

ALVARO MARI JUNIOR

**GEOESTATÍSTICA NA AVALIAÇÃO DE TEOR DE CLOROFILA EM AVEIA
PRETA**

Botucatu

2018

ALVARO MARI JUNIOR

**GEOESTATÍSTICA NA AVALIAÇÃO DE TEOR DE CLOROFILA EM AVEIA
PRETA**

Tese apresentada à Faculdade de Ciências Agronômicas da Unesp Campus de Botucatu, para obtenção do título de Doutor em Agronomia (Irrigação e drenagem).

Orientador: Hélio Grassi Filho

Coorientadora: Célia Regina Zimback

Botucatu

2018

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO DA INFORMAÇÃO - DIRETORIA TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - UNESP - FCA - LAGEADO - BOTUCATU (SP)

M332g Mari Junior, Alvaro, 1988-
 Geoestatística na avaliação de teor de clorofila em
 aveia preta / Alvaro Mari Junior. - Botucatu: [s.n.],
 2018
 74 p.: ils. color., grafs. color., tabs.

 Tese (Doutorado)- Universidade Estadual Paulista, Fa-
 culdade de Ciências Agrônômicas, Botucatu, 2018
 Orientador: Hélio Grassi Filho
 Coorientadora: Célia Regina Zimback
 Inclui bibliografia

 1. Aveia. 2. Imagens de satélite. 3. Geoprocessamento.
 4. Correlação de dados. 5. Aprendizado de máquinas. 6.
 Clorofila. I. Grassi Filho, Hélio. II. Zimback, Célia
 Regina. III. Universidade Estadual Paulista "Júlio de
 Mesquita Filho" Câmpus de Botucatu). Faculdade de Ciên-
 cias Agrônômicas. IV. Título.

Elaborada por Ana Lucia G. Kempinas - CRB-8:7310

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte"

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: GEOESTATÍSTICA NA AVALIAÇÃO DE TEOR DE CLOROFILA EM AVEIA PRETA

AUTOR: ALVARO MARI JUNIOR

ORIENTADOR: HÉLIO GRASSI FILHO

COORDINADORA: CÉLIA REGINA LOPES ZIMBACK

Aprovado como parte das exigências para obtenção do Título de Doutor em AGRONOMIA (IRRIGAÇÃO E DRENAGEM), pela Comissão Examinadora:



Prof. Dr. HÉLIO GRASSI FILHO
Depto de Solos e Recursos Ambientais / Faculdade de Ciências Agronômicas de Botucatu - UNESP



Profa. Dra. MARIA HELENA MORAES SPINELLI
Depto de Solos e Recursos Ambientais / Faculdade de Ciências Agronômicas de Botucatu, SP



Prof. Dr. PAULO MILTON BARBOSA LANDIM
Depto de Solos e Recursos Ambientais / UNESP - Faculdade de Ciências Agronômicas de Botucatu



Prof. Dr. DIEGO AUGUSTO DE CAMPOS MORAES
Depto de Análise e Desenvolvimento de Sistemas / Faculdade Eduvale de Avaré



PROF. DR. ALINE AZEVEDO NAZARIO
ÁGUA E SOLO/ REÚSO DE ÁGUA NA AGRICULTURA / CENTRO UNIVERSITÁRIO ADVENTISTA DE SÃO PAULO - ENG. COELHO

Botucatu, 08 de maio de 2018

AGRADECIMENTOS

A Deus, antes de qualquer coisa;

A meus pais Alvaro Mari e Edemari Mari por tudo que fizeram por mim até aqui e que ainda farão;

A meu irmão Angelo Gabriel Mari pela parceria;

A minha esposa Ana Claudia Cabral Mari por, além de me fazer feliz, suportar os momentos finais de tensão na elaboração desta Tese;

A minha filha Camile Cabral Mari, por existir em minha vida;

Aos professores Dr. Hélio Grassi Filho e Dra. Célia Regina Lopes Zimback pela orientação, confiança e amizade;

Aos meus amigos, Diego Moraes, Roberto Figueiras e Victor Costa Leda, pela grande ajuda que me prestaram em meus estudos;

Aos amigos Jannaylton, Elvis, Leandro e Marcos pelo auxílio em campo nas primeiras coletadas de dados;

A Faculdade de Ciências Agronômicas da UNESP e ao Programa de Irrigação e Drenagem pela oportunidade;

A Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) pela concessão da bolsa de estudos;

Grato!

RESUMO

A atual demanda por informações precisas e de baixo custo de obtenção causa um aumento na procura por serviços remotos, como o geoprocessamento, a geoestatística e o aprendizado de máquinas. Este trabalho objetiva a obtenção de informações relacionadas a leituras de teores de clorofila, a partir de um banco de dados de campo e por imagens de satélite aplicadas a uma área de cultivo de aveia preta, com baixos custos. Com um banco de dados preciso sobre o teor de clorofila, acredita-se ser possível reduzir custos na implantação e execução de sistemas de irrigação. Diversos autores já realizaram a comparação entre o teor de clorofila e a disponibilidade de água e outros já pontuaram o grande benefício da irrigação sobre diversos cultivares. Com a hipótese de que é possível através da utilização de técnicas de baixo custo a obtenção de valores de teor de clorofila, o trabalho utilizou de técnicas de geoestatística para procurar uma correlação entre os dados físicos e os índices de vegetação gerados a partir de imagens do satélite Landsat 5. Ao encontrar resultados insatisfatórios, partiu-se para um estudo dos dados em treinamento de máquinas e mineração de dados. E após passar por vários processos, o algoritmo *Gradient Boosting Machine (GBM)* possibilitou a geração de uma imagem com valores estimados clorofila bastante semelhante aos reais obtidos em campo.

Palavras-chave: Imagens de satélite. Correlação de dados. Aprendizado de máquinas. Clorofila; Geoestatística

ABSTRACT

The nowadays demand for accurate and low-cost obtained information has increased the demand for remote services, such as geoprocessing, geostatistics and machine learning. This work aims to increase the accuracy of information related to readings of chlorophyll content from a field database and satellite images, applied to a black oat growing area. With a precise database on chlorophyll content, it is believed that it is possible to reduce costs in the implementation and execution of irrigation systems. Several authors have already made the comparison between chlorophyll content and water availability and others have already pointed out the great benefit of irrigation on several cultivars. With the hypothesis that is possible through the use of low cost techniques the obtaining values of chlorophyll content, the work used geostatistics techniques to search for a correlation between the physical data and the vegetation indexes generated from images of Landsat 5 satellite .When finding unsatisfactory results, we started with a study of data in machine training and data mining. And after passing through several processes, the Gradient Boosting Machine (GBM) algorithm allowed the generation of an image with estimated chlorophyll values very similar to those obtained in the field.

Keywords: *Satellite Imagery. Data Correlation. Machine Learnig. Chlorophyll. Geoestatistics*

LISTA DE FIGURAS

Figura 1 – Aveia preta 116 dias após Semeadura.....	20
Figura 2 – Variável regionalizada aleatórias.....	26
Figura 3 – Variograma experimental e modelo matemático ajustado.....	29
Figura 4 – Modelo explanatório de plano de regressão	34
Figura 5 – Exemplo de funcionamento da Random Forest.....	37
Figura 6 – Localização da área de estudo, Brasil, São Paulo, Botucatu...	39
Figura 7 – Etapas metodológicas da pesquisa.....	41
Figura 8 – Malhas amostrais das áreas estudadas.....	42
Figura 9 – Variograma do teor de clorofila (Cascalheira)	50
Figura 10 – Variograma do teor de clorofila (Fazendinha)	51
Figura 11 – Pontos de coleta de campo, repetindo informações do Tiff..	52
Figura 12 – Variograma das informações de NDVI (Cascalheira).....	52
Figura 13 – Variograma das informações de NDVI (Fazendinha).....	53
Figura 14 – Variograma das informações de SAVI (Cascalheira).....	53
Figura 15 – Variograma das informações de SAVI (Fazendinha).....	54
Figura 16 – Validação cruzada Cascalheira (teor de clorofila X NDVI).....	54
Figura 17 – Validação cruzada Fazendinha (teor de clorofila X NDVI).....	55
Figura 18 – Validação cruzada Cascalheira (teor de clorofila X SAVI).....	55
Figura 19 – Validação cruzada Fazendinha (teor de clorofila X SAVI).....	55
Figura 20 – Resultados da Krigagem na área da Fazendinha	56
Figura 21 – Resultados da Krigagem na área da Cascalheira	58
Figura 22 – Gráfico da predição de dados.....	59
Figura 23 – Gráficos das métricas estatísticas no modelo GBM.....	60
Figura 24 – Imagem referente aos dados estimados pelo Gradient Boosting Machine	61
Figura 25 – Comparação do valor do teor de clorofila real e o predito na Cascalheira	62

Figura 26 – Comparação do valor do teor de clorofila real e o predito na Fazendinha	62
Figura 27 – Coeficiente de determinação (Real X Simulado (GBM) na Cascalheira	63
Figura 28 – Coeficiente de determinação (Real X Simulado (GBM) na Fazendinha.....	63

LISTA DE TABELAS

Tabela 1 - Dados físicos inseridos no programa	45
Tabela 2 - Dados espectrais inseridos no programa	45
Tabela 3 - Dados no programa após Band-Ratio	45
Tabela 4 - Dados no programa após exclusão de “colinearidade > 0,98..	46
Tabela 5 - Dados no programa após exclusão do Recursive feature elimination.....	47
Tabela 6 - Estatística descritiva dos dados de campo	59
Tabela 7 - Métricas estatísticas para cada modelo	60

SUMÁRIO

1	INTRODUÇÃO	17
2	REVISÃO DE LITERATURA	19
2.1	Aveia-Preta.....	19
2.2	Demanda hídrica.....	20
2.3	Teor de clorofila.....	21
2.4	Geoprocessamento	22
2.5	Índices de vegetação	23
2.5.1	Índice de vegetação da diferença normalizada (NDVI)	23
2.5.2	Índice de vegetação ajustado ao solo (SAVI)	24
2.6	Geoestatística.....	25
2.6.1	Teoria das variáveis regionalizadas	25
2.6.2	Variograma.....	28
2.6.3	Krigagem ordinária	30
2.7	Geoestatística multivariada.....	31
2.8	Mineração de dados (Data mining)	32
2.8.1	Regressão linear múltipla	33
2.8.2	Rede neural artificial (Artificial neural network)	34
2.8.3	Máquina de vetores de suporte (Support vector machine)	35
2.8.4	Florestas aleatórias (Random Forest)	36
2.8.5	Cubist	37
2.8.6	Aprimoramento gradativo de máquina (Gradient Boosting Machine)	37
3	MATERIAL E MÉTODOS	39
3.1	Descrição da área.....	39
3.2	Material.....	40
3.3	Metodologia	40
3.3.1	Obtenção de dados de campo (Dados físicos).....	41
3.3.2	Obtenção dos índices de vegetação (Imagens de satélite)	42
3.3.3	Análises geoestatística.....	43
3.3.4	Treinamento de máquinas.....	44
3.3.5	Geração de mapa de predição	48

4	RESULTADOS E DISCUSSÃO.....	49
4.1	Dados de campo	49
4.3	Índices de vegetação	49
4.4	Análises geoestatística	50
4.5	Treinamento de máquinas.....	59
4.6	Geração de mapa de predição	61
5	CONCLUSÕES.....	65
	REFERÊNCIAS	66

1 INTRODUÇÃO

A aveia preta é um cultivar com muitas utilidades, uma destas é a alimentação animal, o que a torna ainda mais vantajosa é o número elevado de cortes possíveis a realizar, esta característica é ainda mais importante, já que permite a utilização de uma mesma safra para alimentação de plantéis por um período maior de tempo. Cabe destacar ainda, que por ser uma forrageira de inverno, gera alimento para os animais em uma época do ano em que se costuma ter menos disponibilidade de grãos, quando cultivada acaba reduzindo a pressão sobre o solo causada pela monocultura, já que consome nutrientes diferentes que os cultivos convencionais.

Com os avanços tecnológicos, existentes nas mais diversas áreas de desenvolvimento e pesquisa, foram possíveis o desenvolvimento de técnicas computacionais extremamente úteis, entre elas o geoprocessamento, a geoestatística e o aprendizado de máquina, para diversas áreas de pesquisa, estas técnicas são amplamente utilizadas em estudos agronômicos.

Nesse contexto, tendo em vista a importância destas três técnicas, pode-se destacar a geoestatística por facilitar análises de características relacionadas às variáveis espacialmente dependentes. O que torna possível a verificação de relação entre diversos dados amostrados de solo, plantas, folhas, ou outras variáveis espacialmente dependentes.

Outra técnica que vem ganhando espaço em pesquisas é o aprendizado de máquinas, já que este é capaz de encontrar informações normalmente ocultas em bancos de dados, permitindo análises interessantes e incalculáveis pelas técnicas estatísticas comuns.

O geoprocessamento por sua vez, é uma ferramenta importante para trabalhos com dados georreferenciados, e é através de sua utilização que se executam estudos sobre imagens de satélites, que tornam possível a obtenção de informações importantes sobre uma região, como características geológicas, hidrográficas, e, além disso, no campo espectral é possível definir índices de vegetação, que servem para verificar o comportamento espectral das plantas.

Algumas vezes, quando se possui diversas informações sobre uma região, através dos já citados processos tecnológicos, torna-se possível realizar uma correlação entre diferentes dados, tornando alguma das variáveis melhor definida.

Essa saída mostra-se extremamente útil quando a variável analisada não mostra diretamente uma dependência espacial. Desta forma utiliza-se uma variável que tenha correlação com esta e que mostre dependência espacial.

No contexto da irrigação, vale destacar que existe relação entre o teor de clorofila de uma planta e a disponibilidade hídrica a qual a mesma está exposta, o que evidencia a importância de possuir informações precisas quanto ao teor de clorofila, e como qualquer atividade, informações de baixo custo são preferíveis.

Esta pesquisa foi elaborada com o objetivo de encontrar dados referentes a teores de clorofila em uma área de cultivo de aveia preta utilizando técnicas de baixo custo. Este trabalho tem como hipótese a possibilidade da utilização de técnicas indiretas para obtenção de valores de teor de clorofila, o que viabiliza a utilização destas informações para estudos de irrigação.

Este trabalho objetivou também:

- Analisar os dados de campo em cultivo de aveia preta;
- Estudar a dependência espacial dos dados de clorofila;
- Estudar a correlação dos dados de clorofila com os índices de vegetação;
- Estudar as seguintes técnicas de aprendizado de máquina e analisar seus resultados: Regressão linear múltipla, Redes neurais artificiais, Máquina de vetores de suporte, Floresta aleatória, Cubist e Aprimoramento gradativo de máquina);

2 REVISÃO DE LITERATURA

2.1 Aveia-Preta

A aveia-preta (*Avena strigosa* Schreb) é classificada, de acordo com Sá (1995), como sendo uma gramínea de clima temperado que se adapta bem a diferentes condições climáticas. Ela pode ser utilizada tanto como planta de cobertura, como para alimentação animal e humana, além de algumas vezes ser utilizada para adubação e utilização em efeitos alelopáticos através do efeito de inibição em plantas invasoras, reduzindo desta forma custos com capina.

Durante o período de inverno, a aveia-preta mostra-se uma alternativa bastante viável para forragem (Floss, 1988). Primavesi (2000) afirmou que durante épocas de inverno seco, a aveia pode ser utilizada e desta forma gerando uma diversificação para colheita, além de aproveitar áreas durante a entressafra, devido às suas adaptações, tanto em regiões irrigadas, quanto nas não irrigadas.

Hartwig et al. (2006) citaram que a aveia-preta é muito importante para sistemas agrícolas, pois possui, do ponto de vista de sistemas de manejo, forte potencial de exploração. Fontanelli e Piovezan (1991) reforçaram a importância desta planta para produção de forragem e pastagem, e citaram, ainda, algumas características interessantes da aveia, como a tolerância que possui a acidez, rapidez de formação de cobertura, menor suscetibilidade às moléstias, alto índice de produção de massa seca.

A planta apresenta baixa qualidade industrial, possuindo baixa produção de sementes devido à coloração escura das sementes, a Figura 1 ilustra a planta após o centésimo décimo sexto dia de sua semeadura. Cabe citar, ainda, que a mais importante das características que se pode apontar na aveia forrageira é a possibilidade de números mais elevados de corte, podendo ser cultivada no inverno, o que a torna importante para alimentação animal, já que nesta época existe escassez de pasto (SÁ, 1995).

Figura 1 - Aveia preta 116 dias após semeadura



Fonte: EMBRAPA

2.2 Demanda hídrica

A demanda hídrica de uma cultura é a quantidade de água que a mesma necessita durante o seu ciclo de desenvolvimento, Antunes (2000) e Neto et al. (2001) pontuaram duas variáveis importantes na definição da demanda hídrica de uma cultura sendo elas a evapotranspiração e o coeficiente de cultura (ilustrado comumente com a sigla k_c) e com estas informações é possível estabelecer o momento e quantidade em que a cultura necessita de água.

A evapotranspiração é considerada por Raziei e Pereira (2013) como sendo a mais ativa variável que compreende o ciclo hidrológico, além de ser a componente central do balanço hídrico. Isto a torna um parâmetro extremamente importante em estudos ambientais e para projetos de irrigação (KUMAR et al., 2008). Teixeira e Lima Filho (1997) lembraram que a evapotranspiração possui uma relação direta com as condições climáticas, como temperatura e velocidade do vento.

Já o coeficiente de cultura (K_c) como citado por Medeiros et al. (2004), é um parâmetro que envolve tanto os fatores fisiológicos da planta como ambientais segundo os autores ainda que seja preferencial que seja determinado um valor para a região de estudo. O valor de K_c varia ao longo do desenvolvimento fisiológico da planta. Oliveira et al. (2007) descreveram essa variação como sendo baixa no estabelecimento da cultura e, à medida que a planta desenvolve-se, tende a aumentar até estabilizar. Este coeficiente pode ser considerado como sendo a relação entre a evapotranspiração potencial e a de referência (SOUZA et al., 2011).

Vale lembrar que uma importante ferramenta para o atendimento das demandas hídricas da cultura é o manejo de irrigação, que pode ser utilizado tanto para evitar escassez ou excessos de água (GOMES, 2005).

Uma forma de suprir a demanda hídrica é através de irrigação e sua importância é descrita pela EMBRAPA (2010), já que esta técnica foi utilizada ao longo dos séculos, e tem como principal finalidade a disponibilização de água para os cultivares, visando à produção de forma adequada. A técnica foi aprimorada ao longo dos anos e existem diferentes métodos de irrigação. A empresa destacou ainda a importância do manejo adequado em função dos diversos sistemas de irrigação existentes. Vale salientar que em condições de excesso de umidade no solo, ocorre a redução na disponibilidade de oxigênio para a planta e esta redução acarreta em distúrbios generalizados, afetando, principalmente, a capacidade de absorção de água e nutrientes pela planta (GLINSKI; STEPNIEWSKI, 1986).

Lugão et al. (2011) afirmaram que a irrigação pode ser utilizada para aumentar a produtividade da aveia preta, o que causaria aumento na oferta de forragem, tendo em vista que um dos fatores mais limitantes ao cultivo da aveia é o déficit hídrico. Tal aumento teria por consequência efeito positivo na produção de leite. Os autores, em seu experimento realizado no Noroeste do estado do Paraná, observaram um incremento de até 70% na massa seca. Com a utilização das técnicas de irrigação é possível cultivar a aveia preta em todas as regiões de São Paulo, além de gerar perfeito enchimento dos grãos quando o manejo é feito de forma adequada (PRIMAVESI, 2000).

Para García-Sánchez et al. (2007), o estresse hídrico pode reduzir o teor de clorofila, já que este causa a redução das atividades fotoquímicas da planta.

2.3 Teor de clorofila

Dwyer et al. (1995) afirmaram a existência de uma relação entre o índice de esverdeamento de uma planta e seu teor de clorofila. Alguns outros autores estudam a possibilidade de realizar diagnose nutricional, de diferentes espécies de plantas, a partir da leitura do clorofilômetro (GODOY et al., 2007; DEMOTES-MAINARD et al., 2008).

O baixo teor de clorofila, para Hendry e Price (1993), pode representar estresse ambiental. Streit (2005) concordou, e incrementou afirmando que é devido à instabilidade do pigmento (clorofila), pH, além de outras variáveis que podem também causar a degradação da clorofila, fazendo com que apresentem baixo teor.

O equipamento utilizado para fazer leituras de teor de clorofila, conhecido como clorofilômetro, é um medidor portátil capaz de quantificar a clorofila de uma planta instantaneamente, é um importante equipamento já que os métodos tradicionais envolvem a destruição de amostras de tecido e trabalho extenuante para extração e quantificação da clorofila (ARGENTA, 2001).

O clorofilômetro vem sendo amplamente utilizado em trabalhos científicos para calcular a necessidade de adubação nitrogenada em diversas culturas, como pode ser observados nos trabalhos de Bredemeier (1999) e Argenta et al. (2001).

Para a cultura da aveia preta, o valor médio para o teor de clorofila, de acordo com Pacheco (2012) é de 41,4 unidades SPAD.

2.4 Geoprocessamento

Silva (2003) definiu geoprocessamento como representado por qualquer tipo de processamento de dados com georreferência, como a cartografia, sensoriamento remoto e Sistemas de Informação Geográfica (SIGs).

Os SIGs, de acordo com Cavallari (2007), são ferramentas que permitem uma verificação de dados computacionais geográficos (tanto quantitativa quanto qualitativamente), cuja utilização possibilita o gerenciamento de recursos e aplicação de técnicas que embasadas em diagnósticos georreferenciados. Veiga e Silva (2004) incrementaram a definição de SIGs, como sendo ferramentas que podem manipular feições geográficas e também bancos de dados através de relação espacial.

É importante também destacar o sensoriamento remoto, que Rees (1990) conceitua como sendo a obtenção de dados sobre um alvo sem entrar em contato com o mesmo. Definições mais elaboradas descrevem o sensoriamento remoto como sendo a ciência de se obter informações sobre uma área ou fenômeno sem contato direto com o alvo do estudo, ou seja, adquirindo estas informações por sensores ou dispositivos remotamente (LILLESAND E KIEFER, 1994).

2.5 Índices de vegetação

A cobertura vegetal na imagem, é um produto gerado através de sensoriamento remoto, possui uma aparência que é obtida a partir de um complexo processo envolvendo muitos fatores e parâmetros ambientais. Esta aparência pode variar não somente pelas características da vegetação, mas pode também sofrer influência de parâmetros diferentes, como espelhamento atmosférico, radiação, arquitetura foliar, umidade do solo, sombras e etc. (PONZONI, 2001).

A utilização de diferentes índices de vegetação como o NDVI (Índice de Vegetação da Diferença Normalizada) e o SAVI (Índice de Vegetação Ajustado ao Solo), de acordo com Jensen (2009) pode ser uma forma de facilitar a modelagem das características biofísicas de plantas como área foliar e cobertura do solo. Cabe salientar também a importância do espectro eletromagnético infravermelho, já que nele existem contidas informações referente à evapotranspiração das plantas.

Para Boratto e Gomide (2013), a modelagem dos índices de vegetação tem como base o comportamento da reflectância da vegetação na região do visível, podendo-se dizer, portanto, que a reflectância será menor quando a absorção pelas plantas for maior (processo que ocorre através dos pigmentos fotossintetizantes). Outra informação que pode ser observada é a de que quanto mais adensada for a vegetação de uma região, maior será o índice de reflectância já que existem mais camadas de folhas para refletir.

2.5.1 Índice de vegetação da diferença normalizada (NDVI)

Ramos et al. (2010) citaram o NDVI como sendo um modelo que resulta da combinação de níveis de reflectância em imagens de satélites sendo as bandas utilizadas para a elaboração deste índice o vermelho visível e o infravermelho próximo.

O valor obtido no NDVI é um indicador para a quantidade e condição de uma vegetação estudada. Os valores para o NDVI variam de -1 a 1, sendo que as áreas com água ou cobertas por nuvem aparecem com valores inferiores a 0 (BORATTO, 2013).

A Equação 1 que é a utilizada para o cálculo do NDVI foi apresentada por Rouse et al. (1973) onde:

$$NDVI = \frac{\text{Infra Vermelho próximo} - \text{Vermelho}}{\text{Infra Vermelho próximo} + \text{Vermelho}} \quad (1)$$

Moreira (2003) explicou que a região visível compreende toda radiação eletromagnética da faixa espectral que compreende valores entre 400 e 700 nm. O autor afirmou que grande parte da radiação incidente sobre o dossel é absorvido. Para a região infravermelho o comprimento de onda varia entre 700 nm e 1 mm e divide-se em infravermelho próximo, médio e distante com os comprimentos de onda entre 700 e 1300nm; 1300 a 2500 nm e 2500 a 1 mm, respectivamente.

2.5.2 Índice de vegetação ajustado ao solo (SAVI)

Boratto (2013) definiu o SAVI como sendo um índice que, de forma diferente do que ocorre com o NDVI, considera os efeitos de solo exposto sobre a imagem obtida, ou seja, de certa forma serve como ajuste do NDVI em ocasiões onde não existe cobertura completa do solo. Existe uma variável na equação, fator de ajuste, que serve para modelar diferentes coberturas de solo. Huete (1988) afirmou que os valores devem ser escolhidos entre: 0,25, 0,50, e 1 que representam respectivamente: vegetação densa, vegetação com densidade intermediária e vegetação com baixa densidade. Caso seja adotado o valor 0, o NDVI e o SAVI apresentarão valores idênticos.

A Equação 2, conforme apresentada por Huete (1988), é utilizada para a obtenção do SAVI de uma imagem:

$$SAVI = \frac{(1+FDA) + (\text{Infra Vermelho próximo} - \text{Vermelho})}{FDA + \text{Infra Vermelho próximo} + \text{Vermelho}} \quad (2)$$

onde: FDA é o fator de ajuste.

O valor para FDA utilizado na maioria dos trabalhos é de 0,5, de acordo com Boratto (2013).

Pelo fator de cobertura da área de estudo, o valor adotado para o FDA foi 0,5 também.

2.6 Geoestatística

A geoestatística, de acordo com Andriotti (2002), faz duplo uso de dados obtidos, sendo a primeira utilização no processo de estimação de autocorrelação espacial, e em seguida, o uso para predições. O autor pontua, ainda, a relação desta técnica com outras técnicas estatísticas, já que assim como elas, trata-se de um conceito probabilístico.

Borrungh et al. (1994) lembraram que pesquisadores de solos preocupavam-se com os problemas relacionados à variação espacial e temporal que ocorreram no solo desde o começo do século. Eles afirmaram, ainda, que foi aproximadamente entre os anos 1950 e 1960 que com os avanços teóricos na estatística espacial, foi percebido o potencial destas ferramentas na avaliação da variabilidade do solo.

Para autores como Souza et al. (2004), Trangmar et al. (1985) e Vieira et al. (1983), quando dados de atributos de solo são analisados por estatística clássica, as consequências das diferenças espaciais são ignoradas, porém, quando utilizada, a geoestatística (através da análise de variogramas e krigagem) é capaz de não só analisar a dependência espacial como também realizar a interpolação de atributos do solo através do processo de krigagem. Através desse processo, torna-se possível a verificação de um atributo quanto à existência ou não de uma estrutura espacial (SOUZA et al., 2006).

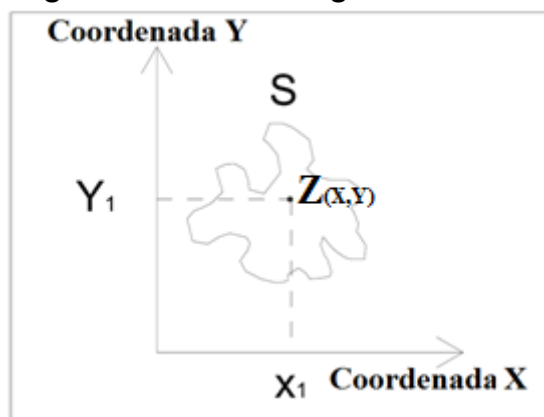
A função da geoestatística aparece quando existe uma variação, entre valores observados, em função da direção e distância em que ocorre o afastamento entre dois pontos de medida. Pode-se sugerir então que existam consistentes diferenças em toda a área estudada, ou seja, não possam ser consideradas como constantes. Segundo Andriotti (2002), algumas das técnicas de geoestatística são: a criação de variogramas (para a descrição e modelagem de padrões espaciais), a krigagem (para obtenção de valores não amostrados) e a variância de krigagem (para estimar incertezas associadas aos valores obtidos pela krigagem).

2.6.1 Teoria das variáveis regionalizadas

Genú (2004), de forma interessante, explicou que na teoria das variáveis regionalizadas, um ponto $Z(x)$ pode ser considerado uma variável aleatória, ou seja, pode assumir diferentes valores em Z , e estes se modificam em função de x dentro

da região definida por S , desta forma geral, o ponto $Z(x)$ é representado por pares de coordenadas $(X_1$ e $Y_1)$, como pode ser observado na Figura 2.

Figura 2 - Variável regionalizada aleatórias



Adaptado de Genú (2004).

Isaak e Srivastava (1989) descreveram as variáveis regionalizadas aleatórias como algo em que a dependência entre os diferentes valores do conjunto pode ser especificada por mecanismos probabilísticos. Cabe citar então que o conjunto de variáveis $Z(x)$, quando medido em toda a área S pode ser considerado uma função aleatória $Z(x)$.

Journel e Huijbregts (1978) afirmaram que, em sua interpretação probabilística, $Z(x)$ é uma realização de uma função aleatória $Z(x)$, apenas quando atende completamente ou parcialmente a lei de distribuição de probabilidade. O fato de que na prática cada um dos possíveis pontos de x possui apenas uma realização $Z(x)$ torna o número de pontos amostrais finito e, isto por sua vez, faz com que na maioria das vezes seja impossível deduzir a distribuição de $Z(x)$. Por decorrência desta dificuldade, torna-se necessária a aplicação de certas hipóteses, comumente conhecidas por hipóteses de estacionaridade, que envolvem vários graus de homogeneidade espacial (GENU, 2004).

Trangmar et al. (1985) definiram a hipótese de estacionaridade de primeira ordem como sendo aquela que apresenta valor constante em toda a área durante o momento de primeira ordem da distribuição da função aleatória $Z(x)$. Genú (2004) expressa pela Equação 3 a definição da função de estacionaridade.

$$E[Z(x)] = E[Z(x + h)] = m \quad (3)$$

onde: m representa a média entre os valores amostrados.

Genú (2004) continuou afirmando que se for tomado um vetor h de separação entre dois pontos, o qual apresenta módulo e direção, para qualquer h tem-se a Equação 4:

$$E [Z(x) - Z(x + h)] = 0 \quad (4)$$

A autora afirmou que considerando a diferença entre as duas variáveis aleatórias como sendo também uma variável aleatória, é possível precisar que o primeiro momento desta variável aleatória pode ser considerado zero. Já estacionaridade de segunda ordem, pode ser definida quando, além de atender a de primeira ordem, a função aleatória possui para cada possível par de valores $[Z(x) Z(x+h)]$, além de existir covariância, é apenas dependente da separação h, o que leva a Equação 5.

$$Cov_{(x,x+h)} = E [Z_{(x)} * Z_{(x+h)}] - [E[Z_{(x)}]][E[Z_{(x+h)}]] \quad (5)$$

O segundo momento da variável aleatória é dado pela diferença entre o quadrado da variância e o quadrado da covariância dos valores como descrito na Equação 6. A função de variância é dada como a metade desta diferença.

$$2\gamma_{(x,x+h)} = E\{[Z_{(x)} - Z_{(x+h)}]^2\} \quad (6)$$

Journel e Huijbregts (1978) destacaram que a covariância e a variância são (dentro da hipótese de estacionaridade de segunda ordem) ferramentas equivalentes na caracterização da autocorrelação entre duas variáveis, separadas por “h”, sendo que a covariância expressa a semelhança entre os valores e a variância o afastamento relativo entre eles. A estacionaridade de segunda ordem garante a existência de covariância e, por consequência, uma variância finita. O fato desta hipótese não poder ser satisfeita em análises de alguns fenômenos físicos, faz com que algo menos restritivo tenha que ser aplicado, trata-se da hipótese intrínseca

(VIEIRA, 2000). Em linhas gerais, Genú (2004) destacou que a hipótese intrínseca requer apenas a estacionaridade do variograma, não sendo obrigatório a variância ser finita, ou seja, para todo vetor h , “a variância da diferença $[Z(x)-Z(x+h)]$ é finita e independente da posição na região, tornando-a dependente apenas do valor de h ”, sendo assim expressa pela Equação 7.

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

2.6.2 Variograma

Para Vieira (2000) e Carvalho et al. (2002), na geoestatística, o variograma é uma ferramenta básica para estimar valores, ferramenta que ilustra o grau de dependência espacial entre diversas amostras e, desde que estejam em um mesmo campo experimental, tal ação é capaz de resultar em mapas para análise de variabilidade espacial além de permitir, através de krigagem, a definição de valores não amostrados, (SALVIANO, 1996).

A estimativa através de variograma experimental apresenta resultados mais precisos quando utilizados em menores distâncias entre os pontos, ou seja, mais próximos à origem, quando ocorre o ajuste pode ser feito através de tentativa e erro utilizando-se de um modelo que normalmente se embasa na sua aparência visual (JOURNAL ; HUIBREGTS, 1978)

Em análise a diversos trabalhos da área de solo, McBratney e Webster (1986) afirmaram que o procedimento de ajuste dos parâmetros foi definido por escolha dos autores, ou seja, de forma subjetiva, o que resultou em variogramas estimados visualmente. Trabalhos que não se utilizaram de subjetividade destacaram a utilização dos métodos de quadrados mínimos ordinários e ponderados.

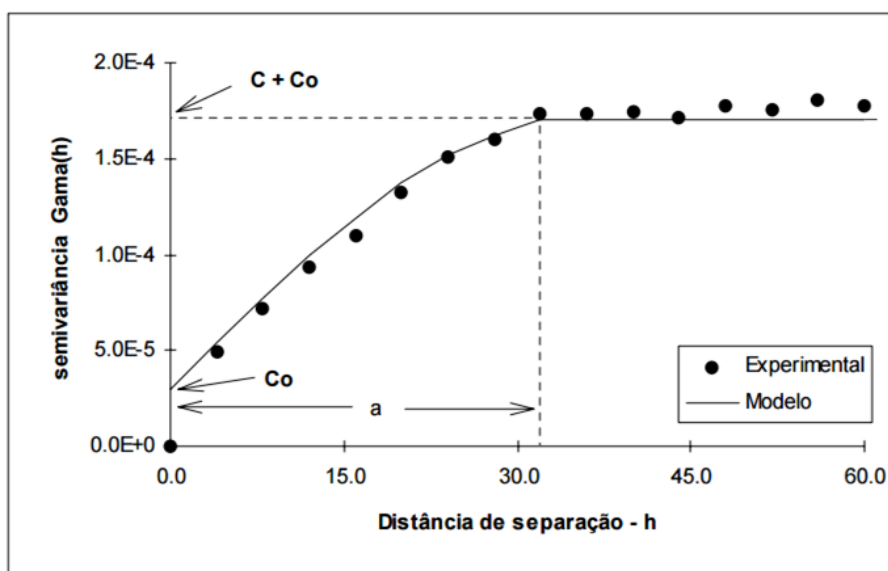
Silva (1988) explicou como através do processo de autocorrelação é possível estimar a dependência entre amostras vizinhas no espaço, quando as amostragens são feitas apenas em uma direção. Quando as amostras variam em duas coordenadas, o variograma mostra-se o instrumento apropriado para tais estimativas.

Genú (2004) afirmou que a função variância $\gamma(h)$ é definida como sendo a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados por uma distância h , o que poderia ser descrito pela Equação 7.

Souza (1999) mostrou que o variograma é apresentado no gráfico das variáveis $\gamma(h)$ e h e, através dele, é possível estimar a variância entre todas as combinações de pares de pontos, o que permite verificar qual o grau de dependência espacial da variável analisada e assim definir parâmetros para a estimativa de pontos não amostrados. Camargo (1997) explicou que os valores de $\gamma(h)$ aumentam a medida que h se eleva, isso ilustra que a dependência espacial entre pontos próximos é maior que em pontos distantes.

Os parâmetros do variograma, como apresentados por Genú (2004) são definidos por modelagem matemática nos gráficos, são apresentados na Figura 3: a pepita (C_0), que é ilustrada como o valor de γ quando $h=0$; o alcance da dependência espacial, ou apenas alcance, que a distância (a) até onde existe aumento dos valores de h e quando tais valores se estabilizam existe o patamar ($C+C_0$), cujo valor normalmente representa um valor similar à variância dos dados, que quando existe, pode ser obtida pela soma do efeito pepita com a variância estrutural (C).

Figura 3 - Variograma experimental e modelo matemático ajustado.



Fonte: Genú (2004).

Genú (2004) pontuou que quando a distância entre as amostras é menor do que o valor do alcance elas são consideradas espacialmente dependentes, quando maiores são consideradas aleatórias, ou independentes, na Figura 3, isso ocorre no

patamar, onde inclusive, pode-se dizer que a variância dos dados se torna constante quando comparada com a distância entre as amostras. Silva (1988) destacou a importância de saber o alcance de uma amostra, já que é através dele que se torna possível determinar qual é a distância entre pontos que separa um conjunto de dados espacialmente dependentes dos aleatórios.

2.6.3 Krigagem ordinária

Tendo conhecimento do variograma de uma variável e caso confirmada a dependência espacial dos dados, é possível fazer a interpolação dos valores de qualquer coordenada dentro do campo de estudo com variância mínima (VIEIRA, 2000). Este método é denominado krigagem e utiliza dados amostrais da variável na região em conjunto com propriedades do variograma. Esta técnica resulta em mapas de isolinhas ou de superfície que ilustram o comportamento da variável na região (GENÚ, 2004).

De acordo com alguns autores, pode-se definir a krigagem ordinária como sendo um método da geoestatística univariada, algumas pesquisas mostram sua eficiência na interpolação de dados (CARVALHO e ASSAD, 2005; CARVALHO ET AL., 2012).

Isaak e Srivastava (1989) citaram a cokrigagem que é uma extensão da krigagem, porém, com utilização de duas variáveis. Esta pode ser utilizada quando existe uma correlação espacial entre estas variáveis estudadas. Os autores ainda explanaram a utilidade desta extensão, já que várias vezes este método permite que subamostragem de variáveis não ocasione a perda dos dados, utilizando a correlação de uma segunda variável e assim cobrindo pontos amostrais inexistentes de uma primeira variável.

Li e Heap (2011) alegaram que um fator que influi na qualidade da predição da krigagem é a configuração de amostragem, sendo que as técnicas mais adotadas são as amostragens regulares e as aleatórias. A primeira cobre de forma sistemática a área de estudo, esta técnica apresenta dificuldade em prever variações em distâncias curtas, e a segunda apresenta menor exatidão, porém o espaçamento entre pontos costuma ser menor que o de grade regular. Para definir qual amostragem apresentava o melhor ganho de desempenho uma pesquisa foi elaborada, comparando o quão significativo o padrão de amostragem era na

definição de informação pelo interpolador da krigagem ordinária, e foi constatado que não era significativa a variação de padrão amostral (ENGLUND ET AL, 1992).

A krigagem ordinária é considerada um excelente interpolador local, já que possui como princípio a estimativa ótima “não enviesada nos pontos não amostrados”, isso quando adotadas as propriedades dos variogramas. (ISAACS E SRIVASTAVA, 1989; SILVA JUNIOR ET AL, 2013).

De acordo com Lundgren et al (2017) para a verificação de possíveis erros na estimativa da krigagem adota-se o uso da validação cruzada. Esta técnica permite comparar os valores previstos com os amostrados, desta forma é possível verificar se os dados previstos possuem alto grau de confiança (ISAACS E SRIVASTAVA, 1989).

2.7 Geoestatística multivariada

Vieira (2000) explicou que na ciência do solo, várias são às vezes em que existe correlação espacial entre duas variáveis e, que quando isso ocorre, é possível estimar valor de uma delas utilizando-se de informações de ambas (expressas no variograma cruzado), esta técnica é denominada cokrigagem.

Genú (2004) apresentou a cokrigagem como sendo um procedimento que permite estimar diversas variáveis regionalizadas em conjuntos, utilizando-se como base a correlação espacial que existe entre os dados. Esta técnica é de certa forma uma extensão multivariada da krigagem (quando um único valor é substituído por um vetor de valores dentro de uma área).

É bastante evidente o uso da cokrigagem em casos onde se possuem informações de duas ou mais variáveis de um mesmo ponto amostral e apresentam correlação. Landim et al. (2002) afirmaram que esta técnica é útil em casos em que existe uma subamostragem de alguma variável e maior quantidade de dados das demais.

De forma semelhante à krigagem, a cokrigagem deve possuir uma variância mínima e não ser tendenciosa.

Da mesma forma que quando utilizando apenas uma variável, porém são consideradas mais variáveis obtidas dentro de uma região S e nas mesmas coordenadas, desta forma estabelecem-se duas ou mais variáveis medidas em uma

mesma coordenada. Cabe salientar, porém, que o variograma cruzado embora se assemelhe com o variograma de uma variável, possui algumas diferenças. O alcance do variograma cruzado representa a distância máxima de dependência espacial. O patamar aproxima-se do valor de covariância entre as duas variáveis, isso se o patamar existir. Uma informação importante é que caso a correlação entre as variáveis seja inversa, o variograma cruzado apresentará valores abaixo de zero, desta forma apresentando-se negativo (GENÚ, 2004).

2.8 Mineração de dados (Data mining)

Cardoso e Machado (2008) explicaram que a mineração de dados é uma técnica que se encontra dentro da metodologia de descoberta em banco de dados, que por sua vez tem como função gerar, armazenar e distribuir conhecimento bem como a utilização do mesmo, ou seja, gerenciar conhecimento através da tecnologia.

Uma definição melhor elaborada diz que mineração de dados é o processo incomum de identificar, em banco de dados, padrões válidos e novos dados, que tendem a ser úteis. (FAYYAD et al. 1996).

Decker e Focardi (1995) já haviam anteriormente definido o data mining como uma forma de encontrar relações matemática e/ou lógica, sendo normalmente complexas, ou padrões em um banco de dados. King (2003) complementou explicando que a mineração de dados é uma técnica capaz de encontrar relações escondidas em grande conjunto de dados. Esta técnica é considerada multidisciplinar já que faz uso de tecnologia de bancos de dados e reconhecimento de padrões, além de diversas outras áreas da ciência da computação e estatística. Esta área de estudo é abordada em diferentes trabalhos em várias áreas de pesquisa (CARDOSO E MACHADO, 2008).

De uma forma mais prática, Cardoso e Machado (2008) explicaram que o objetivo desta técnica é descobrir de forma semi ou totalmente automática, informações ocultas dentro de um banco de dados e ainda citaram que as ferramentas as quais esta técnica faz uso são análises estatísticas e técnicas das áreas de inteligência artificiais, técnicas baseadas em regra entre outras

Algumas técnicas de mineração de dados serão abordadas neste trabalho.

2.8.1 Regressão linear múltipla

A regressão linear múltipla é um processo que envolve ao menos três variáveis, embora exista apenas uma variável dependente, sendo as demais explanatórias, ou independentes. Este tipo de análise pode ser considerado um avanço das técnicas de regressão linear simples, e de forma semelhante, objetiva definir uma equação que possa determinar, através da predição, valores para uma variável analisada. Sendo que as variáveis explanatórias são utilizadas visando o incremento da precisão, diferentemente da análise de regressão linear simples (STEVENSON, 2001).

Stevenson (2001) explicou ainda que a este tipo de análise use como base as equações dos mínimos quadrados para gerar a regressão, o autor afirmou ainda que para este tipo de cálculo, no ponto de vista prático, é interessante o emprego de computadores. O autor ainda definiu como equação para este tipo de regressão o seguinte:

$$y_c = a + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (9)$$

Sendo:

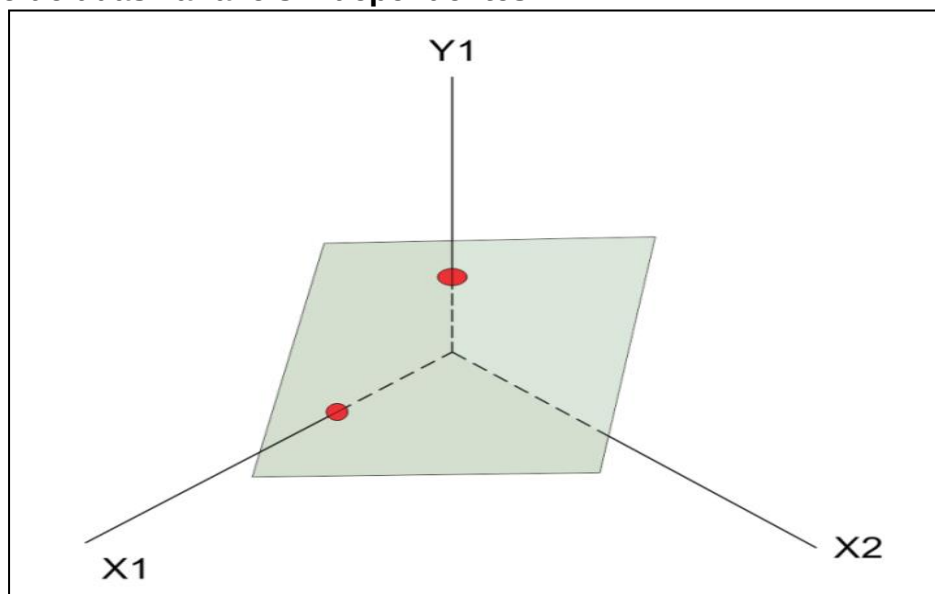
a = Intercepto – y

b_1 = Coeficientes angulares

k = número de variáveis independentes

Cabe salientar também que, de acordo com Stevenson (2001), que ao contrário da regressão simples, que gera uma equação de uma reta, a regressão múltipla tem como resultado uma superfície de mais alto grau, e o número de variáveis independentes podem resultar em um hiperplano, cuja complexidade não permite representação gráfica (diferentemente do que ocorre ao se trabalhar com até três variáveis). O autor explicou ainda que, como os conceitos são os mesmos, pode-se ilustrar este plano de forma simplificada quando se utiliza apenas duas variáveis independentes, na forma que a figura 4 ilustra:

Figura 4 - Modelo explanatório de plano de regressão gerado pela utilização de duas variáveis independentes



Fonte: Alvaro Mari Junior

Desta forma, os dados ficam dispersos em torno de um plano, de forma distinta da regressão simples, onde os dados se agrupam em torno de uma linha. Stevenson (2001) afirmou ainda em seu livro, que dentre as diversas áreas onde se torna útil a análise de regressão múltipla, podem ser citados estudos agronômicos.

2.8.2 Rede neural artificial (Artificial neural network)

A Artificial neural network (ANN) de acordo com Lippmann (1987) é uma teia de diversos elementos computacionais não lineares, porém altamente interconectados que trabalham paralelamente, e se organizam em padrões similares a uma rede neural biológica, conseguindo adquirir conhecimento através da experiência.

Jayas et al. (2000) apontaram que uma das desvantagens da utilização da ANN, é o fato de não existir metodologias teóricas para definir uma estrutura ideal para a rede, desta forma impossibilitando a definição de número de camadas ocultas e neurônios ideais, já que estas são as partes responsáveis pela capacidade de aprendizado e classificação da ANN. Os autores comparam também as redes com caixas pretas, devido à impossibilidade de compreender como a máquina chegou à conclusão ilustrada como resposta.

O treinamento das redes para que possam atingir certas respostas, pode levar várias horas podendo variar de acordo com fatores como, número de neurônios e

amostras, além das características presentes no algoritmo utilizado para o treino, vale dizer que após a rede estar treinada, o tempo que leva para processar e classificar as informações é reduzido, se tornando rápido o processamento (MITCHELL, 1997).

Jayas et al. (2000) define como sendo um eficiente sistema para aprendizado, para uma rede neural, o algoritmo possuir múltiplas camadas e utilizando a retropropagação do erro como regra de aprendizado, tendo como base uma série de exemplos.

Redes neurais também já foram utilizadas por Roush et al. (1996) onde os autores fizeram a comparação entre os resultados preditos por uma ANN e diagnósticos laboratoriais. Concluíram que algumas doenças em frangos de corte (ascite) podem ser identificadas por redes neurais. Os autores aprimoraram o estudo e foram capazes de concluir em um novo estudo que modelos de ANN podem melhorar diagnóstico da doença (ROUSH et al. 1997).

2.8.3 Máquina de vetores de suporte (Support vector machine)

De acordo com Dos Santos e Gomes (2002) a support vector machine (SVM) é uma técnica utilizada para aprendizagem de maquinas. Tem como matriz a teoria da aprendizagem estatística e a o aperfeiçoamento matemático. A autora afirma ainda que esta técnica é amplamente utilizada nas mais diversas áreas de pesquisa.

Vapnik (1995) citou que os algoritmos de aprendizagem de maquina objetivam encontrar os limites de decisão de forma a encontrar uma separação ótima entre classes através da redução dos erros. O autor citou ainda que o SVM é uma técnica computacional de aprendizado que pode ser utilizada para reconhecer padrões. Diversos trabalhos dentro da área de sensoriamento remoto, podendo ser citadas as pesquisas de Brown et. al (2000) e Melgani e Bruzzone (2004), vêm fazendo uso da SVM e obtendo sucesso na aplicação.

Nascimento et al. (2009) em um trabalho que objetivou analisar a viabilidade da utilização do SVM para classificação de imagens concluiu que, através da utilização desta ferramenta, é possível reduzir esforços em edição matricial, tornando o processo de classificação mais operacional. Os resultados apresentaram uma acurácia de 75% nos dados analisados.

2.8.4 Florestas aleatórias (Random Forest)

Breiman (2001) e Neto (2014) em trabalhos utilizando Random Forest (RF) explicaram que este é um algoritmo classificador que tem como recurso, para a classificação, o uso de árvores de decisão que permite uma melhor análise de dados passados. A RF, porém, é mais avançado que a árvore de decisão já que ele tem como meta a criação de diversas destas árvores através do uso de um subconjunto de dados aleatoriamente escolhidos (tendo como base o conjunto original de dados) e sem excluir nenhum dos atributos além de fazer uso de bootstrap, ou seja, melhores resultados devido à reposição de dados.

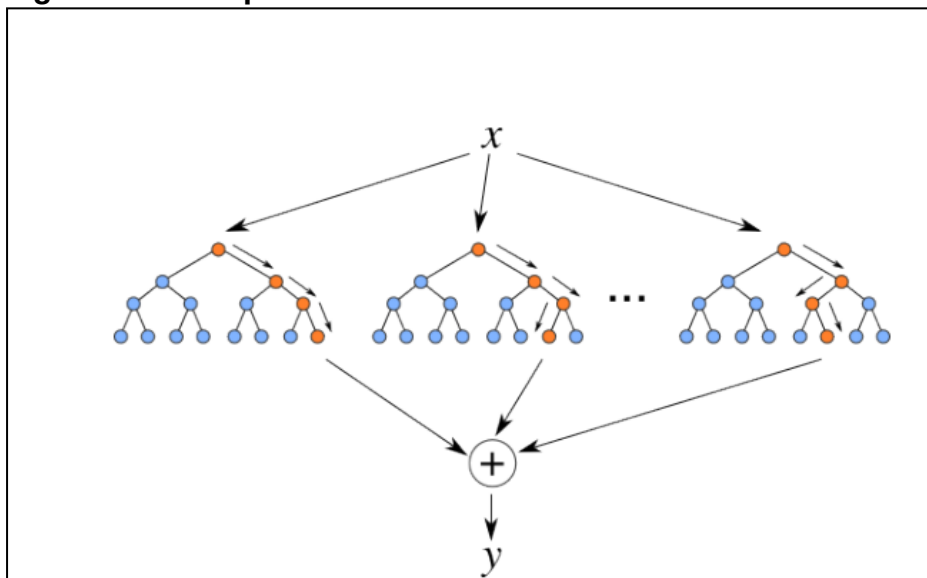
O nome da Random Forest se originou pelo fato de se tratar de um conjunto de árvores de decisões, e seu funcionamento se explica com a separação do grupo original de dados em diversos subconjuntos (cada grupo de subconjunto pode ser considerado como uma árvore de decisão) e a partir daí novas árvores são geradas a partir da ruptura de cada subconjunto. Com diversas árvores de decisão o algoritmo se torna capaz de analisar em qual das ramificações obteve melhor aprendizado para a solução do problema proposto, ou seja, o algoritmo deve escolher em qual subconjunto houve maiores benefícios antes de tomar a decisão. A RF na próxima etapa classifica os atributos chave, esta classificação possui um grau de relevância, que é diretamente afetada por dois fatores: a semelhança entre mais de uma árvore de decisão, em linhas gerais, quanto mais singular uma árvore de decisão for, melhor se comportará em relação à resposta do problema e a precisão da resposta, explica-se, já que o fato de uma árvore de decisão encontrar respostas precisas, esta também é mais bem avaliada pelo algoritmo (NETO, 2014).

Lorenzetti e Telöcken (2016) afirmaram que a RF se destaca das demais técnicas de treinamento de máquinas pois além de ser capaz de trabalhar comparando mais de uma árvore de decisão, apresenta boa precisão, não sofre por sobreajuste (Overfitting), ou seja, raramente se ajusta bem ao modelo sem mostrar boas previsões de novos dados, e por classificar de forma totalmente aleatória exclui a necessidade de intervenção humana.

A Figura 5 ilustra de forma sucinta o funcionamento da RF, mostrando desde a partição do primeiro grupo de dados em diversas árvores de decisão até o resultado obtido após a análise do algoritmo, em relação à árvore de decisão fonte das regras

mais exatas, já que cada uma das árvores pode ter verificado padrões importantes na tomada da decisão mais precisa.

Figura 5 - Exemplo de funcionamento da Random Forest



Fonte: Lorenzetti e Telöcken (2016)

2.8.5 Cubist

De acordo com Rulequest - Research (2010) o Cubist é uma ferramenta que tem como uma das funções a geração de modelos de predição, usa como base sistemas baseados em regras. Podem ser utilizados para resolução de problemas de regressão. De forma sucinta, o Cubist cria regras de forma caótica, e estas são organizadas pelo algoritmo em ordem crescente dos valores médios preditos.

A companhia Australiana explanou também que os modelos gerados pelo Cubist são modelos lineares conhecidos como piecewise, e que o algoritmo pode construir diversos tipos de modelos, entre eles os compostos, que podem tornar mais precisas as predições. Outra característica desta ferramenta é que ela combina uma árvore modelo (calculando com uso de regra) com métodos baseados em aprendizado por memória (RULEQUEST – RESEARCH, 2010).

2.8.6 Aprimoramento gradativo de máquina (Gradient Boosting Machine)

De acordo com Natekin e Knoll (2013) os Gradient Boosting Machines, ou GBM, são um conjunto de técnicas de aprendizado de máquinas e que vêm demonstrando

muito sucesso em escala global em diversas áreas de pesquisa. Estas técnicas são extremamente configuráveis e se adaptam para cada uso específico. Em linhas gerais a metodologia da GBM consiste em uma estratégia construtiva de formação de conjuntos, essa técnica tem como principal objetivo inserir e agrupar novos modelos sequencialmente. Cada iteração específica, um novo modelo de aprendizado é treinado levando em consideração todos os erros aprendidos pelo método até então (considerando as etapas anteriores).

Em busca de criar uma conexão com os quadros estatísticos, há a derivação de uma fórmula baseada em aprimoramento gradativo. Esta fórmula por sua vez somada aos seus modelos é conhecida por GBM. Este quadro estatístico garante justificativas essenciais para parâmetros e estabelece uma base metodológica para futuros modelos de aprimoramento gradativo de máquinas (FREUND E SCHAPIRE, 1997; FRIEDMAN ET AL., 2000; FRIEDMAN, 2001).

O processo de aprendizado da GBM consiste em ajustes no novo modelo que garantem uma estimativa mais precisa para a variável resposta. O princípio do algoritmo é a construção de novos aprendizados base, para ser correlacionado ao máximo com gradientes negativos da função de perda, associando assim com o conjunto todo (NATEKIN E KNOLL, 2013). Os autores afirmaram também que esta função de perda aplicada, embora possa ser arbitrária, apresentou melhores resultados quando adotada a função do quadrado dos erros. O procedimento então, não resulta em diversos erros de ajuste, porém, os autores explicaram que a escolha da função de perda pode ser definida pelo próprio pesquisador, já que além de existir uma grande variedade para esse tipo de função exista uma grande possibilidade de aprimoramentos que podem ser feitos partindo do tipo de estudo a ser elaborado.

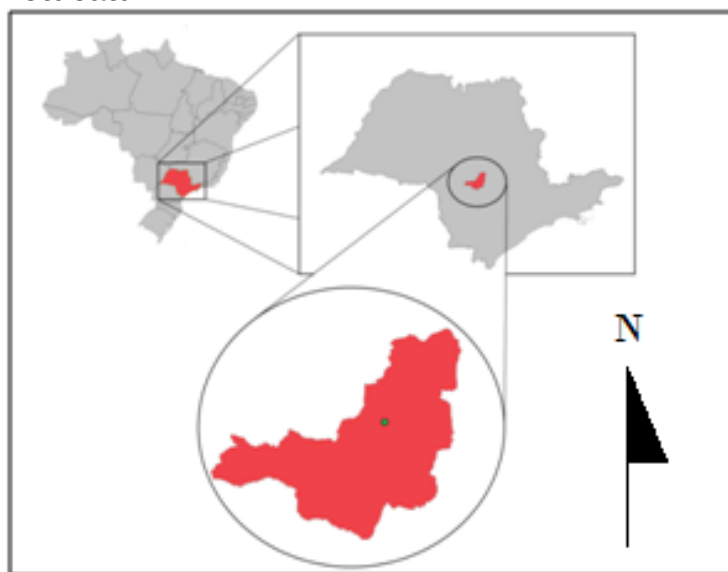
Diversos autores, como Bissacco et al. (2007) Pittman e Brown (2011) e Johnson e Zhang (2012), utilizaram-se das ferramentas da GBM devido à sua alta flexibilidade e potencial para sua personalização nas tarefas, podendo se adaptar de acordo com dados de entrada. Porém deixam claro que esta grande liberdade que existe acaba tornando a escolha da função de perda uma questão de tentativa e erro, embora a técnica seja relativamente simples de se executar, o que garante a possibilidade da realização de um único experimento com diferentes modelos de ajuste. A GBM apresentou sucesso considerável tanto em aplicações práticas quanto no aprendizado de máquinas nas pesquisas elaboradas pelos autores.

3 MATERIAL E MÉTODOS

3.1 Descrição da área

A coleta dos dados foi feita em duas áreas distintas para a confirmação da correlação entre os dados. A primeira área denominada como Fazendinha e a segunda como Cascalheira, ambas em sistemas de plantio diferentes, o que buscou verificar se há variação nos dados com relação às diferentes condições de manejo. As áreas fazem parte da Fazenda Experimental Lageado, na Faculdade de Ciências Agrônômicas, UNESP em Botucatu – São Paulo. A Figura 6 mostra a localização das áreas de estudo.

Figura 6 - Localização da área de estudo, Brasil, São Paulo, Botucatu



A região que compreende as áreas localiza-se em aproximadamente 22°51'S e 48°25'W com uma altitude média de 786 metros, e o solo, classificado de acordo com a Embrapa (2006), é um Nitossolo Vermelho. Porém a caracterização mais precisa em Cascalheira seria Argissolo devido às suas características observadas a campo. A temperatura média anual da região é de 20,6 °C e a precipitação média anual é de 1.314 milímetros. A classificação de Köppen (1948) aponta a região como sendo temperada chuvosa úmida (Cwa), ou seja, possui verões quentes e invernos secos.

A aveia preta foi semeada no processo de rotação de culturas, sendo a cultura presente no solo antes da semeadura o milho.

Tanto na Cascalheira quanto na Fazendinha, o milho vem sendo cultivado há diversos anos. Porém na Cascalheira a forma de plantio foi sistema convencional, ou seja, escarificação seguida de aração e gradagem. Já na Fazendinha há mais de dez anos o sistema de plantio direto foi implementado, restringindo a movimentação do solo à linha de semeadura, e tráfego de máquinas (no trato e colheita da cultura). Ambas as áreas receberam aplicação de 200 Kg/ha de uréia, e ainda a Fazendinha recebeu adubação no plantio anterior de 280 Kg/ha, ao tempo que Cascalheira recebeu 300 Kg/ha

3.2 Material

Para a obtenção dos dados foram utilizados GPS de navegação e clorofilômetro no mês de Julho de 2011.

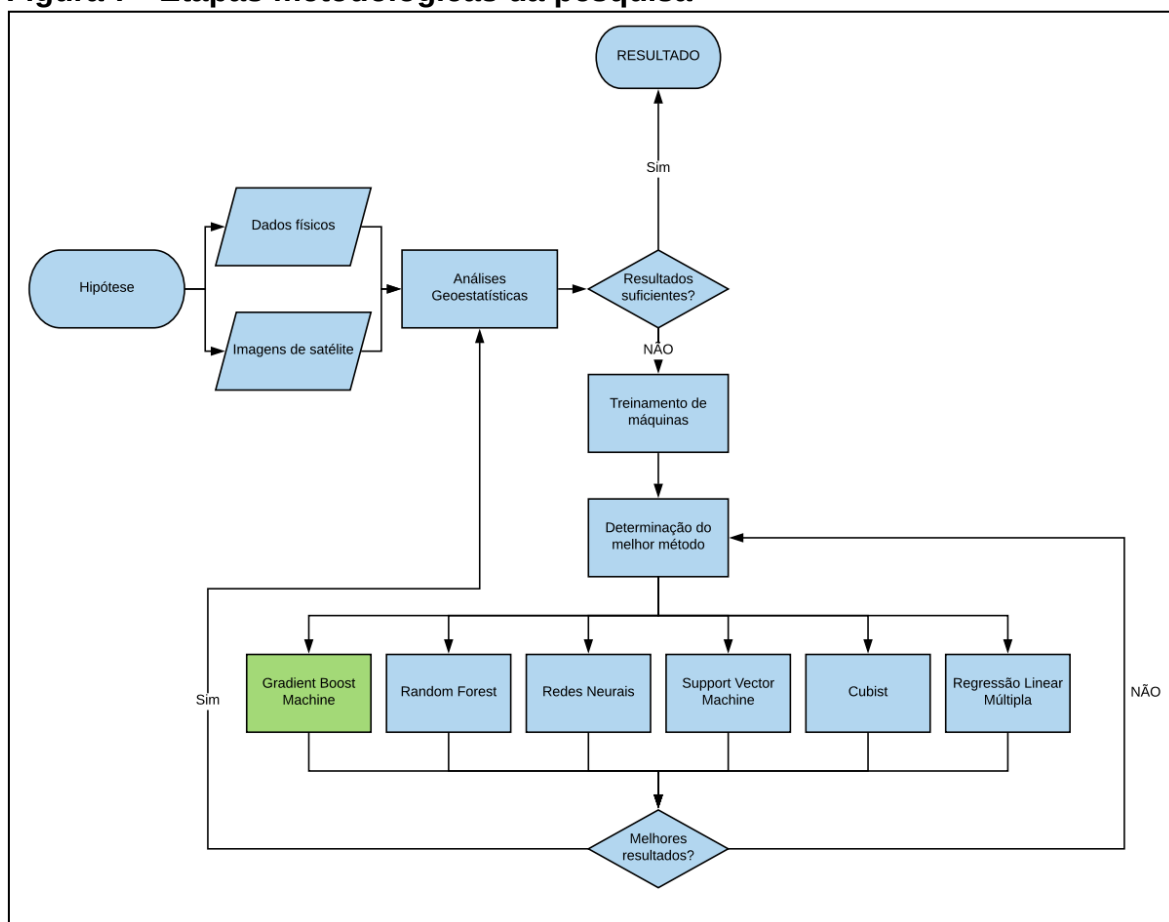
Para a elaboração dos índices de vegetação foram utilizadas imagens LANDSAT-5 disponibilizadas gratuitamente no site da United States Geological Survey (USGS). As imagens foram tratadas e os índices foram calculados no programa QGIS.

As análises geoestatísticas foram realizadas com o uso do programa GS+.

Para a execução da etapa de treinamento de máquinas e geração de mapa de predição foi utilizado o programa R com o pacote de modelagem estatística.

3.3 Metodologia

A figura 7 ilustra, de forma sucinta, as etapas realizadas durante a pesquisa através de um fluxograma.

Figura 7 - Etapas metodológicas da pesquisa

Fonte: Alvaro Mari Junior

Cada etapa será detalhada a seguir para melhor compreensão de cada técnica utilizada.

3.3.1 Obtenção de dados de campo (Dados físicos)

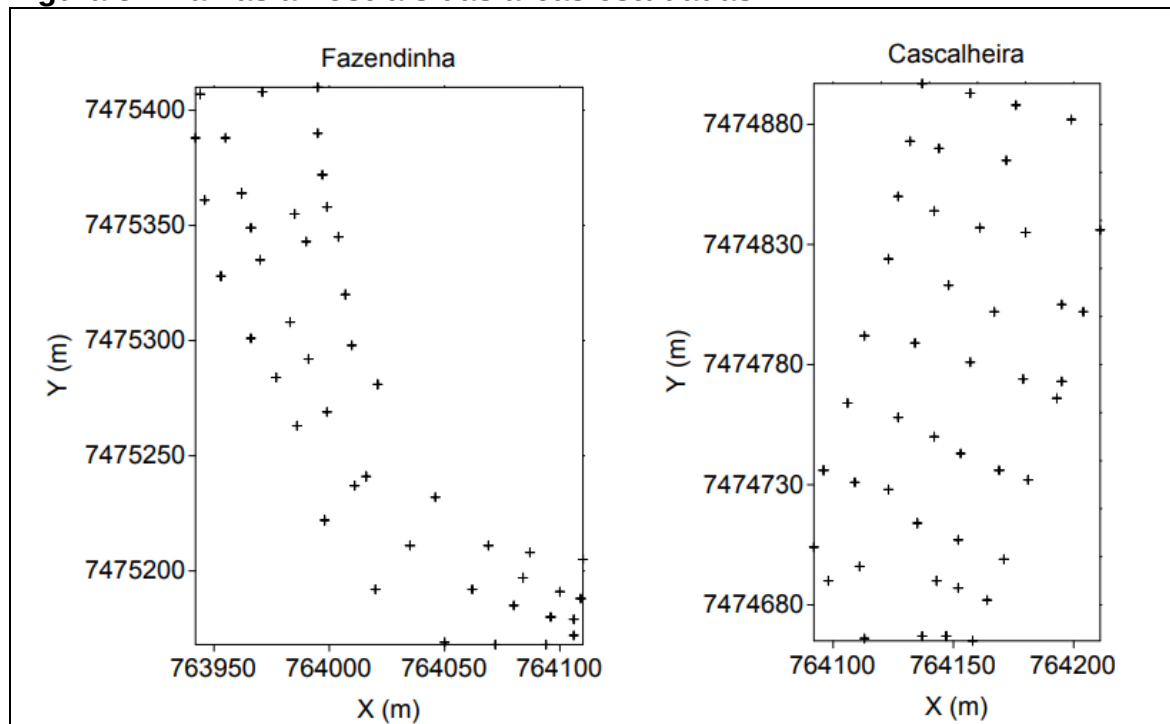
As duas áreas com diferentes manejos foram semeadas com a utilização de maquinário agrícola, uma delas em regime de plantio direto na área denominada Fazendinha e na área conhecida como Cascalheira em plantio convencional.

A coleta de dados a campo foi realizada com o auxílio de um GPS de navegação, desta forma foram obtidas as coordenadas precisas dos 45 pontos amostrais em cada uma das áreas, totalizando 90 pontos.

Em cada um dos 90 pontos, foram feitas as coletas dos dados do teor de clorofila com auxílio de um clorofilômetro portátil, estas coletas ocorreram no período

de floração plena. As malhas amostrais, de ambas as áreas, podem ser observadas na figura 8.

Figura 8 - Malhas amostrais das áreas estudadas



Nos mesmos pontos de coleta foram retiradas amostras deformadas de solo da camada de 0 a 20 centímetros, para análise granulométrica do solo seguindo a metodologia da pipeta proposta pela Embrapa (1997).

Após obtenção do resultado das análises, todos os dados foram organizados em uma planilha, para a facilitação da análise de correlação entre as variáveis.

3.3.2 Obtenção dos índices de vegetação (Imagens de satélite)

O NDVI e SAVI foram obtidos através do georeferenciamento das imagens obtidas junto ao USGS, LANDSAT-5, que devido a um número maior de pesquisas realizadas utilizando este tipo de imagem optou-se pelo seu uso ao invés das imagens do LANDSAT 8, seguida de álgebra dos arquivos Tiffs, e extração das informações através da ferramenta Point Sampling Tool, com uso do programa Q-Gis, utilizando as coordenadas dos pontos amostrais de campo. Em seguida, a malha de amostragem foi adensada, coletando informações referentes aos índices de vegetação em pontos próximos às áreas experimentais, na tentativa de refinar as

informações de teor de clorofila, que foram na etapa de análises geoestatística, correlacionada com estes dados do NDVI e SAVI.

Para a elaboração destes índices de vegetação, foram obtidas imagens para cada mês em que a aveia estava em campo, ou seja, junho, julho, agosto e setembro de 2011. A partir das imagens foram calculados o NDVI e o SAVI utilizando-se das metodologias propostas por Rouse et al. (1973) e Huete (1988), cabe salientar ainda que as imagens passaram por correção atmosférica antes da realização do SAVI. O processo de cálculo dos índices de vegetação foi realizado no software Q-GIS que através de suas ferramentas, permitiu extrair a informação de NDVI e SAVI de cada uma das 90 coordenadas com exatidão. Os dados foram inseridos na mesma planilha dos dados de campo, e permitiram desta forma verificar a correlação entre os dados.

Após a elaboração de mapa de cada mês para ambos os índices de vegetação, todos foram analisados para definir que aqueles que coincidiram com a data em que foi feita a obtenção dos dados de clorofila apresentaram correlação com os dados de teor de clorofila enquanto os de períodos divergentes não apresentaram semelhança nenhuma com os dados de campo. Pode ser afirmado portanto que os dados das características das plantas estavam em um mesmo momento tanto para os dados de clorofila quanto reflectância das imagens.

3.3.3 Análises geoestatística

A primeira etapa foi análise exploratória, pois é um importante processo a ser efetuado na busca para conhecer e resumir a variável de estudo. Este processo foi realizado para a determinação do melhor método estatístico para tratar os dados. Foi feita uma análise descritiva prévia para facilitar a compreensão do banco de dados, visualizando desta forma suas variações e distribuições, o que permitiu uma interpretação mais precisa, evitando assim a utilização de dados inconsistentes (SILVA, 2014).

A partir destas análises, foram definidos os valores de média, variância, desvio padrão, coeficiente de variação, valores mínimos e máximos, coeficiente de assimetria e coeficiente de curtose.

Em seguida e com posse das informações referentes aos dados de campo e as informações dos índices de vegetação, os dados foram organizados em planilhas no

programa GS+7, onde as análises geoestatística de cada área foram realizadas, de início considerando apenas a variável referente ao teor de clorofila e, posteriormente, correlacionando os dados com os dados dos índices de vegetação.

O variograma utilizado foi o isotrópico, já que este é amplamente utilizado para detecção de anisotropia (SILVA, 2014).

Para verificar o grau de dependência espacial utilizou-se o índice de dependência espacial (IDE), que como descrito por Zimback (2001), é a razão entre a componente estrutural e o patamar. Desta forma através da análise dos variogramas foi possível verificar até que distância amostral os pontos eram influenciados pela sua proximidade.

Foram gerados variogramas, do teor de clorofila individualmente e correlacionando-o com os dados de campo obtidos com o clorofilômetro, esta etapa foi realizada para ambas as áreas experimentais.

Quanto à krigagem, o uso do GS+7 tornou possível a elaboração da interpolação dos dados, bem como a validação cruzada que teve por objetivo a verificação dos valores quanto à semelhança com os dados reais, para assim observar se a interpolação conseguiu prever com eficiência os valores dos dados analisados.

Outro processo que foi realizado com o uso de geoestatística foi a análise dos pontos coletados, na imagem de predição de clorofila (que será explicado na próxima etapa), comparando-os com os valores reais obtidos com o uso do clorofilômetro.

3.3.4 Treinamento de máquinas

O processo de treinamento de máquinas, por se tratar de um processo complexo, foi dividido em várias etapas, sendo a primeira a entrada dos dados, para que o processo encontrasse melhores resultados, a malha amostral de ambas as áreas foi unida (o que pode ter afetado de forma negativa a pesquisa, já que são áreas embora apresentem as mesmas características de clima, se diferem em tipo de solo), gerando assim uma malha de 90 pontos amostrais. Destes pontos, foram inseridos para o programa os dados de campo mostrados na tabela 1:

Tabela 1 - Dados físicos inseridos no programa

Dados		
Coordenada X	Coordenada Y	Teor de Clorofila

Informações espectrais também foram inseridas como dados de entrada, todas as bandas do satélite foram utilizadas para o treinamento de máquina, a tabela 2 ilustra quais foram essas bandas:

Tabela 2 - Dados espectrais inseridos no programa

Dados		
B1 (Azul)	B2 (Verde)	B3 (Vermelho)
B4 (Infravermelho próximo)	B5 (Infravermelho médio)	B6 (Infravermelho termal)
	B7 (Infravermelho distante)	

Após os dados inseridos no programa, o algoritmo Band-Ratio foi efetuado, trata-se de uma técnica que aprimora o contraste entre feições utilizando-se da divisão da refletância de cada pixel em uma banda, com as demais. Quando o algoritmo findou, o número de variáveis alterou-se, ficando os dados demonstrados na tabela 3.

Tabela 3 - Dados no programa após Band-Ratio

Dados		
Coordenada X	Coordenada Y	Teor de clorofila
B1	B2	B3
B4	B5	B6
B7	B1-B2	B1-B3
B1-B4	B1-B5	B1-B6
B1-B7	B2-B3	B2-B4
B2-B5	B2-B6	B2-B7
B3-B4	B3-B5	B3-B6
B3-B7	B4-B5	B4-B6
B4-B7	B5-B6	B5-B7
	B6-B7	

Os dados representados por uma relação entre duas bandas (BX-BY) foram gerados pelo Band-Ratio, que fez uma análise comparativa entre elas, duas a duas,

gerando novos dados, e estes foram adicionados ao banco de dados como novas variáveis.

Em seguida, para reduzir dados semelhantes, que responderiam ao problema por caminhos extremamente parecidos, o que ocasionaria a explicação de um mesmo comportamento, um algoritmo de exclusão de dados foi executado, procurando dados com semelhança maior que 98%. Desta forma, houve um peneiramento destes dados, reduzindo as variáveis no interior do programa para as apresentadas na tabela 4.

Tabela 4 - Dados no programa após exclusão de “colinearidade > 0,98”

Dados		
Teor de clorofila	B1	B2
B4	B6	B1-B2
B1-B3	B1-B4	B1-B6
B2-B3	B2-B4	B2-B5
B2-B6	B2-B7	B3-B4
B3-B5	B3-B6	B3-B7
B4-B5	B4-B7	B5-B6
B5-B7	B6-B7	

Após a remoção destes dados sobressalentes, foi executado um algoritmo visando à eliminação de dados de menor importância para o processamento da resposta, após a função Recursive feature elimination (pacote CARET) o banco de dados foi mais uma vez reduzido, passando a serem os dados descritos na tabela 5. Cabe salientar que nesta etapa, o valor de Teor de clorofila passou a ser definido como variável a ser explicada, e as demais variáveis consideradas independentes (buscando explicar o comportamento da variável primária).

Tabela 5 - Dados no programa após exclusão do Recursive feature elimination

Dados		
Teor de clorofila	B1	B2
B4	B6	B1-B2
B1-B3	B1-B4	B1-B6
B2-B3	B2-B4	B2-B6
B2-B7	B3-B4	B3-B5
B3-B6	B3-B7	B4-B5
B4-B7	B5-B6	B6-B7

Nesta etapa, 80% de todos os dados foram inseridos no algoritmo como dados de treinamento, enquanto os demais 20% utilizados para a validação do modelo.

Na sequência, o software executou seis diferentes técnicas de mineração de dados: Regressão linear múltipla (LM), Máquina de vetores de suporte (SVM), Redes neurais artificiais (ANN), Floresta aleatória (RF), Cubist, e Aprimoramento gradativo de maquina (GBM). Depois de finalizado o processo, os parâmetros observados, para a tomada de decisão referente ao melhor resultado apresentado por cada uma das técnicas, foram as métricas estatísticas: Coeficiente de determinação (r^2), Raiz do erro médio quadrático (RMSE), Índice Nash-Sutcliffe Efficiency (NSE) erro médio de viés (MBE) e erro médio absoluto (MAE) (KENNEDY AND NEVILLE, 1986; NASH AND SUTCLIFFE, 1970).

As equações que descrevem essas métricas são respectivamente as ilustradas nas equações, 10, 11, 12, 13 e 14

$$r^2 = \frac{(\sum (Pi - \bar{P})(Oi - \bar{O}))^2}{(\sum (Pi - \bar{P})^2 \sum (Oi - \bar{O})^2)} \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (Oi - Pi)^2} \quad (11)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Oi - Pi)^2}{\sum_{i=1}^n (Oi - \bar{O})^2} \quad (12)$$

$$MBE = \frac{1}{n} \sum (Pi - Oi) \quad (13)$$

$$MAE = \frac{1}{n} \sum |Pi - Oi| \quad (14)$$

Onde “P” são os valores preditos enquanto “O” são os observados (obtidos a campo).

3.3.5 Geração de mapa de predição

Quando definido, através da análise dos resultados referentes aos parâmetros observados, a técnica que apresentou melhores resultados, iniciou-se novamente um processo no software R, porém como a máquina já estava treinada, foi adotada a Gradient Boosting Machine como ferramenta, as imagens novamente foram inseridas como dados de entrada, e agora com a ciência de quais são as variáveis que são consideradas importantes para a predição, (o algoritmo treinado e validado com 80% dos dados e validado com os outros 20%), o processo foi executado durante 10 ciclos, considerando a cada ciclo um cenário diferente. E deste processo foi obtido o mapa de predição de clorofila.

4 RESULTADOS E DISCUSSÃO

4.1 Dados de campo

Ao serem analisados os dados de campo, notou-se o fato das distâncias entre os pontos amostrais serem pequenas, tendo pontos a menos de 10 metros de distância uns dos outros, isto apenas se apresentou como um problema devido aos pixels de 30 metros encontrados nas imagens LANDSAT 5. Com uma verificação minuciosa, os outliers foram suprimidos para encontrar resultados aceitáveis.

A tabela 6 mostra os resultados da análise descritiva dos dados referentes ao teor de clorofila e granulometria da malha amostral.

Tabela 6 - Estatística descritiva dos dados de campo (90 pontos)

Variável	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Teor de						
Clorofila	31,3	39,15	42,45	42,19	45,60	53,20
(SPAD)						
Argila	371	522,5	565	576	606,8	779
(g/Kg)						
Areia	47	77	88	89,81	101,75	136
(g/Kg)						
Silte	162	306,2	349	334,2	376	522
(g/Kg)						

Esta análise foi importante para se ter uma melhor compreensão sobre o comportamento dos dados.

4.2 Índices de vegetação

Os índices de vegetação de todos os meses em que a aveia preta estava no campo foram calculados, porém, depois de algumas análises dos dados, principalmente no que se refere à correlação dos mesmos, notou-se que os índices do mesmo mês em que foi feita a coleta das informações referentes aos teores de clorofila eram dados que se mostraram coerentes, isto se deu provavelmente pelo

fato de que tanto as imagens anteriores quanto as posteriores à coleta dos dados em campo, não conseguiriam explicar o comportamento da clorofila, já que esta pode variar de acordo com a precipitação, desta forma, a clorofila pode ter variado entre eventos climáticos, anteriores e posteriores a coleta dos dados. Por isso foi importante o estudo de todas as imagens, para que se concluísse o fato de as informações geradas por elas, no que se refere ao teor de clorofila na aveia preta, conseguem ter certo grau de precisão apenas quando estimam valores de clorofila de épocas similar àquela de captura da imagem pelo satélite na região.

4.3 Análises geoestatística

As figuras 9 e 10 ilustram, respectivamente, as configurações e o variograma de Cascalheira e da Fazendinha, analisando exclusivamente a variável de teor de clorofila.

Figura 9 - Variograma do teor de clorofila (Cascalheira)

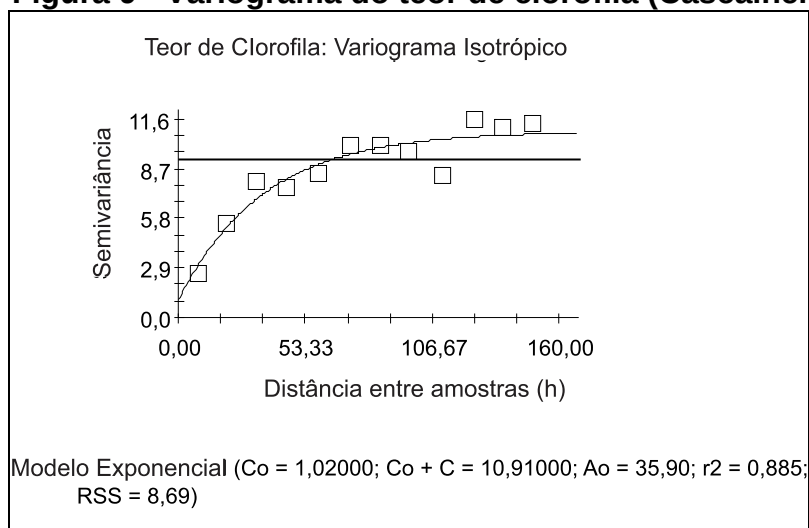
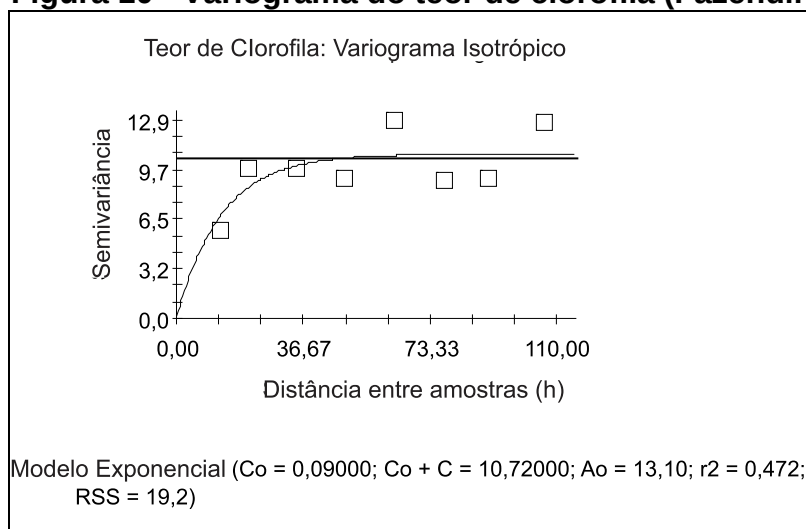
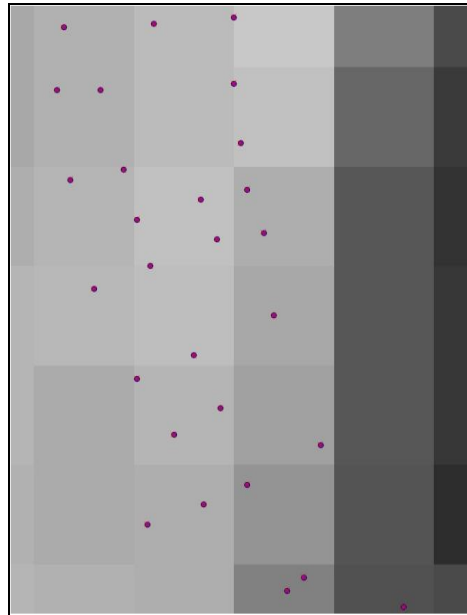


Figura 10 - Variograma do teor de clorofila (Fazendinha)

Pode-se notar que o variograma na área da Cascalheira apresentou melhor ajuste, com menor efeito pepita e uma distribuição de pontos capazes de afirmar a grande dependência espacial dos dados de teor de clorofila, porém o mesmo não ocorreu com o variograma da área da Fazendinha, o que acaba levantando um questionamento, sobre o porquê não houve uma dependência tão clara para essa segunda área experimental.

Depois de gerado o variograma de ambas as áreas sem a utilização de covariáveis, iniciou-se a etapa de comparação entre os dados referentes ao teor de clorofila e os índices de vegetação. Uma informação importante que não pode ser deixada de lado é o fato de que os pixels de 30m das bandas do Landsat influenciaram nas análises dos índices de vegetação, já que diversos foram os pontos encontrados dentro de um mesmo pixel do arquivo Tiff. O que pode ser mais bem compreendido observando a figura 11, que é uma captura de tela mostrando a sobreposição de pontos amostrais na imagem utilizada.

Figura 11 - Pontos de coleta de campo, repetindo informações do Tiff



Pode-se observar, portanto, que uma malha amostral com pontos mais afastados seria, em âmbito de imagens de Landsat, melhor analisável. Mantendo em mente que a distância máxima entre os pontos amostrais também deve ser respeitada.

Nesta etapa foram gerados os variogramas dos índices de vegetação. Para o NDVI, obtiveram-se variogramas vistos nas figuras 12 e 13, que mostram respectivamente os variogramas de Cascalheira e Fazendinha utilizando os dados de NDVI como covariável.

Figura 12 - Variograma das informações de NDVI (Cascalheira)

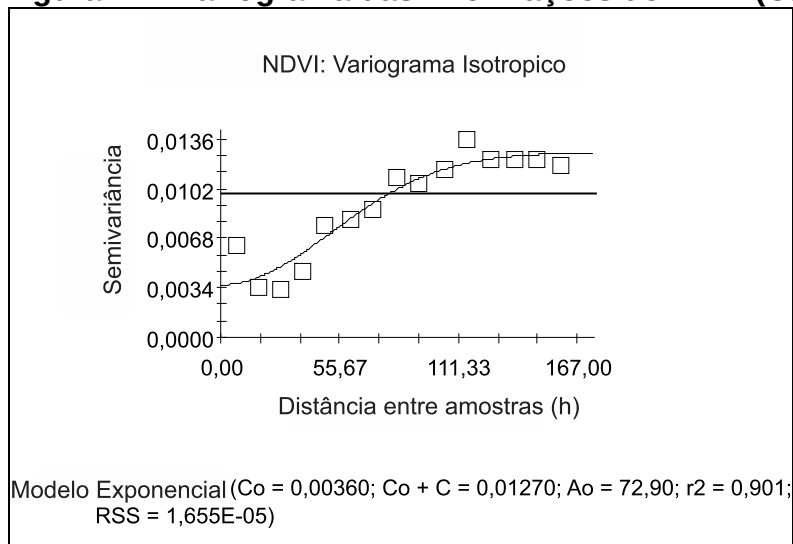
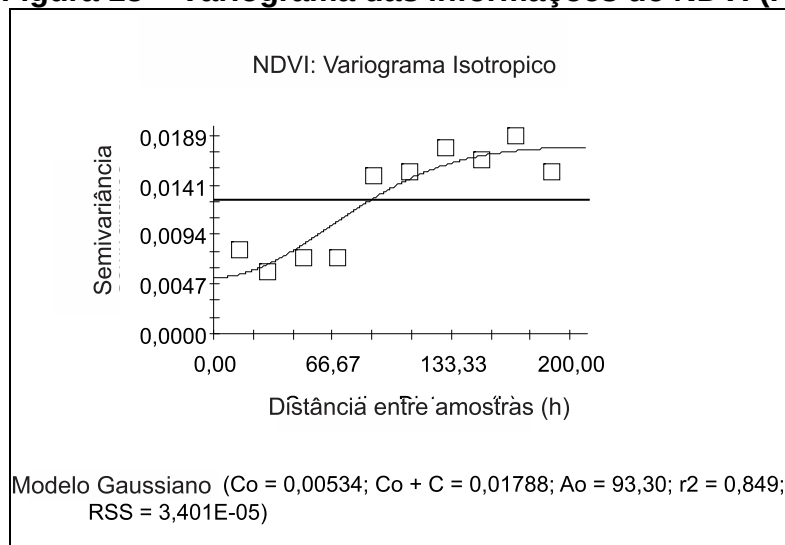


Figura 13 – Variograma das informações de NDVI (Fazendinha)

Embora o comportamento de ambos os variogramas tenham sido semelhantes, é possível perceber que, de forma similar ao que ocorreu no variograma da Fazendinha sem utilização de covariável (figura 10), há quantidade reduzida de pontos na nuvem de variância nesta área experimental, existe a possibilidade de que a ausência dos pontos suprimidos tenha ocasionado tal escassez, que causou como consequência esse efeito no variograma, agravado ainda pelo tamanho dos pixels, que já foi explicado.

Na utilização do índice SAVI observaram-se os seguintes variograma, ilustrados nas figuras 14 e 15.

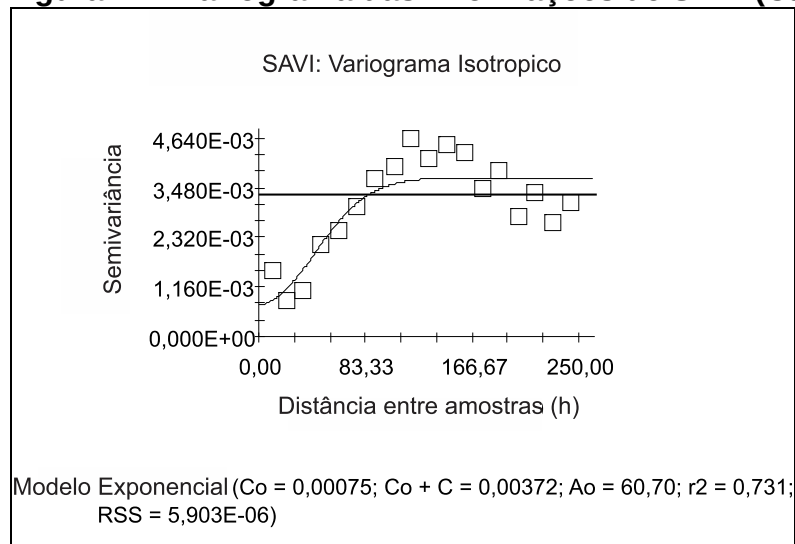
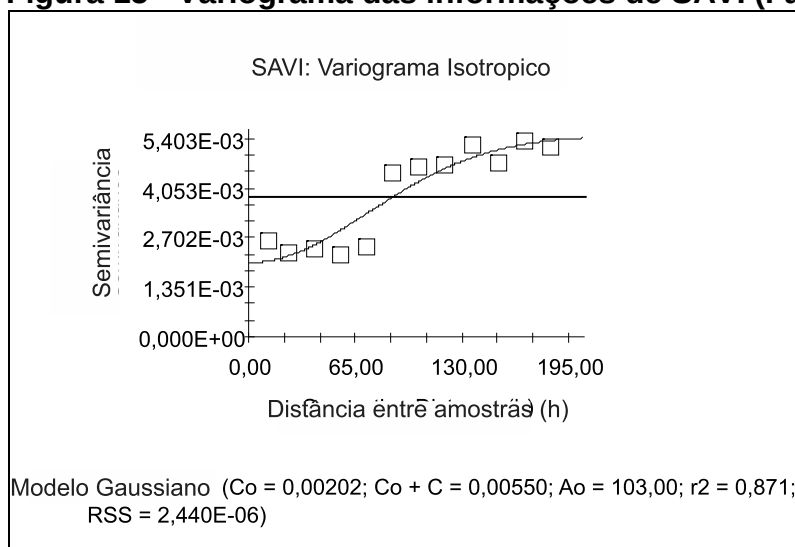
Figura 14 - Variograma das informações de SAVI (Cascalheira)

Figura 15 - Variograma das informações de SAVI (Fazendinha)

O variograma da Fazendinha não apresentou um alcance claro, e pode ser definido como uma pepita pura, existe a possibilidade de tal fato ter ocorrido pelo número reduzido de pixels do cálculo do SAVI encontrados dentro desta área experimental, o mesmo não ocorreu sobre os dados da outra área.

As validações cruzadas, para as krigagens utilizando os dois índices de vegetação foram realizadas e as figuras 16 e 17 mostram esta validação em função do NDVI, enquanto as figuras 18 e 19 mostram em função do SAVI. Em ambas às comparações entre as duas áreas é possível notar a diferença no coeficiente de regressão entre elas, pode-se perceber então que as amostras obtidas na Cascalheira apresentaram r^2 mais elevado quando comparado à Fazendinha.

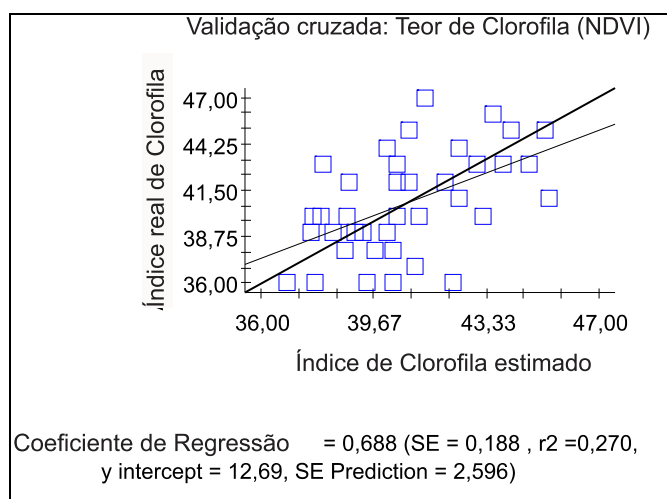
Figura 16 - Validação cruzada Cascalheira (teor de clorofila X**NDVI)**

Figura 17 - Validação cruzada Fazendinha (teor clorofila X NDVI)

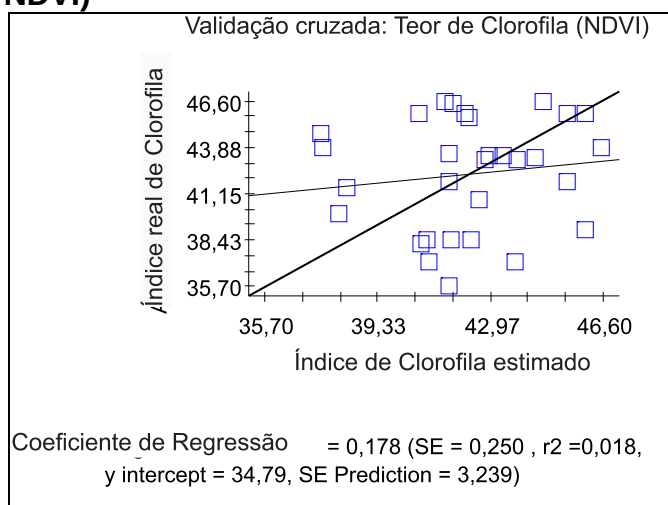


Figura 18 - Validação cruzada Cascalheira (teor de clorofila X SAVI)

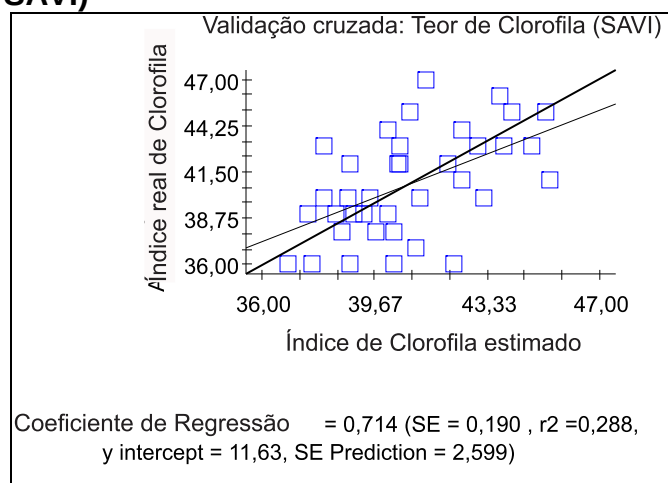
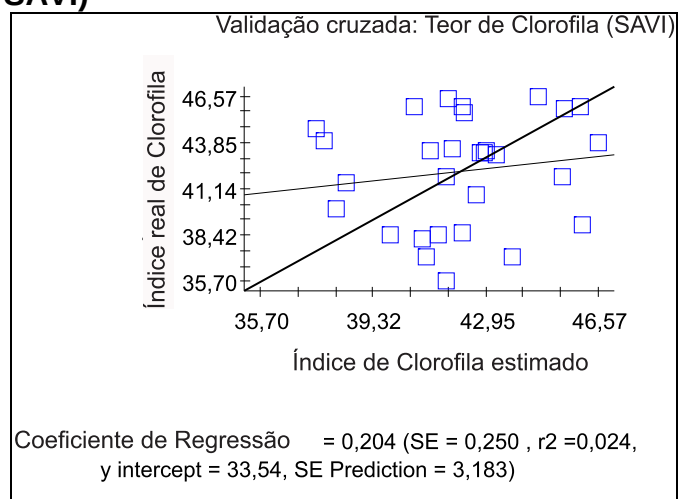
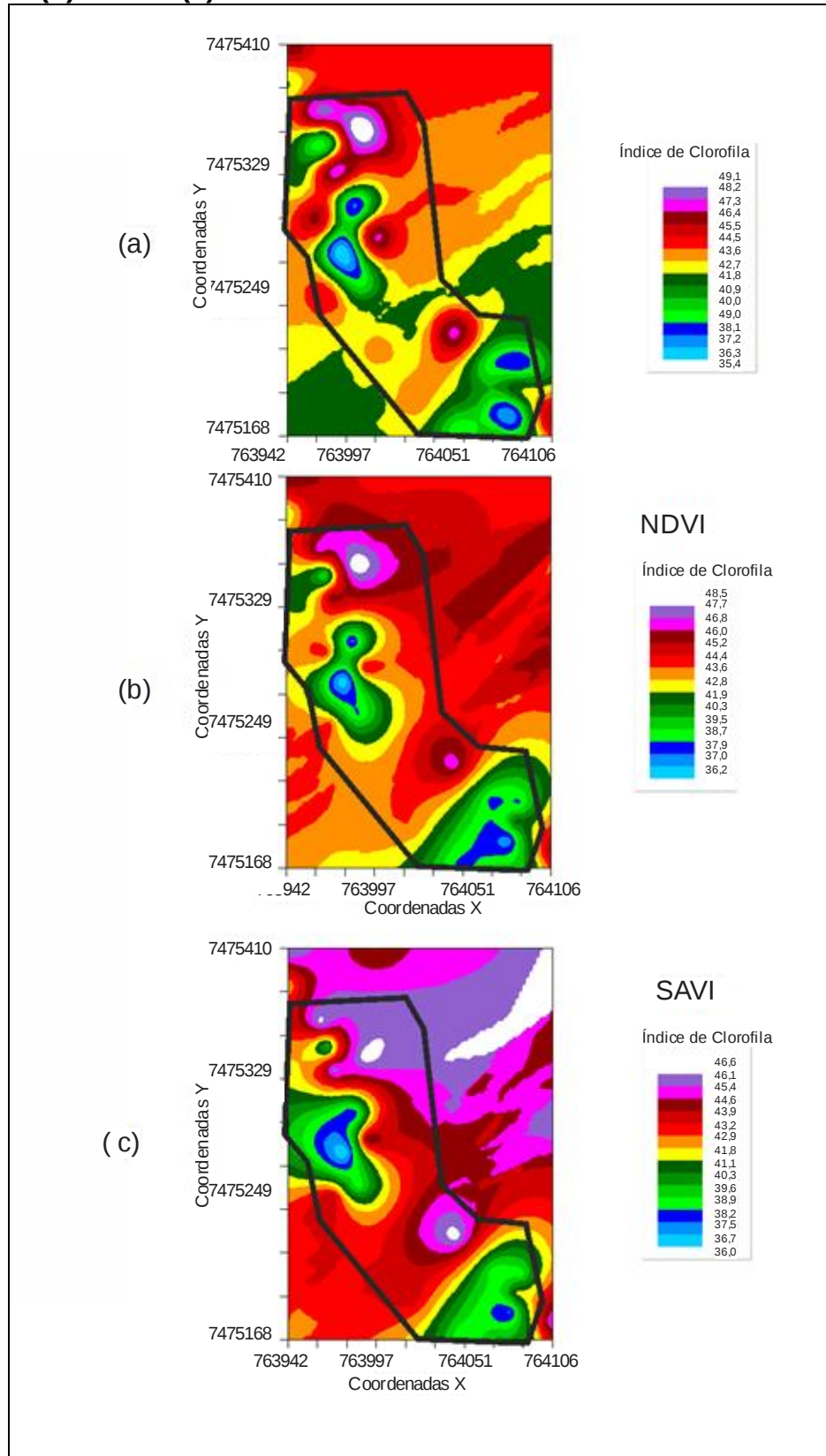


Figura 19 - Validação cruzada Fazendinha (teor de clorofila X SAVI)



Na figura 20 podem ser observados os resultados das Krigagem na Fazendinha.

Figura 20 - Resultados da Krigagem na área da Fazendinha utilizando-se de apenas da variável teor de clorofila isolada (a) e cookrigagem utilizando NDVI (b) e SAVI (c)

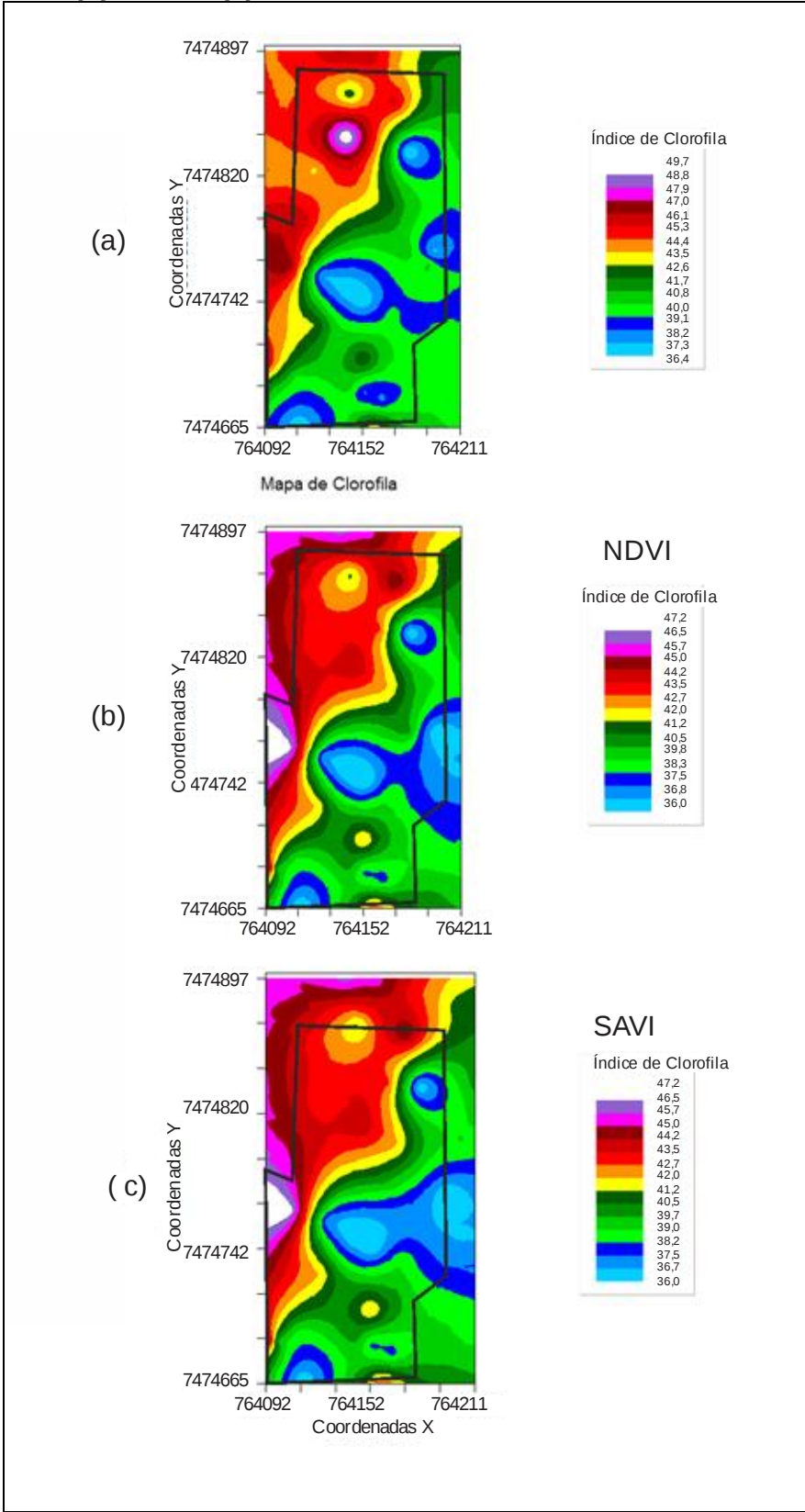


Pode ser observado que as áreas no entorno do limite da área de estudo possuem mais informações quando se utiliza os índices de vegetação, o que ocorreu devido ao adensamento de pontos que foi feito sobre o NDVI e o SAVI, porém essas informações não contribuem com a predição de clorofila, já que não se existia informação da variável primária para fora da área de estudo sendo assim, os dados externos à área de estudo foram desconsiderados.

Para a área experimental da Cascalheira os resultados da Krigagem e Cokrigagem, podem ser observados na figura 21, que de forma similar à área anterior, foi feita utilizando dados de clorofila (independente) e também com o uso do NDVI e SAVI como covariáveis.

O principal efeito que pôde ser observado é a suavização das informações quando a krigagem foi feita com auxílio dos índices de vegetação como covariáveis, isso acontece naturalmente pelo uso da krigagem, não podendo ser vinculado à correlação. Destaca-se o fato de existirem poucas diferenças efetivas entre as feições observadas em todas as três, Krigagem e Cookrigagens desta área.

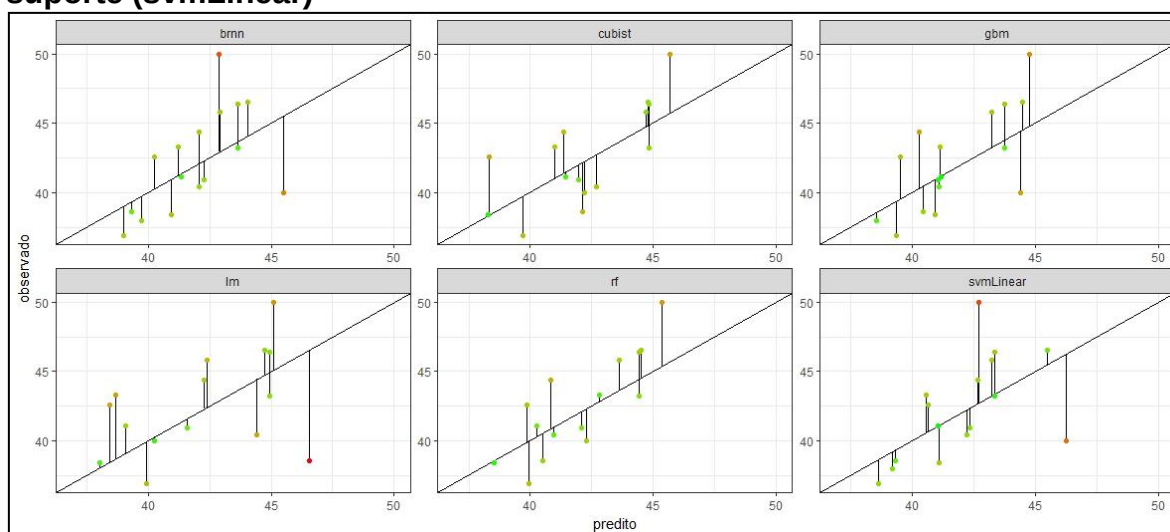
Figura 21 - Resultados da Krigagem na área da Cascalheira utilizando-se de apenas da variável teor de clorofila isolada (a) e cookrigagem utilizando NDVI (b) e SAVI (c)



4.4 Treinamento de máquinas

Este item da pesquisa apontou resultados bastante significativos na predição do teor de clorofila pelo uso de imagens. Os gráficos de predição para cada uma das técnicas de mineração de dados utilizadas podem ser observados na Figura 22, onde se pode notar que a dispersão dos dados foi menor no modelo gbm, o que fez com este fosse a técnica utilizada para a confecção do mapa de predição do teor de clorofila.

Figura 22 - Gráfico da predição de dados, valores preditos (eixo x) por valores observados (eixo y) de cada uma das técnicas adotadas: Redes neurais artificiais (brnn), Cubist, Aprimoramento gradativo de maquina (gbm), Regressão linear múltipla (lm), Floresta aleatória (rf), e Máquina de vetores de suporte (svmLinear)



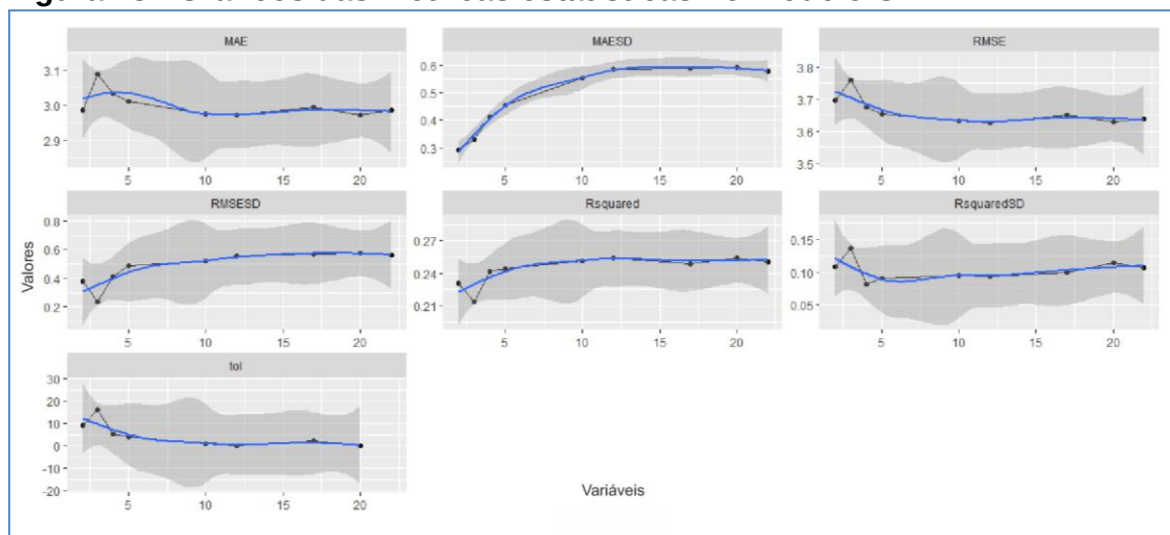
Os dados organizados na tabela 7 mostram o resultado de cada uma das métricas estatísticas de cada um dos modelos de mineração de dados utilizados, sendo eles: Coeficiente de determinação (r^2), Raiz do erro médio quadrático (RMSE), Índice Nash-Sutcliffe Efficiency (NSE) erro médio de viés (MBE) e erro médio absoluto (MAE).

Tabela 7 - Métricas estatísticas para cada modelo

Modelo	r^2	RMSE	NSE	MBE	MAE
lm	0,4079	3,1266	0,2916	-0,4800	2,6321
brnn	0,5692	2,4278	-0,3346	-0,3879	1,9630
svmLinear	0,4007	3,0021	0,1366	-0,7116	2,5406
rf	0,5774	2,4085	-0,1853	-0,5065	2,0108
cubist	0,2240	3,8010	-0,0818	-0,8870	3,2284
gbm	0,6360	2,2291	0,2884	-0,6002	1,7458

Analisando a tabela 7, pode-se fazer a seguinte interpretação, através do valor médio do MBE, observando o fato de estar positivo ou negativo pode-se definir se os valores estão sendo superestimados ou subestimados, os valores de RME e MAE, mostram a variação do erro enquanto o r^2 nos mostra o encaixe dos dados na predição.

Depois de analisados estes dados, optou-se método do Aprimoramento Gradativo de máquina (gbm) que este apresentou melhor coeficiente de determinação. Para facilitar a visualização das métricas estatísticas no modelo, foram criados gráficos conforme pode ser observado na figura 23.

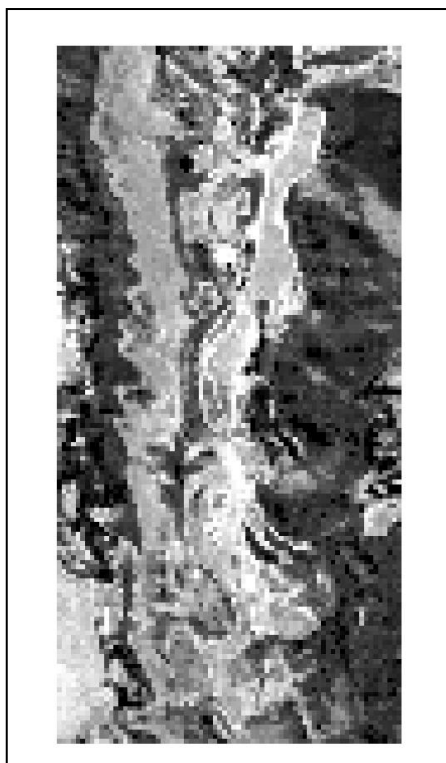
Figura 23 - Gráficos das métricas estatísticas no modelo GBM

Com o modelo selecionado, o programa R gerou, através do treinamento já inserido na máquina e adotando a predição proposta pelo Gradient Boosting Machine, um mapa de predição para os valores do teor de clorofila.

4.5 Geração de mapa de predição

Através da análise dos dados considerados importantes pelo treinamento de máquina (Tabela 5), o Modelo de aprimoramento gradativo de máquina (gbm), estudou todas as combinações dos dados e gerou o mapa de predição de valores do teor de clorofila que pode ser observado na Figura 24.

Figura 24 - Imagem referente aos dados estimados pelo Gradient Boosting Machine



A imagem, por ser um Tiff, foi prontamente inserida no banco de dados, onde a malha amostral foi utilizada para fazer sobreposição dos dados, e de forma similar ao feito com os índices de vegetação foi feita a coleta dos dados do teor de clorofila predito.

Em uma planilha foram analisados, para as coordenadas da coleta de campo, os valores do teor de clorofila medido e o predito para ambas as áreas, as figuras 25 e 26 ilustram na forma de um gráfico de radar, as informações verificadas.

Figura 25 - Comparação do valor do teor de clorofila real e o predito na Cascalheira

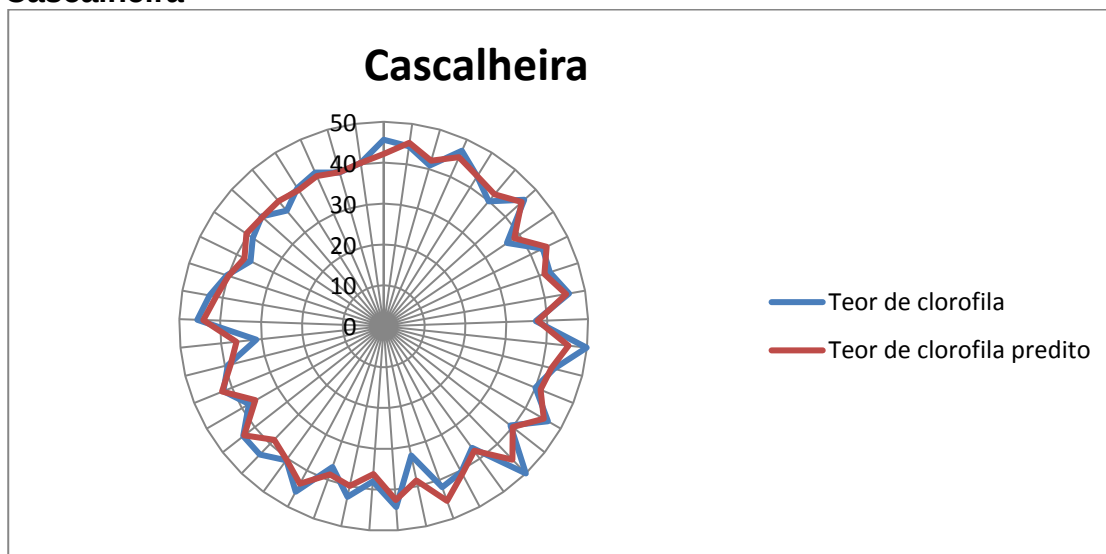
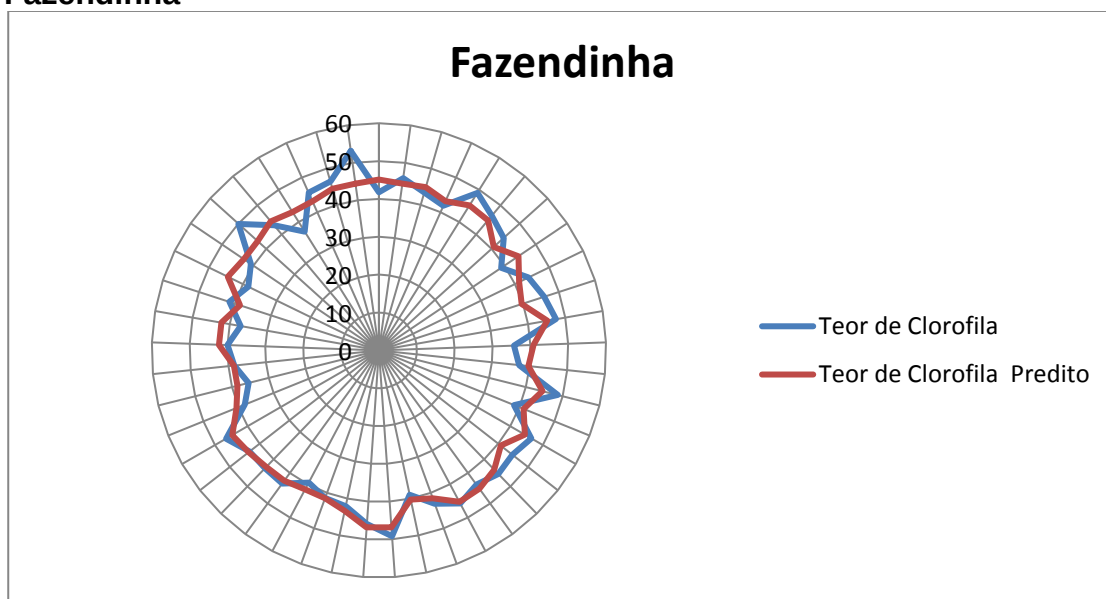


Figura 26 - Comparação do valor do teor de clorofila real e o predito na Fazendinha



Embora visualmente os dados sejam similares, uma análise estatística foi gerada para verificar o coeficiente de determinação (r^2). Os resultados podem ser observados nas Figuras 27 e 28.

Figura 27 – Coeficiente de determinação (Real X Simulado (GBM) na Cascalheira

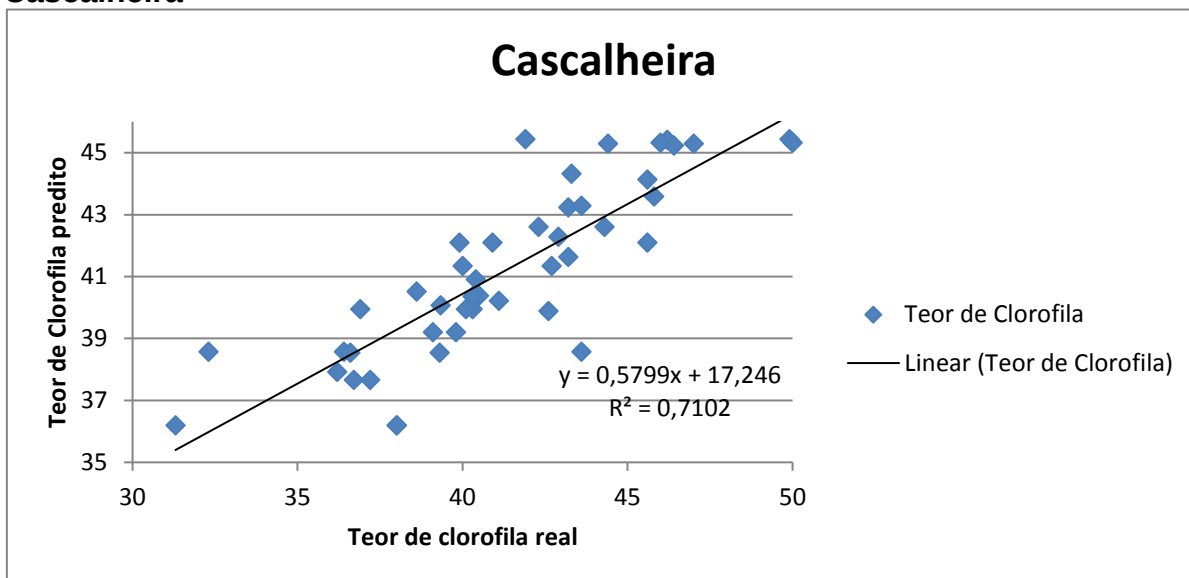
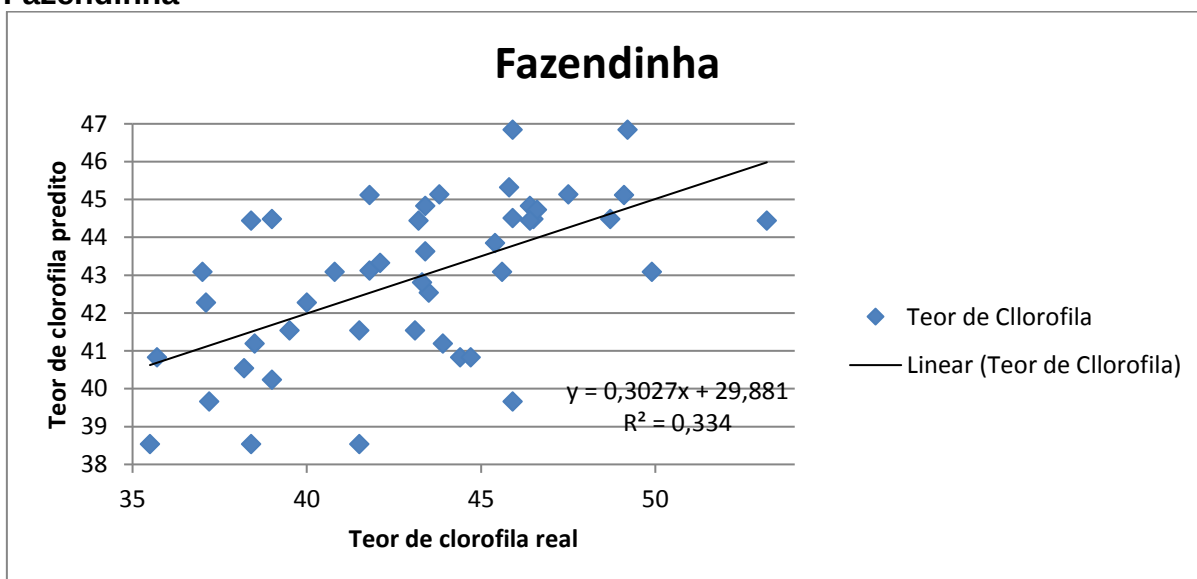


Figura 28 – Coeficiente de determinação (Real X Simulado (GBM) na Fazendinha



Observando os dados pode ser notado que o modelo conseguiu um coeficiente de determinação de 0,70 para a área de Cascalheira, o que pode ser considerado um resultado positivo, porém o mesmo não ocorreu com relação aos valores da Fazendinha.

Existe a possibilidade de que os dados referentes ao teor clorofila obtidos na área da Fazendinha tenha sofrido alguma influência em sua etapa de obtenção e/ou tabulação, pois em todas as etapas desta pesquisa, incluindo até mesmo os variogramas, alguma incoerência nos dados foi percebida. Não se pode descartar

também a possibilidade de que a união entre os 45 pontos amostrais de duas diferentes áreas tenham sido a responsável pelo baixo coeficiente de determinação na Fazendinha.

Porém é possível afirmar, que para a área da Cascalheira o modelo conseguiu prever os valores do teor de clorofila com uma precisão considerável, muito acima do previsto para a Fazendinha.

5 CONCLUSÕES

Os dados de teor de clorofila são dados que possuem dependência espacial, porém, nesta pesquisa, a correlação foi maior em uma das áreas quando comparado à outra, possivelmente devido ao tamanho dos pixels da imagem de satélite e a concentração da malha amostral, que teve por consequência diversos pontos amostrais com a mesma informação dos índices, cabe citar que a repetição de dados não foi a mesma nas duas áreas.

O treinamento de máquinas e a mineração de dados são ferramentas que se mostram muito eficientes na obtenção de respostas computacionais, demonstrando, nesta pesquisa, resultados mais relevantes que os apresentados por geoestatística.

Os valores de teor de clorofila estimados pelo Gradient Boosting Machine na área experimental de Cascalheira mostram resultados bastante satisfatórios, o que deixa claro que novas pesquisas utilizando esta ferramenta, na predição do teor de clorofila pode sim reduzir os custos com obtenção de dados de campo, referentes ao teor de clorofila, e por consequência facilitar decisões que usem como insumo esta informação, como por exemplo, necessidade de irrigação.

Cabe destacar, porém, que algumas outras variáveis devem ser estudadas para confirmar esta hipótese, como a profundidade do solo, variáveis climáticas, fertilidade do solo e a própria física do solo. Para só então existir uma confirmação concreta da eficiência da técnica.

REFERÊNCIAS

- ANDRIOTTI, J. L. S. **Notas de Geoestatística**. Acta Geologica Leopoldensia XXV(55) 3-14, 2002
- ARGENTA, G. **Monitoramento do nível de nitrogênio na planta como indicador da adubação nitrogenada em milho**. Porto Alegre, Universidade Federal do Rio Grande do Sul, 2001. 112p. Tese de Doutorado.
- ARGENTA, G; SILVA, P. R. F; BORTOLINI, C. G; FORTHOFER, E. L; STRIEDER, M. L. **Relação da leitura do clorofilômetro com os teores de clorofila extraível e de nitrogênio na folha de milho**. Revista Brasileira de Fisiologia Vegetal, n.13 p.158-167, 2001.
- ANTUNES, R. C. B. **Determinação da evapotranspiração e influencia da irrigação e da fertirrigação em componentes vegetativos, reprodutivos e nutricionais do café arábica**. Viçosa: UFV, 2000. 165p. (Dissertação de Mestrado em Meteorologia Agrícola) – Universidade Federal de Viçosa, 2000.
- BISSACCO, A; YANG, M. H; AND SOATTO, S. **Fast human pose estimation using appearance and motion via multi-dimensional boosting regression**, in IEEE Conference on Computer Vision and Pattern Recognition, CVPR'07. Minneapolis, MN, 2007.
- BORATTO, I. M. P; GOMIDE, R. L. **Aplicação dos índices de vegetação NDVI, SAVI e IAF na caracterização da cobertura vegetativa da região Norte de Minas Gerais**. Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR, Foz do Iguaçu, PR, Brasil, 13 a 18 de abril de 2013, INPE.
- BREDEMEIER, C. **Predição da necessidade de nitrogênio em cobertura em trigo e aveia**. Porto Alegre, Universidade Federal do Rio Grande do Sul, 1999. 101p. Dissertação de Mestrado.
- BREIMAN, L. **Random Forest**. Machine Learning, 45, 5–32, 2001. Disponível em: < <https://link.springer.com/content/pdf/10.1023%2FA%3A1010933404324.pdf> >
- BROWN, M; LEWIS, H. G; GUNN, S. R. **Linear spectral mixture models and support vector machines for remote sensing**. IEEE Transactions on geoscience and remote sensing, v. 38, n. 5, p. 2346-2360. 2000
- BURROUGH, P. A.; BOUMA, J.; YATES, S. R. **The state of the art in pedometrics**. Geoderma, Amsterdam, v. 62, n. 1/3, p. 311-326, 1994.
- CAMARGO, E. C. G. **Desenvolvimento, implementação e teste de procedimentos geoestatísticos (krigeagem) no sistema de processamento de informações georreferenciadas (Spring)**. São José dos Campos, 1997. 123p. Dissertação (Mestrado) - Instituto Nacional de Pesquisas Espaciais.

CAMPOS, S.; PISSARRA, T. C. T.; RODRIGUES, F. M.; BARROS, Z. X.; RIBEIRO, F. L. **Análise multivariada de microbacias em relação ao tipo de solo**. Irriga, v.15, p. 208-216, 2010.

CARDOSO, O. N. P.; MACHADO, R. T. M. **Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras**. Revista de Administração. Pública, Rio de Janeiro, v.42, n.3, p.495-528, 2008. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-76122008000300004&lng=en&nrm=iso>

CARVALHO, J. R. P. de; ASSAD, E. D. **Comparação de interpoladores espaciais univariados para precipitação pluvial anual no Estado de São Paulo**. Campinas: Embrapa Informática Agropecuária. 2002, 6p. Comunicado Técnico.

CARVALHO, J. R. P. de; ASSAD, E. D. **Análise espacial da precipitação pluviométrica no Estado de São Paulo: comparação de métodos de interpolação**. Engenharia Agrícola, v.25, p.377-384, 2005.

CARVALHO, J. R. P. de; VIEIRA S. R.; GREGO C. R.; **Comparação de métodos para ajuste de modelos de variograma da precipitação pluvial anual**. Revista Brasileira de Engenharia Agrícola e Ambiental, v.13, n.4 p.443-448, 2009

CARVALHO, J. R. P. de; ASSAD, E. D.; PINTO, H. S. **Interpoladores geoestatísticos na análise da distribuição espacial da precipitação anual e de sua relação com altitude**. Pesquisa Agropecuária Brasileira, v.47, p.1235-1242, 2012.

CAVALLARI, R. L. **A importância de um Sistema de Informações Geográficas no Estudo de Microbacias Hidrográficas**. Revista Científica Eletrônica de Agronomia, v.6, n. 11, 2007.

DECKER, K.; FOCARDI, S. **Technological overview: a report on data mining**. CSCS – Swiss National Supercomputing Center, Relatório Técnico, Zurique, 1995. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.6800&rep=rep1&type=pdf>>

DEMOTES-MAINARD, S; BOUMAZA, R.; MEYER, S.; CEROVIC, Z. G. **Indicators of nitrogen status for ornamental woody plants based on optical measurements of leaf epidermal polyphenol and chlorophyll contents**. Scientia Horticulturae, v. 115, p. 377-385, 2008.

DOS SANTOS, E. M. E GOMES, H. M. **A Comparative Study of Polynomial Kernel SVM Applied to Appearance-Based Object Recognition**. In: Lee SW., Verri A. (eds) **Pattern Recognition with Support Vector Machines**. Lecture Notes in Computer Science, vol 2388. Springer, Berlin, Heidelberg, 2002.

DWYER, L.M.; ANDERSON, A.M.; MA, B.L.; STEWART, D.W.; TOLLENAAR, M. e GREGORICH, E. **Quantifying the nonlinearity in chlorophyll meter response to corn leaf nitrogen concentration**. Canadian Journal of Plant Science, 1995.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. **Manual de Métodos de Análise de Solo**. Centro Nacional de Pesquisa de Solos, 2 ed. Rio de Janeiro, 1997. 212 p.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. 2.ed. Rio de Janeiro, Embrapa, 2006. 374 p.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA – EMBRAPA-Semiárido. **Sistemas de Produção**. 2010. Versão online. Disponível em:<<https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Melancia/SistemaProducaoMelancia/irrigacao.htm>>

ENGLUND, E.; WEBER, D.; LEVIANT, N. **The effects of sampling design parameters on block selection**. Mathematical Geology, v.24, p.329-343, 1992.

FAYYAD, U. M; PIATETSKY-SHAPIRO, G; PADHRAIC, S. **From data mining to knowledge discovery: an overview**. In: Advances in knowledge discovery and data mining. California: AAAI/The MIT, p.1-34, 1996

FLOSS, E. L. **Manejo forrageiro de aveia (*Avena sp.*) e azevem (*Lolium sp.*)** Simpósio sobre manejo de pastagem, 1988. Piracicaba. Anais... Piracicaba: FEALQ, 1988.

FONTANELLI, R.S.; PIOVEZAN, A.J. **Efeitos de cortes no rendimento de forragem e grãos de aveia**. Pesquisa Agropecuária Brasileira, Brasília, 26(5): p.691-697, maio 1991.

FREUND, Y. E SCHAPIRE, R. **A decision-theoretic generalization of on-line learning and an application to boosting**. Journal of Computer and system sciences. 55, 119-139, 1997.

FRIEDMAN, J. H. **Greedy boosting approximation: a gradient boosting machine**. The Annals of Statistics 29, 1189–1232, 2001. Disponível em: <<https://statweb.stanford.edu/~jhf/ftp/trebst.pdf>>

FRIEDMAN, J; ASTIE, T; TIBSHIRANI, R. **Additive logistic regression: a statistical view of boosting**. Annals of Statistics 28, 337-407, 2000. Disponível em: <<https://projecteuclid.org/euclid.aos/1016218223#info>>

GARCÍA-SÁNCHEZ, F.; SYVERTSEN, J.P.; GIMENO, V.; BOTÍA, P.; PEREZ-PEREZ, J.G. **Responses to flooding and drought stress by two citrus rootstock seedlings with different water-use efficiency**. Physiologia Plantarum, Copenhagen, v.130, p. 532-542, 2007.

GENU, A. M. **Geoestatística multivariada. Departamento de ciências exatas**, 2004. 17p. Apostila. Disponível em: <http://www.lce.esalq.usp.br/tadeu/aline.pdf>>

GLINSKY, J.; STEPNIIEWSKI, W. **Soil aeration and its role for plants**. CRC Press Inc, Florida. 228 p. 1986.

GODOY, L. J. G.; SOUTO, L.S.; FERNANDES, D.M.; VILLAS BÔAS, R.L. **Uso do clorofilômetro no manejo da adubação nitrogenada para milho em sucessão a pastagem de Brachiaria decumbens**. Ciência Rural, v.37, n.1, p.38-44, 2007.

GOMES, E. P; TESTEZLAF, R. **Manejo de irrigação na tomaticultura-de-mesa**. Universidade Estadual de Campinas, 2005. Disponível em: <
http://tudosobretomate.com.br/publicacoes/textos/text_12.pdf >

HARTWIG, I.; CARVALHO, F. I. F.; OLIVEIRA, A. C.; SILVA, J.A.G.; LORENCETTI, C; BENIN G.; VIEIRA, E. A.; BERTAN, I. ; SILVA, G. O. ; VALÉRIO, I. P.; SCHMIDT, D. J. M. **Correlações fenotípicas entre caracteres agronômicos de interesse em cruzamentos dialélicos de aveia branca**. Revista Brasileira de Agrociência, Pelotas, v. 12, n. 3, p. 273-278, jul-set, 2006.

HENDRY, G.A.F.; PRICE, A.H. **Stress indicators: chlorophylls and carotenoids**. In: Hendry, G.A.F.; Grime, J.P. (eds), *Methods in Comparative Plant Ecology*, p. 148-152. London, Chapman & Hall, 1993.

HENRIQUES, A. B.. **A moderna agricultura no final do século XIX em São Paulo: algumas propostas**. História [online]. 2011, vol.30, n.2, pp. 359-380.

HUETE, A. R. **A Soil-Adjusted Vegetation Index (SAVI)**. Remote sensing of environment 25:295-309, 1988.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An Introduction to Applied Geostatistics**. New York: Oxford University Press, 561 p. 1989

JAYAS, D. S.; PALIWAL, J.; VISEN, N. S. **Multi-layer neural networks for image analysis of agricultural products**. Journal of Agricultural Engineering Research, Londres, v.77 n.2, p.119-128, 2000

JENSEN, J. R. (1949) **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres** / traduzido por José Carlos Neves Epiphany et al. São José dos Campos. SP. 2009

JOHNSON, R. E ZHANG, T. **Learning Nonlinear Functions Using Regularized Greedy Forest**. Relatório técnico. arXiv:1109.0887. 2012. Disponível em: <
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0020583>>

JOURNEL, A.G.; HUIJBREGTS, Ch.J. **Mining Geostatistics**. London: Academic Press, 1978. 600 p.

KENNEDY, J. B. E NEVILLE, A. M. **Basic Statistical Methods for Engineers and Scientists**, 3rd Edition, Harper and Row: New York. 1986

KING, D. **Numerical machine learning**. Georgia: Tech College of Computing.2003
Disponível em: <www.cc.gatech.edu/kingd/datamine/datamine.html>

KÖPPEN, W. 1948. **Climatologia: com um estúdio de los climas de la tierra**.
Publications In: 104p. 1948.

KUMAR, M.; BANDYOPADHYAY, A.; RAGHUWANSHI, N.S.; SINGH, R.
Comparative study of conventional and artificial neural network based ETo estimation models. Irrigation Science, v.26, p.531-545, 2008.

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.

LI, J.; HEAP, A.D. **A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors**.
Ecological Informatics, v.6, p.228-241, 2011.

LILLESAND, T. M. E KIEFER, R. W. **Remote sensing and image interpretation**. 3 ed. Wiley & Sons, Nova Iorque. 1994

LIPPMANN, R.P. **An introduction to computing with neural nets**. IEEE Acoustics Speech and Signal Processing Magazine, Washington, v.2, n.4, p.4-22, 1987.

LORENZETT, C. D. C; TELÖCKEN, A. V. Estudo **Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão**. II Simpósio de Pesquisa e Desenvolvimento em Computação. Cruz Alta, RS. 2016.

LUGÃO, S. M. B.; ALMEIDA, E. L. D.; COSTA, A. L.; SOARES, T. M. D. M.; GOBBI, K. F.; BETT, V. **Lâminas de água na produção de aveia preta para forragem, no Noroeste do Paraná**. IAPAR, Belém – PA, Julho de 2011.

LUNDGREN, W. J. C.; DA SILVA, J. A. A.; FERREIRA, R. L. C. A **Precisão da Estimativa do Erro da Krigagem pela Validação Cruzada**. Floresta e Ambiente 2017;

LUTZENBERGER, J. A.. **O absurdo da agricultura**. *Estud. av.* [online]. 2001, vol.15, n.43, pp. 61-74.

McBRATNEY, A.B.; WEBSTER, R. **Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates**. Journal of Soil Science, v.37, p.617-639, 1986.

MEDEIROS, G. A.; ARRUDA, F. B.; SAKAI, E. **Relações entre o coeficiente de cultura e cobertura vegetal do feijoeiro: Erros envolvidos e análises para diferentes intervalos de tempo**. Acta Scientiarum, v.26, p.513-519, 2004

MELGANI, F; BRUZZONE, L. **Classification of Hyperspectral Remote Sensing Images with Support Vector Machines**. IEEE Transactions on Geoscience and Remote Sensing, vol. 42, No. 8, 2004.

MITCHELL, T.M. **Machine learning**. McGraw Hill Companies Inc; 1997. 414p.

MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. 2.ed. Viçosa - MG: UFV, 2003. 307 p.

NASCIMENTO, R. F. F; DE ALCÂNTARA, E. H; KAMPEL, M; STECH J. L; NOVO, E. M. L. M; FONSECA L. M. G; **O algoritmo Support Vector Machines (SVM): avaliação da separação ótima de classes em imagens CCD-CBERS-2**. Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 2009.

NASH, J. E. E SUTCLIFFE, J. V. **River flow forecasting through conceptual models, Part I - A discussion of principles**, J. Hydrol., 10, 282–290, 1970

NATEKIN, A. E KNOLL A. **Gradient boosting machines, a tutorial**. Frontiers in neurorobotics. Alemanha. Online. Disponível em:
< <https://www.frontiersin.org/articles/10.3389/fnbot.2013.00021/full>>

NETO, A. C. F; MANTOVANI, E. C; RENA, A. B; MOURA, B. R; SIMÃO, F. R e FACCIOLI, G. G. **Determinação da evapotranspiração de dois cultivares de café arábica na fase de maturação dos frutos**. In: II Simpósio de Pesquisa de Cafés do Brasil, 2001.

NETO, C. D. G. **Potencial de técnicas de mineração de dados para o mapeamento de áreas cafeeiras**. INPE, São José dos Campos, 2014.

NOVO, E. M. L. M. **Sensoriamento Remoto: Princípios e aplicações**. São Paulo: Edgar Blucher, 1988. 308p.

OLIVEIRA, L. F. C; OLIVEIRA, R. Z; BORGES, L. B; WEHR, T. R; BONOMO, R. **Coeficiente de cultura e relações hídricas do cafeeiro, cultivar Catucaí, sob dois sistemas de manejo da Irrigação**. Pesq. Agropec. Trop. 37(3): 154-162, set. 2007.

PACHECO, F. P. **Água residuária de suinocultura aplicada em cobertura na cultura do milho com cultivo seqüencial de aveia preta**. Dissertação, Cascavel. 2012

PITTMAN, S. J. E BROWN, K. A. **Multi-scale approach for predicting fish species distributions across coral reef seascapes**, PLOS ONE 6(5): ed. 20583. 2011. Disponível em: <<https://doi.org/10.1371/journal.pone.0020583>>

PONZONI, F. J. **Comportamento Espectral da Vegetação**. In: MENESES, P. R., NETTO, J. S. M. (org) Sensoriamento remoto, reflectância dos alvos naturais. Brasília – DF: Editora Universidade de Brasília - UNB, Embrapa Cerrados, p 157-199, 2001.

PRIMAVESI, A. C; RODRIGUES A. A; GODOY, R. **Recomendações técnicas para o cultivo de aveia**. Embrapa Pecuária Sudeste. Boletim de pesquisa.

RAIJ, B. V.; ANDRADE, J. C.; CANTARELLA, H.; QUAGGIO, J. A. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas: Instituto Agronômico, 2001

RAMOS, R. R. D; LOPES, H. L; MELO JUNIOR, C. F; CANDEIAS, A. L. B; SIQUEIRA FILHO, J. A. **Aplicação do índice da vegetação por diferença normalizada (NDVI) na avaliação de áreas degradadas e potenciais para unidades de conservação**. III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação Recife - PE, 27-30 de Julho de 2010 p. 001 – 006

RAZIEI, T; PEREIRA, L.S. **Estimation of ETo with Hargreaves-Samani and FAO-PM temperature methods for a wide range of climates in Iran**. Agricultural Water Management, v.121, p.1-18, 2013.

REES, W. G. **Physical principles of remote sensing. Topics in remote sensing**. 247 p. Cambridge University Press, Cambridge, 1990

ROUSE, J.W.; HAAS, R.H.; SCHELL, J.A.; DEERING, D.W. **Monitoring vegetation systems in the great plains with ERTS**. In: Third ERTS Symposium, Proceedings, NASA SP-351, NASA, Washington, DC, v. 1, p. 309-317, 1973.

ROUSH, W. B; CRAVENER, T.L; KIRBY, Y.K; WIDEMAN JR, R.F. **Probabilistic neural network prediction of ascites in broilers based on minimally invasive physiological factors**. Poultry Science; 76:1513-1516. 1997

ROUSH, W. B; KIRBY, Y. K; CRAVENER, T. L; WIDEMAN JR, R. F. **Artificial neural network prediction of ascites in broilers**. Poultry Science; 75:1479-1487, 1996.

RULEQUEST RESEARCH. **Data mining with cubist**. Disponível em:<
<http://www.rulequest.com/cubist-info.html>>

SÁ, J. P. G., 1995. **Utilização da aveia na alimentação animal**. Londrina: IAPAR, 1995, 20p.

SALVIANO, A.A.C. **Variabilidade de atributos de solo e de Crotalaria juncea em solo degradado do município de Piracicaba-SP**. Piracicaba, 1996. 91p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

SILVA, A. D. B. **Sistema de Informações Geo-referenciadas: conceitos e fundamentos**. Campinas: UNICAMP, 2003. 240 p.

SILVA, A. F. da. **Análise multivariada de dados espaciais na classificação interpretativa de solos**. Tese de doutorado. 2014 Universidade Estadual Paulista – FCA. Botucatu, 2014. 96 p. Tese de Doutorado.

SILVA, A. P. **Variabilidade espacial de atributos físicos do solo**. Piracicaba, 1988, 105p. Doutorado

SOUSA, I. F. de; FACCIOLLI, G. NETTO, A. O. SILVA, A. A. G; QUEIROZ, M. G; **Estudo do coeficiente de cultura do centro da Região Agreste do estado do Sergipe**. In: XVII Congresso Brasileiro de Agrometeorologia. SESC – Centro de Turismo de Guarapari/ES, 2011.

SOUZA, D. H. de. **Modelagem evolutiva do PLanalto Sul de Minas através de correlações entre análise morfométrica da rede de drenagem e termocronologia por traços de fissão em apatitas**. Rio Claro, 2013. 170p. Dissertação (Mestrado) - UNESP Rio Claro

SOUZA, L. C. de. **Variabilidade espacial da salinidade de um solo aluvial no semi-árido paraibano**. Campina Grande, 1999. 77p. Dissertação (Mestrado) - Universidade Federal da Paraíba

SOUZA, Z. M; CAMPOS, M. C. C; CAVALCANTE, I. H. L; MARQUES JUNIOR, J. CESARIN, L. G; SOUZA, S. R. **Dependência espacial da resistência do solo à penetração e do teor de água do solo sob cultivo contínuo de cana-de-açúcar**. Ciência Rural, Santa Maria, v.36, n.1, p.128-134, jan-fev, 2006.

SOUZA, Z. M; MARQUES JUNIOR, J; PEREIRA, G. T. **Variabilidade espacial da estabilidade de agregados e matéria orgânica em solos de relevos diferentes**. Pesquisa Agropecuária Brasileira, v.39, p.491-499, 2004.

STEVENSON, W. J; **Estatística: Aplicada à administração**. Harbra, 2001

STREIT, N. M; CANTERLE, L. P; CANTO, M. W; HECKTEUR, L. H. H. H. **As Clorofilas**. Ciência Rural, Santa Maria, v.35, n.3, p.748-755, mai-jun, 2005.

TEIXEIRA, A. H. de C.; LIMA FILHO, J. M. P. **Relações entre o índice de área foliar e radiação solar na cultura da videira**. Revista Brasileira de Agrometeorologia, Santa Maria, v.5, n.2, p.143-146, 1997.

TRAGMAR, B; YOST, R. S; UEHARA, G. **Application of geostatistics to spatial studies of soil properties**. Advances in Agronomy, New York, v.38, n.1, p.45-93, 1985.

UNITED STATES GEOLOGICAL SURVEY – USGS. Earth explorer. 2016. Disponível em:< <https://earthexplorer.usgs.gov/> > . Acesso em: 10 set. 2016.

VAPNIK, V. **The Nature of Statistical Learning Theory**. New York: Springer-Verlag, 1995.

VEIGA, T. C. & SILVA, J. X. da. **Geoprocessamento aplicado à identificação de áreas potenciais para atividades turísticas: O caso do Município de Macaé – RJ**. In: SILVA, J. X. & ZAIDAN, R. T. (Org.) Geoprocessamento & Análise ambiental. Rio de Janeiro: Bertrand Brasil, p.179-215, 2004.

VIEIRA, S. R; HATFIELD, J. L; NIELSEN, D. R; BIGGAR, J. W. **Geostatistical theory and application to variability of some agronomical properties**. Hilgardia, Berkeley, v.51, n.1, p.1-75, 1983.

VIEIRA, S. R. **Geoestatística em estudos de variabilidade espacial do solo**. In: Novais, R. F. de; Alvarez, V. V. H.; Schaefer, (ed.), C. E. G. R. Tópicos em Ciência do Solo, Sociedade Brasileira de Ciência do Solo, v.1, p.1-54, 2000.

ZIMBACK, C. R. L. **Análise espacial de atributos químicos de solos para fins de mapeamento da fertilidade do solo**. 2001. 114f. Tese (Livre-Docência em Levantamento do solo e Fotopedologia), Faculdade de Ciências Agronômicas, Universidade Estadual Paulista, Botucatu, 2001.