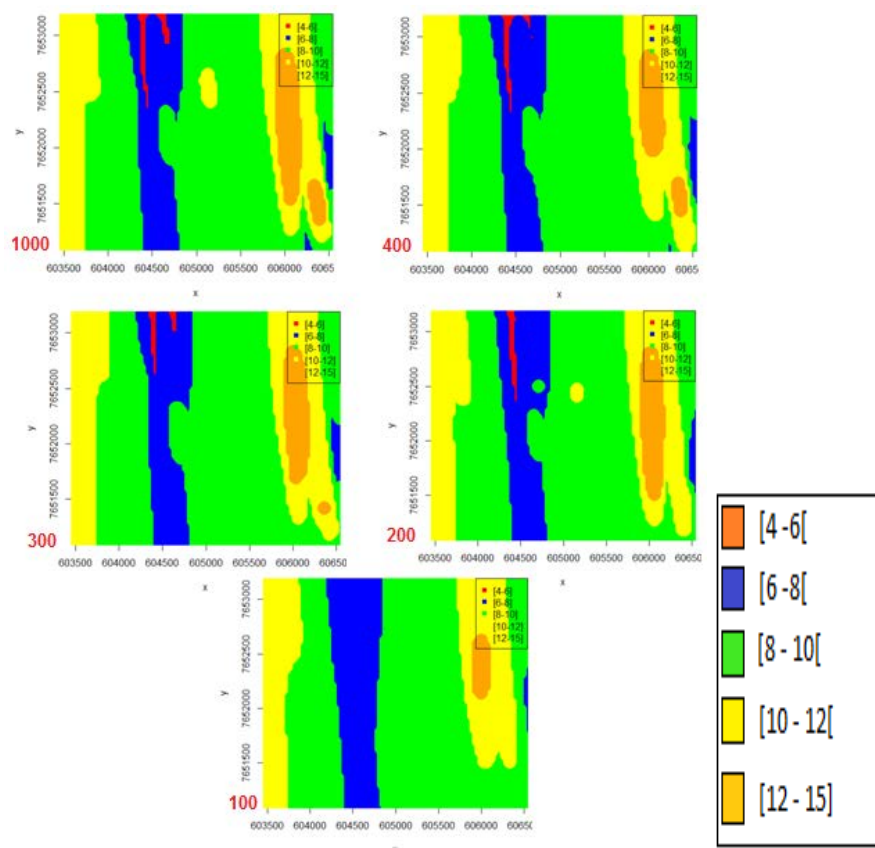


Figura 7- Mapa de classes da Krigagem Ordinária das amostras de tamanhos 1000, 400, 300, 200 e 100.

Para a validação dos resultados, calculou-se o EMQ e o Índice Kappa. O número de classes usadas no cálculo do Índice Kappa foi o mesmo, 5.

A Tabela 8 apresenta os resultados do EMQ e do Índice Kappa da inferência espacial utilizando cada amostra.

Tabela 8 - Resultados do EMQ e Índice Kappa para a inferência espacial utilizando as amostras simples.

	Amostra 1000	Amostra 400	Amostra 300	Amostra 200	Amostra 100
EMQ	0,008	0,059	0,062	0,084	0,239
Índice Kappa	0,960	0,930	0,920	0,900	0,800

A partir da Tabela 8, pode-se dizer que em todos os casos o resultado da inferência espacial da clorofila-a foi satisfatório, sendo que a amostra de 1000

observações teve uma leve vantagem em relação as demais amostras, como esperado. Os resultados do Índice Kappa mostraram que a concordância entre as estimativas obtidas através das amostras e da população total varia de satisfatória à excelente.

- Amostragem Aleatória Estratificada

Como foi dito anteriormente, foram utilizados 3 estratos para o cálculo da amostragem estratificada, sendo que a quantidade de dados concentrados em cada estrato eram: 20% no estrato 1, 50% no estrato 2 e 30% no estrato 3. Portanto, independentemente do tamanho da amostra, a amostra tinha elementos dos três estratos nessas porcentagens.

Com as amostras obtidas, deu-se início ao processo de inferência espacial construindo o mapa variográfico para cada amostra (Figura 48). Nota-se que novamente o fenômeno é isotrópico em todos os casos, pelo mesmo motivo já citado anteriormente. Sendo, portanto, não necessário o ajustamento de um modelo teórico, podendo-se aplicar a Krigagem Ordinária diretamente.

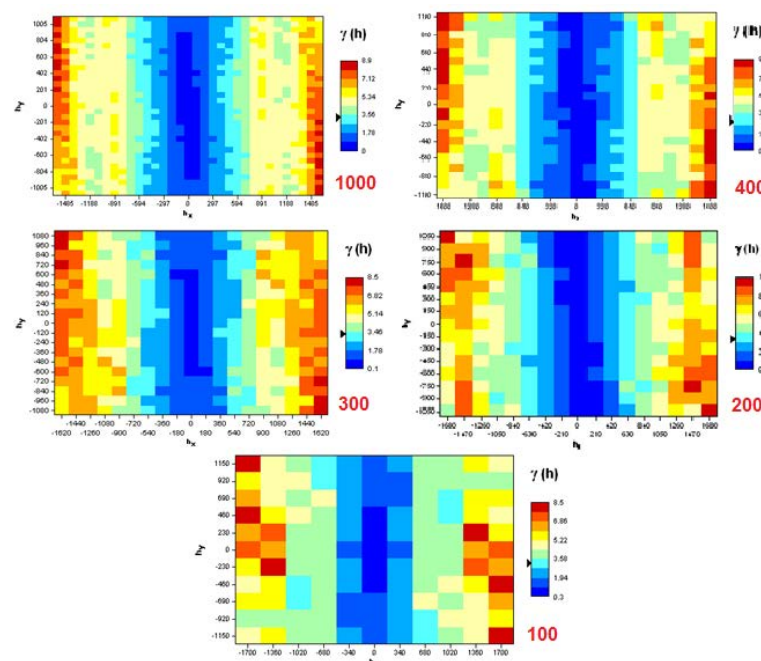


Figura 48- Mapas variográficos das amostras estratificadas de tamanhos 1000, 400, 300, 200 e 100.

Após a análise feita a partir dos mapas variográficos, deu-se início à inferência dos dados de clorofila-a por meio da Krigagem Ordinária. Para isso, novamente foi gerada no *software* R uma máscara com resolução de 30 metros, gerando 7384 estimativas. Os resultados das krigagens utilizando cada amostra estão na Figura 49.

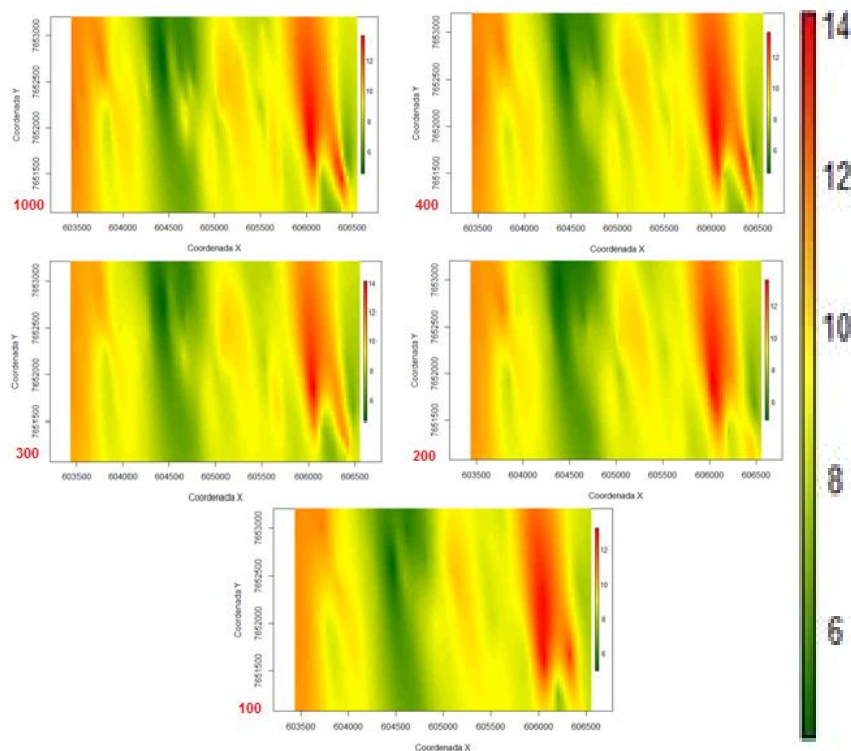


Figura 49 - Krigagem Ordinária da clorofila-a utilizando as amostras estratificadas de tamanhos 1000, 400, 300, 200 e 100.

Nota-se que os mapeamentos da Figura 49 são bem semelhantes entre si, com pequenas variações no centro das imagens.

Antes de validar os resultados, construíram-se novamente os mapas de classes das cinco Krigagens Ordinárias (Figura 50). Eles apresentaram as mesmas características dos mapas construídos com as amostras simples, com diferenças mínimas entre eles no canto direito.

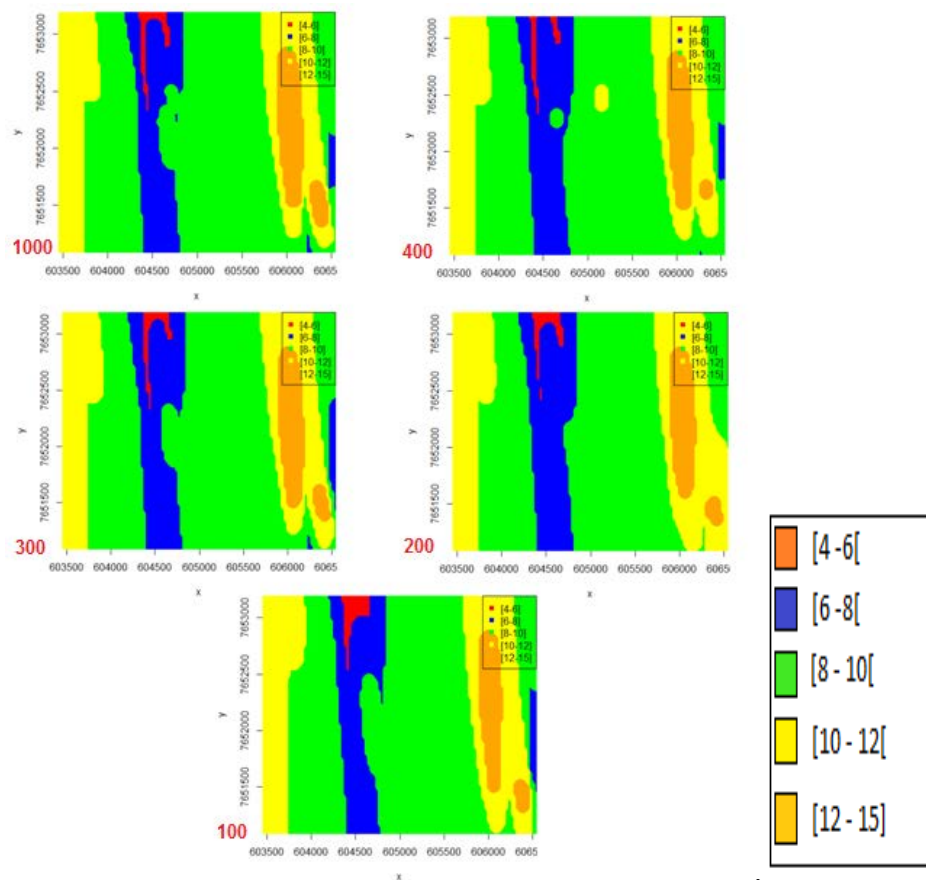


Figura 50- Mapa de classes da Krigagem Ordinária das amostras estratificadas de tamanhos 1000, 400, 300, 200 e 100.

Para avaliar os resultados das inferências espaciais calcularam-se o EMQ e o Índice Kappa, apresentados na Tabela 9. Todos os valores de EMQ foram baixos e os resultados do Índice Kappa mostraram que os resultados foram satisfatórios para as amostras de tamanhos 100 e 200, e excelentes para as amostras de tamanhos 300, 400 e 1000. Novamente, as amostras de tamanhos maiores apresentaram resultados ligeiramente melhores que as amostras menores.

Tabela 9 - Resultados do EMQ e Índice Kappa para a inferência espacial utilizando as amostras estratificadas.

	Amostra 1000	Amostra 400	Amostra 300	Amostra 200	Amostra 100
EMQ	0,011	0,052	0,061	0,127	0,185
Índice Kappa	0,970	0,950	0,940	0,870	0,860

- Amostragem Sistemática

Inicialmente, foram obtidos os valores de k (intervalo entre os elementos amostrais = $\frac{N}{n}$) para cada amostra (Tabela 10). Assim, selecionou-se aleatoriamente um número entre 0 e k , que corresponde ao id do primeiro elemento da amostra. As unidades amostrais seguintes foram obtidas somando-se o valor de k ao valor do id do ponto anterior, até que completasse as observações da amostra.

Tabela 10 - Resultados das estimativas através da AS para as amostras de clorofila-a.

N	n	K	Média	$\bar{v}(\bar{Y})$
5884	1000	5,884	8,948	2,309
5884	400	14,710	8,959	2,751
5884	300	19,613	9,008	2,342
5884	200	29,420	8,997	2,399
5884	100	58,840	8,992	2,245

Considerando que a média das predições é 8,961 pode-se dizer a partir dos valores das médias que as amostras são bem representativas. Com as amostras obtidas, iniciou-se o processo de inferência espacial com a construção dos mapas variográficos, dispostos na Figura 51.

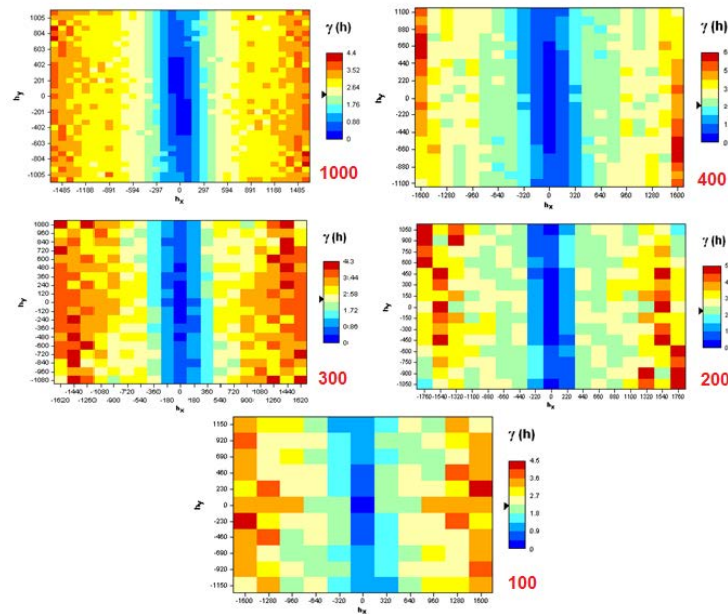


Figura 51 - Mapas variográficos para as amostras sistemáticas de tamanhos 1000, 400, 300, 200 e 100.

Como ocorrido na Amostragem Simples e na Amostragem Estratificada das predições, em todas as amostras obtidas pela Amostragem Sistemática os dados são isotrópicos. Por isso, a próxima etapa seria a inferência dos dados de clorofila-a por meio da Krigagem Ordinária. Para isso, foi usada a mesma máscara com resolução de 30 metros, gerando 7384 estimativas. Os resultados das krigagens utilizando cada amostra estão na Figura 52.

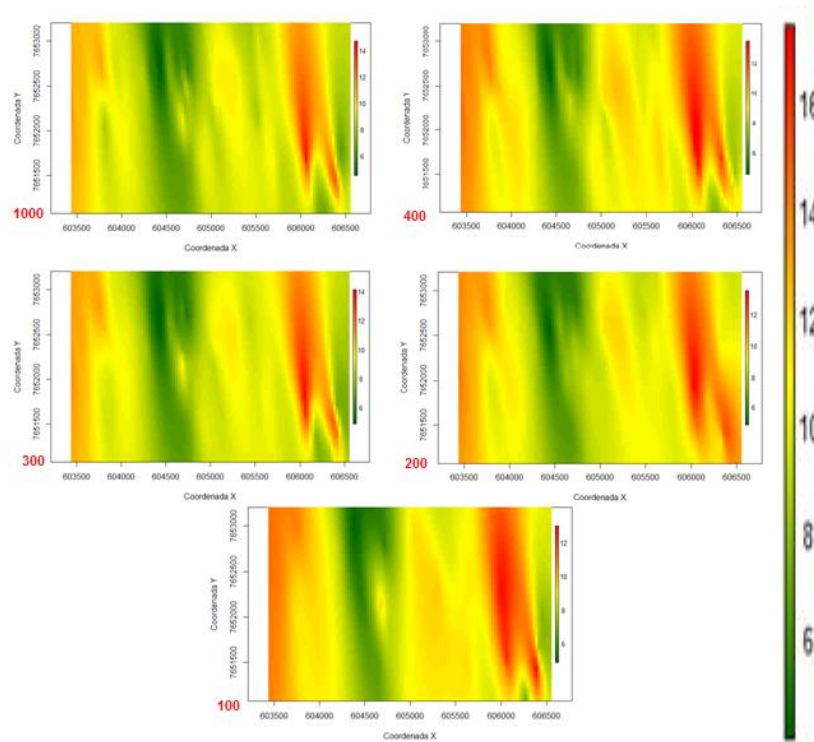


Figura 52- Krigagem Ordinária da clorofila-a utilizando as amostras sistemáticas de tamanhos 1000, 400, 300, 200 e 100.

Novamente os mapeamentos da Figura 52 são bem semelhantes entre si e se aproximam bastante do mapa gerado com todos os dados (Figura 41).

Em seguida, construíram-se no *software* R os mapas de classes das Krigagens Ordinárias das cinco amostras (Figura 53). Eles apresentaram as mesmas características dos mapas construídos com as amostras simples e estratificadas, com pequenas variações no canto direito.

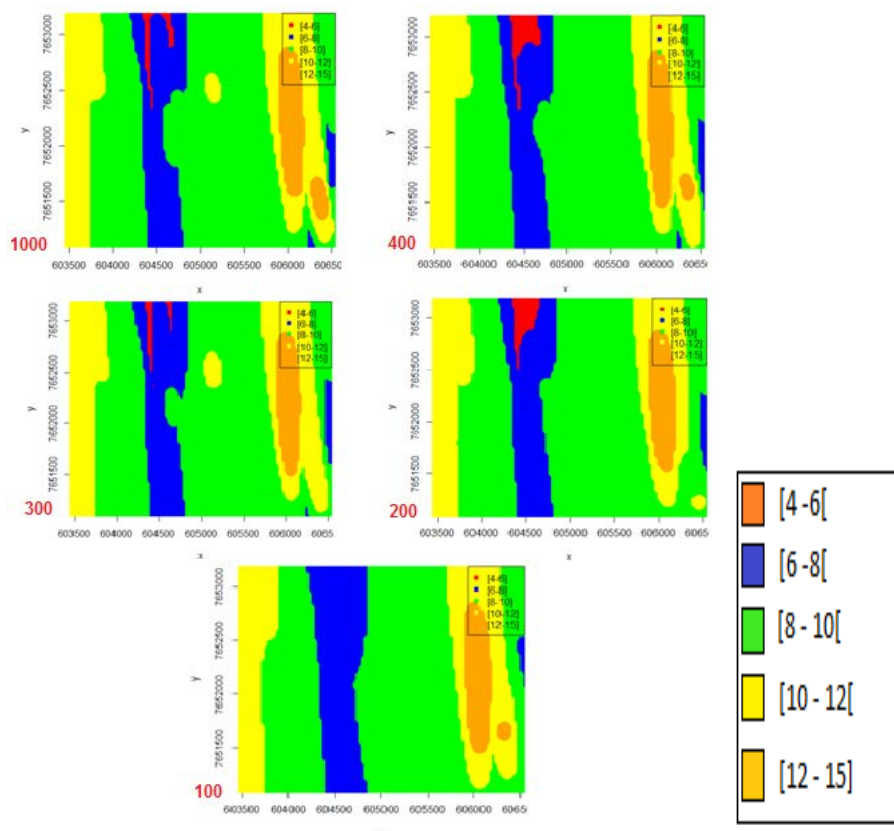


Figura 53- Mapa de classes da Krigagem Ordinária das amostras sistemáticas de tamanhos 1000, 400, 300, 200 e 100.

Por fim, calcularam-se o EMQ e o Índice Kappa para validar os resultados das inferências espaciais (Tabela 11). Nota-se que todas as amostras apresentaram boa classificação, porém o EMQ e o Índice Kappa resultante da amostra de tamanho 1000 foram os mais satisfatórios, comprovando o que já foi dito anteriormente: quanto maior a amostra, menor o erro.

Tabela 11 - Resultados do EMQ e Índice Kappa para a inferência espacial utilizando as amostras sistemáticas.

	Amostra 1000	Amostra 400	Amostra 300	Amostra 200	Amostra 100
EMQ	0,010	0,038	0,042	0,076	0,170
Índice Kappa	0,980	0,940	0,930	0,890	0,840

Para uma análise geral, a Tabela 12 apresenta os resultados do EMQ e do Índice Kappa para todas as inferências espaciais realizadas.

Tabela 12 - Resultados do EMQ e Índice Kappa para todos os casos.

Tipo de amostragem	Tamanho da amostra	EMQ	Índice Kappa
Aleatória (AA)	1000	0,008	0,96
	400	0,059	0,93
	300	0,062	0,92
	200	0,084	0,90
	100	0,239	0,80
Estratificada (AE)	1000	0,011	0,97
	400	0,052	0,95
	300	0,061	0,94
	200	0,127	0,87
	100	0,185	0,86
Sistemática (AS)	1000	0,010	0,98
	400	0,038	0,94
	300	0,042	0,93
	200	0,076	0,89
	100	0,170	0,84

Nota-se que de modo geral todas as amostras apresentaram boa classificação, porém as amostras maiores apresentaram resultados ligeiramente melhores que as amostras menores em todos os casos, assim como a amostragem simples mostrou-se menos eficiente que os outros tipos de amostragem utilizados. Além disso, a Amostragem Sistemática forneceu os melhores resultados na inferência espacial ao definir o menor EQM e o maior Índice Kappa.

4.3 Amostragem pontual

Como foi dito anteriormente, Utsumi (2012) realizou uma amostragem de concentração de clorofila-a em 70 pontos selecionados aleatoriamente em um trecho do Reservatório de Nova Avanhandava. O delineamento amostral foi definido previamente através de levantamentos exploratórios do reservatório e de imagens Landsat, selecionando áreas com variações nas características do ambiente aquático e florescimento de algas.

Dos 70 pontos coletados, apenas 59 foram utilizados para esta análise, pois 11 foram considerados inutilizáveis devido à instabilidade de leitura do aparelho ou proximidade com alguma superfície.

Primeiramente, como nas outras análises realizadas, separou-se uma pequena parte do conjunto de dados para a validação dos resultados da inferência espacial, neste caso foram separados 9 observações e 50 foram utilizadas para o cálculo da Krigagem Ordinária.

Inicialmente, para analisar melhor o comportamento da variável em estudo, fez-se o variograma omnidirecional com direção 0° e tolerância angular de 90° (Figura 54). Em seguida, para auxiliar no processo de modelagem, construiu-se um mapa variográfico em pixels, que mostra simultaneamente os variogramas em todas as direções. Como pode-se observar na Figura 55, a variável clorofila-a possui um comportamento anisotrópico, ou seja, não varia uniformemente em todas as direções, por isso, torna-se necessário a modelagem da anisotropia.

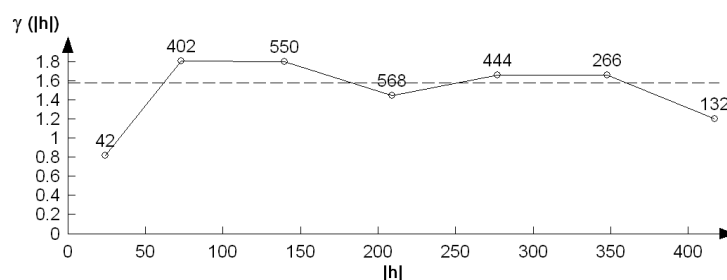


Figura 54- Semivariograma omnidirecional.

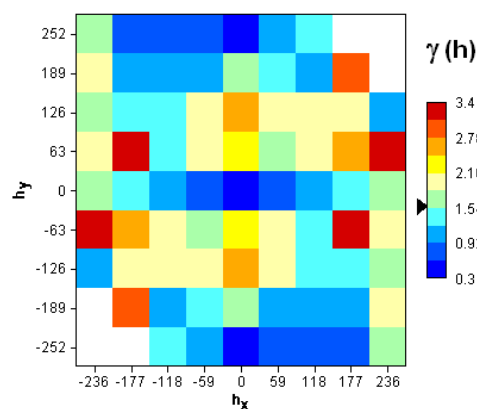


Figura 55- Mapa variográfico dos dados pontuais.

Considerando como 0° o sentido Leste-Oeste e as demais direções deslocando em sentido horário, nota-se que as direções anisotrópicas são 45° e 90°.

Para um estudo mais detalhado, fez-se os variogramas nas direções 0°, 45°, 90° e 135° com tolerância angular de 45 ° e 6 lags de comprimento 60.

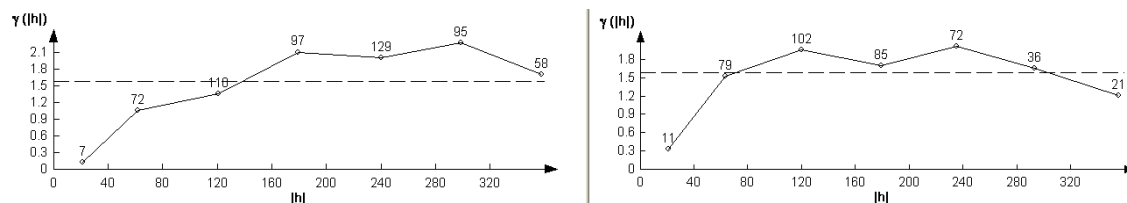


Figura 56- Semivariogramas das direções 0° e 45° para a amostra de dados pontuais.

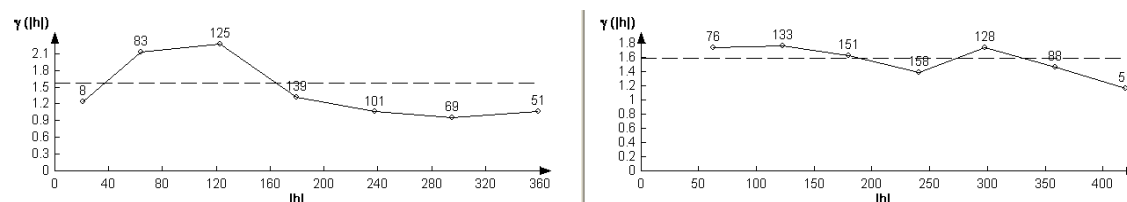


Figura 57- Semivariogramas das direções 90° e 135° para a amostra de dados pontuais.

Com base nos variogramas acima, percebe-se que as direções 90° e 45° representam os eixos de menor e maior continuidade, respectivamente. Portanto, a modelagem da anisotropia será feita baseada nessas duas direções. A Figura 58 apresenta o ajuste de um modelo teórico para cada uma das direções. Em ambas as direções, utilizou-se o modelo Exponencial e o valor de 0,224 para o efeito pepita.

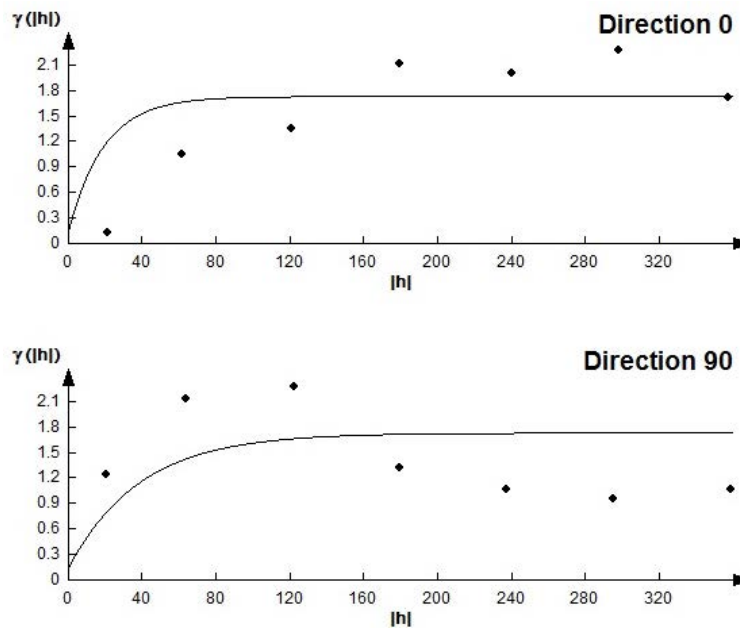


Figura 58- Modelagem da anisotropia para a amostra dos dados pontuais.

O tipo de anisotropia encontrado foi a anisotropia geométrica, em que a forma e o patamar dos semivariogramas são iguais e apenas o alcance é diferente. As duas direções, 90° e 45°, possuem patamar 1,6 e alcances igual à 120 e 240 respectivamente. Sendo assim, ao dividir o valor dos alcances nas direções de menor e maior continuidade, obteve-se o valor da Razão de Anisotropia igual a 0,5.

Após a modelagem, a inferência dos dados de clorofila-a por meio da Krigagem Ordinária foi realizada através do *software* R. Para isso foi gerada uma máscara com resolução de 5 metros, resultando em um total de 84 linhas e 90 colunas, com isso foram geradas 7560 estimativas. O resultado da krigagem está disposto na Figura 59.

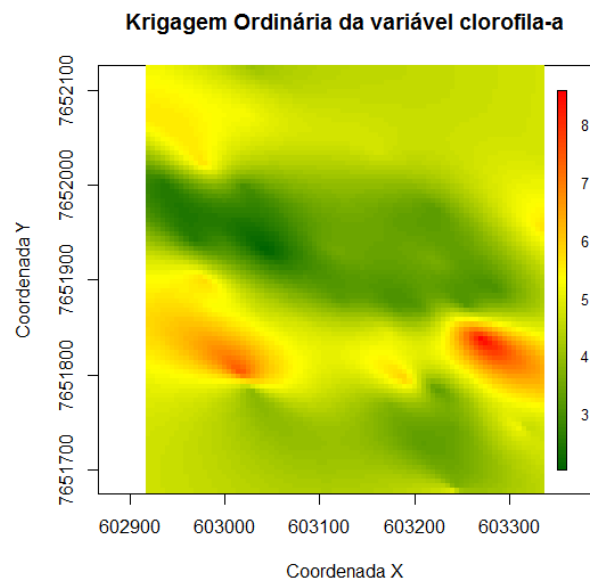


Figura 59 - Interpolação espacial por Krigagem Ordinária da amostra de dados pontuais de clorofila-a.

Também construiu-se o mapa de variância do erro de estimação (Figura 60), verificando que a variância varia entre 0,0017 (verde escuro) e 0,533 (vermelho). Percebe-se que a variância tende a aumentar nos locais onde não foram amostrados, como a parte esquerda inferior.

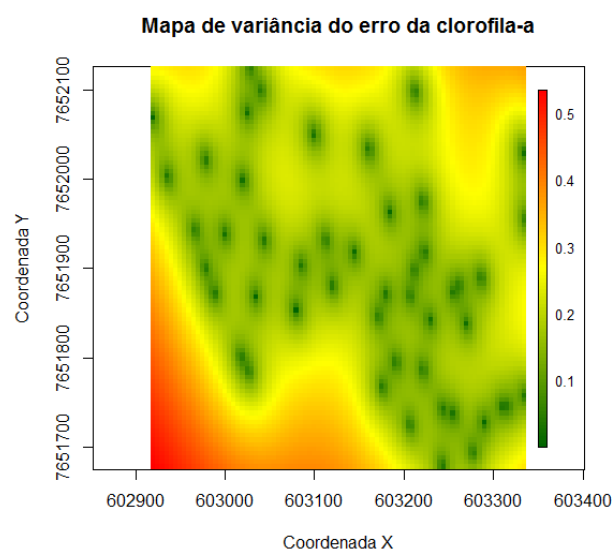


Figura 60- Mapa de variância do erro para a amostra dos dados pontuais.

Feito isso, gerou-se um fatiamento das predições em 3 classes, de 2 à 4, de 4 à 6 e de 6 à 8,5 ppb. A Figura 61 apresenta o resultado dessa divisão em classes da

Krigagem Ordinária para a toda área de estudo. Pode-se perceber que em quase toda a área as concentrações de clorofila-a variam de 2 a 6 ppb, com apenas duas pequenas regiões onde os valores são maiores.

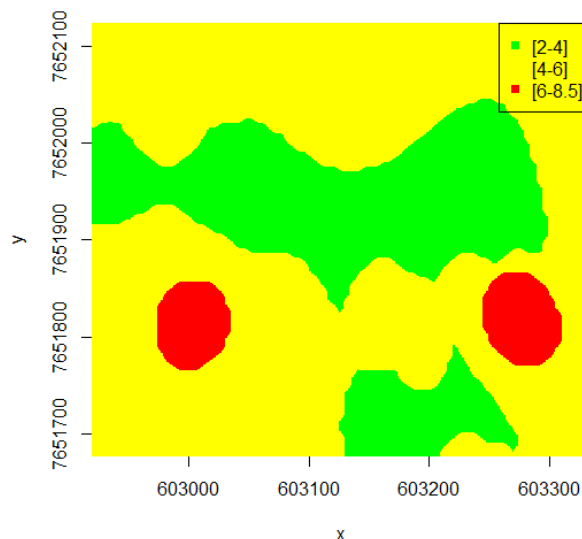


Figura 61- Mapa de classes da Krigagem Ordinária dos dados pontuais.

Para validar os resultados foi calculado o Erro Médio Quadrático (EMQ), dando resultado igual a 0,70, valor considerado satisfatório e inferior ao obtido por Utsumi (2012) em seu estudo (1,55). Também, verificou-se pela Figura 60, que a variância do erro de estimação neste caso foi bem menor, o valor máximo foi de 0,70 e nos estudo de Utsumi (2012) foi de 3,50. Tudo isso levando em consideração que foram utilizadas 10 observações amais neste estudo.

5 CONCLUSÕES E RECOMENDAÇÕES

O objetivo principal deste trabalho foi alcançado, sendo possível analisar o impacto que diferentes disposições espaciais dos elementos amostrais podem causar nos resultados da inferência espacial da clorofila-a.

Foram analisados três tipos de dados, dados em transecto, malha de dados simulados e dados pontuais. Nos dados em transecto, a partir da redução do conjunto de dados através da amostragem sistemática, percebeu-se que uma amostra menor seria suficiente para representar o conjunto de dados, uma vez que ao comparar as médias e as variâncias de todas as amostras com as da população, seus valores eram muito semelhantes. Além disso, ao analisar as distribuições espaciais das amostras percebeu-se que mesmo diminuindo o tamanho da amostra, o formato do transecto não sofreu grandes alterações.

De acordo com o EMQ e o Índice Kappa, a amostra de tamanho 400 obteve uma leve vantagem em relação às outras amostras. Porém, de forma geral, os resultados não foram satisfatórios, talvez porquê ao coletar dados em transectos são deixados muitos espaços sem coletar, isso faz com que a variância do erro de estimação seja alta nessas regiões.

Com os dados simulados, o intuito era imaginar uma situação real com uma malha cheia de dados e aplicar os três métodos de amostragem com diferentes tamanhos de amostras, e com isso analisar a influência da escolha do método e do tamanho nos resultados da inferência espacial. Os resultados do EMQ e do Índice Kappa mostraram que, de modo geral, a Amostragem Sistemática forneceu os melhores resultados na inferência espacial. Isso porquê vários autores, entre eles Olea (1984) apud Englund (1988) e Yamamoto e Landim (2013), já confirmaram que na presença de autocorrelação espacial, a amostragem sistemática sobre um gride produz resultados mais eficientes.

Em ambas as análises, com os dados em transecto e com os dados simulados, notou-se a influência do tamanho da amostra nos resultados, visto que, em termos de resultados do EMQ e do Índice Kappa, as amostras maiores forneceram melhores resultados nas inferências espaciais. Porém, mesmo para amostras menores as

estimativas foram consideradas satisfatórias. Na análise do conjunto menor, foi possível perceber a importância do tamanho da amostra em uma pesquisa, pois um pequeno aumento de 20% no tamanho da amostra utilizado por Utsumi (2012), de 40 para 50, gerou uma melhora considerável nos resultados ao comparar os resultados do EMQ e da variância do erro de estimação.

Os resultados obtidos neste trabalho não podem ser generalizados para qualquer pesquisa, pois cada conjunto de dados e cada variável têm suas particularidades, portanto os resultados da inferência espacial podem variar.

Recomenda-se para os próximos estudos testar outros métodos de inferência espacial além da Krigagem Ordinária, para verificar se ao mudar o método de predição as conclusões sobre os métodos de amostragem também mudariam. Seria interessante também aplicar os métodos estatísticos de amostragem em outro conjunto de dados distinto, que não seja em transecto, e com tamanhos de amostras diferentes, com isso seria possível verificar se a amostragem sistemática também seria a melhor opção nesse caso.

Também seria interessante aplicar no conjunto de dados em transecto outros métodos de inferência espacial que não sejam geoestatísticos, como por exemplo: inverso do quadrado da distância (IQD), vizinho mais próximo, média simples, média ponderada, etc. Assim, seria possível comparar com os resultados obtidos pela Krigagem Ordinária.

Outra sugestão seria testar outras divisões de estratos para a amostragem estratificadas através de outras técnicas, como Quad-Tri, ou a partir de imagens da própria região.

REFERÊNCIAS

- AES TIÊTE. Apresenta informações sobre o reservatório de Nova Avanhandava. 2013. Disponível em: <<http://www.aestiete.com.br/usinas/Paginas/NovaAvanhandava.aspx>>. Acesso em 06 de out. de 2013.
- ANDRIOTTI, J. L. S. **Fundamentos de estatística e geoestatística**. São Leopoldo: Editora Unisinos, 2003, p. 165.
- ANSELIN, L. **Exploring spatial data with GeoDaTM: a Workbook**. Urbana-Champaign: University of Illinois, 2005. Disponível em <<http://geodacenter.asu.edu/system/files/geodaworkbook.pdf>>. Acesso em 10 jan. 2012.
- ARRAUT, E. M. et al. Estudo do comportamento espectral da clorofila e dos sólidos em suspensão nas águas do lago grande de Curuai (Pará), na época da seca, através de técnicas de espectroscopia de campo. In: **Simpósio Brasileiro de Sensoriamento Remoto**, XII, Anais, 2005. Goiânia: 16-21 Abr 2005, INPE, p. 2447-2456.
- ASSUNÇÃO, R. M. **Estatística Espacial em Epidemiologia, Economia, Sociologia**. 2001. Disponível em <<http://www.est.ufmg.br/leste/publicacoes.htm#2001>>. Acesso em 21 dez. 2013.
- BAILEY, T.C., GATRELL, A.C. **Interactive Spatial Data Analysis**. Essex: Longman Scientific and Technical, 1995.
- BOLFARINE, H., BUSSAB, H. O. **Elementos de Amostragem**. Edgard Blücher, 2005.
- BURROUGH, P. A. **Principles of geographical information systems for land resources assessment**. Oxford: Claredon Press, 1986, 194 p.
- CAEIRO, S.; PAINHO, M.; GOOVAERTS, P.; COSTA, H.; SOUSA, S. Optimal spatial sampling design for mapping estuarine sediment management areas. In: **V AGILE Conference on Geographic Information Science**, 2002, Palma.
- CÂMARA, G. et al. **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA, 2004 (ISBN: 85-7383-260-6). Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/cap3-superficies.pdf>>. Acesso em 12 jan. 2013.
- CAMARGO, E. C. G. **Desenvolvimento, implementação e teste de procedimentos geoestatísticos (Krigagem) no Sistema de Processamento de informações Georreferenciadas (SPRING)**. Dissertação de mestrado (Mestrado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1997.

CAMARGO, E. C. G. et al. **Integração de Geoestatística e Sistemas de Informação Geográfica: Uma necessidade.** São José dos Campos: INPE, 2000.

CAMARGO, E. C. G. **Curso: Geoprocessamento para Projetos Ambientais.** Capítulo 5- Geoestatística: Fundamentos e aplicações. São José dos Campos: INPE, 1996. Disponível em: <http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/>. Acesso em: 22 fev. 2014.

CASTRO, F. S. et al. Avaliação do desempenho dos diferentes métodos de interpoladores para parâmetros do balanço hídrico climatológico. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.14, n.8, 2010.

CETESB (São Paulo). Companhia de Tecnologia de Saneamento Ambiental. **Relatório de qualidade das águas interiores do Estado de São Paulo 2003.** CETESB, São Paulo, 273 p., 2004. Disponível em: <http://www.mpsp.mp.br/portal/page/portal/cao_urbanismo_e_meio_ambiente/bibliotec_a_virtual/bv_informativos_tecnicos/Relat%C3%B3rio%20Anual.pdf>. Acesso em: 10 jan. 2015.

CETESB (São Paulo). **Manual de cianobactérias planctônicas: legislação, orientações para o monitoramento e aspectos ambientais.** CETESB, São Paulo; CARVALHO M. C. et al., 47 p., 2013. Disponível em: <<http://www.cetesb.sp.gov.br/userfiles/file/agua/aguas-superficiais/manual-cianobacterias-2013.pdf>>. Acesso em: 02. jan. 2015.

CICERELLI, R. E. **Estudo da ocorrência de cianobactérias em ambiente aquático continental por meio da inferência espacial do pigmento ficocianina.** 2013. 165 f. Tese (Doutorado em Ciências Cartográficas) Faculdade de Ciência e Tecnologia - UNESP, Presidente Prudente.

COCHRAN, W. G. **Sampling Techniques.** 3.ed. New York: John Wiley & Sons, 1977.

ENGLUND, E. J. **Spatial Autocorrelation: Implications for sampling and estimation.** U.S. Environmental Protection Agency, Las Vegas, 1988. Disponível em: <www.epa.gov/esd/cmb/research/papers/ee107.pdf>. Acesso em: 06 jan. 2015.

FERREIRA, M. C. **Inferência espacial de clorofila-a por redes neurais artificiais aplicadas a imagens multiespectrais e medidas tomadas *in situ*.** 2011. 96 f. Dissertação (Mestrado em Ciências Cartográficas) - Universidade Estadual Paulista, Faculdade de Ciências e Tecnologia, Presidente Prudente.

FERREIRA, T. F.; FRAGOSO, C. R. J. MARQUES, D. M. Efeito da Liberação de Nutrientes por Plantas Aquáticas sobre a Dinâmica de Estados Alternativos da Comunidade Fitoplânctônica em um Lago Raso Subtropical. **Revista Brasileira de Recursos Hídricos**, v. 13,p. 151-160, 2008.

GENÚ, A. M. **Geoestatística Multivariada**. Piracicaba: Universidade de São Paulo, 2004. Disponível em: <<http://www.lce.esalq.usp.br/tadeu/aline.pdf>>. Acesso em: 18 jan. 2014.

HAINING, R. **Spatial data analysis in the social and environmental sciences**. Cambridge: University Press, 1990, p. 432.

HALL, D. O.; RAO, K. K. **Fotossíntese**. São Paulo: EPU, 1980, v.10, p. 89.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989.

KIRK, J. T. O. **Light and photosynthesis in aquatic ecosystems**. 2 ed. Cambridge: University Press, 1994.

KOTZ, S.; JOHNSON, N. L.; READ, C. B. **Encyclopedia of statistical sciences**. New York: John Wiley & Sons, 1983, v. 4, p. 352.

LANDIM, P. M. B. **Análise estatística de dados geológicos**. São Paulo: Editora UNESP, 1998.

LANDIM, P.M.B., STURARO, J.R., MONTEIRO, R. C. **Exemplos de aplicação da cokrigagem**. DGA, IGCE, UNESP/Rio Claro, Lab. Geomatemática, Texto Didático 09, 17 pp. 2002. Disponível em:<<http://www.rc.unesp.br/igce/aplicada/DIDATICOS/LANDIM/cokrigagem.pdf>>.

LANDIS, J. R., KOCH, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v.33, p.159-75, 1977.

LIU, Y.; CHEN, Y.; CHENG, J. A comparative study of optimization methods and conventional methods for sampling design in fishery independent surveys. **ICES Journal of Marine Science**, v. 66, p. 1873–1882, 2009.

LUIZ, A. J. B.; EPIPHANIO, J. C. N. Amostragem por pontos em imagens de sensoriamento remoto para estimativa de área plantada por município. In: **XSimpósio Brasileiro de Sensoriamento Remoto**, Anais, 2001, Foz do Iguaçu.

LUIZ, A. J. B. **Estatísticas agrícolas por amostragem auxiliadas pelo sensoriamento remoto**. 2003. 112 f. Tese (Doutorado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos.

MARTINI, et al. Uso de sensoriamento remoto orbital para avaliação da distribuição espacial de clorofila-a na Lagoa da Conceição – Florianópolis-SC. **Eng. San. Ambiental**, v.11, n.4, p. 318-324, 2006.

MSPC – INFORMAÇÕES TÉCNICAS. Apresenta informações sobre Probabilidade e Estatística. 2008. Disponível em: <http://www.mspc.eng.br/matm/prob_est320.shtml>. Acesso em: 5 out. 2013.

NETO, J. C. P. C. **Determinação do tamanho da amostra**. 2004. Disponível em: <http://fesppr.br/~centropesq/Calculo_do_tamanho_da_amostra/Tamanho%20da%20Amostra%20-%201.pdf>. Acesso em: 15 dez. 2014.

OLEA, E. A. Sampling Design Optimization for Spatial Functions. **Mathematical Geology**, v.16, n.4, p. 369-389, 1984.

PAN, X. et al. Remote sensing of pic phytoplankton distribution in the northern South China. **Sea Remote Sensing of Environment**, v. 128, p. 162– 175, 2013.

PEREIRA, A. C. F. et al. Amostragem em corpos d'água: definição de elementos amostrais, posicionamento e coleta de dados “in situ”. In: **II Simpósio Brasileiro de Geomática**, 2007, Presidente Prudente.

PERROCA, M. G.; GAIDZINSKI, R.R. Avaliando a confiabilidade interavaliadores de um instrumento para classificação de pacientes – coeficiente Kappa. **Revista da Escola de Enfermagem da USP**, v. 37, n.1, p. 72-80, 2003.

PINTO, A. F.; VON SPERLING, E. Desenvolvimento da metodologia para determinação da clorofila “a” através de medidas contínuas da fluorescência emitida pela comunidade fitoplanctônica. In: **XX Congresso Brasileiro de Engenharia Sanitária e Ambiental**. 1999, Rio de Janeiro.

PINTO, et al. **Métodos para Estimação de Reprodutividade de Medidas: Estatística Kappa (K)**. Faculdade de Medicina do Porto. 2014. Disponível em: <<http://users.med.up.pt/joakim/intromed/estatisticakappa.htm>>. Acesso em: 12 jan. 2014.

POVOA, et al. Rotinas para Análises Geoestatísticas Utilizando a Linguagem R: um exemplo com dados agro-ambientais. In: **II Simpósio de geoestatística aplicada em ciências agrárias**, 2011, Botucatu.

ROTTA, L. H. S.; FERREIRA, M. S.; IMAI, N. N.; TACHIBANA, V. M. Análise espacial de macrófitas submersas no reservatório de Porto Colômbia. In: **III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**, 2010, Recife.

ROSSONI, D. F. **Análise de variância para experimentos com dependência espacial**. 2011. 108 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG, Brasil, 108p.

SAMIZAVA, T. M.; IMAI, N. N.; ROTTA, L. H. S.; FERREIRA, M. S.; GALO, M. L. B. T.; ROCHA, R. R. A.; ENNES, R.. Proposta de delineamento amostral para levantamento de medidas de variáveis limnológicas e de dados espectrorradiométrico

sem planície de inundação. In: **II Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**, 2008, Recife.

SANTOS, P. S.; EPIPHANIO, J. C. N. Aprimoramento do método de amostragem simples utilizado pelo Projeto Geosafras para estimativa municipal de área plantada com soja. In: **Anais XIII Simpósio Brasileiro de Sensoriamento Remoto**, 2007, Florianópolis.

SHEAFER, R. L.; MENDEKALL, W.; OTT, L. **Elementary survey sampling**. 3. ed. Boston: PWS Publishing Company, 1986.

SILVA, S. U. **Análise Estatística Espacial dos Óbitos por Câncer de Traquéia, Brônquios, Pulmão e Estômago Registrados para o Estado de São Paulo de 1995 a 2010**. 2011. Trabalho de Conclusão de Curso (Graduação em Estatística) – Faculdade de Ciência e Tecnologia, Universidade Estadual Paulista Júlio de Mesquita Filho, Presidente Prudente.

THOMPSON, S. K. **Sampling**. 2. ed. New York: Wiley, 2002.

TURNER DESIGNS. **Cyanobacteria detection in water using *in vivo* fluorometry**. 2004. Disponível em: <http://www.turnerdesigns.com/t2/doc/appnotes/algae.html>. Acesso em: 8 de abril de 2014.

UTSUMI, A. G. **Inferência espacial de cianobactérias a partir da fluorescência in vivo do pigmento ficocianina**. 2012. 110f. Dissertação (Mestrado em Ciências Cartográficas) - Universidade Estadual Paulista, Faculdade de Ciências e Tecnologia, Presidente Prudente.

WANG, J. F. et al. A review of spatial sampling. **Spatial Statistics - Elsevier**, 2012. doi:10.1016/j.spasta.2012.08.001.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for environmental scientists**. England: John Wiley & Sons, 2007.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: conceitos e aplicações**. 1. ed. São Paulo: Oficina de Textos, 2013.