

**UFRRJ**

**INSTITUTO DE AGRONOMIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
AGRONOMIA CIÊNCIA DO SOLO**

**TESE**

**Eficiência de Diferentes Interpoladores no  
Mapeamento de Atributos Físico Hídricos do Solo  
em Área de Produção Agroecológica**

**Martin de Oliveira Freire**

**2017**



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
INSTITUTO DE AGRONOMIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA  
CIÊNCIA DO SOLO**

**EFICIÊNCIA DE DIFERENTES INTERPOLADORES NO  
MAPEAMENTO DE ATRIBUTOS FÍSICO HÍDRICOS DO SOLO EM  
ÁREA DE PRODUÇÃO AGROECOLÓGICA**

**MARTIN DE OLIVEIRA FREIRE**

*Sob a Orientação do Professor*

**Marcos Bacis Ceddia**

*e Co-orientação da Professora*

**Michele Duarte Menezes**

Tese submetida como requisito parcial  
para obtenção do grau de **Doutor**, no  
Programa de Pós-Graduação em  
Agronomia, Área de Concentração em  
Ciência do Solo.

Seropédica, RJ  
Fevereiro de 2017

Universidade Federal Rural do Rio de Janeiro  
Biblioteca Central/Seção de Processamento Técnico

Ficha catalográfica elaborada  
Com os dados fornecidos pelo(a) Autor(a)

F866e	<p>Freire, Martin de Oliveira, 1971- Eficiência de Diferentes Interpoladores no Mapeamento de Atributos Físico Hídricos do Solo em Área de Produção Agroecológica/Martin de Oliveira Freire. – Seropédica, RJ, 2017. 74 f.: il.</p> <p>Orientador: Marcos Bacis Ceddia. Tese (Doutorado). – – Universidade Federal Rural do Rio de Janeiro, Pós-Graduação em Agronomia – Ciência do Solo, 2017.</p> <p>1. Geoestatística. 2. Interpoladores. 3. Atributos físico-hídricos do solo. I. Ceddia, Marcos Bacis, 1968-, orient. II. Universidade Federal Rural do Rio de Janeiro. III. Título.</p>
-------	---

**O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.**

<p>É permitida a cópia parcial ou total desta Tese, desde que seja citada a fonte.</p>
--

**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
INSTITUTO DE AGRONOMIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA - CIÊNCIA DO SOLO**

**MARTIN DE OLIVEIRA FREIRE**

Tese submetida como requisito parcial para obtenção do grau de **Doutor**, no Programa de Pós-Graduação em Agronomia, área de Concentração em Ciência do Solo.

TESE APROVADA EM 27/02/2017.

---

Marcos Bacis Ceddia. Dr. UFRRJ  
(Orientador)

---

Marcos Gervasio Pereira. Dr. UFRRJ

---

Helena Saraiva Koenow Pinheiro Dra. UFRRJ

---

David Villas Boas de Campos Dr. Embrapa Solos

---

Ednaldo da Silva Araújo Dr. Embrapa Agrobiologia

## **AGRADECIMENTOS**

Ao Programa de Pós Graduação em Ciência do Solo pela oportunidade de cursar o doutorado e desenvolver esse trabalho de tese.

Ao CNPq pela bolsa de pesquisa.

À UFRRJ por todas as oportunidades que me foram oferecidas, desde o início da graduação, passando pelas experiências profissionais de docência e pelo atual emprego como Engenheiro Agrônomo.

Aos meus amigos da pós pelas inúmeras ajudas prestadas.

À minha amiga Andréa da Silva Gomes, pela grande ajuda nos trabalhos.

Aos professores Marcos Gervasio Pereira e Lúcia Helena dos Anjos pela presença constante e pelo incentivo nos momentos mais difíceis nos últimos anos.

À minha co orientadora Michele, pelas conversas e auxílios à distância.

Ao meu orientador Marcos Bacis Ceddia, pela orientação, ensino, incentivo e amizade sempre presentes.

À minha família.

E, em especial, para minhas queridas Laura, Luiza e Ruth, que sempre tiveram muita paciência comigo nos últimos anos. Sem vocês, nada disso faria sentido.

## RESUMO

FREIRE, Martin de Oliveira. **Eficiência de diferentes interpoladores no mapeamento de atributos físico hídricos do solo em área de produção agroecológica.** 74 f. 2017. Tese (Doutorado em Agronomia, Ciência do Solo). Instituto de Agronomia, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ. 2017.

Neste trabalho foram testados diferentes interpoladores para estimar atributos físico hídricos do solo numa área de pesquisa e produção agroecológica (SIPA) localizada no Município de Seropédica/RJ. Os atributos interpolados foram Capacidade de Campo (CC), Água Facilmente Disponível (AFD), Água Disponível Total (ADT), Água retida a 80 KPa (80 KPa) e Ponto de Murcha Permanente (PMP). Esses atributos foram obtidos de um banco de dados com valores em 3 profundidades: 0-10 cm, 10-20 cm e 20-30 cm. Os métodos de interpolação utilizados foram Krigagem Ordinária (OK), Co Krigagem Isotópica (CK) e Krigagem com Regressão (RK). Os resultados obtidos com OK foram comparados com os obtidos por CK e RK para avaliar se há melhoria na acurácia dos resultados. A CK não apresentou resultados viáveis para validação em todas as combinações, apresentando pouca melhoria em relação à OK. A RK também não resultou em ganho de acurácia com os atributos testados. Para a diminuição do número de amostras os atributos 80 KPa e AFD foram interpolados com OK e RK. Não houve diminuição significativa da acurácia dos resultados com a diminuição para as proporções de 90%, 80%, 70%, 60% e 50% dos dados. Sendo possível interpolar os atributos com 50% do total do banco de dados utilizado.

**Palavras chave:** Geoestatística. Interpoladores. Atributos físico-hídricos do solo.

## ABSTRACT

FREIRE, Martin de Oliveira. **Efficiency of different methods for mapping of soil physical hydric attributes an agroecology integrated production system.** 74p. 2017. Thesis (Doctorate in Agronomy, Soil Science). Instituto de Agronomia, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ. 2017.

Different methods of interpolation was tested for soil physical hydric properties in a research area of agroecology production (SIPA) in Seropédica / RJ. The attributes tested was Field Capacity (FC), Easy Water Available (EWA), Available Water Capacity (AWC), Water at Tension of 80 KPa (80 KPa) and Wilting Point (WP). This attributes came from a dataset with values in three different depths: 0-10 cm, 10-20 cm e 20-30 cm. Ordinary Kriging (OK) and CoKriging (CK) methods was used as interpolators. The results obtained with OK were compared with those obtained by CK and KR to evaluate if there was improvement in the accuracy of the results. CK did not present viable results for validation in all combinations, showing little improvement over OK. The RK also did not result in gain of accuracy with the attributes tested. For the decrease of the number of samples the attributes 80 KPa and AFD were interpolated with OK and KR. There was no significant decrease in the accuracy of the results with the decrease for the proportions of 90%, 80%, 70%, 60% and 50% of the data. It is possible to interpolate the attributes with 50% of the total database used.

**Keywords:** Geostatistics. Interpolators. Soil physic hydric attributes.

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Localização do SIPA (VILLELA, 2007). .....	4
<b>Figura 2.</b> Representação gráfica do semivariograma. Os pontos azuis representam o semivariograma experimental e a linha tracejada o modelo teórico.....	7
<b>Figura 3.</b> Grade de amostragem de textura, Densidade do Solo (Ds); Capacidade de Campo (CC); Água facilmente disponível (AFD); Ponto de Murcha permanente (PMP); Porosidade de Aeração (PA); Porosidade Total (PT) (VILELLA, 2007). .....	11
<b>Figura 4.</b> Fluxograma resumido da sequência das análises realizadas.....	13
<b>Figura 5.</b> Distribuição espacial dos pontos de 80 KPa 0-10 cm de acordo com a porcentagem dos dados utilizados para interpolação. ....	16
<b>Figura 6.</b> Distribuição espacial dos pontos de AFD 10-20 cm de acordo com a porcentagem dos dados utilizados para interpolação. ....	17
<b>Figura 7.</b> Distribuição espacial dos pontos de validação para os interpoladores. ....	18
<b>Figura 8.</b> Variáveis geomorfológicas derivadas a partir do modelo digital de elevação. MDE – Modelo Digital de Elevação; WI – Índice Topográfico de umidade.....	21
<b>Figura 9.</b> Variáveis geomorfológicas derivadas a partir do modelo digital de elevação. MRVBF - Índice multiresolução do nivelamento do fundo de vales ; MRRTF - Índice multiresolução de nivelamento de topo de paisagens; AACN - Altitude Acima da Rede de Canais. ....	23
<b>Figura 10.</b> Semivariogramas da elevação e dos atributos físicos na profundidade 0-10 cm. A- CC; B-80 KPa; C - PMP; D - AFD; E - ADT; F - DS; G - DP; H - Areia; I - Silte; J - Argila; K - PT; L - Aeração; M - Elevação. Entre parênteses estão os parâmetros do semivariogramas.....	26
<b>Figura 11.</b> Semivariogramas dos atributos físicos na profundidade 10-20 cm. A- CC; B-80 KPa; C - PMP; D - AFD; E - ADT; F - DS; G - DP; H - Areia; I - Silte; J - Argila; K - PT; L - Aeração. Entre parênteses estão os parâmetros do semivariogramas. ....	28
<b>Figura 12.</b> Semivariogramas dos atributos físicos na profundidade 20-30 cm. A- CC; B-80 KPa; C - PMP; D - AFD; E - ADT; F - DS; G - DP; H - Areia; I - Silte; J - Argila; K - PT; L - Aeração. Entre parênteses estão os parâmetros do semivariogramas. ....	29
<b>Figura 13.</b> Semivariogramas de CC 20-30 cm (A) e atributo Argila (B) com semivariogramas cruzado (C); semivariogramas de 80 KPa 0-10 cm (D) e atributo argila (E) com semivariogramas cruzado (F). Entre parênteses estão os parâmetros do semivariogramas. ....	32
<b>Figura 14.</b> Semivariogramas de 80 KPa 10-20 cm (A e D), atributos Argila (B) e Silte (E) com seus respectivos semivariogramas cruzado (C e F). Entre parênteses estão os parâmetros do semivariogramas. ....	33
<b>Figura 15.</b> Semivariogramas de PMP 10-20 cm (A) e silte (B) e respectivo semivariogramas cruzado (C); semivariogramas de PMP 20-30 cm (D) e argila (E) e respectivo semivariogramas cruzado (F). Entre parênteses estão os parâmetros do semivariogramas. ....	34
<b>Figura 16.</b> Semivariogramas de AFD 10-20 cm (A e D) e atributos CNBL (B) e Elevação (E) com seus respectivos semivariogramas cruzados (C e F). Entre parênteses estão os parâmetros dos semivariogramas.....	35



<b>Figura 17.</b> Semivariogramas de AFD 20-30 cm (A e D) e atributos CNBL (B) e areia (E) com seus respectivos semivariogramas cruzados (C e F). Entre parênteses estão os parâmetros dos semivariogramas. ....	36
<b>Figura 18.</b> Semivariogramas de ADT 10-20 cm (A) e atributo argila (B) com respectivo semivariograma cruzado (C). E semivariogramas de ADT 20-30 cm (D e E) e atributo argila (F). Entre parênteses estão os parâmetros dos semivariogramas.....	37
<b>Figura 19.</b> Silte (G) com seus respectivos semivariogramas cruzados (H e I). Entre parênteses estão os parâmetros dos semivariogramas. ....	38
<b>Figura 20.</b> Gráfico com os resultados de RMSE para 80 KPa 0-10 cm e AFD 10-20 cm obtidos por OK e RK nas seis proporções usadas do banco de dados (Eixo X em % do total de dados).....	48

## ÍNDICE DE TABELAS

<b>Tabela 1.</b> Estatística descritiva das covariáveis numéricas do terreno. ....	22
<b>Tabela 2.</b> Parâmetros dos modelos dos semivariogramas dos atributos físicos do solo nas três profundidades do solo. ....	25
<b>Tabela 3.</b> Resultado da validação da Krigagem Ordinária dos atributos principais. ....	30
<b>Tabela 4.</b> Validação dos resultados viáveis obtidos por co krigagem. ....	39
<b>Tabela 5.</b> Melhora relativa da Co Krigagem (CK) comparada à Krigagem Ordinária (OK). .	40
<b>Tabela 6.</b> Modelos de regressão para as variáveis de interesse na profundidade de 0-10 cm.	41
<b>Tabela 7.</b> Modelos de regressão para as variáveis de interesse na profundidade de 10-20 cm. ....	41
<b>Tabela 8.</b> Modelos de regressão para as variáveis de interesse na profundidade de 20-30 cm. ....	41
<b>Tabela 9.</b> Validação dos resultados viáveis obtidos por Krigagem com Regressão (RK). X: não foi possível completar a análise. ....	42
<b>Tabela 10.</b> Melhora relativa (MR) da Krigagem com Regressão (RK) comparada à Krigagem Ordinária (OK). ....	43
<b>Tabela 11.</b> Validação da OK para redução de dados de 80 KPa de 0-10 cm. ....	44
<b>Tabela 12.</b> Validação da OK para redução de dados de 80 KPa de 10-20 cm. ....	44
<b>Tabela 13.</b> Melhora relativa (MR) da OK para redução dos pontos da variável 80 KPa 0-10 cm, comparada com resultados da OK 100%. ....	44
<b>Tabela 14.</b> Melhora relativa (MR) da OK para redução dos pontos da variável AFD 10-20 cm, comparada com resultados da OK 100%. ....	45
<b>Tabela 15.</b> Modelos de regressão para 80 KPa na profundidade de 0-10 cm com diferentes proporções do banco de dados. ....	45
<b>Tabela 16.</b> Modelos de regressão para AFD na profundidade de 10-20 cm com diferentes proporções do banco de dados. ....	46
<b>Tabela 17.</b> Validação da RK para redução de dados de 80 KPa de 0-10 cm. ....	46
<b>Tabela 18.</b> Validação da RK para redução de dados de AFD de 10-20 cm. ....	46
<b>Tabela 19.</b> Melhora relativa (MR) da RK para redução dos pontos da variável 80 KP a 0-10 cm, comparada com RK 100%. ....	47
<b>Tabela 20.</b> Melhora Relativa (MR) da RK para redução dos pontos da variável AFD 10-20 cm, comparada com RK 100%. ....	47
<b>Tabela 21.</b> Comparação entre Krigagem Ordinária (OK) e Krigagem com regressão (RK) para estimar a variável 80 KPa 0-10 cm com redução na porcentagem do banco de dados. ....	47
<b>Tabela 22.</b> Comparação entre Krigagem Ordinária (OK) e Krigagem com Regressão (RK) para estimar a variável AFD 10-20 cm com redução na porcentagem do banco de dados. ....	48

## SUMÁRIO

1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA .....	2
2.1. O Sistema Integrado de Produção Agroecológica (SIPA), Breve Histórico e Principais Dados Físico-Hídricos Existentes.....	3
2.2. Métodos Interpoladores e Amostragem.....	5
2.2.1. Interpoladores em geral.....	5
2.2.2. Krigagem.....	5
2.2.3. Cokrigagem.....	8
2.2.4. Krigagem com regressão.....	9
3. MATERIAL E MÉTODOS.....	11
3.1. Área de Estudo e os Dados Físico-Hídricos Existentes.....	11
3.2. Atributos do Relevo.....	12
3.3. Definição da Proporção de Dados a Serem Utilizados para Desenvolvimento e Validação dos Algoritmos.....	12
3.4. Krigagem, Cokrigagem e Krigagem com Regressão dos Atributos Físico-Hídricos.....	13
3.5. Redução no Número de Amostras Necessárias para Interpolação dos Atributos Físico-Hídricos.....	15
4. RESULTADOS E DISCUSSÃO.....	19
4.1. Covariáveis do Relevo.....	19
4.2. Organização e Estatística Descritiva dos Conjuntos de Treinamento e Validação.....	24
4.3. Avaliação Geoestatística dos Atributos Físico-Hídricos.....	24
4.4. Co Krigagem.....	31
4.5. Krigagem com Regressão.....	40
4.6. Diminuição da Proporção dos Dados de Treinamento.....	43
5. CONCLUSÕES.....	49
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	50
7. ANEXOS.....	56

## 1. INTRODUÇÃO

O conhecimento dos atributos físicos do solo é fundamental para o bom manejo do mesmo. Dentre eles, os atributos físico-hídricos são essenciais para escolha das culturas que se pretende cultivar. Tanto para a escolha da localização espacial quanto para fornecer subsídios que permitam uma irrigação com a melhor eficiência no uso de água. A determinação dos atributos físico hídricos relacionados ao armazenamento de água no solo é trabalhosa (EMBRAPA, 1997) e a busca por métodos que permitam mapear esses atributos de forma acurada é muito importante. As técnicas de mapeamento digital, usando diferentes modelos de predição são alternativa para obter mapas acurados e precisos desses atributos.

O presente trabalho se propõe estudar a variabilidade espacial de cinco diferentes atributos relacionados à disponibilidade de água no solo numa área de pesquisa e produção agroecológica em Seropédica. Para tanto foi usado banco de dados com cinco diferentes medições de disponibilidade de água em três diferentes profundidades, de 0-10 cm, 10-20 cm e 20-30 cm. As medições correspondem a Capacidade de Campo, Água retida a 80 KPa, Água Facilmente Disponível, Água Disponível Total e Ponto de Murcha Permanente.

A variabilidade espacial desses atributos físicos-hídricos pode apresentar relação com outros atributos físicos do solo e atributos do relevo. Além de estudar a variabilidade espacial desses atributos através da Krigagem, também foi avaliada (quantificada) essa relação com os demais atributos do solo. A quantificação dessas relações foi feita com técnicas geoestatísticas multivariadas e híbridas (Cokrigagem e Krigagem com Regressão). A proposta é que essas técnicas permitam o mapeamento desses atributos físico-hídricos dos solos utilizando um menor número de amostras sem que haja perda significativa na acurácia dos mapas gerados.

A importância de estudar a redução no número de amostras está diretamente relacionada ao custo e esforço necessário para obter esses dados. De acordo com alguns autores, são necessários pelo menos 100 pontos para que se obtenha uma boa interpolação (Webster & Oliver, 1992). A proposta desse trabalho é avaliar como a acurácia dos resultados de diferentes interpoladores se comporta com a redução gradual do número de amostras.

A hipótese do estudo é de que é possível manter a acurácia e precisão de diferentes métodos de interpolação geoestatísticas diminuindo o número de amostras usadas na análise.

O objetivo geral foi avaliar o desempenho de diferentes interpoladores geoestatísticos puros (Krigagem ordinária e Cokrigagem) e híbridos (Regression Kriging) para o mapeamento de atributos físico-hídricos do solo (Capacidade de Campo, Água retida a 80 kPa, Ponto de Murcha Permanente, Água Facilmente Disponível e Água Disponível Total) que demandem menos amostras de solo e que explorem mais eficientemente a informação de covariáveis do relevo.

No desenvolvimento do estudo foram exploradas as seguintes questões sobre o mapeamento de atributos do solo:

- Quantas amostras são necessárias para estimar atributos físico hídricos do solo?
- Quais variáveis auxiliares de fácil obtenção são viáveis para aperfeiçoar a estimativa de atributos físico hídricos?
- Em quais situações o uso da krigagem com regressão e a co krigagem permite ganho de acurácia, comparado com krigagem ordinária.

Os objetivos específicos foram: i) quantificar as relações espaciais entre atributos físico-hídricos dos solos e do relevo, testando diferentes combinações de número e distribuição espacial de amostras de solo; e ii) explorar o potencial de covariáveis do relevo para aperfeiçoar a interpolação geoestatística (cokrigagem e regression kriging).

## 2. REVISÃO DE LITERATURA

A qualidade do solo está relacionada à capacidade de um solo de manter uma boa produtividade vegetal, e animal, ao longo do tempo (KARLEN et al., 1997). Uma forma de quantificar a qualidade do solo é estudar as funções do solo através de determinados indicadores. Sejam eles físicos, químicos ou biológicos (KARLEN et al., 1997).

Como exemplo de propriedades físicas do solo que afetam diretamente a produtividade agrícola temos a disponibilidade de água, aeração, resistência à penetração e temperatura do solo (LETEY, 1985). De acordo com o autor, a disponibilidade de água seria o fator mais importante e os demais seriam afetados pela quantidade de água presente no solo. Além disso, a densidade do solo e a textura podem ser usados como fatores indiretos, pois influenciam na relação entre disponibilidade de água e aeração com a resistência à penetração (LETEY, 1985). Seguindo esse raciocínio, um solo com boa qualidade seria aquele que possa fornecer uma boa quantidade de água às plantas, sem que houvesse impedimento à penetração das raízes e sem problemas para aeração das mesmas.

A utilização de água na irrigação sempre deve ser pensada sob a ótica de evitar o desperdício (MEDICI et al., 2014). Sistemas de irrigação mais eficientes no uso da água são uma demanda atual, considerando o desperdício de água que certos sistemas ainda causam. Dessa forma, a determinação dos atributos hídricos é uma etapa importante para qualquer projeto de irrigação

Através da determinação dos valores de Capacidade de Campo (CC), Ponto de Murcha Permanente (PMP), Água Facilmente Disponível (AFD) e Água Disponível Total (ADT) é possível propor manejos mais eficientes quanto a irrigação. Esses atributos irão indicar a disponibilidade hídrica do solo. Já outros, como porosidade de aeração (PA) e resistência à penetração (RP) são importantes para avaliar as trocas gasosas e reações de oxidação no solo, bem como o grau de compactação/adensamento que interfere no desenvolvimento radicular das plantas, respectivamente (LETEY, 1985, CEDDIA et al, 2014).

Um problema para a elaboração de mapas detalhados desses atributos é o custo e o tempo necessário para obter uma grande quantidade de dados desses atributos. Dessa forma a utilização de técnicas de Mapeamento Digital de Solos (MDS) é uma opção extremamente interessante para a obtenção de mapas detalhados que contenham informações relevantes e úteis para tomada de decisão sobre os manejos a serem implantados numa determinada área.

A determinação exata dos valores de CC e PMP são objeto de algumas discussões na literatura (REICHARDT & TIMM, 2004; VAN LIER, 2010). O mais comum é que esses valores sejam referentes à quantidade de água retida no solo de 10 a 30 kPa (CC) e 1500 kPa (PMP). A ADT refere-se ao valor da diferença entre CC e PMP. Esta diferença pode ser chamada de capacidade hídrica, capacidade de água disponível, disponibilidade hídrica ou água disponível no solo (REICHARDT, 1990).

Utset et al. (2000), em solos de Cuba, usaram diferentes técnicas de determinação de densidade de solo e CC para predição desses atributos. A krigagem apresentou os melhores resultados quando comparado com a predição baseada nos mapas de solos. Villela (2007) trabalhou utilizando o conceito de Água Facilmente Disponível (AFD) em área de pesquisa dentro da EMBRAPA Agrobiologia (SIPA). O atributo apresentou dependência espacial, e pode ser interpolada por krigagem ordinária. Os resultados desse trabalho serviram de base para criar o mapa de qualidade físico hídrica da área do SIPA, através de algoritmo e de software que foram desenvolvidos no estudo (CEDDIA et al., 2014).

Farias (2008) trabalhou com dados do PROJIR (Projeto de Irrigação e Drenagem do Norte Fluminense) utilizando o conceito de ADT. A determinação da ADT foi a partir de valores já determinados de CC e PMP. No estudo dos semivariogramas foi observada dependência espacial desse atributo, o que permitiu a elaboração de mapas por krigagem.

Apesar de serem relativamente menos frequentes os trabalhos de MDS aplicados com esses atributos físico-hídricos do solo, existem alguns trabalhos com condutividade hidráulica e retenção de água (FARIAS, 2008), que também são muito importantes para o estudo da dinâmica da água no solo. Assim como trabalhos que utilizam funções de pedotransferência (GIVI et al., 2004; PIEDALLU et al., 2011) para determinação de CC e PMP.

Ainda são necessários mais estudos para determinar qual das diferentes técnicas geoestatísticas deve ser escolhida de forma a otimizar (aumentar a acurácia e reduzir o número de amostras) o trabalho de mapeamento de atributos hídricos do solo.

## **2.1. O Sistema Integrado de Produção Agroecológica (SIPA), Breve Histórico e Principais Dados Físico-Hídricos Existentes**

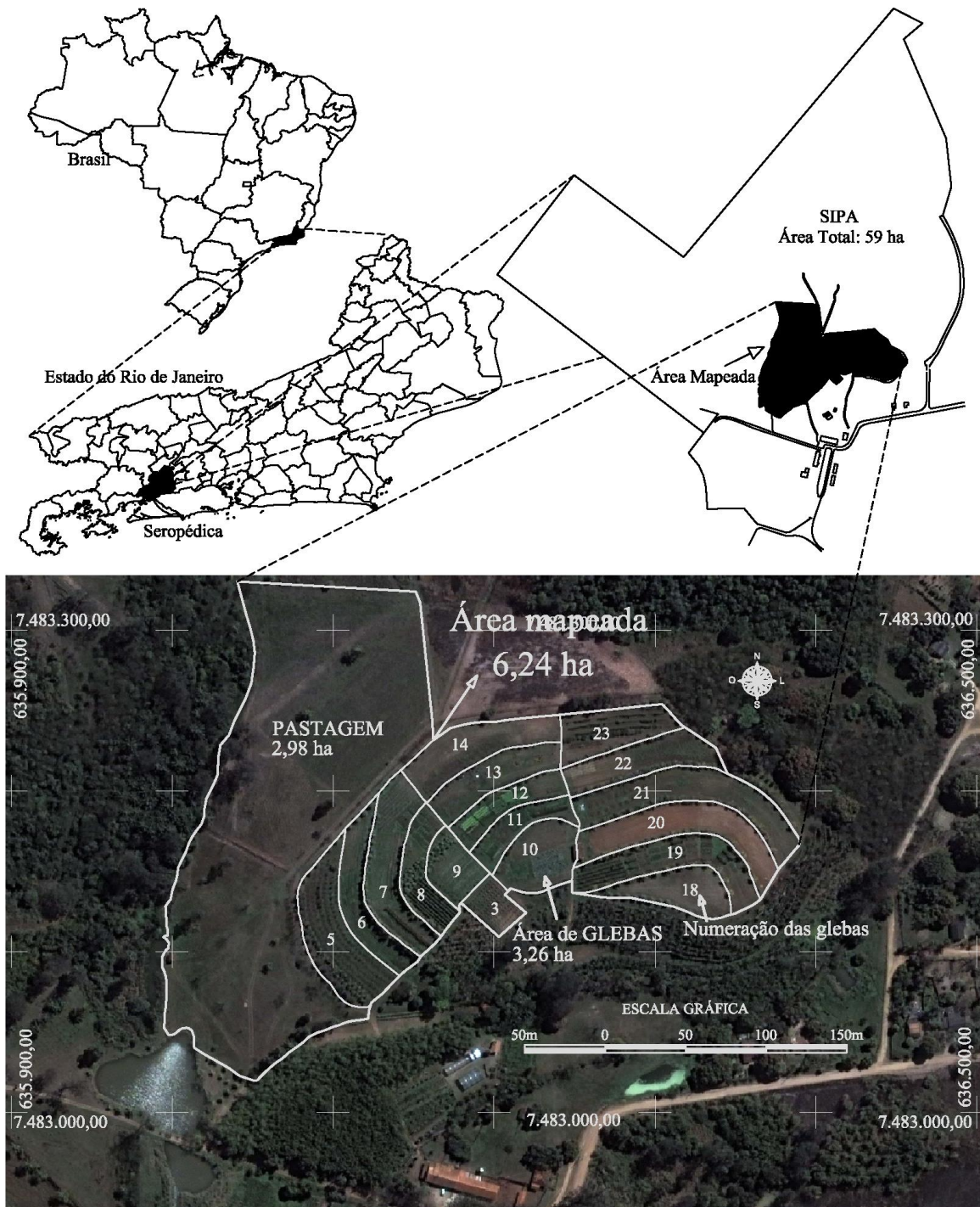
O SIPA é uma área de pesquisa em produção agrícola usando técnicas agroecológicas, o qual foi criado em 1993 através de um trabalho conjunto da UFRRJ, PESAGRO-RIO, EMBRAPA Agrobiologia e EMBRAPA Solos (ALMEIDA et al, 2003; ABBOUD et al., 2005). A área está localizada no Município de Seropédica, possui um total de 59 ha, porem apenas uma parte, 6,24 ha, é utilizada como campo de produção e experimentação (Figura 1). Nessa área foram desenvolvidas diversas teses e pesquisas científicas, sendo uma referência dentro do estudo de agroecologia no Brasil e no mundo.

Dentre os diferentes estudos conduzidos na área, destacam-se aqueles desenvolvidos por Villela (2007), Mata (2012), Ceddia et al. (2009) e Ceddia et al. (2014), os quais foram realizados visando aferir tanto a qualidade do solo quanto prover uma melhor caracterização dos atributos do solo. Esses trabalhos empregaram a krigagem ordinária para gerar mapas de variabilidade espacial de atributos físico-hídricos, químicos e da qualidade do solo, culminando na criação de um algoritmo e software denominado Qualifis (CEDDIA et al., 2014). O algoritmo combina diferentes atributos do solo para geração de mapas de qualidade físico-hídrica do solo para o uso agrícola.

Esses trabalhos de mapeamento e monitoramento são importantes para que se possa, ao longo do tempo, avaliar como alguns atributos do solo se comportam em relação aos diferentes manejos e práticas adotados no local. Porém, existe o problema em relação a forma como o monitoramento deve ser realizado para cada atributo.

A coleta de dados físico hídricos de solos é um processo oneroso, tanto considerando o trabalho de campo quanto o de laboratório para geração desses valores. Para a geração de mapas a partir de diferentes interpoladores, sempre são sugeridos o uso de extensos bancos de dados, o que implica em trabalhos de coleta intensos. Porém não há clareza na literatura sobre o tamanho ideal, ou o espaçamento e tipo de grade amostral necessário para a obtenção de mapas de boa qualidade. Dependendo da escala e da resolução que se propõe, o número de amostras e o espaçamento entre estas pode sofrer grande alteração.

Uma vez que se obtenha mapa, a partir dos atributos coletados, surge a necessidade de sua atualização periódica. Caso sejam informações que possam ser usadas para monitorar a qualidade do solo ao longo do tempo, esse intervalo de atualização de dados deve ser o mais regular possível. O problema é o custo desse monitoramento. Por isso torna-se necessário pesquisar como diminuir o número de amostras necessárias sem que se perca a qualidade e acurácia do mapa gerado.



**Figura 1.** Localização do SIPA (VILLELA, 2007).

A proposta desse projeto foi estudar como a qualidade do interpolador escolhido para diferentes atributos físico hídricos do solo varia quando se diminui o número de pontos amostrais utilizados. E também se o produto final da interpolação apresenta melhorias com a utilização de variáveis predictoras auxiliares. Os resultados foram gerados através de técnicas geoestatísticas (OK, CK e RK). Os resultados desse estudo podem apontar em grande economia de recursos financeiros e logísticos no processo de monitoramento de atributos físico hídricos utilizados para avaliar a qualidade de solos.

## 2.2. Métodos Interpoladores e Amostragem

### 2.2.1. Interpoladores em geral

Todo estudo que se propõe a mapear a variabilidade espacial de um atributo qualquer, após as medições de campo e tabulações de dados, passa pela etapa de interpolação de dados (LI & HEAP, 2008). A interpolação é um processo de estimativa de valores entre pontos amostrados, de forma a aferir valores em pontos não medidos dentro de uma determinada área. A busca comum em qualquer processo de interpolação é que o interpolador tenha elevada acurácia e precisão, o que leva à busca contínua de qual a metodologia é melhor aplicada para cada caso. De uma forma geral, os métodos de inferência podem ser classificados em três categorias: estatísticos (ex. árvore de regressão e regressão múltipla); geoestatísticos (ex. krigagem ordinária e cokrigagem); híbridos (ex. krigagem com regressão e árvore de regressão combinada com krigagem) (LI & HEAP, 2008; ZHU & LIN, 2010).

Quase todos os métodos preditores, incluindo as formas mais simples de krigagem, podem ser vistos como médias ponderadas dos dados. Assim, a equação geral dos estimadores pode ser escrita como (WEBSTER & OLIVER, 2001):

$$z^*(x_0) = \sum_{i=1}^N \lambda_i z(x_i), \quad (\text{eq-1})$$

em que:

$z(x_0)$  - é o ponto a ser predito;

$z(x_i)$  - são os pontos medidos ( $i = 1, 2, 3, 4 \dots N$ );

$\lambda_i$  - são os pesos associados a cada ponto medido

Analisando a fórmula acima, constata-se que a grande diferença entre os métodos de predição é a forma como os pesos  $\lambda_i$  são atribuídos a cada um dos pontos medidos na estimativa de um ponto qualquer. A seguir são apresentados alguns métodos de interpolação (LI & HEAP, 2008).

### 2.2.2. Krigagem

A krigagem é uma coleção de técnicas de regressão linear generalizada para minimizar estimativa da variância definida a partir de modelo para uma covariância (DEUSTCH & JOURNAL, 1998). Trata-se de interpolador que não apresenta tendência e possui variância mínima. Essa abordagem foi inicialmente desenvolvida por Daniel Krige, que usou o método para prever localização de minas de ouro na África do Sul (OLIVER & WEBSTER, 2014). Na década de 1960 o matemático francês G. Matheron formalizou o método criando a teoria das variáveis regionalizadas (KERRY & OLIVER, 2007). A krigagem é uma média móvel ponderada, onde os pesos atribuídos são variáveis de acordo com a variabilidade espacial que será expressa no semivariograma (VIEIRA, 2000). O que torna a krigagem um interpolador ótimo é a maneira como os pesos são determinados e distribuídos (equação abaixo).

$$z^*(x_0) = \sum_{i=1}^N \lambda_i z(x_i) \quad (\text{eq-2})$$

em que:

$z^*(x_0)$  - valor a ser estimado;

$N$  - número de valores medidos,  $z(x_i)$ , envolvidos na estimativa; e.



$\lambda_i$  – são os pesos associados a cada valor medido,  $z(x_i)$ , de acordo com o resultado do semivariograma.

O uso da técnica de krigagem para inferir pontos não amostrados em propriedades do solo é inicialmente apresentada nos trabalhos de Burges & Webster (1980a, 1980b), Webster & Burges (1980), Burges et al. (1981), Vieira et al. (1981), Yost et al. (1982) e Braga e Druck (1993). Esses autores mostraram a boa adequação dos interpoladores geoestatísticos ao estudo de dados de solos e consideraram que a krigagem apresenta uma estrutura teórica desejável, condições favoráveis de implementação e uma estabilidade às variações amostrais.

Como a krigagem é um interpolador geoestatístico, algumas considerações são importantes na escolha deste método de interpolação. A geoestatística objetiva o estudo dos fenômenos que variam no tempo e/ou no espaço, oferecendo uma relação de ferramentas determinísticas e estatísticas que possibilitam entender e modelar a variabilidade espacial (DEUTSCH & JOURNAL, 1998). Todos os conceitos de geoestatística têm suas bases teóricas em funções e variáveis aleatórias. Uma função aleatória, pelo fato de ser contínua, pode ser submetida a uma grande gama de hipóteses, sem as quais a dedução das equações é impossível. Quando se têm algumas realizações de certa variável aleatória, e deseja-se estimar valores para os locais não amostrados, se faz necessário à introdução de uma restrição de que a variável regionalizada seja estacionária estatisticamente.

De acordo com Vieira (2000), uma variável regionalizada é considerada estacionária se os momentos estatísticos da variável aleatória  $Z(x_i+h)$ , média e variância, forem os mesmos para qualquer vetor  $h$ . A existência de estacionariedade dá a oportunidade de repetir um experimento mesmo que as amostras sejam coletadas em pontos diferentes, porque todas são consideradas pertencentes a populações com os mesmos momentos estatísticos (VIEIRA, 2000). Segundo Deutsch & Journel (1998), a estacionariedade é uma propriedade para que o modelo possa ser aplicado, e assim torna-se mais uma decisão do que uma hipótese de trabalho. A decisão de estacionariedade pode mudar com a escala do trabalho e de acordo com a disponibilidade dos dados.

Para a geoestatística o variograma é a peça fundamental no estudo e modelagem. Porque tanto a descrição quanto o entendimento da variabilidade espacial quantitativa serão realizados a partir do variograma modelado, sendo a base para se fazer previsões a respeito do fenômeno estudado (MCBRATNEY & WEBSTER, 1986). O variograma é uma função matemática (eq-3) definida para representar e quantificar o nível de dependência espacial que existe entre duas variáveis aleatórias locais. Desta forma, pode ser aplicado em diferentes tipos de análises. Estas duas variáveis podem ser: o mesmo atributo em duas posições diferentes (autovariograma), dois atributos diferentes medidos na mesma posição (covariância das variáveis) ou dois atributos diferentes medidos em duas posições (variograma cruzado).

$$2\gamma(h) = \frac{1}{n} \sum_1^n \{Z(x_i) - Z(x_i + h)\}^2$$

(eq-3)

Em que:

$2\gamma(h)$  – é a variância em função do vetor  $h$ ;

$\sum$  – valor esperado;

$Z(x_i)$  – valor da variável na posição  $x_i$ ; e

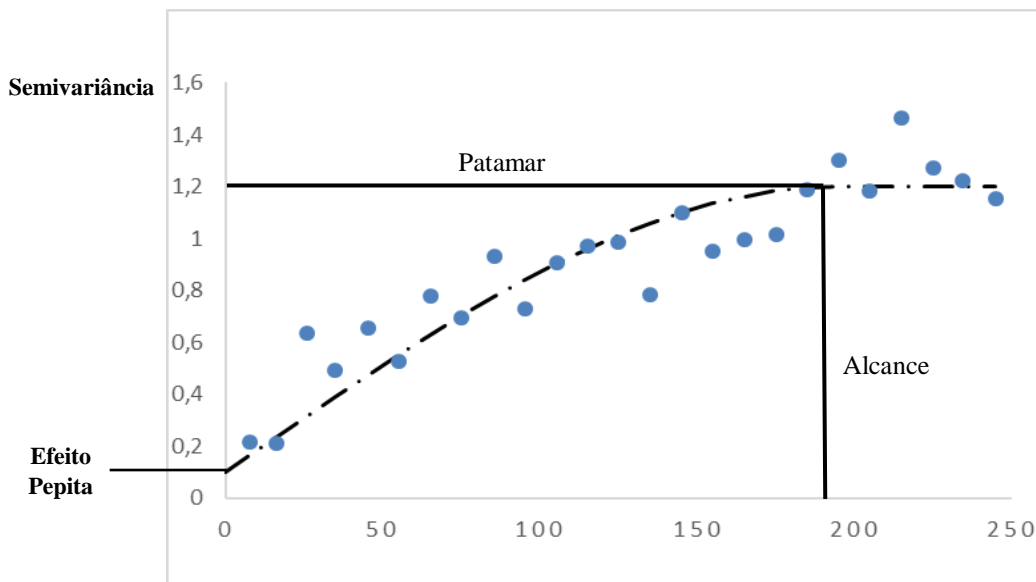
$Z(x_i+h)$  - valor da variável na posição  $x_i+h$ .

O semivariograma (eq-4) é a função anterior dividida por 2, função  $\gamma(h)$ . O fator 2 é utilizado com a finalidade de cancelamento e simplificação da equação e porque a quantidade mais frequentemente utilizada é  $\gamma(h)$  e não  $2\gamma(h)$ .

$$\gamma(h) = \frac{1}{2n} \sum_1^n \{Z(x_i) - Z(x_i + h)\}^2$$

(eq-4)

O semivariograma pode ser representado como um gráfico da função de dependência espacial em relação à distância (Figura 2). Observa-se que o valor da semivariância aumenta com o aumento da distância entre os pontos.



**Figura 2.** Representação gráfica do semivariograma. Os pontos azuis representam o semivariograma experimental e a linha tracejada o modelo teórico.

O semivariograma experimental é calculado a partir dos dados observados e a partir dele são associados diferentes modelos teóricos de semivariogramas. Esses modelos teóricos são determinados por tipo e pelos seus parâmetros numéricos associados. Entre os diferentes tipos podemos ter modelos linear, esférico, gaussiano, exponencial e potencia. Os parâmetros numéricos utilizados são chamados de efeito pepita, patamar e alcance. No exemplo da figura acima (Figura 2), o semivariograma experimental é representado por pontos azuis, enquanto que o modelo teórico (esférico) é a linha tracejada em preto. O efeito pepita revela a descontinuidade do semivariogramas ( $\gamma(h)$ ) para distâncias menores do que a menor distância entre as amostras. Esse efeito pode ser causado por variabilidade em pequena escala, por erro de amostragem, erro de determinação ou por algum limite do equipamento (VIEIRA, 2000). Ou seja, o efeito pepita é a porção do semivariograma devido ao acaso. À medida que a distância entre os pontos ( $h$ ) aumenta,  $\gamma(h)$  também irá aumentar até chegar num valor no qual se estabiliza, caso haja dependência espacial entre as variáveis. Esse valor numérico é denominado patamar. Se denomina semivariância estrutural a parte do semivariograma entre o efeito pepita e o patamar, que corresponde a parte da semivariância que é efetivamente modelada. O valor numérico do patamar é aproximadamente igual à variância dos dados,  $\text{Var}(z)$ , se houver dependência espacial. Através do patamar também pode-se estimar a variabilidade do fenômeno.

O alcance refere-se à distância no qual  $\gamma(h)$  atinge o patamar e representa a distância máxima de dependência espacial (VIEIRA, 2000). Dentro dos limites do alcance considera-se que as variáveis são correlacionadas umas às outras. Isto permite que se façam interpolações para espaçamentos menores do que os amostrados. Devido a isto o alcance é utilizado para avaliar o grau de continuidade espacial dos dados. Para distâncias maiores que o alcance, as medições apresentam distribuição espacial aleatória e podem ser consideradas independentes entre si e a estatística clássica pode ser aplicada sem restrições (VIEIRA, 2000).

Conhecido o semivariograma da variável, e havendo dependência espacial entre as amostras, é possível fazer interpolação nos intervalos entre as amostragens e criar mapas de isolinhas ou de superfície através da krigagem (Vieira, 2000).

A quantidade de amostras é um dos fatores mais determinantes na determinação da acurácia desse método (SCHLOEDER et al., 2001), especialmente em áreas com paisagens mais complexas. Extensivas amostragens nem sempre são viáveis devido a restrições quando ao custo e tempo da obtenção das mesmas (ZHU & LIN, 2010). Webster & Oliver (1992) reportam que é necessário mais que 100 pontos para a estimação de um variograma regular, enquanto que Kerry & Oliver (2007) mostraram que 50 pontos, em distância apropriada uns dos outros, foi adequado. E ainda, não apenas a quantidade de pontos, mas também o espaçamento entre amostras e o grau de dependência espacial são relevantes na acurácia da interpolação. Uma interpolação mais precisa, com uma menor quantidade de amostras, consiste ainda numa importante questão (ZHU & LIN, 2010). Assim como o papel de co-variáveis ambientais exaustivamente levantadas, que podem contribuir para uma maior acurácia de interpolação com reduzida quantidade de pontos amostrados.

### 2.2.3. Cokrigagem

Existem diferentes variações da krigagem que podem ser utilizadas dentro do mapeamento digital para melhorar a acurácia na predição de variáveis (LI & HEAP, 2008). Entre essas variações a utilização de outra variável, ou mais de uma, são utilizadas nos métodos conhecidos como co krigagem e krigagem com regressão.

A cokrigagem é uma variação multivariada da krigagem, na qual uma informação auxiliar é incorporada na estimação dos valores em locais não medidos através da correlação espacial da variável primária  $Z_1$  com o atributo secundário  $Z_v$  (GOOVAERTS, 1998). Semivariogramas experimentais e cruzados, tanto da variável principal quanto das auxiliares, são calculados e modelados na forma de modelos lineares de coregionalização (GOULARD & VOLTZ, 1992). O estimador da cokrigagem ( $Z_{cok}^*(u)$ ) para um ponto não medido é:

$$Z_{cok}^*(u) = \sum_{a=1}^{n(u)} \lambda_a(u)[Z_1(u)] + \sum_{v=1}^{n(v)} \lambda_v(u)[Z_v(u)]$$

Onde:

$$\sum_{a=1}^{n(u)} \lambda_a = 1$$

E:

$$\sum_{v=1}^{n(v)} \lambda_v = 0$$

A cokrigagem não necessita que a informação secundária esteja disponível em todos os pontos onde  $u$  necessita ser estimado. A influência da informação secundária na estimativa

do valor Z depende dos seguintes aspectos: da correlação entre a variável principal e as secundárias; da continuidade espacial dos atributos; e da densidade amostral e configuração espacial das variáveis principal e secundárias (SIMBAHAN et al, 2006). Quando todas as variáveis estão medidas nas mesmas coordenadas, a cokrigagem é chamada de isotópica. Quando as variáveis não estão medidas nas mesmas coordenadas, a cokrigagem é chamada de heterotópica. Caso algumas posições não sejam compartilhadas pelas variáveis, por exemplo quando a variável alvo tem menor número de amostras, o método é denominado cokrigagem heterotópica parcial. Utilizando a cokrigagem heterotópica parcial é aconselhável, sempre que possível, que a inferência do variograma cruzado ou do modelo de função da covariância seja feita com base na parcela do conjunto de dados que for isotópica (WACKERNAGEL, 2003).

A partir dos variogramas das variáveis principais e secundária, é construído um variograma cruzado que também será usado para a interpolação dos dados. O método apresenta, como restrição, a necessidade que esses dois variogramas e o variograma cruzado sejam do mesmo modelo teórico e tenham o mesmo alcance (NIELSEN & WENDROTH, 2003).

A co krigagem pode usar diferentes variáveis auxiliares, uma de cada vez, para predição. Mas ela foi desenvolvida para situações onde essas variáveis não estão em grandes quantidades (KNOTTERS et al., 1995). E apresenta a desvantagem de requerer muito tempo para a modelagem simultânea dos variogramas e variogramas cruzados de diferentes combinações com todas as covariáveis disponíveis (HENGL et al., 2007).

#### **2.2.4. Krigagem com regressão**

O termo krigagem com regressão (regression kriging – RK) foi cunhado por ODEH et al. (1994) como sendo um método que combina regressão linear múltipla de variáveis auxiliares com krigagem. Na krigagem com regressão (RK) primeiramente se desenvolve um modelo de regressão com variáveis auxiliares (Regressão linear simples ou múltipla). Assim, gera-se um mapa do atributo do solo (também denominado de tendência). Em seguida, extrai-se os resíduos (diferenças entre os valores medidos e estimados) e utiliza-se os resíduos da regressão para fazer uma krigagem. O próximo passo é somar o mapa de regressão com o de resíduos interpolados utilizando krigagem (HENGL et al., 2007).

De acordo com Li & Heap (2008) diferentes variações de krigagem com regressão existem na literatura e foram propostas e testadas ao longo do tempo (ODEH et al., 1994; ODEH et al., 1995; HENGL et al., 2007; MINASNY & MCBRATNEY, 2007; LI & HEAP, 2008). Esses autores sistematizaram essas variações classificando-as em seis diferentes tipos, que são apresentados a seguir:

a) Krigagem com regressão tipo A, chamada de krigagem combinada com regressão linear. É feita uma regressão linear com os dados e, em seguida, é realizada a krigagem desses valores;

b) Krigagem com regressão tipo B, chamada krigagem com valores estimados. (tambem denominada em ingles como “kriging with a guess field”). Essa forma envolve uma primeira etapa com regressão, então são krigados os valores preditos e os resíduos separadamente e, no final, ambos os resultados são somados para obter a predição final;

c) Krigagem com regressão tipo C, pode ser chamada de “regressão com resíduos de krigagem “simples”, krigagem sem tendências, krigagem e cokrigagem residual modificada ou krigagem residual. É semelhante ao tipo B, mas somente se faz a krigagem dos resíduos e então soma-se esses valores aos preditos pela regressão para se obter a predição final;

d) Krigagem com regressão tipo D. Similar ao tipo C, mas usa o método dos quadrados mínimos generalizados. Faz-se o cálculo da regressão por esse método, os resíduos da regressão são interpolados usando krigagem simples, e a estimativa final é a soma do

resultado dessa krigagem com os valores obtidos na regressão;

e) Krigagem com regressão tipo E é uma mistura do tipo C e A. O ponto chave dessa variação é que tanto os pontos medidos quanto os preditos são tratados como equivalentes no processo de interpolação;

f) Krigagem com regressão tipo F é similar ao tipo C, porém usa um modelo aditivo generalizado (MAG) para calcular a regressão. O MAG é ajustado aos dados, então os resíduos são interpolados usando krigagem ordinária, e as estimativas finais são a soma dos resíduos krigados e a superfície gerada pelo MAG.

As técnicas híbridas com krigagem incluem correlações com atributos ambientais, e têm se tornado populares recentemente, devido à possibilidade de se utilizar informações secundárias, o que tem frequentemente resultado em previsões locais mais acuradas (GOOVAERTS, 1999; MCBRATNEY et al., 2000).

Essa técnica permite o uso de métodos complexos de regressão, incluindo modelos lineares generalizados. Apesar de apresentar certas vantagens, esse método ainda não é amplamente usado em geoestatística. Estudos têm mostrado que krigagem com regressão apresentou melhor performance que a krigagem ordinária, co-krigagem e as regressões lineares múltiplas (ODEH et al., 1995; HERBST et al., 2006; SUMFLETH & DUTTMANN, 2008; ZHU & LIN, 2010; SUN et al., 2012), onde os resultados são mais detalhados e as previsões têm apresentado elevada acurácia (HENGL et al., 2004). No Brasil ainda é relativamente menor os estudos que empregam técnicas geostatísticas que permitam o uso de co-variáveis ambientais. Principalmente quanto ao emprego de potenciais co-variáveis ambientais adequadas às condições específicas de solo, paisagem e manejo.

De acordo com Bishop & McBratney (2001), mesmo quando uma variável auxiliar fracamente correlacionada está disponível, os métodos híbridos ainda apresentam melhor performance que os métodos geostatísticos mais genéricos, como a krigagem ordinária. Zhu & Lin (2010) estudaram os efeitos do tamanho da amostra, estrutura espacial e co-variáveis ambientais para alguns atributos do solo, e concluíram que, quando a estrutura espacial não pode ser bem capturada pelas observações baseadas em pontos ou quando uma relação mais forte existe entre o atributo do solo e a variável auxiliar ( $R^2 > 0,6$ ), a krigagem com regressão apresentou maior acurácia na interpolação de atributos do solo. Caso contrário, a krigagem ordinária apresenta melhores resultados (KRAVCHENKO & ROBERTSON, 2007; ZHU & LIN, 2010). A regressão linear é a mais comumente usada para modelar atributos de solo. No entanto, a relação entre solos e co-variáveis ambientais não é necessariamente linear, e pode ser desconhecida ou apresentar ruídos (HENGL et al., 2004).

Na krigagem com regressão, a regressão pode ser feita de diferentes maneiras, tanto através de modelos lineares generalizados quanto através de modelos não lineares, os quais podem possibilitar a inclusão de mais variáveis auxiliares. Uma estrutura genérica para krigagem com regressão é proposto por Hengl et al. (2004).

Algumas limitações são apontadas que podem fazer com que os resultados da krigagem com regressão sejam de baixa qualidade (HENGL et al., 2007): qualidade dos dados; baixo número de amostras; estimativa confiável da estrutura covariância/correlação; extrapolação fora da área amostrada; preditores com relação desbalanceada com a variável a ser interpolada. Apesar de apresentarem potencial para melhorar as previsões, CK e RK são métodos mais complexos. Portanto deve-se sempre levar em consideração esse esforço extra na hora de avaliar a aplicação de cada método (CEDDIA et al., 2017).

### 3. MATERIAL E MÉTODOS

#### 3.1. Área de Estudo e os Dados Físico-Hídricos Existentes

A área de estudo consiste num espaço destinado à pesquisa e prática da agroecologia. Está localizada em área pertencente à EMBRAPA Agrobiologia no município de Seropédica, Rio de Janeiro, entre as coordenadas 43°40'00" e 43°40'10" de longitude oeste e 22°44'00" e 22°45'00" de latitude sul. A área com 59 ha é designada como Sistema Integrado de Pesquisa em Produção Agroecológica (SIPA), sendo conduzida em parceria entre a EMBRAPA, PESAGRO-RJ e a Universidade Federal Rural do Rio de Janeiro desde 1993. A área possui classificação climática do tipo Aw segundo Köppen, com clima quente e úmido, com temperaturas elevadas no verão e quentes no inverno. A vegetação nativa é composta por floresta tropical subcaducifólia nas partes mais elevadas do relevo, e na parte baixa, campo higrófilo de várzea.

Do total da área, 6,24 ha foram amostrados e os solos foram mapeados, sendo 3,26 ha manejados sob sistema de rotação em policultivos distribuídos em glebas, e 2,98 ha destinados a pastagem. Nessa área foi realizada uma campanha de campo para coletar dados de diferentes atributos do solo.

Essa campanha gerou um banco de dados com diferentes atributos do solo, e esses dois bancos de dados foram usados nesse projeto. O banco de dados é resultado de uma coleta em grade irregular realizada em 2001 e é composto de 122 pontos (Figura 3). São dados dos seguintes atributos: textura, densidade do solo (DS); Capacidade de Campo (CC); Água facilmente disponível (AFD); Ponto de Murcha Permanente (PMP); Porosidade de Aeração (PA); Porosidade Total (PT); Carbono Orgânico Total (C-org). Esses dados estão representados em três profundidades: 0-10 cm; 10-20 cm; 20-30 cm (Vilella, 2007).

Os dados CC, 80 KPa e PMP foram obtidos através de medição direta usando o extrator de Richard (EMBRAPA, 1997), com tensões de 10 KPa, 80 KPa e 1500 KPa respectivamente. O valor do atributo AFD foi obtido através da diferença entre CC e 80 KPa. A ADT foi obtida através da diferença entre CC e PMP.



**Figura 3.** Grade de amostragem de textura, Densidade do Solo (Ds); Capacidade de Campo (CC); Água facilmente disponível (AFD); Ponto de Murcha permanente (PMP); Porosidade de Aeração (PA); Porosidade Total (PT) (Vilella, 2007).

### **3.2. Atributos do Relevo**

Para a geração do Modelo Digital de Elevação (MDE) foram usados 122 pontos planialtimétricos coletados com um GPS com correção diferencial (DGPS – Trimble – GeoExplorer modelo 3) com precisão submétrica. Esses pontos foram coletados durante a campanha de campo em 2001. O MDE foi obtido a partir da ferramenta TopoToRaster do ArcGIS 10.0 da ESRI, o qual possui resolução espacial de 5X5 metros.

Esse MDE foi utilizado para a geração das covariáveis do terreno. Essas covariáveis foram obtidas através dos softwares SAGA GIS 2.0.6 (BÖHNER et al., 2010) e ArcGIS spatial analyst (ESRI, 2010). As covariáveis obtidas foram: Declividade (SLOPE); Perfil de Curvatura (PROFILE); Plano de Curvatura (PLAN); Altitude acima da rede de canais (AACN); Índice topográfico de umidade (WI); índice de multiresolução de fundo de vales (MRVBF); Posição relativa da declividade (RSP); Índice de multiresolução de planicidade de topo (MRRTF); nível de base da rede de canais (CNBL); distância vertical da rede de canais (VDN) (MOORE et al., 1993; GALLANT & DOWLONG, 2003; WANG et al., 2010).

Essas covariáveis do terreno foram utilizados como variáveis preditoras auxiliares dos atributos físico-hídricos de interesse.

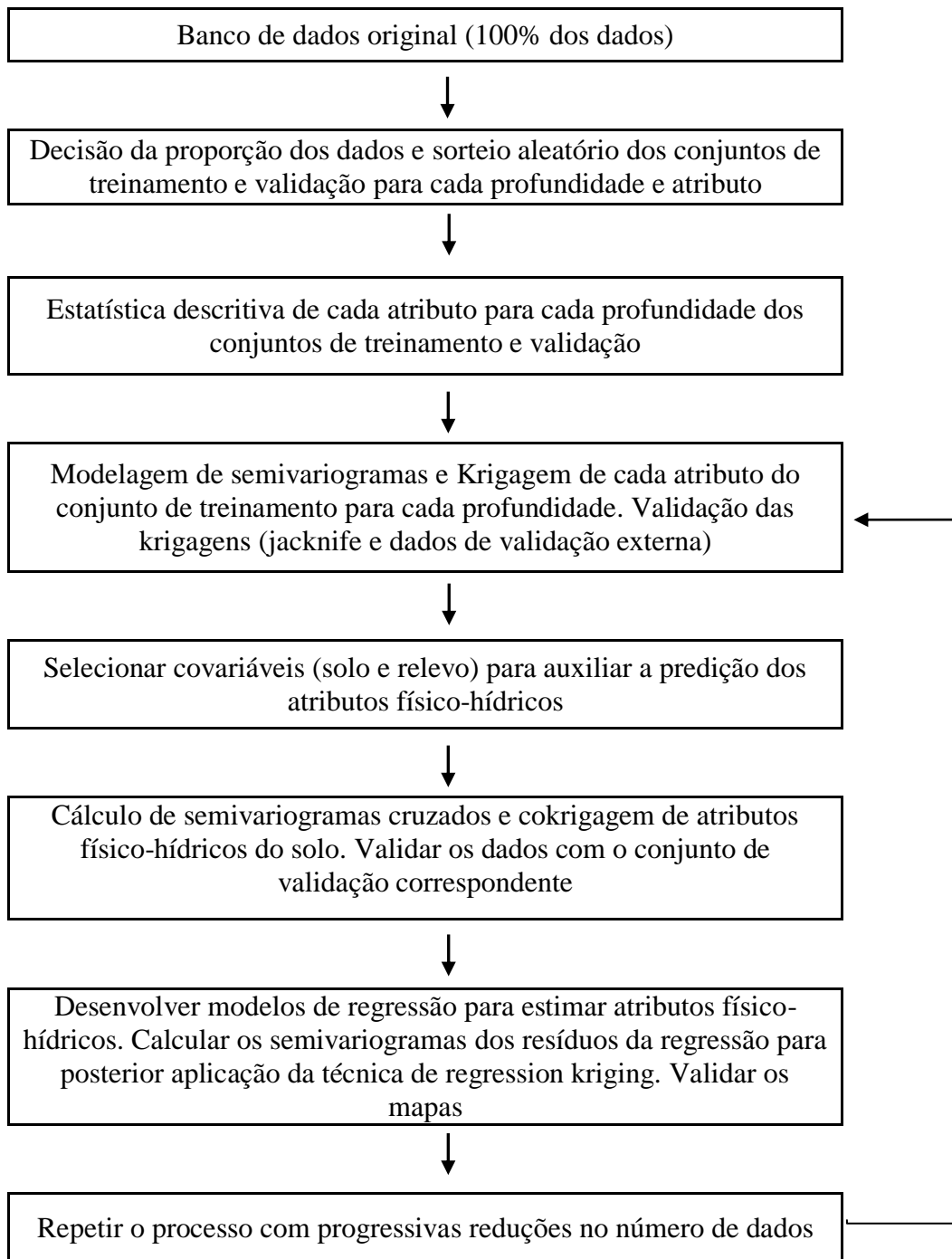
### **3.3. Definição da Proporção de Dados a Serem Utilizados para Desenvolvimento e Validação dos Algoritmos**

A partir da base de dados inicial foi necessário a determinação de um conjunto de dados para servir de treinamento e para validação. Nessa primeira etapa precisou ser determinada qual a proporção de dados seria usada para utilizar os diferentes métodos de interpolação. Foram criadas três diferentes proporções para ver se haveria diferença de comportamento estatístico dos dados com diferentes proporções.

Cada conjunto possui uma proporção diferente de dados para ser usado como treinamento e validação. As combinações de proporção de dados para treinamento e validação usados foram 80/20, 75/25 e 70/30, respectivamente. Esses valores correspondem à percentagem de dados utilizados, a partir do dataset original, para realização, respectivamente, da krigagem dos atributos (treinamento) e para a validação posterior dos mapas gerados (validação).

Para cada atributo foi realizada uma análise estatística descritiva para avaliar se a distribuição de frequência dos dados foi alterada significativamente com a diminuição do número de amostras. Dessa forma, cada atributo foi estudado separadamente para cada profundidade nas três proporções de forma a determinar a proporção definitiva a ser usada para as interpolações.

Na Figura 4 está representado, de forma simplificada, a sequência geral do processo que foi seguido de acordo com a metodologia proposta.



**Figura 4.** Fluxograma resumido da seqüência das análises realizadas.

### 3.4. Krigagem, Cokrigagem e Krigagem com Regressão dos Atributos Físicos-Hídricos

As variáveis escolhidas como principais para serem preditas nas interpolações são: capacidade de campo (CC), Ponto de Murcha Permanente (PMP), água retida a 80 KPa (80 KPa), Água disponível total (ADT) e água facilmente disponível (AFD).

Após a determinação das proporções dos dados para os conjuntos de treinamento e validação foi feita a análise estatística descritiva dos dados e a obtenção dos semivariogramas dos atributos de solo usado o software Geostat.



Em seguida foram estudadas as correlações entre os atributos principais a serem medidos com os demais atributos e covariáveis do terreno, de forma a elencar quais serão usados para a co krigagem e para a krigagem com regressão.

A etapa seguinte consistiu na interpolação das variáveis principais através de Krigagem Ordinária (OK), cokrigagem (CK) e krigagem com regressão (RK). Para interpolação com OK e CK foi usado o software Geostat. A modelagem das equações usadas na RK foi desenvolvida através de pacotes do programa R. Posteriormente os softwares Geostat e ArcGis foram utilizados para realizar a krigagem dos resíduos da regressão e a soma desses resíduos com o resultado da regressão, gerando o resultado final. Os resultados da CK e RK foram comparados aos gerados por OK para avaliar o ganho ou não de performance obtido. Para extração dos valores para validação dos resultados, foi usado o software ArcGis.

Os resultados foram comparados através dos índices de erro médio de predição (Mean Error - ME), raiz quadrada do erro médio (Root Mean Square Error - RMSE) e Erro Médio Absoluto (Mean Absolut Error - MAE) (ISAACS & SRIVASTAVA, 1989; BISHOP et al., 2015; HENGL et al., 2004; ODEH et al., 1995). Esses índices foram comparados de forma a permitir avaliar como cada escolha de metodologia para gerar os mapas altera na qualidade dos mesmo em termos de acurácia e precisão. O erro médio de predição (Mean Error - ME) foi calculado comparando-se os valores estimados ( $\hat{z}(s_j)$ ) com os dados de validação  $z^*(s_j)$ :

$$ME = \frac{1}{l} \sum_{j=1}^l [\hat{z}(s_j) - z^*(s_j)]$$

O Erro Médio Absoluto (MAE):

$$MAE = \frac{1}{l} \sum_{j=1}^l |\hat{z}(s_j) - z^*(s_j)|$$

E a raiz quadrada do erro médio (RMSE):

$$RMSE = \sqrt{\frac{1}{l} \sum_{j=1}^l [\hat{z}(s_j) - z^*(s_j)]^2}$$

Onde  $l$  é o número de pontos destinados a validação. O ME mede a tendência ou viés da predição, e o RMSE e o MAE medem a acurácia (ISAACS & SRIVASTAVA, 1989).

Para avaliar as diferentes performances dos interpoladores avaliou-se a melhoria relativa (MR). Esse índice foi usado para comparar os resultados da co-krigagem e krigagem com regressão em relação à krigagem ordinária. Esse índice se baseia na utilização dos diferentes valores obtidos com RMSE a partir de diferentes métodos e interpolação (SIMBAHAN et al., 2006; SILVA Jr, 2013):

$$MR = \frac{RMSE_{krigagem\ ordinária} - RMSE_{melhor\ método}}{RMSE_{krigagem\ ordinária}} * 100$$

Esse índice permite mostrar o aumento, ou não, da acurácia dos mapas comparando as diferentes técnicas de interpolação. O resultado é expresso em porcentagem, sendo que valores positivos indicam aumento da acurácia, pela diminuição do valor do RMSE. Quando o resultado é negativo, significa que o novo método gerou um mapa com pior acurácia. Dessa

forma pode-se avaliar se o aumento do esforço na geração desses mapas, de acordo com cada método de interpolação, é justificado. Caso haja aumento de acurácia, isso significará que há um ganho real da qualidade dos mapas gerados com a utilização dos diferentes interpoladores.

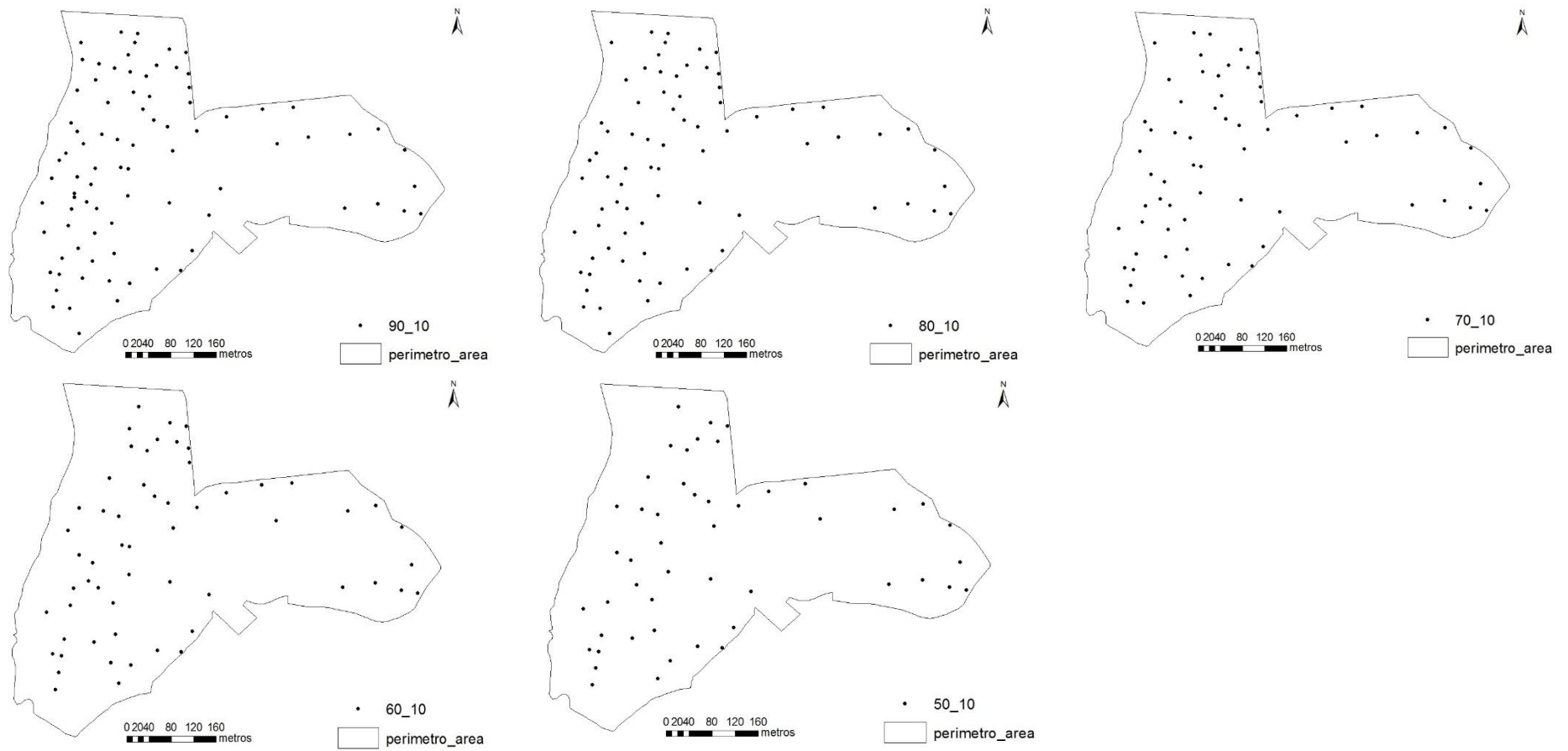
### **3.5. Redução no Número de Amostras Necessárias para Interpolação dos Atributos Físico-Hídricos**

Para testar como a acurácia dos resultados se comporta ao reduzir o número de amostras, foram sorteadas parcelas cada vez menores do banco de dados original de treinamento. Nesses novos conjuntos de dados foram aplicados os métodos de interpolação de Krigagem (OK) e Krigagem com regressão (RK). Conjuntos de dados contendo 90%, 80%, 70%, 60%, 50% relativos ao conjunto de dados usado para treinamento alimentaram os modelos de interpoladores e os resultados foram comparados quanto à acurácia.

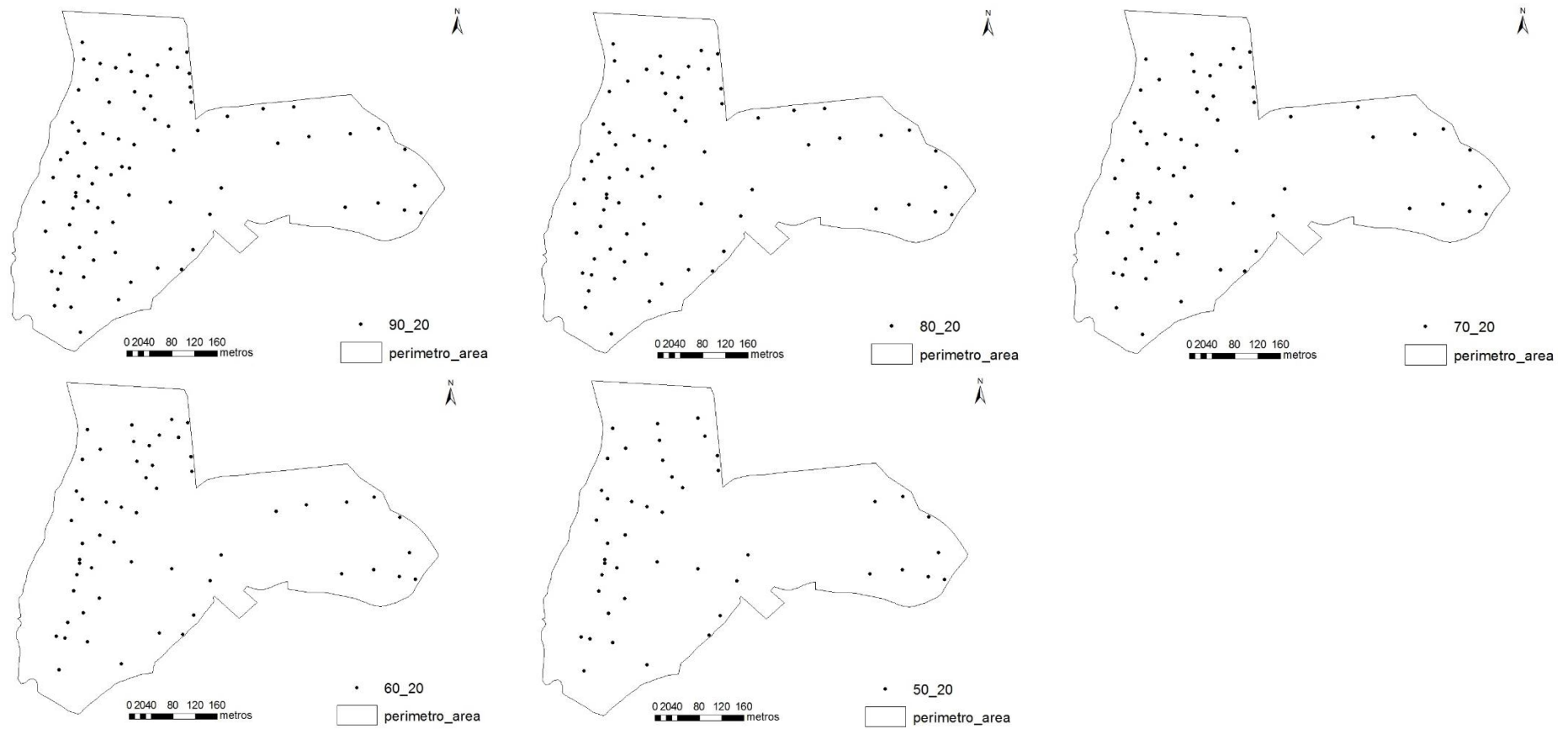
Para essa etapa foram escolhidas as variáveis 80 KPa, na profundidade 0-10 cm, e AFD na profundidade 10-20 cm. Essas variáveis foram as que obtiveram ganho em precisão na comparação dos resultados de RK com OK.

A avaliação da acurácia dos resultados foi efetuada comparando os resultados de RMSE obtidos ao se utilizar o conjunto total de treinamento para interpolação com OK e RK. O índice MR foi usado para as comparações, tanto de redução de número de dados, quanto para comparar os resultados de OK e RK em cada proporção. A retirada dos pontos de treinamento, ao reduzir a porcentagem dos dados para a interpolação, foi feita de forma completamente aleatória.

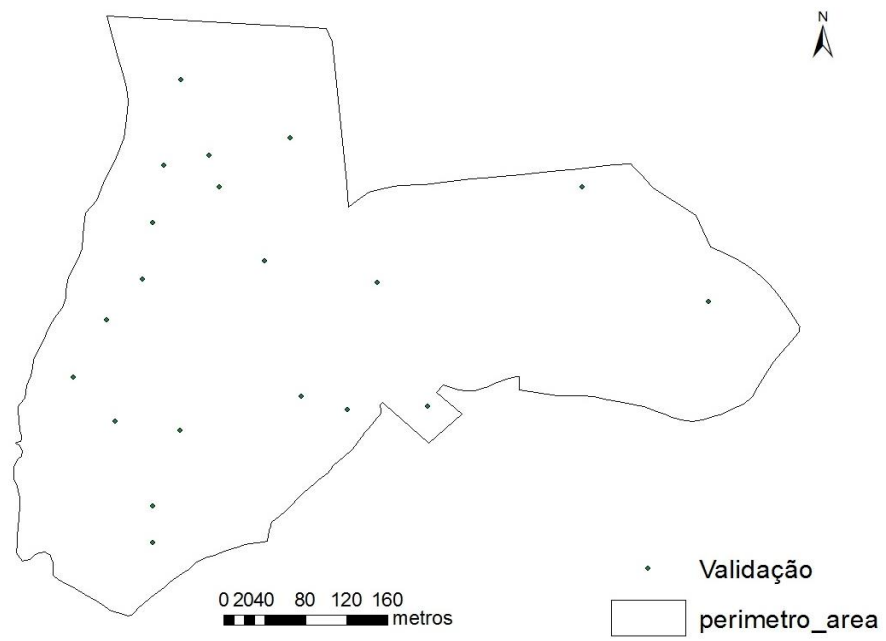
Nas Figuras 5 e 6 estão representadas as distribuições espaciais dos pontos nas diferentes proporções utilizadas para as duas variáveis. Na Figura 5 está representada 80 KPa 0-10 cm e na Figura 6 AFD 10-20 cm. Na Figura 7 está a distribuição espacial dos pontos usados para validação dos resultados.



**Figura 5.** Distribuição espacial dos pontos de 80 KPa 0-10 cm de acordo com a porcentagem dos dados utilizados para interpolação.



**Figura 6.** Distribuição espacial dos pontos de AFD 10-20 cm de acordo com a porcentagem dos dados utilizados para interpolação.



**Figura 7.** Distribuição espacial dos pontos de validação para os interpoladores.

## 4. RESULTADOS E DISCUSSÃO

### 4.1. Covariáveis do Relevo

O Modelo Digital de Elevação (DEM) e as covariáveis derivadas a partir dele são apresentadas nas Figuras 8 e 9. A estatística descritiva das mesmas está na Tabela 1.

A altitude da área de estudo, varia de 26 a 38 metros acima do nível do mar (Figura 8a). A Declividade (SLOPE) consiste no gradiente de elevação do terreno (Figura 8b). Quanto maior o valor dessa variável, espera-se que maior seja a susceptibilidade à erosão e o arraste de material. Os valores gerados variam de 0,01 a 0,06. O Plano de curvatura (PLAN) (Figura 8c) é calculado transversal ao sentido do declive e mede a convergência ou divergência do terreno, bem como a concentração de água na paisagem (MOORE et al., 1993). O conceito do plano de curvatura é relacionado com a taxa de variação da declividade ao longo das curvas de nível e mede a tendência da água convergir, ou divergir, quando atravessa o terreno. Os valores negativos são referentes aos locais onde a água irá convergir ao escoar no terreno, enquanto que os valores positivos refere-se aos locais onde a água irá divergir. Valores próximos a zero correspondem à partes mais planas do terreno (SILVA, 2013). Na área os valores variam de -0,02 a 0,03.

O perfil de curvatura (PROFILE) (Figura 8d) é a direção do máximo declive e é portanto importante para fluxos de água e transporte de sedimentos. Valores positivos correspondem a terrenos convexos e valores negativos a terrenos côncavos. Terrenos mais planos tem valores perto de zero. Os valores no mapa gerado (Figura 8d) seguem a mesma tendência dos encontrados para o Plano de Curvatura, tendo a mesma amplitude de valores, de -0,02 a 0,03.

O Índice topográfico de umidade (WI - *topographic wetness index*) é usado para caracterizar zonas de saturação de água superficial e o conteúdo de água nos solos. Indica as áreas onde a água deve se acumular no terreno (Figura 8e). É calculado como  $\ln(a/\tan\beta)$ , onde  $a$  – área de drenagem por unidade de comprimento da curva de nível cruzada pelo escoamento, e  $\beta$  a tangente do declive local. Os maiores valores desse índice indicam os locais onde há uma maior probabilidade de condições de saturação de água no solo. Geralmente são encontradas nas partes mais baixas e convergentes, podendo ser associados a solos de baixa condutividade hidráulica ou áreas de pequeno declive (BEVEN & WOOD, 1983). Esse índice pode ser usado para identificar os padrões de fluxo de água na paisagem (SUMFLETH & DUTTMAN, 2008). Apresenta variação de 1,97 a 8,2.

A covariável Posição Relativa na declividade (Relative Slope Position – RSP) (Figura 8f) varia de 0 a 1, fazendo referência à altura relativa do ponto no modelo digital do terreno. Esse índice foi utilizado para alimentar os modelos de RK, servindo como uma das covariáveis. Mas não entrou em nenhum dos modelos finais de regressão.

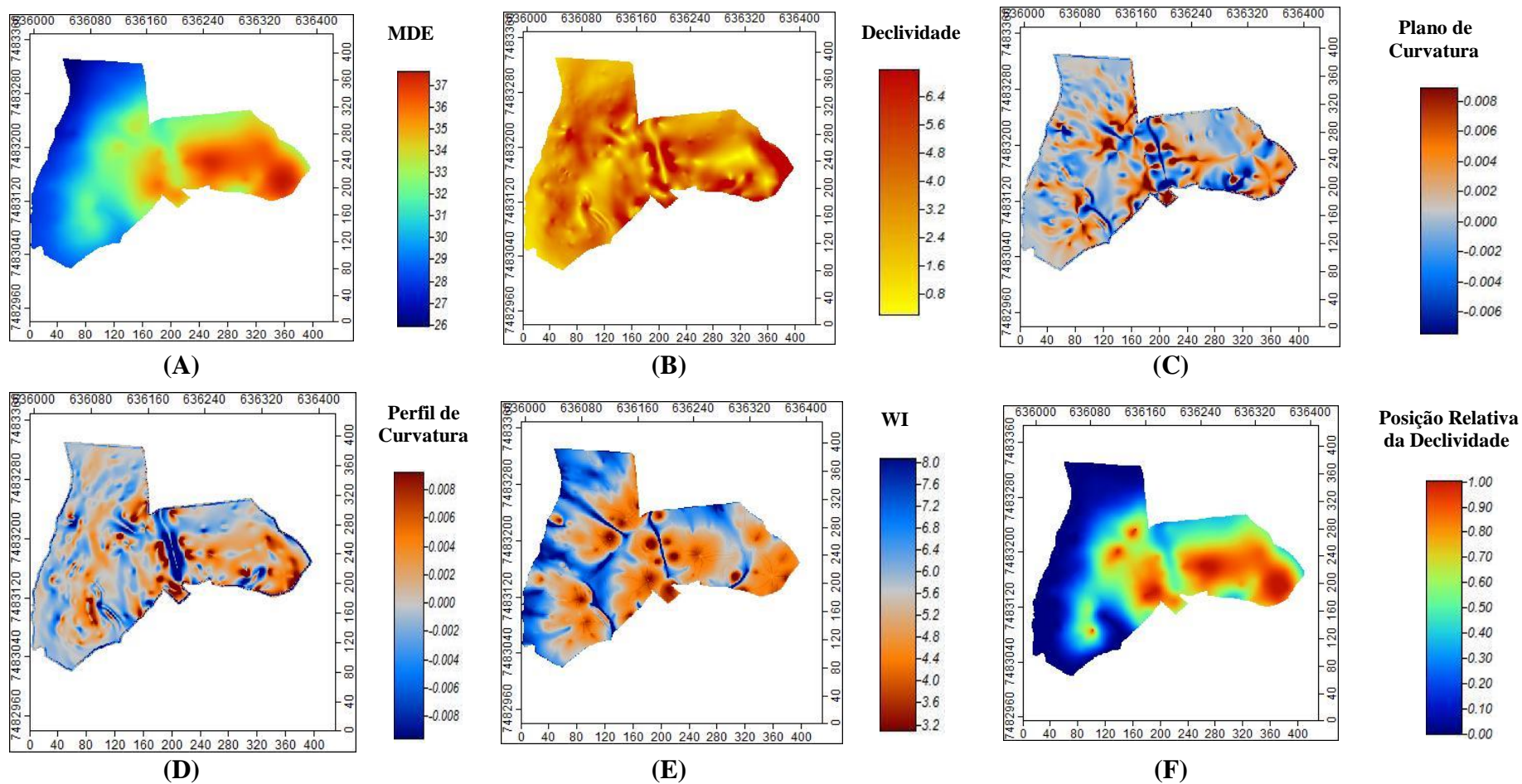
O Índice multiresolução do nivelamento do fundo de vales (MRVBF - *Multiresolution Index of Valley Bottom Flatness*) (Figura 9a) define e distingue os fundos de vale de encostas em diferentes escalas e combina diferentes paisagens em um único índice (WANG et al., 2010). Este pode ser útil nas correlações realizadas, uma vez que este índice é usado para mapear áreas de deposição na paisagem (GALLANT & DOWLONG, 2003), como as de ocorrência de Planossolos ou Gleissolos. De acordo com Mckenzie & Gallant (2007), os valores inferiores a 0,5 são considerados superfícies de erosão e valores superiores a 0,5 de deposição. Na área de estudo os valores variam de 0,02 a 2,94. O outro índice, MRRTF (Índice multiresolução de nivelamento de topo de paisagens, *Multiresolution ridge top flatness index*, Figura 9b) é similar ao anterior, porém indica as partes altas do terreno (GALLANT & DOWLING, 2003). Ambos indicam no mapa as áreas onde se espera maior

deposição de material, porém a interpretação dos valores é inversa. Apresentou valores de 0,02 a 3,95.

A Altitude Acima da Rede de Canais (AACN - *altitude above the channel network*)(Figura 9c) descreve a distância vertical entre cada ponto do grid e a rede de drenagem. Os maiores valores correspondem no mapa aos pontos mais altos, em relação à rede de drenagem e os menores os mais próximos. Isso significa que os menores valores correspondem à locais de deposição de material e acúmulo de água, variando de 0 a 3,09.

O índice CNBL (Channel Network Base Level – Altura base da rede de canais) é calculado de acordo com a distância vertical do nível do canal de base da rede local com o ponto no terreno (PRATES, 2010). Pode ser usado para previsão de profundidade de solos hidromórficos, e associado ao AACN pode indicar a energia potencial da água, evidenciando áreas com maior potencial de erosão. Apresenta variação de 26,2 a 30,3 no terreno.

Essas covariáveis numéricas do terreno foram utilizadas como variáveis secundárias tanto para variografia na CK quanto para gerar as equações de regressão múltipla usadas para RK.



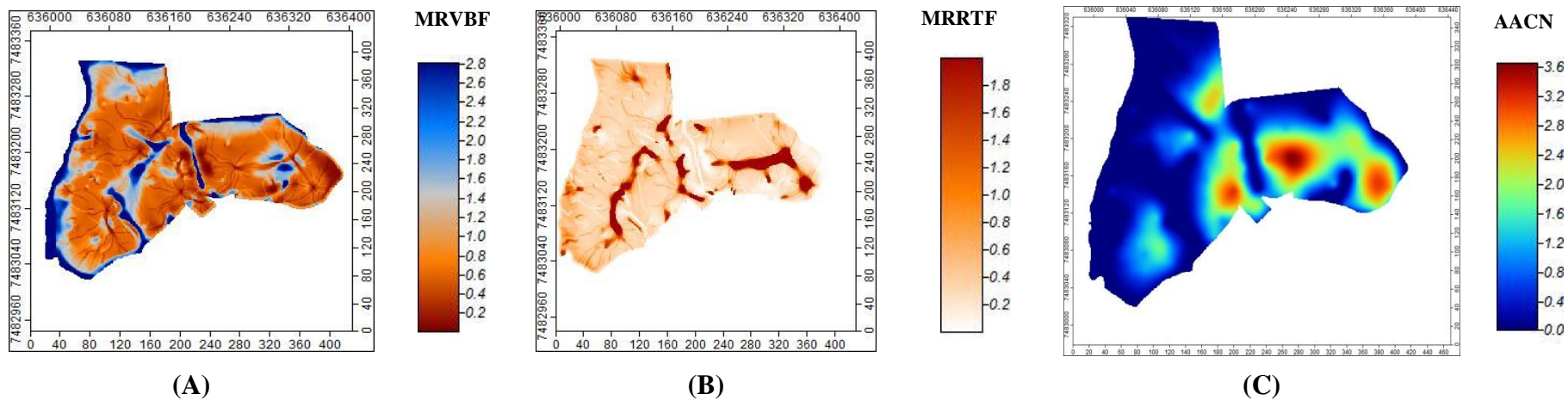
**Figura 8.** Variáveis geomorfológicas derivadas a partir do modelo digital de elevação. MDE – Modelo Digital de Elevação; WI – Índice Topográfico de umidade.



**Tabela 1.** Estatística descritiva das covariáveis numéricas do terreno.

Nome	Número	Média	Variância	Desvio Padrão	CV	Mínimo	Máximo	Assimetria	Curtose
AACN	87	0,64	0,55	0,74	116,90	0,00	3,09	1,41	1,70
CNBL	87	29,03	0,93	0,96	3,32	26,22	30,30	-0,71	0,15
MDE	87	30,96	6,44	2,54	8,19	26,23	37,58	0,52	-0,13
MRRTF	87	0,57	0,56	0,75	132,00	0,02	3,95	3,09	9,77
MRVBF	87	0,79	0,34	0,58	73,30	0,02	2,94	1,37	2,32
PLAN	87	0,00	0,00	0,01	533,80	-0,02	0,03	0,67	2,19
PROFILE	87	0,00	0,00	0,01	1421,00	-0,02	0,03	0,69	2,94
RSP	87	0,34	0,08	0,28	82,54	0,00	1,00	0,58	-0,71
SLOPE	87	0,07	0,00	0,02	35,39	0,01	0,12	0,07	-0,14
VDN	87	3,84	4,46	2,11	54,98	0,00	9,03	0,41	-0,43
WI	87	5,39	2,15	1,47	27,18	1,97	8,02	-0,21	-0,63

Num – Número de pontos; CV – Coeficiente de Variação; AACN - Altitude Acima da Rede de Canais; CNBL - Altura base da rede de canais; MDE - Modelo Digital de Elevação; MRRTF - Índice multiresolução de nivelamento de topo de paisagens; MRVBF - Índice multiresolução do nivelamento do fundo de vales; PLAN - Plano de curvatura; PROFILE – Perfil de curvatura; RSP - ; SLOPE - Declividade; VDN - ; WI - Índice Topográfico de umidade.



**Figura 9.** Variáveis geomorfológicas derivadas a partir do modelo digital de elevação. MRVBF - Índice multiresolução do nivelamento do fundo de vales ; MRRTF - Índice multiresolução de nivelamento de topo de paisagens; AACN - Altitude Acima da Rede de Canais.

## **4.2. Organização e Estatística Descritiva dos Conjuntos de Treinamento e Validação**

Com o objetivo de determinar a proporção do conjunto de treinamento/validação foi feita uma avaliação sobre o comportamento estatístico das variáveis principais. Essa avaliação teve como objetivo identificar se a mudança na proporção dos dados, de 80/20 até 70/30 por exemplo, interferia no comportamento estatístico, mantendo a normalidade ou não dos dados. A importância dessa etapa era para definir se a maior proporção de dados (80/20) poderia ser usada. A avaliação foi pelos valores de assimetria.

De acordo com os valores de assimetria, os atributos CC, 80 KPa e PMP, nas diferentes proporções, não apresenta alteração de comportamento (Tabelas 1 a 9, Anexo I). Todos apresentam distribuição não normal.

Para AFD, apresentam distribuição normal para a primeira e segunda profundidade e não normal para a terceira, também em todas as combinações (Tabelas 10 a 12, Anexo I). Para ADT (Tabelas 13 a 15, Anexo I) o comportamento estatístico é o mesmo para todas as combinações em todas as profundidades.

Como esses atributos são o objetivo do trabalho, essa avaliação foi feita para determinar se a alteração da proporção causaria alterações no comportamento estatístico das variáveis. Uma das decisões do trabalho foi trabalhar com a não transformação dos dados, e estudar a variabilidade espacial dos mesmos. Isso é parte da decisão do trabalho de interpolação (DEUSTCH & JOURNAL, 1998).

Os demais atributos foram avaliados através do estudo de variografia na etapa seguinte de OK. Suas estatísticas descritivas estão apresentadas nas tabelas 16 a 35 do Anexo I.

O importante, nessa etapa, foi avaliar se haveria mudanças de comportamento das variáveis principais. Considerando que não há diferenças de comportamento estatístico dos atributos nesses 3 conjuntos analisados, optou-se pela escolha da proporção 80/20 para treinamento/validação. Dessa forma podemos utilizar o maior número possível de dados como o conjunto de treinamento dos interpoladores, conjunto esse que será referenciado como 100% dos dados.

## **4.3. Avaliação Geoestatística dos Atributos Físico-Hídricos**

Devido ao comportamento estatístico dos dados, a proporção 80/20 foi escolhida para separar o conjunto original total de dados em conjuntos de treinamento e validação. Nessa etapa, optou-se por não transformar os dados. Usando 80% dos dados originais, como conjunto de treinamento, foram gerados semivariogramas para estudar se existe dependência espacial dos atributos físico hídricos de interesse e se é possível a interpolação dos mesmos. Os atributos de interesse são Capacidade de Campo (CC), água retida a 80 KPa (80 KPa), Ponto de Murcha Permanente (PMP), Água Facilmente Disponível (AFD) e Água Disponível Total (ADT). Os valores desses atributos demandam maior esforço e tempo para serem determinados. A determinação é feita utilizando o extrator de Richards (EMBRAPA, 1997), por isso podem ser considerados de mais difícil medição quando comparados com os demais atributos do banco de dados, como por exemplo textura.

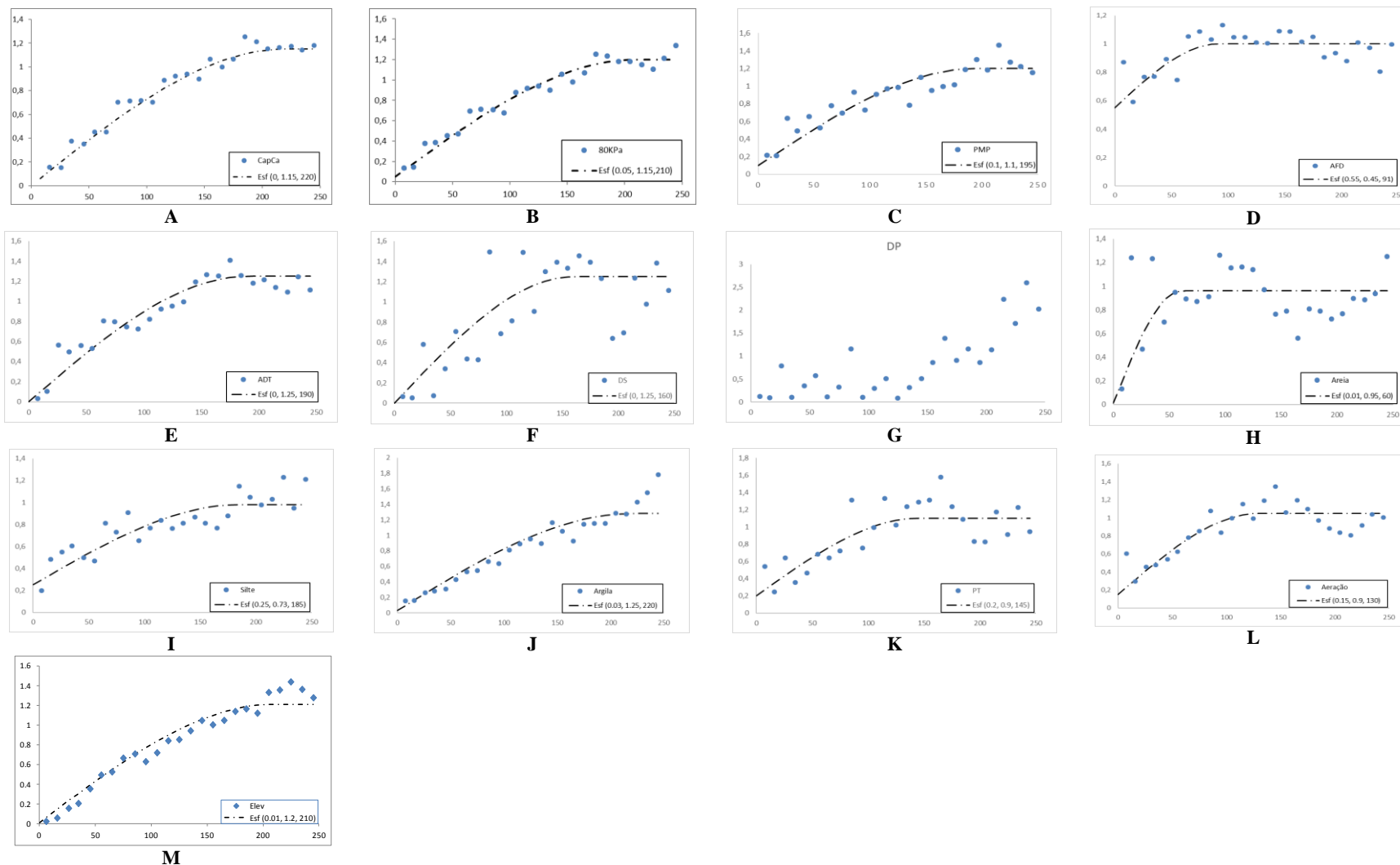
Também foram gerados semivariogramas para os outros atributos para dar subsídio na escolha das variáveis secundárias para as próximas etapas do trabalho. As variáveis secundárias são importantes para serem utilizadas tanto no processo de co krigagem (CK) quanto na krigagem com regressão (RK). Os atributos são: Densidade do Solo (DS), Densidade da Partícula (DP), textura do solo (Areia, Silte e Argila), Porosidade Total (PT) e Aeração do Solo (Porosidade de Aeração). A escolha das variáveis secundárias foi embasada na existência de dependência espacial dos atributos nas diferentes profundidades e na relativa praticidade da obtenção dos mesmos.

Os parâmetros dos modelos dos semivariogramas de todos os atributos, principais e secundários, estão agrupados de acordo com a profundidade do solo na Tabela 2. Os respectivos semivariogramas estão representados nas Figuras 10, 11 e 12. Todos os atributos principais apresentam dependência espacial nas três profundidades, o que permitiu a interpolação de todos por OK.

**Tabela 2.** Parâmetros dos modelos dos semivariogramas dos atributos físicos do solo nas três profundidades do solo.

Prof	Atributo	Modelo	C <sub>0</sub>	C <sub>1</sub>	A (m)	C <sub>0</sub> /(C <sub>0</sub> +C <sub>1</sub> )	
0-10 cm	CC	Esferico	0	1,15	220	0,00	
	80 KPa	Esferico	0,05	1,15	190	4,17	
	PMP	Esferico	0,1	1,1	195	8,33	
	AFD	Esferico	0,55	0,45	91	55,00	
	ADT	Esferico	0	1,25	190	0,00	
	DS	Esferico	0	1,25	160	0,00	
	DP	<b>Sem dependência espacial</b>					
	Areia	Esferico	0,01	0,95	60	1,04	
	Silte	Esferico	0,25	0,73	185	25,51	
	Argila	Esferico	0,03	1,25	220	2,34	
	PT	Esferico	0,2	0,9	145	2,17	
	Aeração	Esferico	0,15	0,9	130	14,29	
	10-20 cm	CC	Gaussiano	0,05	1,4	245	3,45
		80 KPa	Gaussiano	0,1	1,45	250	6,45
PMP		Gaussiano	0,15	1,35	200	10,00	
AFD		Esferico	0,3	0,75	83	28,57	
ADT		Esferico	0	1,2	270	0,00	
DS		Esferico	0,05	1,12	105	4,27	
DP		<b>Sem dependência espacial</b>					
Areia		Gaussiano	0,15	1,22	250	10,95	
Silte		Gaussiano	0,3	1,15	280	20,69	
Argila		Gaussiano	0,2	1,23	230	13,99	
PT		Esferico	0,5	0,6	110	45,45	
Aeração		Esferico	0,45	0,65	140	40,91	
20-30 cm		CC	Gaussiano	0,08	1,2	220	6,25
		80 KPa	Gaussiano	0,15	1,3	240	10,34
	PMP	Gaussiano	0,1	1,4	200	6,67	
	AFD	Esferico	0,3	0,95	185	24,00	
	ADT	Esferico	0	1,05	240	0,00	
	DS	Esferico	0,5	0,6	85	45,45	
	DP	<b>Sem dependência espacial</b>					
	Areia	Esferico	0,01	1,2	220	0,83	
	Silte	Esferico	0,35	0,7	75	33,33	
	Argila	Esferico	0	1,25	250	0,00	
	PT	Esferico	0,5	0,55	60	47,62	
	Aeração	Esferico	0,5	0,5	50	50,00	

C<sub>0</sub> – Efeito Pepita; C<sub>1</sub> – Contribuição; A – alcance (m); C<sub>0</sub>/(C<sub>0</sub>+C<sub>1</sub>) – Efeito Pepita Relativo.



**Figura 10.** Semivariogramas da elevação e dos atributos físicos na profundidade 0-10 cm. A- CC; B-80 KPa; C - PMP; D - AFD; E - ADT; F - DS; G - DP; H - Areia; I - Silte; J - Argila; K - PT; L - Aeração; M - Elevação. Entre parênteses estão os parâmetros do semivariogramas.

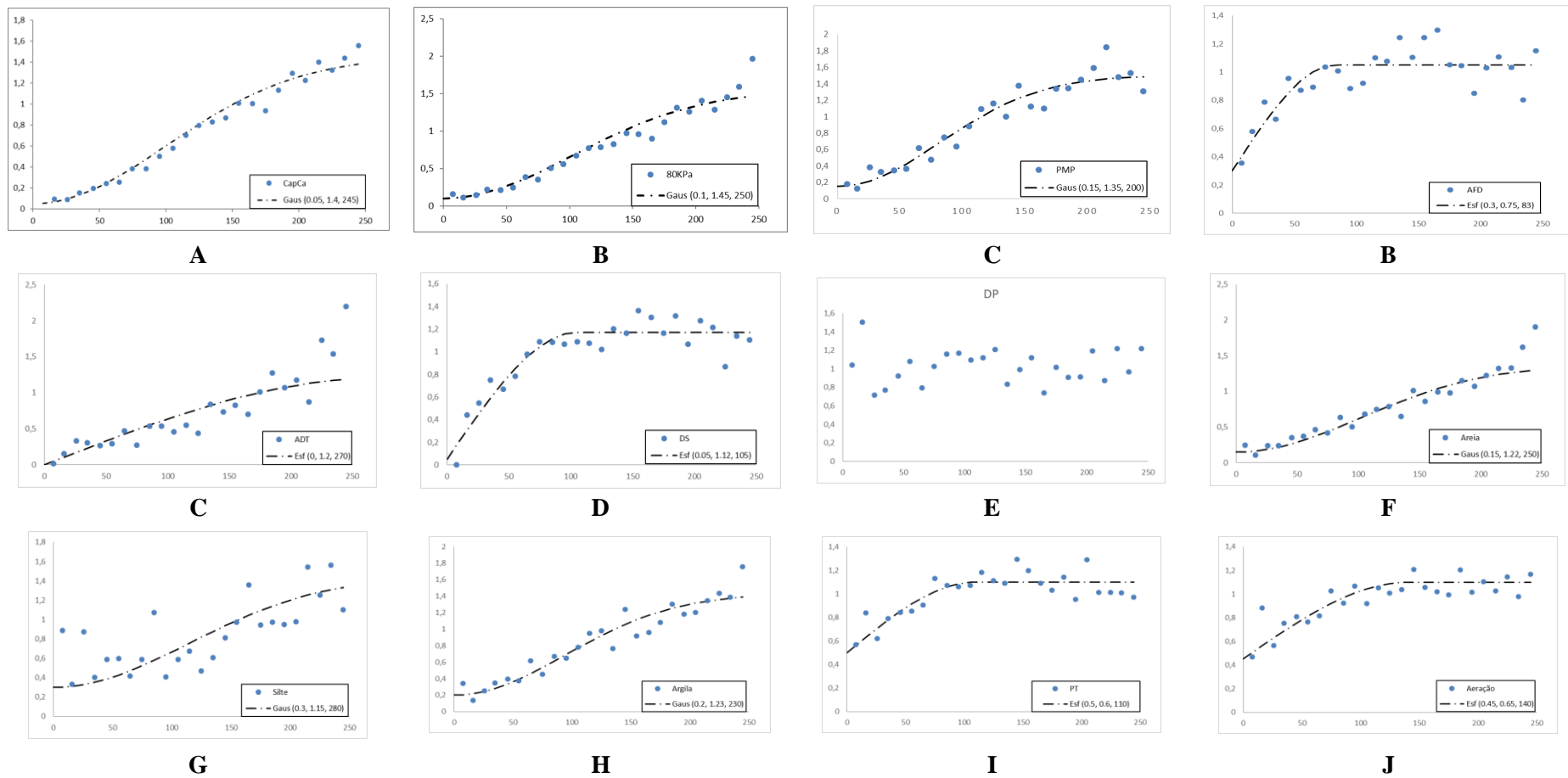
As variáveis principais apresentaram modelo esférico na profundidade 0-10 cm (tabela 10). Na profundidade de 10-20 cm CC, 80 KPa e PMP foram melhor ajustadas para modelo gaussiano, e AFD e ADT apresentam modelo esférico (Tabelas 11 e 12). O efeito pepita relativo (Tabela 2) é relacionado ao efeito de aleatoriedade das variáveis. Não necessariamente indica um efeito de correlação espacial, como proposto por alguns autores (CAMBARDELLA et al., 1994). Como o valor pepita pode ser causado por diferentes efeitos, incluindo erros e variações nas medições, esse valor pode ser melhor entendido como um efeito do acaso.

Como as variáveis principais apresentaram dependência espacial, todas puderam ser interpoladas através da OK (VIEIRA, 2000). O estudo da variografia para as demais variáveis foi para determinação da existência ou não de dependência espacial. O que permitiria que fossem aproveitadas nos outros métodos de interpolação. Somente a variável DP não apresentou dependência espacial.

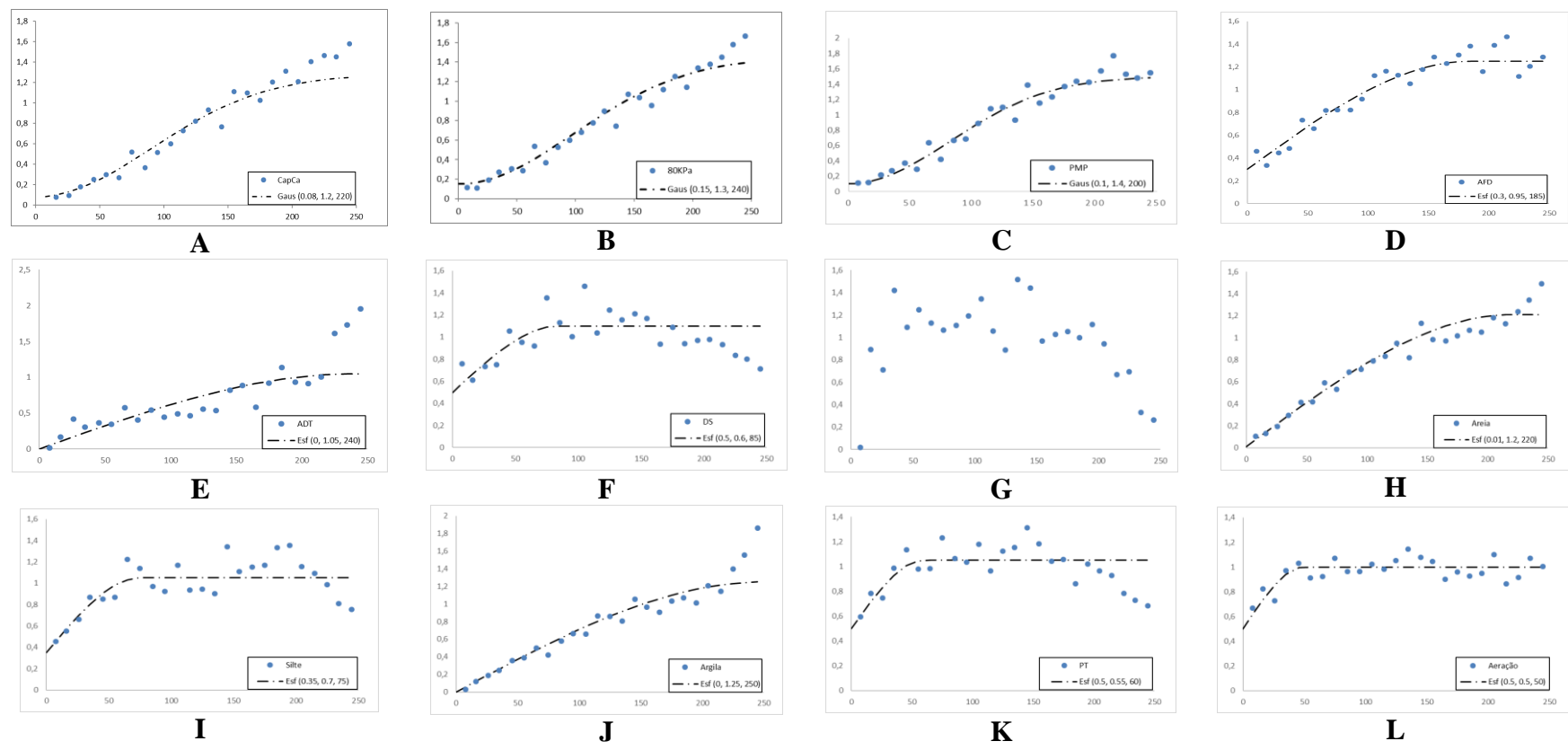
Quando é observado o efeito pepita puro, significa que não há correlação espacial do atributo, não sendo possível interpolar os valores através de Krigagem (VIEIRA, 2000). O modelo gaussiano indica uma função de distribuição de valores mais contínua, sem grandes alterações de valores em pequenas distâncias. Já o modelo esférico sugere uma distribuição mais errática, sendo esperado uma maior variação dos dados em menores distâncias (ISAKS & SRIVASTAVA, 1989; CEDDIA et al., 2009).

Isso significa que as variáveis principais possuem um comportamento espacial com maiores diferenças entre os valores medidos na primeira profundidade, e somente AFD e ADT mantêm esse comportamento com o aumento da profundidade. CC, 80 KPa e PMP, com aumento de profundidade, passam a apresentar uma variação de valores menos errática entre pontos.

Souza et al (2014), trabalhando com amostras de 0-20 cm de profundidade, usaram modelos exponencial e esférico para amostras com 208 e 105 pontos de areia e argila. Os conjuntos de dados utilizados não apresentavam distribuição normal, porém os semivariogramas formavam patamar. A diminuição do número de pontos amostrais nesse trabalho causou perda na acurácia dos mapas gerados através da avaliação do índice RMSE. Ceddia et al (2009), trabalhando com água retida a 10 kPa, 80 kPa e 1500 kPa, com amostras nas profundidades 0-10 cm, 10-20 cm e 20-30 cm, também encontraram dependência espacial para esses atributos. Os modelos encontrados foram diferentes, sendo todos gaussianos para a primeira e terceira profundidade e esférico para a segunda profundidade. Nos modelos deste trabalho, como a área é mais heterogênea, os modelos apresentaram diferenças, indicando um comportamento e dependência espacial diferentes.



**Figura 11.** Semivariogramas dos atributos físicos na profundidade 10-20 cm. A- CC; B-80 KPa; C - PMP; D - AFD; E - ADT; F - DS; G - DP; H - Areia; I - Silte; J - Argila; K - PT; L - Aeração. Entre parênteses estão os parâmetros do semivariogramas.



**Figura 12.** Semivariogramas dos atributos físicos na profundidade 20-30 cm. A- CC; B-80 KPa; C - PMP; D - AFD; E - ADT; F - DS; G - DP; H - Areia; I - Silte; J - Argila; K - PT; L - Aeração. Entre parênteses estão os parâmetros do semivariogramas.



Após o estudo dos variogramas foi realizada a interpolação dos mapas através da OK. O software utilizado nessa etapa foi o Geostat, utilizando os parâmetros dos modelos teóricos gerados (Tabela 2). Todas as variáveis principais foram interpoladas usando OK nas três profundidades.

Para avaliar a eficiência da OK, e permitir a comparação com os outros métodos de interpolação, desses dados, foi feita a validação dos resultados através dos índices propostos anteriormente. Para essa validação foi utilizado o conjunto correspondente a 20% dos dados originais, para cada profundidade, que havia sido previamente separado. Para a validação dos resultados foram utilizados os índices ME (Erro médio), MAE (erro médio absoluto) e RMSE (raiz quadrada do erro médio). Tanto o MAE quanto o RMSE são usados para medir a acurácia do mapa obtido, enquanto o ME é usado para avaliar a precisão do mapa (MILLER et al, 2016; BISHOP et al., 2015). Os valores calculados para atributo principal e em cada profundidade, estão na Tabela 3.

Devido à diferenças na amplitude numérica de cada atributo principal, não se deve fazer a comparação entre os diferentes atributos. Afinal maiores valores numéricos podem refletir em maiores valores de RMSE, por exemplo. Observamos, pela tabela 3, que a precisão aumenta para CC e 80 KPa na profundidade de 10-20 cm quando comparada com as demais. Mas esse comportamento não se repete nas demais variáveis. Para PMP ocorre o contrário, sendo a variável interpolada com mais acurácia na profundidade de 10-20 cm.

O resultado para AFD é uma perda de acurácia do resultado na medida em que aumenta a profundidade. Para ADT ocorre o contrário, sendo observado um aumento na precisão do método na medida em que aumenta a profundidade.

**Tabela 3.** Resultado da validação da Krigagem Ordinária dos atributos principais.

<b>Krigagem</b>	<b>ME</b>	<b>MAE</b>	<b>RMSE</b>
CC 0-10 cm	0,947	4,085	5,583
CC 10-20 cm	-0,43	2,339	2,797
CC 20-30 cm	-0,17	2,626	4,961
80 KPa 0-10 cm	0,754	3,596	5,206
80 KPa 10-20 cm	-0,336	1,999	2,644
80 KPa 20-30 cm	-0,318	2,458	4,664
PMP 0-10 cm	0,066	2,753	3,905
PMP 10-20 cm	0,744	3,211	5,075
PMP 20-30 cm	-0,797	1,782	3,599
AFD 0-10 cm	0,190	0,603	0,682
AFD 10-20 cm	-0,1	0,809	0,988
AFD 20-30 cm	-0,216	0,695	1,02
ADT 0-10 cm	0,845	5,172	8,024
ADT 10-20 cm	-1,123	3,117	5,519
ADT 20-30 cm	0,676	1,929	3,114

ME – Erro Médio; MAE – Erro Médio Absoluto; RMSE – Raiz Quadrada do Erro Médio.

Esses índices foram utilizados para comparar o ganho, ou perda, de acurácia e precisão tanto com a diminuição do número de amostras quanto em relação à utilização das técnicas híbridas de interpolação. A redução no número de amostras pode levar a uma perda na qualidade dos mapas gerados por interpolação, o que é evidenciado pelo aumento do valor numérico desses índices (SOUZA et al, 2014).

#### 4.4. Co Krigagem

A partir de cada atributo principal foram gerados semivariogramas cruzados com as variáveis secundárias. Em seguida foram ajustados os modelos das variáveis secundárias e dos variogramas cruzados de forma a atender à exigência do método. As variáveis secundárias escolhidas nessa etapa foram os dados de composição granulométrica, elevação e as covariáveis numéricas do terreno geradas a partir do MDE. Das covariáveis numéricas do terreno, MRRTF e MRVBF não apresentaram dependência espacial e não foram usadas para a modelagem. As demais foram testadas e utilizadas na análise.

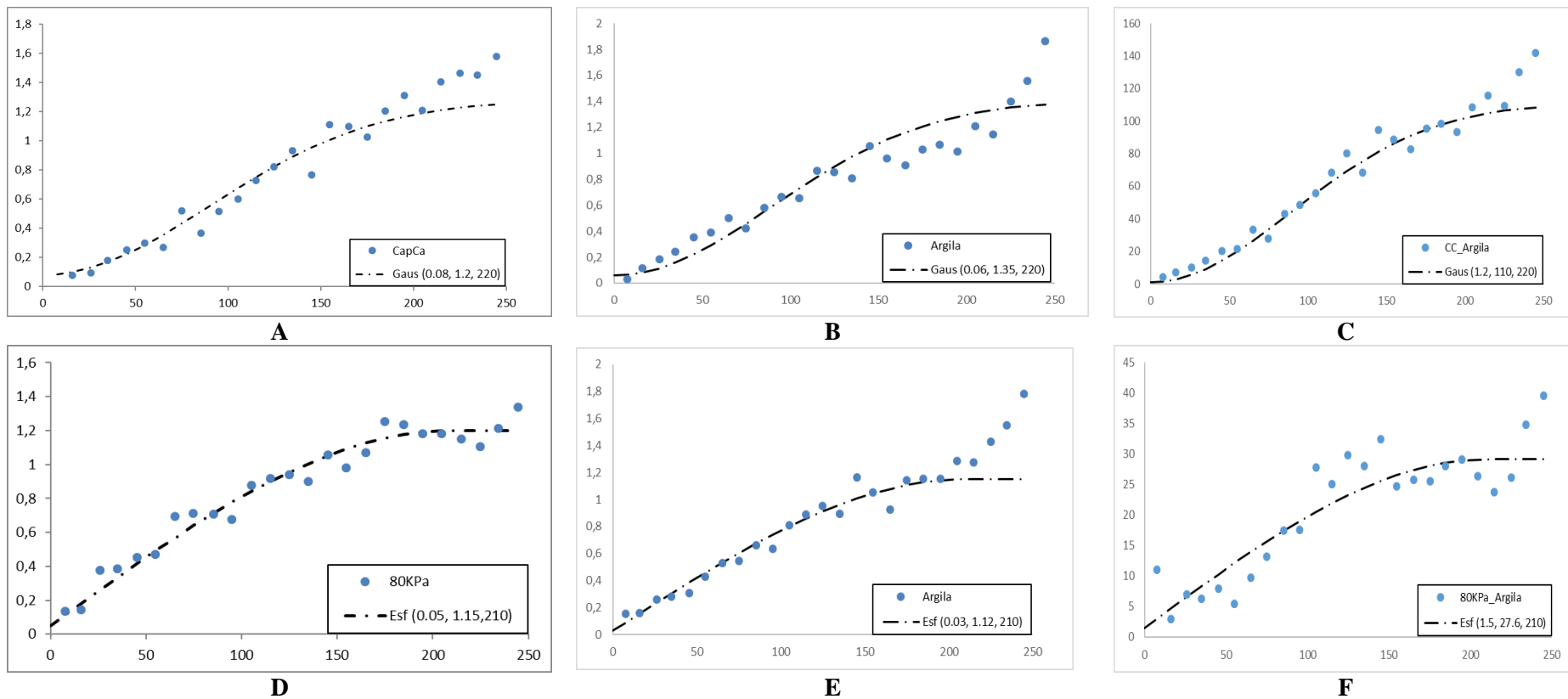
Os semivariogramas cruzados foram calculados usando rotinas do programa Geostat. De forma a atender as exigências do método (NIELSEN & WENDROTH, 2003), os modelos das variáveis secundárias e do variograma cruzado foram escolhidos de acordo com o modelo ajustado da variável principal. A escolha das variáveis secundárias é relacionada com a facilidade de obtenção desses dados, quando comparado à variável principal (MILLER et al., 2016). Na cokrigagem isotópica, usada nesse estudo, tanto a variável principal quanto a secundária precisam estar representadas no mesmo ponto (WACKERNAGEL, 2003).

Uma observação importante sobre esse método é o fato de que nem sempre é possível realizar a cokrigagem. O método é trabalhoso pois, para atender as exigências, todos os semivariogramas precisam apresentar o mesmo modelo e mesmo alcance que a variável que se deseja interpolar. Isso se aplica à variável principal, que se deseja interpolar, à variável secundária e ao semivariograma cruzado das duas (NIELSEN & WENDROTH, 2003; WACKERNAGEL, 2003). Essas exigências do modelo não necessariamente permitem gerar semivariogramas viáveis para todas as combinações de variáveis. Em algumas situações não é possível realizar o cálculo da cokrigagem por não conseguir que a variável principal, a secundária e o variograma cruzado atendam essas condições, como observado nos resultados abaixo.

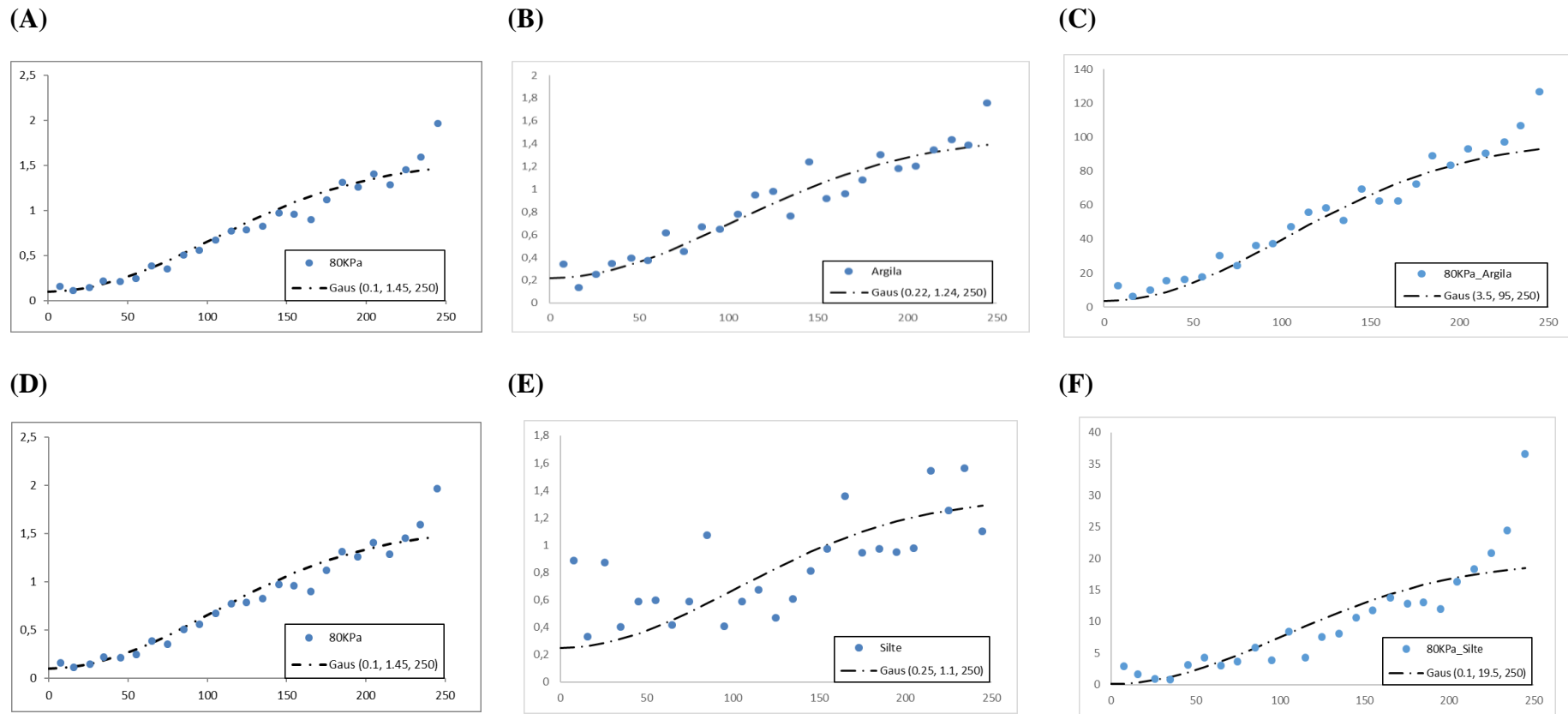
As possíveis combinações obtidas neste trabalho estão representadas nas Figuras 13 a 19, com os semivariogramas das variáveis principais, variáveis secundárias e semivariogramas cruzados.

Podem ocorrer mudanças na dependência espacial dos atributos do solo em diferentes profundidades (CEDDIA et al, 2009) e na correlação entre as diferentes variáveis. Por isso nem sempre um determinado atributo irá usar, necessariamente, as mesmas variáveis secundárias para auxiliar na predição através da cokrigagem, em todas as profundidades.

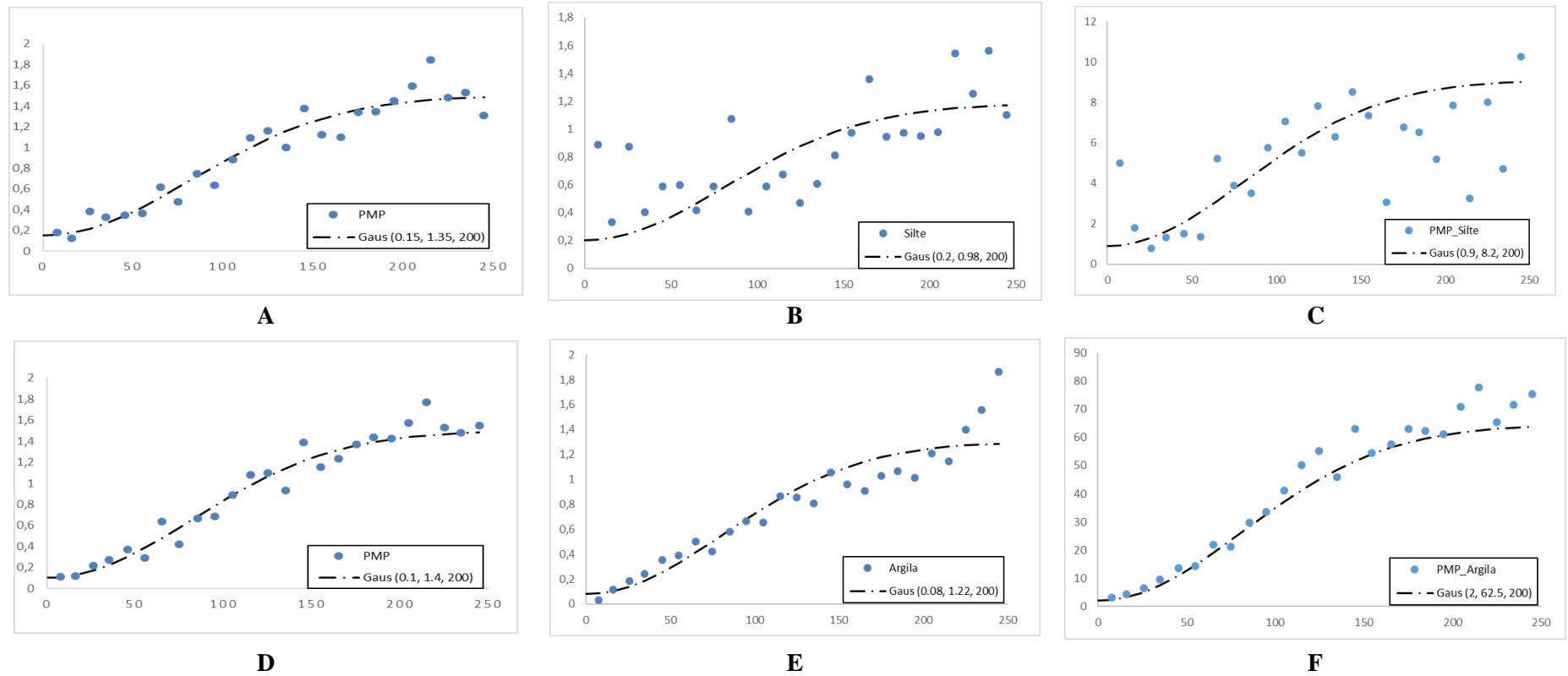
Após a seleção e escolha das variáveis secundárias, através dos semivariogramas cruzados, foi utilizado o software Geostat para interpolar os valores através da cokrigagem isotópica. Porém a interpolação através da CK apresentou problemas no momento da confecção dos mapas em algumas das combinações selecionadas. Dessa forma nem todos os resultados obtidos com os variogramas cruzados puderam ser validados para ser comparados com os resultados da OK. Somente puderam ser validados as combinações que estão apresentadas na Tabela 4. Os demais resultados gerados possuem muitas falhas, inviabilizando a etapa de validação. Por isso não foi possível proceder na comparação através da Melhora Relativa (MR) de todas as combinações que apresentaram semivariogramas cruzado.



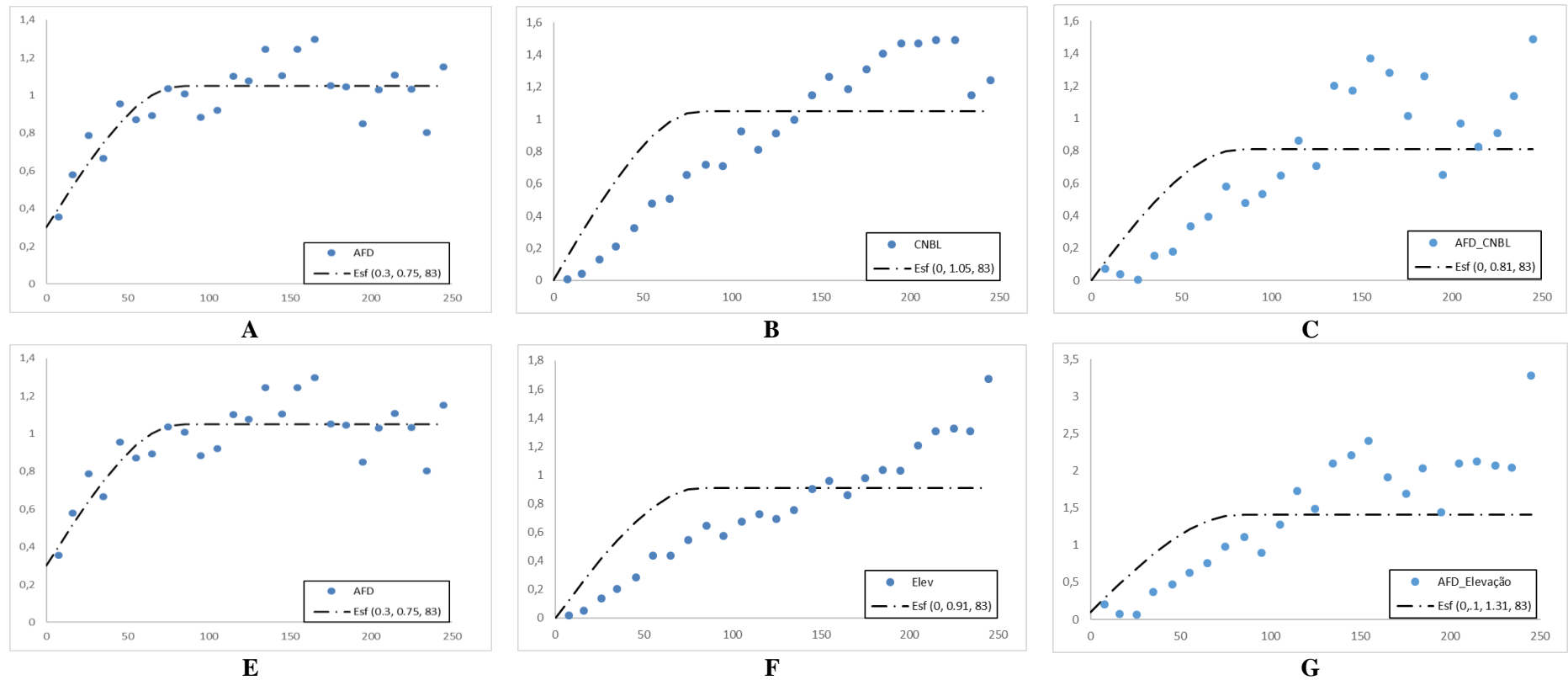
**Figura 13.** Semivariogramas de CC 20-30 cm (A) e atributo Argila (B) com semivariogramas cruzado (C); semivariogramas de 80 KPa 0-10 cm (D) e atributo argila (E) com semivariogramas cruzado (F). Entre parênteses estão os parâmetros do semivariogramas.



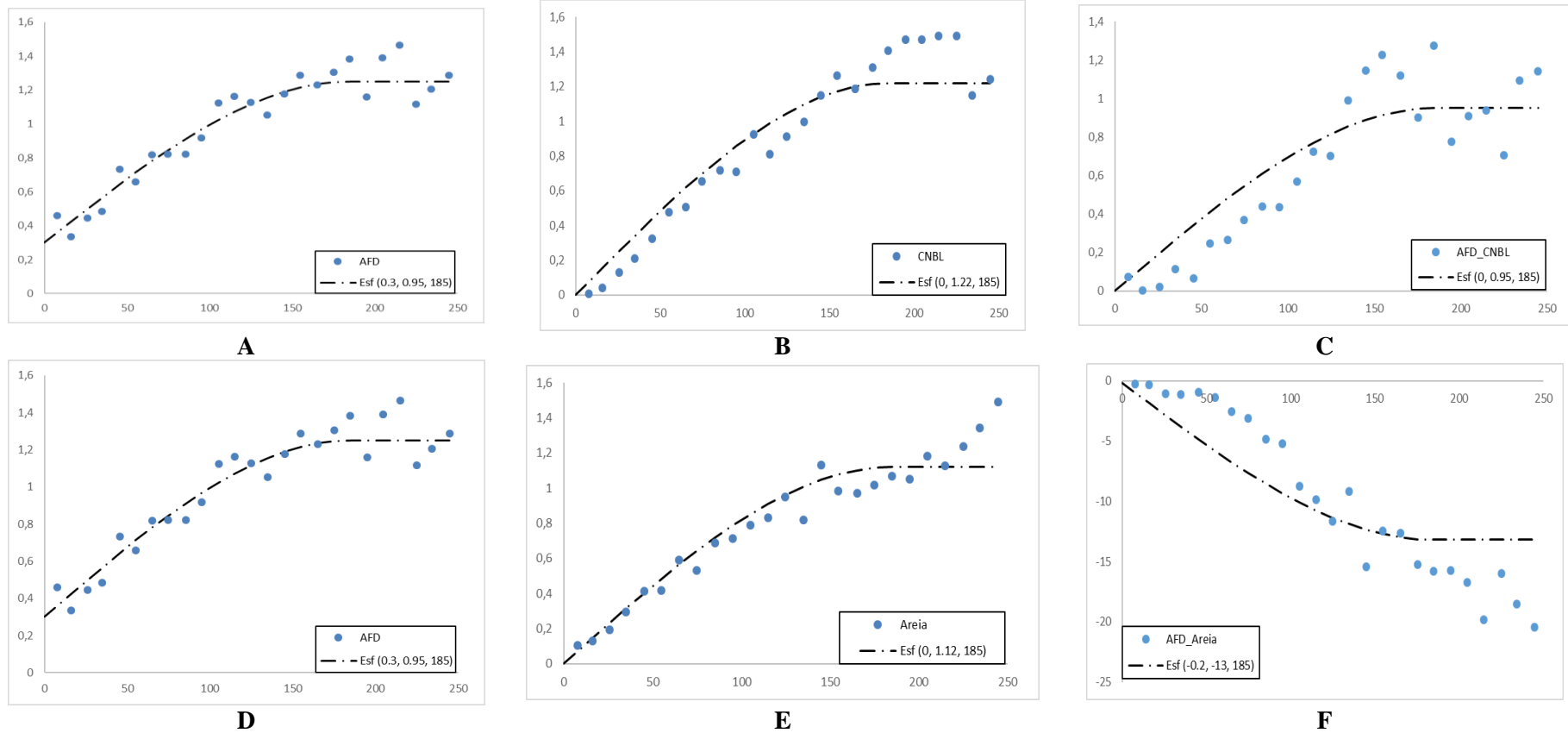
**Figura 14.** Semivariogramas de 80 KPa 10-20 cm (A e D), atributos Argila (B) e Silte (E) com seus respectivos semivariogramas cruzado (C e F). Entre parênteses estão os parâmetros do semivariogramas.



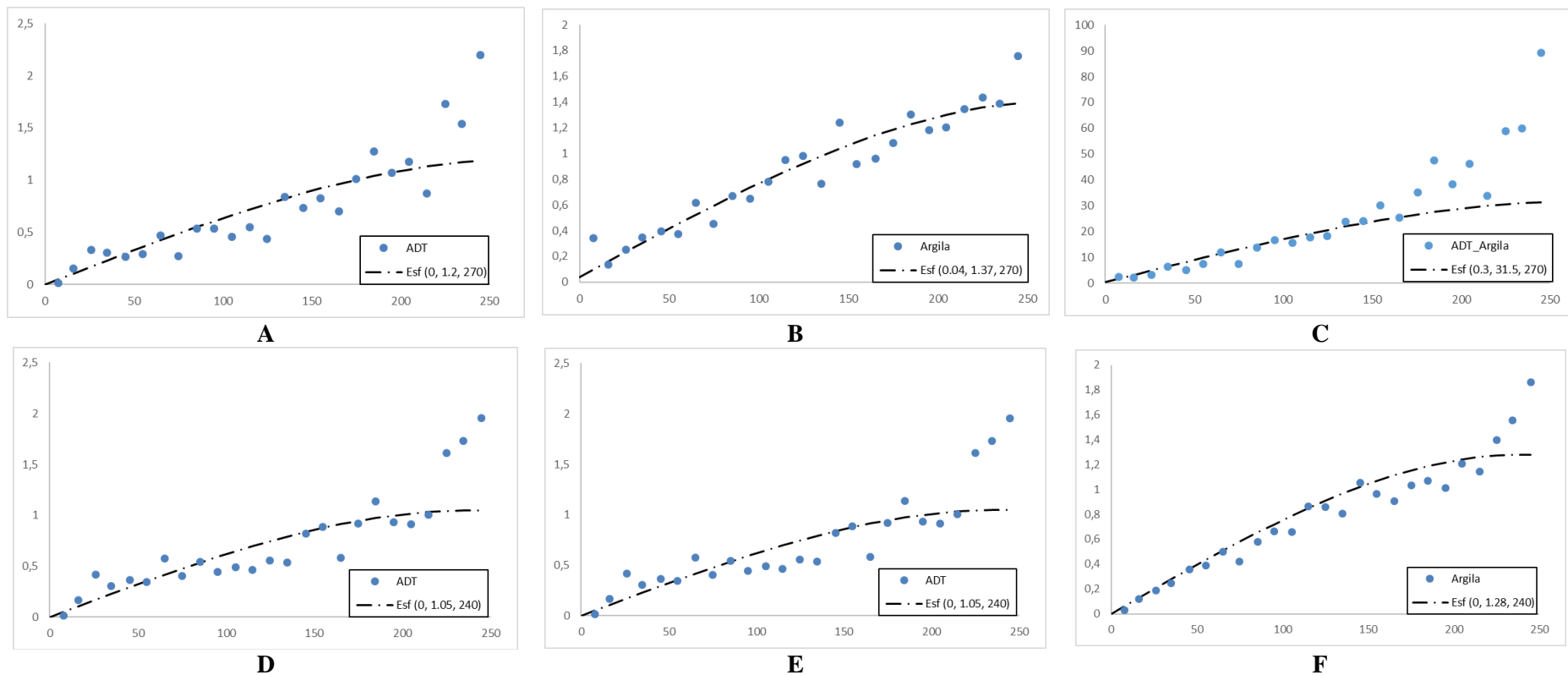
**Figura 15.** Semivariogramas de PMP 10-20 cm (A) e silte (B) e respectivo semivariogramas cruzado (C); semivariogramas de PMP 20-30 cm (D) e argila (E) e respectivo semivariogramas cruzado (F). Entre parênteses estão os parâmetros do semivariogramas.



**Figura 16.** Semivariogramas de AFD 10-20 cm (A e D) e atributos CNBL (B) e Elevação (E) com seus respectivos semivariogramas cruzados (C e F). Entre parênteses estão os parâmetros dos semivariogramas.

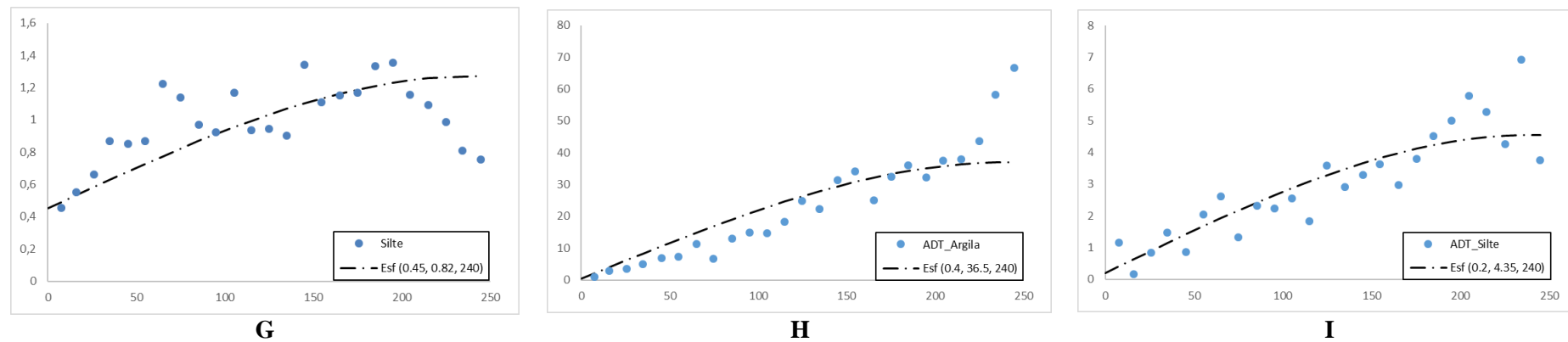


**Figura 17.** Semivariogramas de AFD 20-30 cm (A e D) e atributos CNBL (B) e areia (E) com seus respectivos semivariogramas cruzados (C e F). Entre parênteses estão os parâmetros dos semivariogramas.



**Figura 18.** Semivariogramas de ADT 10-20 cm (A) e atributo argila (B) com respectivo semivariograma cruzado (C). E semivariogramas de ADT 20-30 cm (D e E) e atributo argila (F). Entre parênteses estão os parâmetros dos semivariogramas.





**Figura 19.** Silte (G) com seus respectivos semivariogramas cruzados (H e I). Entre parênteses estão os parâmetros dos semivariogramas.

Até o presente momento não foi identificada a razão da falha da subrotina do software utilizado, que gerou áreas vazias dentro dos mapas impossibilitando a validação. Por isso ainda não foi possível comparar se houve ou não aumento na acurácia e precisão em todas as combinações testadas. Nem ainda pode-se apontar quais variáveis secundárias contribuem para aumentar ou diminuir a precisão dos interpoladores utilizados. É necessária essa comparação pois o método é trabalhoso e envolve muitas modelagens de semivariogramas quando comparado à OK e não necessariamente há ganho de acurácia nos mapas gerados (KNOTTERS et al, 1995; LIAO et al 2011).

**Tabela 4.** Validação dos resultados viáveis obtidos por Cokrigagem.

Co Krigagem	ME	MAE	RMSE
CC_argila 20-30 cm	-0,528	2,725	4,874
80 KPa_argila 0-10 cm	-0,772	5,119	8,605
80 KPa_argila 10-20 cm	-0,591	2,011	2,738
80 KPa_silte 10-20 cm	-0,646	3,295	4,806
PMP_silte 10-20 cm	0,807	3,458	5,318
PMP_argila 20-30 cm	-1,379	2,498	4,079
AFD_CNBL 10-20 cm	-0,054	0,77	0,959
AFD_Elevação 10-20 cm	0,019	0,901	1,079
AFD_CNBL 20-30 cm	-0,211	0,699	1,04
AFD_areia 20-30 cm	-0,276	0,737	1,019
AFD_silte 20-30 cm	-0,264	0,74	1,075
ADT_argila 10-20 cm	-1,168	3,231	5,627
ADT_argila 20-30 cm	-0,863	1,969	3,176
ADT_silte 20-30 cm	0,593	1,906	2,972

ME – Erro Médio; MAE – Erro Médio Absoluto; RMSE – Raiz Quadrada do Erro Médio.

Somente foi utilizado o índice MR nas combinações de CK que foram possíveis de serem validadas. De forma geral não se observou ganhos na precisão da interpolação usando a CK, quando comparada à OK (Tabela 5). Na maioria das combinações onde foi possível ser feita a validação, observa-se que a OK obteve um desempenho superior. Isso é esperado devido à pouca interferência da aleatoriedade nos resultados (CAMBARDELLA et al. 1994), pois nesses casos a utilização de variáveis auxiliares não significa ganho de acurácia e precisão em relação à OK (KRAVCHENKO & ROBERTSON, 2007).

Considerando o esforço que é exigido pelo método, ele não seria interessante para áreas similares, como método de predição para essas variáveis.

**Tabela 5.** Melhora relativa da Co Krigagem (CK) comparada à Krigagem Ordinária (OK).

Atributos	RMSE OK	RMSE CK	MR (%)
CC_argila 20-30 cm	4,961	4,874	1,753679
80 KPa_argila 0-10 cm	5,206	8,605	-65,29
80 KPa_argila 10-20 cm	2,644	2,738	-3,55522
80 KPa_silte 10-20 cm	2,644	4,806	-81,77
PMP_silte 10-20 cm	5,075	5,318	-4,78818
PMP_argila 20-30 cm	3,599	4,079	-13,337
AFD_CNBL 10-20 cm	0,988	0,959	2,935223
AFD_Elevação 10-20 cm	0,988	1,079	-9,21053
AFD_CNBL 20-30 cm	1,02	1,04	-1,96078
AFD_areia 20-30 cm	1,02	1,019	0,098039
AFD_silte 20-30 cm	1,02	1,075	-5,39216
ADT_argila 10-20 cm	5,519	5,627	-1,95688
ADT_argila 20-30 cm	3,114	3,176	-1,99101
ADT_silte 20-30 cm	3,114	2,972	4,560051

ME – Erro Médio; MAE – Erro Médio Absoluto; RMSE – Raiz Quadrada do Erro Médio.

#### 4.5. Krigagem com Regressão

Os modelos de regressão foram desenvolvidos através de pacotes do software R. O método escolhido para ajuste das covariáveis para a regressão múltipla foi o stepwise. Esse método seleciona os preditores ajustando o modelo através de substituições das variáveis, enquanto se ajusta a equação. O critério de escolha é feita através do índice AIC (Akaike's Information Criterion) (PAN, 2001), que o software usa como referência para avaliar os modelos gerados. Em seguida foi feita a correção dos preditores escolhidos retirando as variáveis que apresentam multicolinearidade (FOX & MONETTE, 1992). A retirada dessas variáveis é feita através do cálculo do valor do VIF (PEREIRA et al., 2014), também pelo R, usando o valor 5 como limite para o descarte do preditor. Por último foi realizada uma ANOVA para manter nos modelos somente as variáveis que apresentaram significância com a variável principal. O script básico utilizado no software R está apresentado no Anexo II.

Como variáveis preditoras foram usadas outras variáveis físicas do solo e as covariáveis numéricas do terreno geradas a partir do MDE da área. Não foram usadas como preditoras na modelagem as demais variáveis principais, por serem todas correlacionadas fisicamente, o que causaria um efeito de multicolinearidade. Por exemplo, na determinação do modelo para CC, as variáveis PMP, AFD, ADT e 80 KPa não foram utilizadas na modelagem. Também não foram utilizadas as variáveis Aeração, que é derivada do valor de CC, e a DP, que não apresenta relação direta com as propriedades físico hídricas do solo.

Os modelos de regressão são utilizados para gerar os mapas de regressão através do software ArcGIS. Usando a ferramenta Map Algebra, são gerados os mapas a partir dos quais serão retirados os resíduos para krigagem. A krigagem dos resíduos foi realizada através do software Geostat. O mapa final de cada variável principal é obtido usando novamente o ArcGis, somando o mapa krigado dos resíduos com o mapa obtido pela regressão. Somente as variáveis que atendem todas essas etapas podem ser usadas até o final do processo.

Os modelos desenvolvidos estão apresentados na Tabela 6 para profundidade de 0-10 cm, na Tabela 7, profundidade de 10-20 cm e Tabela 8, profundidade de 20-30 cm.

**Tabela 6.** Modelos de regressão para as variáveis de interesse na profundidade de 0-10 cm.

Variável Principal	Modelo de Regressão	R <sup>2</sup> <sub>adj</sub>
CC	-36,433 - 9,2 DS + 0,36Silte + 0,57 Argila + 1,899 CNBL + 1,06 MRVBF	0,827
ADT	-48,06 - 7,911 DS + 0,309 Argila + 2,131 CNBL + 2,15 MRVBF - 190,584 Profile	0,489
PMP	5,007 + 0,382 Silte + 0,247 Argila - 0,557 WI	0,508
AFD	2,183 + 0,058 Silte + 0,077 Argila	0,193
80 KPa	-29,616 - 7,69 DS + 0,292 Silte + 0,534 Argila + 1,5 CNBL + 0,818 MRVBF - 125,826 Profile	0,842

**Tabela 7.** Modelos de regressão para as variáveis de interesse na profundidade de 10-20 cm.

Variável Principal	Modelo de Regressão	R <sup>2</sup> <sub>adj</sub>
CC	-40,715 + 0,457 Silte + 0,583 Argila + 1,502 CNBL	0,894
ADT	40,837 - 14,286 DS + 0,302 Argila - 0,442 PT + 1,693 AACN + 145,23 Profile	0,615
PMP	4,864 + 11,221 DS - 0,34 Areia + 0,31 PT	0,612
AFD	-13,898 + 0,032 Argila + 0,586 CNBL	0,223
80 KPa	6,579 + 0,345 Silte + 0,661 Argila - 0,195 PT + 89,116 Profile	0,929

**Tabela 8.** Modelos de regressão para as variáveis de interesse na profundidade de 20-30 cm.

Variável Principal	Modelo de Regressão	R <sup>2</sup> <sub>adj</sub>
CC	2,772 + 0,808 Argila - 1,286 AACN	0,874
ADT	-14,547 + 1,225 Elevação - 15,625 DS + 42,942 Slope + 0,95 WI	0,482
PMP	-72,739 - 6,04 DS + 0,723 Silte + 0,141 AACN + 2,906 CNBL	0,49
AFD	-5,843 - 6,141 DS + 0,086 Silte - 0,0004 Argila + 0,644 CNBL	0,352
80 KPa	0,321 + 0,737 Argila - 1,236 AACN	0,903

Para CC, a variável argila está presente como preditora nas 3 profundidades, e silte em duas, mostrando a importância da textura na determinação desse atributo, ambas apresentando correlação positiva. As covariáveis de terreno não se repetem nas 3 profundidades, com exceção do CNBL nas duas primeiras profundidades, que é substituído por AACN na terceira profundidade. Essas duas covariáveis numéricas podem ser usadas para indicar energia potencial da água, gerando, portanto, informações similares quanto ao comportamento da água no terreno (PRATES et al., 2012). Porém tem correlação diferente com CC, sendo CNBL positiva e AACN negativa.

Para ADT a argila aparece como preditora somente nas duas primeiras profundidades, e DS aparece, com correlação negativa, nas três profundidades. As covariáveis de terreno mudam de acordo com a profundidade. Somente o Profile aparece em duas profundidades, porém com a combinação com as demais preditoras apresenta correlação negativa na primeira e positiva na segunda. Isso pode ser explicado pela mudança no sinal do intercepto, causada pela escolha do modelo via stepwise.

O atributo PMP apresenta relação com textura nas 3 profundidades e usa diferentes covariáveis como preditoras na primeira e na terceira profundidade. Na segunda profundidade o modelo utiliza somente DS, Areia e PT como estimadores, não usando nenhuma covariável numérica do terreno.

Para AFD somente CNBL aparece como preditor na segunda e terceira profundidade. O atributo argila aparece como preditor nas 3 profundidades e silte na primeira e terceira. Esse modelo não colocou como preditoras mais covariáveis do terreno, o que pode explicar ter os menores valores de  $R^2$ .

O atributo 80 KPa apresenta modelos mais simples, utilizando menos preditores, na medida em que aumenta a profundidade. O número das covariáveis numéricas do terreno diminuem na medida em que aumenta a profundidade, mas não há redução dos valores de  $R^2$ . Ao contrário, na profundidade 10-20 cm apresenta o maior valor de  $R^2$  de todos os modelos gerados.

Apesar da grande variação das diferentes variáveis preditoras, pode-se observar a maior presença dos atributos de textura, e a covariável de terreno que mais aparece relacionada com as variáveis ambientais é a CNBL. Mello et al. (2007) encontrou resultados similares relacionados à importância da textura na determinação de atributos físico hídricos do solo.

As variáveis que puderam ser interpoladas, até o final, através da RK estão apresentadas na Tabela 9, junto com os índices de precisão (ME e MAE) e acurácia (RMSE). As variáveis as quais não puderam ser modelados os resíduos, não permitiram chegar no resultado final do método.

Os resultados de RK para CC pioram com o aumento da profundidade. Para 80 KPa o melhor resultado foi para a profundidade de 10-20 cm. Nessa profundidade o  $R^2$  apresentou o maior resultado, podendo indicar que as variáveis preditoras foram melhor escolhidas pelo stepwise. Para a PMP só foi possível validar para a profundidade de 20-30 cm. Nas outras profundidades não foi possível terminar a análise. Para as variáveis AFD e ADT a análise não foi possível na profundidade 0-10 cm. Comparando os valores de RMSE, a AFD apresenta perda de acurácia com aumento de profundidade enquanto ADT apresenta comportamento contrário.

**Tabela 9.** Validação dos resultados viáveis obtidos por Krigagem com Regressão (RK).

RK	ME	MAE	RMSE
CC 0-10 cm	X*	X*	X*
CC 10-20 cm	-0,249	2,268	2,889
CC 20-30 cm	-1,449	3,163	5,345
80 KPa 0-10 cm	-0,337	3,732	5,166
80 KPa 10-20 cm	-0,058	2,095	2,751
80 KPa 20-30 cm	-0,98	2,884	5,083
PMP 0-10 cm	X*	X*	X*
PMP 10-20 cm	X*	X*	X*
PMP 20-30 cm	-1,122	2,145	3,912
AFD 0-10 cm	X	X	X
AFD 10-20 cm	0,008	0,823	0,963
AFD 20-30 cm	-0,309	0,803	1,115
ADT 0-10 cm	X*	X*	X*
ADT 10-20 cm	-0,264	3,843	5,409
ADT 20-30 cm	-0,442	2,583	3,254

X\* Não foi possível completar a análise. ME – Erro Médio; MAE – Erro Médio Absoluto; RMSE – Raiz Quadrada do Erro Médio.

Na Tabela 10 estão os resultados do índice MR para comparar os resultados da OK com a RK. De modo geral, não houve ganho de precisão na interpolação das variáveis usando o método de RK quando comparado com a OK. Para a primeira profundidade, somente 80 KPa apresentou ganho de acurácia. Para a profundidade de 10-20 cm, as variáveis AFD e ADT também apresentaram um pequeno ganho, também comparada com OK. Houve uma pequena perda de acurácia para a profundidade de 20-30 cm para todas as variáveis.

Isso mostra que nem sempre a utilização da RK irá apresentar ganhos significativos em relação à OK. Novamente, como visto antes nos resultados de CK, em casos de forte correlação espacial do atributo (CAMBARDELLA et al., 1994), e em áreas onde não há variações bruscas do terreno, normalmente a OK apresenta melhores resultados (KRAVCHENKO & ROBERTSON, 2007). Em casos onde o atributo não possui alta correlação espacial, e as variáveis preditoras apresentam alta correlação com as principais, se observa um maior ganho de acurácia quando utilizado RK em comparação com OK (ZHU & LIN, 2010). Não foi o que ocorreu nos resultados observados, provavelmente pela alta correlação espacial das variáveis principais (Tabela 2).

**Tabela 10.** Melhora relativa (MR) da Krigagem com Regressão (RK) comparada à Krigagem Ordinária (OK).

Atributos	RMSE OK	RMSE RK	MR (%)
CC 0-10 cm	5,583	X*	X*
CC 10-20 cm	2,797	2,889	-3,2892
CC 20-30 cm	4,961	5,345	-7,7404
80 KPa 0-10 cm	5,206	5,166	0,76834
80 KPa 10-20 cm	2,644	2,751	-4,0469
80 KPa 20-30 cm	4,664	5,083	-8,9837
PMP 0-10 cm	3,905	X*	X*
PMP 10-20 cm	5,075	X*	X*
PMP 20-30 cm	3,599	3,912	-8,6969
AFD 0-10 cm	0,682	X	X*
AFD 10-20 cm	0,988	0,963	2,53036
AFD 20-30 cm	1,02	1,115	-9,3137
ADT 0-10 cm	8,024	X*	X*
ADT 10-20 cm	5,519	5,409	1,99311
ADT 20-30 cm	3,114	3,254	-4,4958

X\* = Não foi possível completar a análise; RMSE – Raiz Quadrada do Erro Médio.

#### 4.6. Diminuição da Proporção dos Dados de Treinamento

Para essa etapa foram escolhidas para interpolação as variáveis 80 KPa, na profundidade 0-10 cm, e AFD na profundidade 10-20 cm. Essas variáveis foram as que apresentaram ganho em precisão ao se comparar os resultados de RK com os resultados de OK (Tabela 10).

Nessa etapa o banco de dados foi reduzido para 90%, 80%, 70%, 60% e 50%, em comparação ao total utilizado nas etapas anteriores. Os métodos utilizados para interpolação foram Krigagem Ordinária (OK) e Krigagem com Regressão (RK). Devido aos problemas do software na confecção de mapas via CoKrigagem (CK), esse método não entrou nessa parte do estudo.

Como o objetivo foi testar a redução no número de dados e avaliar a alteração na acurácia dos resultados, os mesmos índices utilizados anteriormente foram usados para comparar as interpolações. Para OK, nas Tabelas 11 e 12 estão os índices ME, MAE e RMSE comparando todas as proporções utilizadas. Como cada variável apresenta uma diferente

amplitude de valores numéricos, a comparação de uma variável com a outra quanto à eficiência do método fica prejudicada. Como os valores de 80 KPa são maiores numericamente do que os de AFD, não se pode afirmar que por apresentar menores valores de RMSE, a técnica seja mais eficiente para AFD. A comparação deve ser feita entre os resultados das proporções da mesma variável. Assim, ambas variáveis não apresentaram grandes diferenças em relação à precisão e acurácia com a redução da proporção dos pontos usados.

**Tabela 11.** Validação da OK para redução de dados de 80 KPa de 0-10 cm.

Porcentagem dos dados (%)	ME	MAE	RMSE
100	0,754	3,596	5,206
90	0,539	3,56	5,015
80	0,536	3,65	5,172
70	0,579	3,482	4,909
60	0,783	3,783	5,246
50	0,723	3,669	5,555

ME – Erro Médio; MAE – Erro Médio Absoluto; RMSE – Raiz Quadrada do Erro Médio.

**Tabela 12.** Validação da OK para redução de dados de 80 KPa de 10-20 cm.

Porcentagem dos dados (%)	ME	MAE	RMSE
100	-0,1	0,809	0,988
90	0,017	0,777	0,91
80	0,029	0,811	0,953
70	0,091	0,796	0,946
60	-0,028	0,812	0,963
50	-0,028	0,879	1,02

ME – Erro Médio; MAE – Erro Médio Absoluto; RMSE – Raiz Quadrada do Erro Médio.

Nas Tabelas 13 e 14 estão os resultados de melhora relativa das duas variáveis, também apresentadas em cada proporção utilizada. Nesse caso, esses resultados de MR mostram que o pior desempenho ocorre quando se usam 50% dos dados, mas em outras proporções ocorre até um aumento pequeno na eficiência da OK. Isso pode ser explicado pela manutenção da densidade dos pontos amostrais, como visto nas figuras 5 e 6, que mantêm a forte correlação espacial dos dados de cada atributo.

**Tabela 13.** Melhora relativa (MR) da OK para redução dos pontos da variável 80 KPa 0-10 cm, comparada com resultados da OK 100%.

Porcentagem dos dados (%)	ME	MAE	RMSE
100	5,206	5,206	0,000
90	5,015	5,206	3,669
80	5,172	5,206	0,653
70	4,909	5,206	5,705
60	5,246	5,206	-0,768
50	5,555	5,206	-6,704

RMSE – Raiz Quadrada do Erro Médio.

**Tabela 14.** Melhora relativa (MR) da OK para redução dos pontos da variável AFD 10-20 cm, comparada com resultados da OK 100%.

Porcentagem dos dados (%)	ME	MAE	RMSE
100	0,988	0,988	0,000
90	0,91	0,988	7,895
80	0,953	0,988	3,543
70	0,946	0,988	4,251
60	0,963	0,988	2,530
50	1,02	0,988	-3,239

RMSE – Raiz Quadrada do Erro Médio.

Após a OK foi feita a modelagem de regressão para esses dois atributos de forma a aplicar a técnica de RK. A metodologia para o desenvolvimento dos modelos foi a mesma aplicada anteriormente, usando as mesmas variáveis secundárias como preditoras do modelo. Os resultados dos modelos obtidos estão nas tabelas 15 e 16. Os valores de  $r^2$  não apresentam variações muito grandes para as duas variáveis com a redução do número de dados. Ocorre uma pequena redução do  $r^2$  para 80 KPa e um pequeno aumento para AFD. Os valores de  $r^2$  para a modelagem da variável 80 KPa são sempre maiores do que a variável AFD.

**Tabela 15.** Modelos de regressão para 80 KPa na profundidade de 0-10 cm com diferentes proporções do banco de dados.

Porcentagem dos dados (%)	Modelo de Regressão	$R^2_{adj}$
100	-29,616 – 7,69 DS + 0,292 Silte + 0,534 Argila + 1,5 CNBL + 0,818 MRVBF – 125,826 Profile	0,842
90	26,082 – 7,866 DS – 0,542 Areia -0,228 Silte + 1,451 CNBL + 0,845 MRVBF – 149,967 Profile	0,849
80	16,381 – 0,564 Areia – 0,212 Silte + 1,44 CNBL + 0,586 MRVBF – 141,366 Profile	0,825
70	18,687 -0,554 Areia – 0,242 Silte + 1,363 CNBL – 176,815 Plan	0,82
60	-60,372 + 0,261 Silte + 0,519 Argila + 2,194 CNBL - 201,384 Plan	0,801
50	-12,795 – 0,517 Areia – 0,286 Silte + 2,351 CNBL - 203,297 Plan	0,806

Com a redução dos pontos, para 80 KPa 0-10 cm, somente Silte e CNBL aparecem como preditores em todos os modelos. Areia e argila se alternam e Profile é substituído por Plan nas menores proporções.

Para AFD 10-20 cm, os modelos usaram menos preditores, somente dois em cada proporção. CNBL aparece em todas as proporções, indicando ser um bom preditor para esse atributo. Comparando os modelos desenvolvidos para os dois atributos, AFD sempre aparece com menor valor de  $R^2$ .



**Tabela 16.** Modelos de regressão para AFD na profundidade de 10-20 cm com diferentes proporções do banco de dados.

Porcentagem dos dados (%)	Modelo de Regressão	R <sup>2</sup> <sub>adj</sub>
100	-13,898 + 0,032 Argila + 0,586 CNBL	0,223
90	-16,203 + 0,024 Argila + 0,67 CNBL	0,236
80	-26,88 + 0,092 PT + 0,931 CNBL	0,313
70	-26,284 + 0,102 PT + 0,899 CNBL	0,298
60	-17,322 - 4,038 DS + 0,941 CNBL	0,298
50	-18,64 - 3,469 DS + 0,954 CNBL	0,311

Na validação da RK, os resultados para 80 KPa apresentam ganho de acurácia com redução para 50% dos dados (Tabela 17), ao contrário do que ocorreu com OK. As demais proporções apresentam aumento no RMSE, mas não muito evidenciados. Para AFD, com a redução da proporção, todos os valores dos índices usados são maiores quando comparados com 100% dos dados (Tabela 18). O que mostra uma perda de acurácia e precisão para todas as combinações.

**Tabela 17.** Validação da RK para redução de dados de 80 KPa de 0-10 cm.

Porcentagem dos dados (%)	ME	MAE	RMSE
100	-0,337	3,732	5,166
90	-0,327	4,261	5,826
80	-0,705	4,262	5,748
70	-4,47	7,3	5,43
60	-0,996	4,114	5,413
50	-0,686	5,831	4,722

ME – Erro Médio; MAE – Erro Médio Absoluto; RMSE – Raiz Quadrada do Erro Médio.

**Tabela 18.** Validação da RK para redução de dados de AFD de 10-20 cm.

Porcentagem dos dados (%)	ME	MAE	RMSE
100	0,008	0,823	0,963
90	-0,094	1,888	1,578
80	-0,022	1,914	1,466
70	0,053	1,896	1,450
60	-0,127	1,772	1,437
50	-0,173	1,449	1,809

ME – Erro Médio; MAE – Erro Médio Absoluto; RMSE – Raiz Quadrada do Erro Médio.

Ao comparar os resultados de RK das menores proporções com o resultado para 100% dos dados, todos apresentam perda de acurácia. Com exceção para 80 KPa que apresentou uma pequena melhoria comparando 50% dos dados com o conjunto total (Tabela 19). Para a variável AFD (Tabela 20) há perda da eficiência da interpolação com redução dos dados em todas as proporções.

**Tabela 19.** Melhora relativa (MR) da RK para redução dos pontos da variável 80 KPa 0-10 cm, comparada com RK 100%.

Porcentagem dos dados (%)	RMSE RK	RMSE RK (100%)	MR(%)
100	5,166	5,166	0,000
90	5,826	5,166	-12,776
80	5,748	5,166	-11,266
70	5,43	5,166	-5,110
60	5,413	5,166	-4,781
50	4,722	5,166	8,595

RMSE – Raiz Quadrada do Erro Médio.

**Tabela 20.** Melhora relativa (MR) da RK para redução dos pontos da variável AFD 10-20 cm, comparada com RK 100%.

Porcentagem dos dados (%)	RMSE RK	RMSE RK (100%)	MR(%)
100	0,963	0,963	0,000
90	1,578	0,963	-63,863
80	1,466	0,963	-52,233
70	1,450	0,963	-50,571
60	1,437	0,963	-49,221
50	1,809	0,963	-87,850

RMSE – Raiz Quadrada do Erro Médio.

Na comparação entre os dois interpoladores, para a variável 80 KPa, observa-se uma perda na acurácia dos resultados com a diminuição do número de pontos. Porém para o menor conjunto de dados (50%), a RK mostrou ser superior à OK para estimar os valores (Tabela 21). O que significa que nessa proporção as variáveis predictoras selecionadas pelo modelo de regressão melhoraram a interpolação dos resultados.

**Tabela 21.** Comparação entre Krigagem Ordinária (OK) e Krigagem com regressão (RK) para estimar a variável 80 KPa 0-10 cm com redução na porcentagem do banco de dados.

Porcentagem dos dados (%)	RMSE OK	RMSE RK	MR(%)
100	5,206	5,166	0,768
90	5,015	5,826	-16,171
80	5,172	5,748	-11,137
70	4,909	5,43	-10,613
60	5,246	5,413	-3,183
50	5,555	4,722	14,995

MR – Melhora Relativa; RMSE – Raiz Quadrada do Erro Médio.

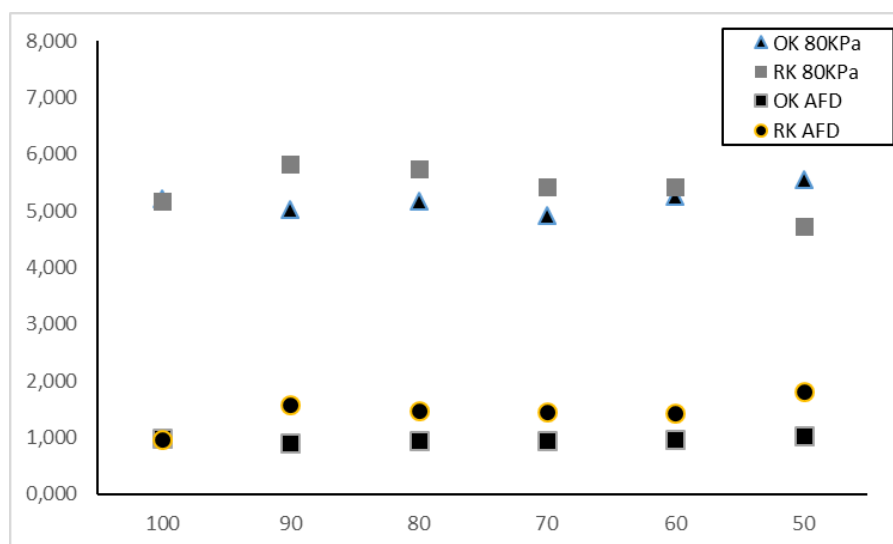
Para a variável AFD, quando comparamos os dois interpoladores, ocorre uma perda na acurácia com a redução do número de pontos (Tabela 22). A OK, com exceção da utilização do conjunto total de dados, é sempre superior à RK. Isso pode ser devido aos baixos valores de  $r^2$  obtidos pelos modelos, com as variáveis predictoras não contribuindo na melhoria da interpolação. Ao contrário do observado para 80 KPa, a utilização de RK não melhorou a qualidade da interpolação dos dados.

**Tabela 22.** Comparação entre Krigagem Ordinária (OK) e Krigagem com Regressão (RK) para estimar a variável AFD 10-20 cm com redução na porcentagem do banco de dados.

Porcentagem dos dados (%)	RMSE OK	RMSE RK	MR(%)
100	0,988	0,963	2,530
90	0,910	1,578	-73,407
80	0,953	1,466	-53,830
70	0,946	1,45	-53,277
60	0,963	1,437	-49,221
50	1,020	1,809	-77,353

MR – Melhora Relativa; RMSE – Raiz Quadrada do Erro Médio.

Na Figura 20 estão os resultados dos valores de RMSE para os dois atributos usando as duas técnicas de interpolação. Como pode ser observado, não há grandes variações na acurácia dos resultados dos interpoladores, para cada variável, com a diminuição do banco de dados. E pode ser observado que não há grande variação nos resultados da OK, usando o RMSE como parâmetro. Mostrando que esse interpolador mantém sua eficácia mesmo com a redução do número de amostras em 50% dos dados originais.



**Figura 20.** Gráfico com os resultados de RMSE para 80 KPa 0-10 cm e AFD 10-20 cm obtidos por OK e RK nas seis proporções usadas do banco de dados (Eixo X em % do total de dados).

Ressalta-se ainda que para AFD, com redução do banco de dados, a OK apresenta melhores resultados. E para 80 KPa, a RK, com a redução dos dados, somente foi superior quando usou-se 100% e 50% dos dados originais, sendo OK melhor nas demais proporções.

Dessa forma, pode-se considerar viável a utilização da OK para interpolar esses atributos usando apenas 50% do banco de dados de treinamento. São necessários mais análises para confirmar que o mesmo resultado pode ser aplicado à todos os demais atributos físico hídricos em todas as profundidades.

## 5. CONCLUSÕES

Usando o conjunto total de dados a OK pôde ser usada para interpolar todas as variáveis principais nas três profundidades amostradas. Todas apresentaram dependência espacial, permitindo a aplicação da técnica.

A CK não pode ser corretamente avaliada devido à problemas na validação das combinações geradas, pois nem todas puderam ser validadas. Nas combinações que foram possíveis realizar a validação, os resultados não foram suficientemente melhores que a OK de forma a justificar todo o esforço em realizar o método. Há necessidade de mais estudos envolvendo outros softwares para resolver esse problema.

A utilização das técnicas de RK também não apresentaram melhoras significativas em relação à OK. Nos resultados viáveis de validação, as variáveis preditoras não foram suficientes para melhorar de forma significativa os resultados das interpolações. Assim as covariáveis ambientais utilizadas neste trabalho não resultaram em melhoras na predição dos resultados. As variáveis que mais apareceram nos modelos foram os atributos de textura e CNBL.

De forma geral, considerando o trabalho envolvendo variografias e modelagem para CK e RK, a OK mostrou-se um melhor interpolador para as variáveis escolhidas na área de estudo.

Em relação à redução do número de amostras, a RK foi superior na interpolação comparada à OK, para a variável 80 KPa na profundidade de 0-10 cm na proporção de 50% dos dados (60 amostras). Porém em todos os demais resultados a OK obteve melhor performance.

Considerando os resultados das duas variáveis, 80 KPa de 0-10 cm e AFD de 10-20 cm, com os interpoladores utilizados, pode-se afirmar que é viável obter resultados consistentes utilizando-se apenas 50% do banco de dados originais através da OK. O que corresponde a 60 medições em 6,24 ha da área de produção agroecológica.

São necessários mais estudos para confirmar se é possível interpolar, com redução do número de amostras e mantendo a acurácia dos resultados, os demais atributos físico hídricos, nas diferentes profundidades, dentro da área estudada.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- ABBOUD, A. C.; RIBEIRO, R. L. D.; ESPINDOLA, J. A. **Doze anos de Fazendinha Agroecológica do km 47: uma parceria entre UFRRJ, Embrapa Agrobiologia e PESAGRO-Rio: Concurso Nacional de Sistematização de Experiências sobre Agroecologia e Agriculturas Alternativas.** Embrapa Agrobiologia, Documentos, 2005. 31p.
- ALMEIDA, D. L.; RIBEIRO, R. L. D.; GUERRA, J. G. M. **Sistema Integrado de Produção Agroecológica: uma experiência de pesquisa em agricultura orgânica.** Documentos 169, Seropédica EMBRAPA Agrobiologia, 37p. 2003.
- BEVEN, K.; WOOD, E. F. Catchment geomorphology and the dynamics of runoff contributing areas. **Journal of Hydrology**, Amsterdam, v. 65, n. 1, p. 139-158, 1983.
- BISHOP, T. F. A.; MCBRATNEY, A. B. A comparison of prediction methods for the creation of field-extent soil property maps. **Geoderma**, Amsterdam, v. 103, n. 1, p. 149-160, Jan. 2001.
- BISHOP, T. F. A.; HORTA, A.; KARUNARATNE, S. B. Validation of Digital Soil Maps at Different Spatial Supports. **Geoderma**, n. 241-242, p. 238-249, 2015.
- BÖHNER, J.; MCCLOY, K.R.; STROBL, J. [Eds.] (2006). **System for an automated geographical analysis.** Disponível em: <<http://www.saga-gis.org>>. Access in: 28 de julho de 2016.
- BURGES, T. M, WEBSTER, R. Optimal interpolation and isarithmic mapping. I. The semivariogram and punctual kriging. **Journal of Soil Science**, v. 31, p. 315-332, 1980a.
- BURGES, T. M, WEBSTER, R. Optimal interpolation and isarithmic mapping. II. Block kriging. **Journal of Soil Science**, v. 31, p. 333-342, 1980b.
- BURGES, T. M., WEBSTER, R., McBRATNEY, A. B. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. **Journal of Soil Science**, v. 32, p. 643-659, 1981.
- BUSSAB, W. O; MORETTIN, P. A. **Estatística Básica.** Editora Atual. Quarta edição. São Paulo. 1981.
- BURROUGH, P. A., Beckett, P. H. T. JARVIS, M. G. The relation between cost and utility in soil survey. **Journal of Soil Science** 22, 368-81. 1971.
- CAMBARDELLA, C. A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F. & KONOPKA, A. E. Field scale variability of soil properties in Central Iowa soils. **Soil Sci. Soc. Am. J.**, 58:1501-1511, 1994.
- CEDDIA, M. B., GOMES, A. S., VASQUES, G. M., PINHEIRO, E. F. M. Soil Carbon Stock and Particle Size Fractions in the Central Amazon Predicted from Remotely Sensed Relief, Multispectral and Radar Data. **Remote Sensing** 9(124), 1-19, 2017.

CEDDIA, M. B., VIEIRA, S. R., VILLELA, A. L. O., MOTA, L. S., ANJOS, L. H. C., CARVALHO, D. F. Topography and Spatial Variability of Soil Physical Properties. **Scientia Agricola** 66 (3), 338-352. 2009.

CEDDIA, M. B., VENTURA, S. D., OLIVEIRA, R. F., VILELLA, A.L.O., VARELLA, C.A.A. An Algorithm for Mapping the Spatial Variability of Soil Physical Quality. **In: Application of Soil Physics in Environmental Analyses. Measuring, Modelling and Data Integration.** Springer, 499 p., 2014.

DEUTSCH, C. V. & JOURNEL, A. G. **GSLIB. Geostatistical Software Library and User's Guide.** Oxford University Press, Second Edition, 369p., 1998.

EMBRAPA. **Manual de Métodos de Análise de Solo.** 2ed, EMBRAPA SOLOS, 212p, 1997.

FARIAS, L. N. **Variabilidade Espacial de Atributos Físico-Hídricos dos Solos da Região Norte Fluminense.** Dissertação (Mestrado em Ciência do Solo) – Universidade Federal Rural do Rio de Janeiro. 65p. 2008.

FOX, J. MONETTE, G. Generalized Collinearity Diagnostics. **Journal of the American Statistical Association**, 48 (417), 178-183, 1992.

GALLANT, J. C.; DOWLING, T. I. A multi-resolution index of valley bottom flatness for mapping depositional areas. **Water Resource Research**, v.39, p.1347, 2003.

GIVI, J. PRASHER, S. O.; PATEL, R. M. Evaluation of Pedotransfer Functions in Predicting the Soil Water Contents at Field Capacity and Wilting Point. **Agricultural Water Management** 70, 83-96. 2004.

GOOVAERTS, P. Ordinary Cokriging Revisited. **Math. Geol.** 30, 21-42, 1998.

GOOVAERTS, P. Geostatistics in soil science: state-of-the-art and perspectives. **Geoderma**, Amsterdam, v. 89, n. 1, p. 1-45, 1999.

GOULARD, M., VOLTZ, M. Linear Correogionalization Model: Tools for Estimation and Choice of Cross-Variogram Matrix. **Math. Geol.** 24 (3), 269-286, 1992.

HENGL, T.; HEUVELINK, G. B. M.; ROSSITER, D.G. About Regression-Kriging: From Equations to Case Studies. **Computers & Geosciences** 33, 1301-1315. 2007.

HENGL, T., HEUVELINK, G. B. M. and STEIN, A. A Generic Framework for Spatial Prediction of Soil Variables based on Regression-Kriging. **Geoderma**, 120: 75-93. 2004.

HERBST, M.; DIEKKRÜ, B.; VEREECKEN, H. Geostatistical co-regionalization of soil hydraulic properties in a micro-scale catchment using terrain attributes. **Geoderma**, Amsterdam, v. 132, n. 1/2, p. 206-221, May 2006.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied Geostatistics.** Oxford, Oxford University Press, 560p, 1989.

KARLEN, D. L.; MAUSBACH, M. J.; DORAN, J. W.; CLINE, R. G.; HARRIS, R. F.;

SCHUMAN, G. E. Soil Quality: A Concept, Definition, and Framework for Evaluation. **Soil Sci. Am. J.** 61:4-10, 1997.

KERRY, R.; OLIVER, M. A. Comparing sampling needs for variograms of soil properties computed by method of moments and residual maximum likelihood. **Geoderma**, v.140, p. 383-396, 2007.

KNOTTERS, M., BRUS, D. J., OUDE VOSHARR, J. H. A Comparison of Kriging, co-kriging and Kriging Combined with Regression for Spatial Interpolation of Horizon Depth with Censored Observations. **Geoderma** 67, 227-246, 1995.

KRAVCHENKO, A. N., ROBERTSON, G. P. Can Topographical and Yield Data Substantially Improve Total Soil Carbon Mapping by Regression Kriging? **Agronomy Journal**, 99, 12-17, 2007.

LETEY, J. Relationship between soil physical properties and crop production. **Adv. Soil Sci.**, v. 1, p. 277-294, 1985.

LI, J. & HEAP, A. D. A Review of Spatial Interpolation Methods for Environmental Scientists. **Geoscience Australia**, Record 2008/23, 137p. 2008.

LIAO, K., XU, S., WU, J., LIN, Q. Cokriging of Soil Cation Exchange Capacity Using the First Principal Component Derived from Soil Physico-Chemical Properties. **Agricultural Sciences in China** 10(8), 1246-1253, 2011.

MATA, M. G. F. **Qualidade do Solo e Avaliação Microeconômica de um Módulo Experimental de Produção Orgânica Intensiva de Hortaliças**. Dissertação (Mestrado em Ciência do Solo) – Universidade Federal Rural do Rio de Janeiro. 75p. 2012.

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

McBRATNEY, A., ODEH, I. O. A., BISHOP, T. F. A., DUNBAR, M. S., SHATAR, T. M. An overview of pedometric techniques for use in soil survey. **Geoderma** 97, 293-327, 2000.

McBRATNEY, A. B; MENDONÇA SANTOS, M. L.; MINASNY, B. On digital soil mapping. **Geoderma**. 117, 3-52. 2003.

McENZIE, N. J. & GALLANT, J. C. Digital Soil Mapping with Improved Environmental Predictors and Models of Pedogenesis. **In:** LAGACHERIE, P.; McBRATNEY, A. B.; VOLTZ, M. *Developments in Soil Science*, v.31, p.327-349, 2007.

MEDICI, L. O.; REINERT, F.; CARVALHO, D. F.; KOZAK, M.; AZEVEDO, R. A. What about keeping plants well watered? **Environmental and Experimental Botany**, 99, 38-42. 2014

MELLO, C. R., GOMES, N. M., SILVA, A. M., JUNQUEIRA Junior, J. A. Modelagem de Atributos Físico Hídricos do Solo numa Bacia Hidrográfica da Região do Alto Rio Grande, MG. **Revista Brasileira de Ciências do Solo** 31, 854-852, 2007.

MILLER, B. A., KOSZINSKI, S., HIEROLD, W.; ROGASIK, H.; SCHRODER, B.; VAN OOST, K., WEHRHAN, M.; SOMMER, M. Towards Mapping Soil Carbon Landscapes: Issues of Sampling Scale and Transferability. **Soil and Tillage Research** vol. 156, p. 194-208, 2016.

MINASNY, B. McBRATNEY, A. B. Spatial prediction of soil properties using EBLUP with the Matern covariance function. **Geoderma**, 140: 324-336. 2007.

MOORE, I. D.; GESSLER, P. E.; NIELSEN, G. A.; PETERSON, G. A. Soils attribute prediction using terrain analysis. **Soil Science Society of American Journal**,. 57(2), 443-452, 1993.

NIELSEN, D. R. & WENDROTH, O. Spatial and Temporal Statistics. Sampling Field Soils and their Vegetation. CATENA VERLAG GMBH, **Geocology Textbook**, 416 p. 2003.

ODEH, I. O. A., McBRATNEY, A. B., CHITTLEBOROUGH, D. J. Spatial Prediction of Soil Properties from Landform Attributes Derived from a Digital Elevation Model. **Geoderma**, 63: 197-214. 1994.

ODEH, I. O. A., McBRATNEY, A. B., CHITTLEBOROUGH, D. J. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. **Geoderma**, 67 (3-4), 215-226. 1995.

OLIVER, M. A., WEBSTER, R. A tutorial Guide to Geostatistics: Computing and Modelling Variograms and Kriging. **Catena** 113, 56-69, 2014.

PAN, W. Akaike's Information Criterion in Generalized Estimating Equations. **Biometrics**, 57, 120-125, 2001.

PEREIRA, G. A., MILANI, L. L., CIRILLO, M. A. Uso de Alguns Estimadores ridge na Análise Estatística de Experimentos em Entomologia. **Revista Ceres**, Viçosa, 61, 338-342, 2014.

PIEDALLU, C.; GÉGOUT, J. C.; BRUAND, A.; SEYNAVE, I. Mapping Soil Water Holding Capacity Over Large Areas to Predict Potential Production of Forest Stands. **Geoderma**, 160, 355-366. 2011.

PLANT, R. E. Site specific management: the application of information technology to crop production. Computers and electronics in agriculture. **Elseviers Science B. V.**, Volume 30, 9-29. 2001.

PRATES, V. **Utilização de Índices para a Representação da Paisagem como apoio para Levantamentos Pedológicos em Ambiente de Geoprocessamento**. Dissertação Mestrado. Pós Graduação em Ciência do Solo, UFPR. 29 p. 2010.

PRATES, V. SOUZA, L. C. P., OLIVEIRA JUNIOR, J. C. Índices para a Representação da Paisagem como apoio para Levantamento Pedológico. **Revista Brasileira de Engenharia Agrícola e Ambiental**, 16(4), 408-414, 2012.

REICHARDT, K. **A Água em Sistemas Agrícolas**. São Paulo: Manole, 1990. 188p.



REICHARDT, K.; TIMM, L. C. **Solo, Planta e Atmosfera: Conceitos, Processos e Aplicações**. Ed. Manole, Barueri-SP, 478p. 2004.

SCHOLOEDER, C. A.; ZIMMERMAN, N. E.; JACOBS, M. J. Comparison of methods for interpolating soil properties using limited data. **Soil Science Society of American Journal**, v. 65, p. 470-479, 2001.

SILVA, G. L. S. **Características Socioeconômicas, edáficas e produtivas das áreas cultivadas com cana-de-açúcar (*Saccharum sp.*) em Paraty**. Dissertação (Mestrado em Ciência do Solo) – Universidade Federal Rural do Rio de Janeiro. 119 p. 2013.

SIMBAHAN, G. C., DOBERMANN, A., GOOVAERTS, P., PING, J. L., HADDIX, M. L. Fine Resolution Mapping of Soil Organic Carbon Based on Multivariate Secondary Data. **Geoderma** 132, 471-489, 2006.

SOUZA, Z. M.; SOUZA, G. S.; MARQUES JUNIOR, J.; PEREIRA, G. T. Numero de Amostras na Análise Geoestatística e na krigagem de mapas de atributos do solo. **Ciência Rural** 44, 261-268, 2014.

SUMFLETH, K.; DUTTMANN, R. Prediction of soil property distribution in paddy soil landscape using terrain data and satellite information as indicators. **Ecological Indicators**, London, v. 8, n. 5, p. 485-501, 2008.

SUN, W.; MINASNY, B.; McBRATNEY, A. Analysis and Prediction of Soil Properties Using Local Regression-kriging. **Geoderma** 171-172, 16-23, 2012.

UTSET, A.; LOPEZ, T.; DIAZ, M. A comparison of soil maps, kriging and a combined method for spatially predicting bulk density and field capacity of ferralsols in the Havana-Matanzas Plain. **Geoderma**, 96, 199-213. 2000.

VIEIRA, S. R., NIELSEN, D. R., BEGGAR, J. W. Spatial variability of field measured infiltration rate. **Soil Science Society of America Journal**, v. 45, p. 1040-1048, 1981.

VIEIRA, S. R. Geoestatística em Estudos de Variabilidade Espacial de Solos. **In: Tópicos avançados em Ciência do Solo**. Sociedade Brasileira de Ciência do Solo. 1-53 p. 2000.

VIEIRA, S. R., CARVALHO, J. R. P., CEDDIA, M. B., GONZÁLEZ, A. P. Detrending non stationary data For Geostatistical Applications. **Bragantia** 69, 1-8, 2010.

VILLELA, A. L. O. **Variabilidade Espacial da Qualidade Físico-Hídrica dos Solos de uma Unidade de Pesquisa em Produção Agroecológica**. Dissertação (Mestrado Em Ciência do Solo) – Universidade Federal Rural do Rio de Janeiro. 44p. 2007.

WACKERNAGEL, H. **Multivariate Geostatistics, an Introduction with Applications**. Springer Verlag, Berlin Heidelberg 3ª ed. 387p, 2003.

WANG, D.; LAFFAN, S. W.; LIU, Y.; WU, L. Morphometric Characterization of Landform from DEMs. **International Journal of Geographical Information Science**, v. 24, p. 305-326, 2010.

WEBSTER, R., BURGESS, T. M. Optimal Interpolation and Isarithmic Mapping of Soil Properties. III. Changing Drift And Universal Kriging. **Journal of Soil Science**, v. 31, p. 505-524, 1980.

WEBSTER, R.; OLIVER, M. Sample adequately to estimate variograms of soil properties. **Soil Science**, Baltimore, v. 43, n. 1, p. 177-192, 1992.

WEBSTER, R. The Development of Pedometrics. **Geoderma**. 62, 1-15. 1994.

WEBSTER, R. & OLIVER, M. A. **Geostatistical for Environmental Scientists**. John Wiley & Sons. West Sussex, PO19 1UD, England. 2001.

YOST, R. S., UEHARA, G., FOX, R. L. Geoestatistical Analysis of Soil Chemical Properties of Large Land Areas. I. Semi-variograms. **Soil Science Society of America Journal**, v. 46, p. 1028-1032, 1982.

ZHU, Q.; LIN, H. S. Comparing Ordinary Kriging and Regression Kriging for Soil Properties in Contrasting Landscapes. **Pedosphere**, 20 (5), 594-606. 2010.

## 7. ANEXOS

### 7.1. Estatística Descritiva dos Atributos de Solo nas Três Profundidades

**Tabela 1.** Estatística Descritiva do Atributo Capacidade de Campo na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	13,98	11,5	8,1	7,7205	59,6072	0,9554	0,0989	36,1	4,1	124
80%	13,8585	11,7	7,8	7,457	56,611	0,9052	0,0632	36,1	4,1	99
20%	14,464	10,4	9,2	8,837	78,102	1,0784	0,1254	36	4,5	25
75%	14,1451	11,8	7,8	7,731	59,778	0,9998	0,2686	36,1	4,5	93
25%	13,487	11,1	9,7	7,7931	60,7338	0,8729	-0,2868	30,1	4,1	31
70%	14,229	11,8	15,5	7,9176	62,6892	0,9981	0,2434	36,1	4,5	87
30%	13,645	10,8	10,2	7,3305	53,7364	0,8364	-0,3799	30	4,1	37

**Tabela 2.** Estatística Descritiva do Atributo Capacidade de Campo na Profundidade de 10 a 20 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	13,12	10,1	4,5	8,2201	67,5704	0,9256	-0,3298	34,1	3,6	122
80%	13,699	10,15	10,1	8,211	67,434	0,8319	-0,5788	32,2	4	98
20%	10,758	7,9	4,5	7,989	63,832	1,5593	2,0301	34,1	3,6	24
75%	13,48	10,1	4,5	8,598	73,93	0,8898	-0,4695	34,1	3,6	92
25%	12,016	9,25	4,2	6,9477	48,271	0,9081	-0,3066	28,2	4	30
70%	13,369	10,25	4,5	8,3393	69,5454	0,9291	-0,2695	34,1	3,6	86
30%	12,525	9,25	4,5	8,0116	64,187	0,9444	-0,455	30	4,4	36

**Tabela 3.** Estatística Descritiva do Atributo Capacidade de Campo na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	12,683	9,3	4,7	8,615	74,2198	0,9105	-0,4222	33,9	3,2	118
80%	12,874	9,4	4,7	8,478	71,891	0,8779	-0,4242	33,9	3,8	95
20%	11,895	7,3	4,2	9,313	86,735	1,1326	-0,1266	31,6	3,2	23
75%	12,614	9,1	4,7	8,629	74,473	0,9701	-0,2969	33,9	3,7	89
25%	12,896	9,9	4,2	8,7184	76,011	0,7721	-0,6776	31,6	3,2	29
70%	13,39	9,7	3,8	8,839	78,1294	0,7739	-0,6704	33,9	3,7	83
30%	11,008	7,5	7,5	7,9288	62,8661	1,3453	0,7806	31,6	3,2	35

**Tabela 4.** Estatística Descritiva da Retenção de Água a 80 kPa na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	10,603	8,1	6,4	6,8449	46,8539	1,0872	0,4182	31,8	2,4	124
80%	10,468	8,2	4,5	6,608	43,671	1,0509	0,4513	31,8	2,4	99
20%	11,136	8	N/D	7,8377	61,4307	1,1703	0,2884	30	2,8	25
75%	10,764	8,7	3,5	6,82	46,513	1,1404	0,6612	31,8	2,5	93
25%	10,119	7,7	3,8	7,009	49,1362	1,0037	-0,0899	26,7	2,4	31
70%	10,726	8,4	3,5	7,033	49,4701	1,1456	0,6084	31,8	2,5	87
30%	10,313	7,8	5,5	6,4639	41,7828	0,9177	-0,2411	24,6	22,2	37

**Tabela 5.** Estatística Descritiva da Retenção de Água a 80 kPa na Profundidade de 10 a 20 cm.

<b>Conjunto de dados</b>	<b>Média</b>	<b>Mediana</b>	<b>Moda</b>	<b>Desvio Padrão</b>	<b>Variância</b>	<b>Assimetria</b>	<b>Curtose</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>Número de dados</b>
100%	9,591	6,8	2,7	7,3233	53,6316	1,0047	-0,158	29,4	1,5	122
80%	10,096	7,35	4	7,373	54,370	0,8718	-0,5052	27,3	1,5	98
20%	7,529	4,9	2,1	6,8789	47,3204	1,8497	3,3636	29,4	1,9	24
75%	9,933	6,65	2,7	7,666	58,782	0,9752	-0,3074	29,4	1,8	92
25%	8,543	6,95	4,8	6,1489	37,8094	0,9267	-0,2571	22,6	1,5	30
70%	9,759	7,1	4	7,3908	54,624	1,0126	-0,1334	29,4	1,5	86
30%	9,191	6,4	5	7,247	52,5202	1,0252	-0,0964	26,9	1,6	36

**Tabela 6.** Estatística Descritiva da Retenção de Água a 80 kPa na Profundidade de 20 a 30 cm.

<b>Conjunto de dados</b>	<b>Média</b>	<b>Mediana</b>	<b>Moda</b>	<b>Desvio Padrão</b>	<b>Variância</b>	<b>Assimetria</b>	<b>Curtose</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>Número de dados</b>
100%	9,199	5,65	2,8	7,6073	57,8723	0,946	-0,3049	29,5	1,4	118
80%	9,376	6,8	2,8	7,440	55,360	0,8785	-0,3611	29,5	1,4	95
20%	8,465	4,8	1,7	8,3984	70,534	1,2820	0,2394	27,2	1,7	23
75%	9,179	6,3	2,8	7,581	57,485	0,9892	-0,1939	29,5	1,6	89
25%	9,258	5	3,7	7,8198	61,1496	0,8667	-0,4858	27,2	1,4	29
70%	9,804	7,1	2,8	7,7727	60,4153	0,7632	-0,6446	29,5	1,6	83
30%	7,762	4,8	2,4	7,1008	50,4224	1,5415	1,4882	27,2	1,4	35

**Tabela 7.** Estatística Descritiva do Atributo Ponto de Murcha Permanente na Profundidade de 0 a 10 cm.

<b>Conjunto de dados</b>	<b>Média</b>	<b>Mediana</b>	<b>Moda</b>	<b>Desvio Padrão</b>	<b>Variância</b>	<b>Assimetria</b>	<b>Curtose</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>Número de dados</b>
100%	6,727	5,3	4,4	4,3187	18,6514	1,5177	2,4704	22,5	1	124
80%	6,665	5,3	4,1	4,237	17,953	1,5442	2,8031	22,5	1,2	99
20%	6,972	5,2	2,8	4,711	22,201	1,4933	2,0076	20,6	1	25
75%	6,748	5,3	4,1	4,217	17,78	1,4594	2,4343	22,5	1	93
25%	6,664	5,1	3,2	4,6805	21,9077	1,7225	3,0711	20,6	1,4	31
70%	6,534	5,5	3,2	3,9509	15,6101	1,4489	2,5826	21,6	1	87
30%	7,181	4,9	4,1	5,11	26,1332	1,4901	1,8646	22,5	1,4	37

**Tabela 8.** Estatística Descritiva do Atributo Ponto de Murcha Permanente na Profundidade de 10 a 20 cm.

<b>Conjunto de dados</b>	<b>Média</b>	<b>Mediana</b>	<b>Moda</b>	<b>Desvio Padrão</b>	<b>Variância</b>	<b>Assimetria</b>	<b>Curtose</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>Número de dados</b>
100%	6,336	4,2	3,9	5,1694	26,723	1,5238	1,7316	23,5	0,8	122
80%	6,611	4,4	3,9	5,3559	28,6857	1,4786	1,5161	23,5	0,8	98
20%	5,216	3,75	2,3	4,2399	17,9771	1,6166	2,3082	17,6	1,1	24
75%	6,581	4,35	2,5	5,4507	29,7109	1,4904	1,5246	23,5	0,8	92
25%	5,586	3,95	2,7	4,1828	17,4963	1,3807	1,0827	16,7	1,3	30
70%	6,54	4,8	3,9	5,3356	28,469	1,5604	1,8809	23,5	1,1	86
30%	5,85	3,65	3,3	4,7854	22,9002	1,3993	1,0627	19,3	0,8	36

**Tabela 9.** Estatística Descritiva do Atributo Ponto de Murcha Permanente na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	6,503	4,1	1,9	5,7641	33,2252	1,1557	0,1657	23,7	1,1	118
80%	6,834	4,6	1,9	5,8633	34,3788	1,0739	0,0134	23,7	1,2	95
20%	5,134	3,4	1,1	5,232	27,3741	1,6869	1,7196	19	1,1	23
75%	6,611	4,1	2,8	5,9364	35,2414	1,1862	0,2067	23,7	1,1	89
25%	6,172	4,1	1,8	5,2844	27,9249	1,0095	-0,2188	19	1,1	29
70%	7,121	4,4	2,8	6,1635	37,9887	0,9798	-0,2856	23,7	1,1	83
30%	5,037	3,4	1,9	4,4235	19,5677	1,5633	1,4117	16,3	1,1	35

**Tabela 10.** Estatística Descritiva do Atributo Água Facilmente Disponível na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	3,378	3,4	2,4	1,4743	2,1737	-0,2037	-0,6392	6,1	0,1	124
80%	3,393	3,4	4,8	1,4931	2,2293	-0,2842	-0,5592	6,1	0,1	99
20%	3,316	3,3	5,1	1,4255	2,0322	0,1667	-0,9054	6	0,9	25
75%	3,381	3,4	2,4	1,5135	2,2908	-0,1848	-0,6495	6,1	0,1	93
25%	3,367	3,3	4,4	1,3736	1,8869	-0,2987	-0,5887	5,6	0,1	31
70%	3,4	3,4	2,4	1,5169	2,3011	-0,2649	-0,5362	6,1	0,1	87
30%	3,327	3,3	4,7	1,3877	1,9259	-0,0393	-0,9635	5,6	0,2	37

**Tabela 11.** Estatística Descritiva do Atributo Água Facilmente Disponível na Profundidade de 10 a 20 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	3,526	3,6	1,8	1,5512	2,4064	0,3424	-0,097	8,8	0,5	122
80%	3,6	3,7	1,8	1,4502	2,1032	-0,0593	-1,0746	6,6	1,1	98
20%	3,225	2,7	1,5	1,9161	3,6715	1,3559	2,2859	8,8	0,5	24
75%	3,542	3,6	1,8	1,601	2,5633	0,4012	0,0201	8,8	0,5	92
25%	3,476	3,65	4,2	1,4119	1,9935	0,037	-0,8969	6,1	1,2	30
70%	3,609	3,7	1,8	1,5982	2,5542	0,3596	0,0901	8,8	0,5	86
30%	3,327	3,15	4	1,4347	2,0586	0,1971	-1,02	6,1	1,1	36

**Tabela 12.** Estatística Descritiva do Atributo Água Facilmente Disponível na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	3,489	3,2	2,4	1,5877	2,5209	0,8008	0,3392	8,2	0,7	118
80%	3,504	3,2	2,5	1,5694	2,4631	0,7163	0,1643	7,7	0,7	95
20%	3,430	3,2	2,4	1,6963	2,8776	1,1742	1,4458	8,2	1,5	23
75%	3,44	3,2	2,4	1,5488	2,3988	0,8949	0,5295	7,7	0,7	89
25%	3,641	3,5	5,4	1,7214	2,9632	0,5723	0,1319	8,2	1,1	29
70%	3,591	3,4	2,4	1,6014	2,5646	0,6692	0,1554	7,7	0,7	83
30%	3,248	2,6	2,2	1,5505	2,4043	1,2056	1,4724	8,2	1,3	35



**Tabela 13.** Estatística Descritiva do Atributo Água Disponível Total na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	7,253	4,75	3,5	6,9006	47,6192	2,1943	4,352	35	1,2	124
80%	7,193	4,8	4,8	6,6652	44,4260	2,0846	3,7904	32	1,2	99
20%	7,488	4,3	4,3	7,9101	62,5702	2,5352	6,2601	35	2,1	25
75%	7,398	4,7	3,3	7,1859	51,6375	2,2205	4,4439	35	1,2	93
25%	6,816	4,8	3,5	6,0515	36,6207	2,0027	3,2219	25,3	1,8	31
70%	7,591	4,7	3,3	7,5193	56,5412	2,1114	3,7231	35	1,2	87
30%	6,456	4,8	4,8	5,1671	26,6991	1,9598	3,2457	22,7	1,6	37

**Tabela 14.** Estatística Descritiva do Atributo Água Disponível Total na Profundidade de 10 a 20 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	6,781	4,65	4,5	5,8197	33,8699	2,3677	5,0491	28,4	1,6	122
80%	7,086	4,8	4,5	5,9204	35,0512	2,2259	4,1800	28,4	1,8	98
20%	5,537	4,2	4,7	5,3243	28,3485	3,4643	14,2443	27,9	1,6	24
75%	6,897	4,65	4,5	6,1089	37,3193	2,3427	4,8009	28,4	1,6	92
25%	6,426	4,6	4,1	4,9031	24,0406	2,3709	5,7296	23,3	2,2	30
70%	6,827	4,7	4,5	5,8112	33,7707	2,4629	5,6765	28,4	1,6	86
30%	6,672	4,25	4,1	5,9212	35,0609	2,2481	4,4208	26,6	1,8	36

**Tabela 15.** Estatística Descritiva do Atributo Água Disponível Total na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	6,181	4,8	3,3	4,7258	22,3335	2,9855	10,0376	28,2	2,1	118
80%	6,038	4,8	3,3	4,1869	17,5304	2,9775	10,9925	27,1	2,1	95
20%	6,769	4,2	4	6,5894	43,4213	2,6352	6,6479	28,2	2,1	23
75%	6,006	4,7	3,3	4,2984	18,477	2,9889	10,7712	27,1	2,1	89
25%	6,717	5,2	5,8	5,9039	34,8571	2,8147	8,2216	28,2	2,1	29
70%	6,271	5,2	3,3	4,383	19,2111	2,8421	9,7988	27,1	2,1	83
30%	5,968	4	3,3	5,5185	30,4545	3,2163	10,7725	28,2	2,1	35

**Tabela 16.** Estatística Descritiva do Atributo Densidade do Solo na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	1,487	1,52	1,53	0,1262	0,0159	-0,7395	0,882	1,73	1,04	123
80%	1,49	1,515	1,6	0,13	0,0169	-0,7819	0,9926	1,73	1,04	98
20%	1,476	1,52	1,52	0,1118	0,0125	-0,6185	0,3010	1,69	1,2	25
75%	1,494	1,52	1,53	0,1317	0,0173	-0,9286	1,3174	1,73	1,04	92
25%	1,466	1,46	1,6	0,1073	0,0115	-0,0273	-1,141	1,67	1,28	31
70%	1,483	1,5	1,46	0,1233	0,0152	-0,5389	0,2957	1,73	1,1	86
30%	1,494	1,53	1,53	0,1341	0,0179	-1,1766	2,3944	1,71	1,04	37

**Tabela 17.** Estatística Descritiva do Atributo Densidade do Solo na Profundidade de 10 a 20 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	1,579	1,585	1,56	0,1036	0,0107	-0,934	1,3679	1,76	1,22	104
80%	1,575	1,58	1,61	0,1069	0,0114	-1,0024	1,4940	1,74	1,22	81
20%	1,549	1,59	1,68	0,0918	0,0084	-0,3995	-0,3700	1,76	1,42	23
75%	1,58	1,58	1,56	0,098	0,0096	-0,7125	1,0225	1,76	1,22	79
25%	1,578	1,61	1,61	0,1219	0,0148	-1,3351	1,9396	1,73	1,24	25
70%	1,571	1,56	1,55	0,0964	0,0093	-0,2805	-0,2999	1,76	1,29	74
30%	1,6	1,615	1,61	0,1188	0,0141	-2,1046	5,0493	1,74	1,22	30

**Tabela 18.** Estatística Descritiva do Atributo Densidade do Solo na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	1,613	1,63	1,68	0,1023	0,0104	-1,4395	3,1883	1,77	1,17	95
80%	1,607	1,63	1,57	0,1048	0,0109	-1,5316	3,5889	1,77	1,17	74
20%	1,634	1,66	1,72	0,0919	0,0084	-0,9173	-0,1279	1,76	1,44	21
75%	1,623	1,63	1,68	0,0931	0,0086	-1,1536	1,3439	1,77	1,35	68
25%	1,588	1,63	1,66	0,1208	0,0145	-1,6457	4,3099	1,76	1,17	27
70%	1,612	1,63	1,68	0,09	0,0081	-0,9633	0,8491	1,77	1,35	65
30%	1,615	1,635	1,72	0,1266	0,016	-1,8127	4,2632	1,77	1,17	30

**Tabela 19.** Estatística Descritiva do Atributo Densidade das Partículas na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	2,53	2,505	2,5	0,1201	0,0144	0,2095	0,4278	2,94	2,26	124
80%	2,53	2,51	2,5	0,1235	0,0152	0,0516	0,2533	2,94	2,26	99
20%	2,529	2,5	2,5	0,1078	0,0116	1,2363	2,0125	2,85	2,38	25
75%	2,531	2,51	2,5	0,1184	0,014	-0,0626	-0,3223	2,85	2,26	93
25%	2,527	2,5	2,5	0,1268	0,016	0,9402	2,8021	2,94	2,27	31
70%	2,524	2,51	2,5	0,1109	0,0123	-0,2927	-0,3412	2,7	2,26	87
30%	2,543	2,5	2,5	0,14	0,0196	0,7207	0,7013	2,94	2,32	37

**Tabela 20.** Estatística Descritiva do Atributo Densidade das Partículas na Profundidade de 10 a 20 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	2,502	2,5	2,56	0,1279	0,0163	-0,4403	0,1287	2,77	2,17	109
80%	2,494	2,5	2,5	0,1327	0,0176	-0,3467	0,0284	2,77	2,17	90
20%	2,540	2,56	2,63	0,0963	0,0092	-0,6601	0,2302	2,7	2,32	19
75%	2,506	2,51	2,56	0,1315	0,0172	-0,3726	-0,1447	2,77	2,17	84
25%	2,488	2,5	2,5	0,1168	0,0136	-0,9472	1,8587	2,7	2,17	25
70%	2,505	2,51	2,56	0,1231	0,0151	-0,6277	0,5425	2,77	2,17	75
30%	2,495	2,5	2,5	0,1397	0,0195	-0,1269	-0,3388	2,77	2,21	34

**Tabela 21.** Estatística Descritiva do Atributo Densidade das Partículas na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	2,536	2,56	2,56	0,1382	0,0191	-0,0997	-0,0378	2,86	2,16	111
80%	2,533	2,555	2,56	0,1423	0,0202	-0,1785	-0,0904	2,86	2,16	88
20%	2,546	2,56	2,56	0,1238	0,0153	0,4955	0,0877	2,78	2,33	23
75%	2,548	2,56	2,56	0,1407	0,0198	-0,0117	-0,2166	2,86	2,16	83
25%	2,5	2,54	2,56	0,1261	0,0159	-0,7442	0,1455	2,7	2,22	28
70%	2,533	2,555	2,56	0,1424	0,0202	-0,1753	0,0781	2,86	2,16	80
30%	2,544	2,56	2,56	0,1288	0,0166	0,224	-0,588	2,78	2,32	31

**Tabela 22.** Estatística Descritiva do Atributo Teor de Areia na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	82,887	86	90	10,3524	107,1741	-0,9352	-0,0504	97	57	124
80%	82,737	86	90	10,4662	109,5426	-0,8915	-0,1676	97	58	99
20%	83,48	87	92	10,075	101,51	-1,1932	0,8997	97	57	25
75%	82,602	86	90	10,1104	102,2204	-0,894	-0,0934	97	58	93
25%	83,741	88	92	11,1772	124,9312	-1,1225	0,3162	96	57	31
70%	82,784	86	90	10,02	100,4011	-0,8121	-0,3432	97	58	88
30%	83,138	87	90	11,2685	126,9802	-1,2002	0,582	96	57	36

**Tabela 23.** Estatística Descritiva do Atributo Teor de Areia na Profundidade de 10 a 20 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	81,95	86	92	12,2473	149,9976	-1,2017	1,2373	97	36	122
80%	81,428	85,5	92	11,9033	141,6907	-1,0716	1,0640	97	36	98
20%	84,083	90	93	13,6251	185,6449	-1,8192	2,9377	95	42	24
75%	81,293	86	90	12,7749	163,1986	-1,1891	1,2122	95	36	92
25%	83,966	86,5	92	10,4005	108,1713	-1,0496	0,0995	97	60	30
70%	81,348	86	91	13,0029	169,0769	-1,2427	1,2477	97	36	86
30%	83,388	87,5	92	10,2435	104,9302	-0,7288	-0,7018	95	62	36

**Tabela 24.** Estatística Descritiva do Atributo Teor de Areia na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	81,55	87,5	91	12,7654	162,9555	-0,861	-0,4069	96	47	120
80%	81,072	85,5	91	12,6359	159,6683	-0,8026	-0,3533	96	47	96
20%	83,458	90	93	13,3741	178,8678	-1,2087	-0,1676	96	56	24
75%	82,111	87	91	12,1085	146,6167	-0,9279	-0,1204	96	47	90
25%	79,866	89	91	14,6563	214,8092	-0,6645	-1,0419	96	49	30
70%	80,411	85	91	13,0523	170,3641	-0,7572	-0,5445	96	47	85
30%	84,314	89	93	11,7617	138,3395	-1,1933	0,2538	96	56	35

**Tabela 25.** Estatística Descritiva do Atributo Teor de Silte na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	6,153	6	7	3,4477	11,8869	1,378	4,394	23	0	124
80%	6,202	6	5	3,5513	12,6118	1,5839	4,8942	23	0	99
20%	5,96	7	7	3,0615	9,3733	-0,0441	-0,3212	12	0	25
75%	5,989	6	7	3,61	13,0324	1,5587	5,0111	23	0	93
25%	6,645	6	5	2,9045	8,4365	0,6482	1,1181	15	1	31
70%	6,1704	6	7	3,2135	10,3269	0,8885	1,1583	17	0	88
30%	6,111	5,5	5	4,0126	16,1015	2,0377	8,1386	23	0	36

**Tabela 26.** Estatística Descritiva do Atributo Teor de Silte na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	6,639	6	3	4,976	24,7614	3,3533	18,1691	40	0	122
80%	6,7142	6	7	4,9283	24,2886	3,4548	20,6514	40	0	98
20%	6,3333	5	3	5,264	27,7101	3,2307	12,9244	28	2	24
75%	6,804	6	3	5,2586	27,6536	3,6226	19,087	40	1	92
25%	6,133	6	2	4,0234	16,1885	0,9915	0,8335	17	0	30
70%	6,755	6	5	5,3576	28,7043	3,6713	19,159	40	0	86
30%	6,361	6	3	3,9724	15,7801	0,9212	0,3313	17	1	36

**Tabela 27.** Estatística Descritiva do Atributo Teor de Silte na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	6	6	8	3,4543	11,9327	0,5921	0,5995	17	0	120
80%	6,25	6,5	8	3,6418	13,2631	0,49709	0,3739	17	0	96
20%	5	5	4	2,3774	5,6521	0,3176	-0,5541	10	1	24
75%	6,077	6	8	3,5418	12,5444	0,7894	0,8517	17	0	90
25%	5,766	6	7	3,2236	10,3919	-0,2709	-0,9289	11	0	30
70%	6,129	6	4	3,305	10,9235	0,3676	0,1751	16	0	85
30%	5,685	5	8	3,8253	14,6336	1,0482	1,6307	17	0	35

**Tabela 28.** Estatística Descritiva do Atributo Teor de Argila na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	11,016	7,5	5	8,8886	79,0078	0,9878	-0,0707	34	0	124
80%	11,121	8	4	8,8953	79,128	0,8993	-0,2288	33	0	99
20%	10,6	7	5	9,0323	81,5833	1,4238	1,0346	34	1	25
75%	11,419	8	4	8,7778	77,0504	0,9348	-0,1994	33	0	93
25%	9,806	6	2	9,2535	85,6279	1,2515	0,7041	34	0	31
70%	11,079	8	5	8,7929	77,3154	0,8869	-0,3155	32	0	88
30%	10,861	7	4	9,2432	85,4373	1,2562	0,6681	34	1	36



**Tabela 29.** Estatística Descritiva do Atributo Teor de Argila na Profundidade de 10 a 20 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	11,5	7,5	4	9,5841	91,8553	1,0521	0,0349	40	0	122
80%	11,9183	8	4	9,4843	89,952	1,0254	0,0406	40	0	98
20%	9,7916	5	3	10,0042	100,0851	1,3298	0,5770	32	0	24
75%	11,989	8	4	10,0225	100,4504	0,9909	-0,1611	40	0	92
25%	10	7	6	8,0601	64,9655	1,1705	0,499	31	0	30
70%	11,976	8	4	10,1076	102,1642	1,043	-0,0387	40	0	86
30%	10,361	7	6	8,2189	67,5515	0,9003	-0,6104	27	0	36

**Tabela 30.** Estatística Descritiva do Atributo Teor de Argila na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	12,475	7	3	10,7531	115,6296	0,9694	-0,2255	41	0	120
80%	12,7083	8	3	10,3983	108,1246	0,9166	-0,1403	41	0	96
20%	11,541	6,5	3	12,2686	150,5199	1,2252	-0,1997	37	1	24
75%	11,866	7	3	9,9867	99,7348	0,9891	-0,1197	40	0	90
25%	14,3	7	3	12,7985	163,8034	0,7991	-0,7967	41	1	30
70%	13,494	8	3	11,0082	121,1815	0,7982	-0,5536	41	0	85
30%	10	6	6	9,8189	96,4117	1,5611	1,6444	37	0	35

**Tabela 31.** Estatística Descritiva do Atributo Porosidade Total na Profundidade de 0 a 10 cm.

<b>Conjunto de dados</b>	<b>Média</b>	<b>Mediana</b>	<b>Moda</b>	<b>Desvio Padrão</b>	<b>Variância</b>	<b>Assimetria</b>	<b>Curtose</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>Número de dados</b>
100%	41,144	41	43,7	5,6619	32,0577	0,3366	0,48	59,2	28,6	123
80%	41,051	40,8	37,9	5,7968	33,6039	0,3966	0,6833	59,2	28,6	98
20%	41,512	41,2	N/D	5,193	26,9677	0,0622	-0,6845	51,9	30,8	25
75%	40,929	40,45	46,7	5,7212	32,7326	0,3312	0,5938	59,2	28,6	92
25%	41,783	41,2	N/D	5,524	30,5147	0,4168	0,3354	56,1	29,8	31
70%	41,188	40,8	43,7	5,4231	29,4102	0,1136	-0,0423	54,8	29,2	86
30%	41,043	41,1	42,8	6,2597	39,1841	0,7098	1,3563	59,2	28,6	37

**Tabela 32.** Estatística Descritiva do Atributo Porosidade Total na Profundidade de 10 a 20 cm.

<b>Conjunto de dados</b>	<b>Média</b>	<b>Mediana</b>	<b>Moda</b>	<b>Desvio Padrão</b>	<b>Variância</b>	<b>Assimetria</b>	<b>Curtose</b>	<b>Máximo</b>	<b>Mínimo</b>	<b>Número de dados</b>
100%	37,351	37,3	34,8	4,4187	19,5255	0,3185	0,6189	51,2	27,4	94
80%	37,257	37,1	34,8	4,6155	21,303	0,3922	0,6628	51,2	27,4	76
20%	37,744	38,1	40,7	3,5534	12,6273	-0,1802	-0,6813	44,3	31,3	18
75%	37,543	37,6	34,8	4,4378	19,6946	0,2924	0,6915	51,2	27,4	72
25%	36,722	35,6	35,2	4,3979	19,3418	0,4325	0,9231	47,9	27,8	22
70%	37,627	37,85	34,8	4,1827	17,4955	0,064	-0,3689	47,7	27,9	66
30%	36,7	36,2	35,6	4,9505	24,5081	0,8425	2,597	51,2	27,4	28

**Tabela 33.** Estatística Descritiva do Atributo Porosidade Total na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	36,147	36,3	36,3	5,2534	27,5984	0,0897	1,4404	55,3	23,1	90
80%	36,31	36,65	38	5,2864	27,9461	0,0572	1,9606	55,3	23,1	70
20%	35,58	36,3	36,7	5,2299	27,3522	0,2104	-0,0884	46,8	26,6	20
75%	36,125	36,3	36,7	4,601	21,1698	-0,4971	0,7329	46,8	23,1	64
25%	36,203	36,65	33,6	6,7005	44,8979	0,5746	1,326	55,3	24,8	26
70%	36,1	36,3	35,2	4,9126	24,1338	-0,5878	0,3152	46,8	23,1	63
30%	36,259	36,3	36,7	6,0748	36,9032	0,9652	2,71	55,3	25,6	27

**Tabela 34.** Estatística Descritiva do Atributo Aeração na Profundidade de 0 a 10 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	27,27	28,6	29,4	8,3026	68,9345	-0,7499	0,265	43,9	6,5	123
80%	27,3275	28,65	32,5	8,0076	64,1218	-0,7413	0,3367	43,9	6,5	98
20%	27,048	28,6	29,4	9,5495	91,1934	-0,7768	0,1319	41	6,7	25
75%	26,925	28,6	29,4	8,1353	66,1836	-0,6999	0,083	41	6,5	92
25%	28,296	29,6	28	8,8386	78,1223	-0,9784	1,1285	43,9	6,7	31
70%	27,215	29,05	31,4	8,2457	67,9925	-0,7662	0,229	41	6,5	86
30%	27,4	28,3	32,5	8,5468	73,0488	-0,7477	0,5535	43,9	6,7	37

**Tabela 35.** Estatística Descritiva do Atributo Aeração na Profundidade de 10 a 20 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	24,992	26,2	32,8	7,5633	57,2042	-0,8007	0,1324	37,1	2,6	91
80%	24,3297	25,6	32,8	7,7884	60,6604	-0,766	-0,0877	36,4	2,6	74
20%	27,8764	29,3	30,3	5,8431	34,1419	-0,3895	-0,2467	37,1	15,6	17
75%	25,0101	26,7	30,3	7,9084	62,5432	-0,8308	0,1362	37,1	2,6	69
25%	24,936	25,25	32,8	6,5296	42,6357	-0,6465	-0,1568	35,1	12,2	22
70%	24,815	26,2	30,3	7,7242	59,6632	-0,7022	0,0147	37,1	2,6	63
30%	25,389	26,4	32,8	7,3102	53,4402	-1,1054	0,8241	34,2	6,3	28

**Tabela 36.** Estatística Descritiva do Atributo Aeração na Profundidade de 20 a 30 cm.

Conjunto de dados	Média	Mediana	Moda	Desvio Padrão	Variância	Assimetria	Curtose	Máximo	Mínimo	Número de dados
100%	23,369	24,7	30,5	9,4231	88,7957	-0,4287	-0,9314	39,5	3,3	89
80%	23,4463	24,7	30,5	9,5778	91,7345	-0,458	-0,9793	37	3,3	69
20%	23,105	24,2	N/D	9,1025	82,8562	-0,3471	-0,5836	39,5	5,8	20
75%	23,012	24,65	30,5	9,8356	96,7407	-0,4275	-1,0032	39,5	3,3	64
25%	24,284	24,7	N/D	8,3921	70,4289	-0,3186	-0,9817	37	7,9	25
70%	22,969	23,35	32,1	9,6038	92,2342	-0,2265	-1,1794	39,5	5,1	62
30%	24,288	26,5	30,5	9,104	82,8841	-1,0028	0,2318	37	3,3	27

## 7.2. Script Básico Utilizado para Modelagem das Variáveis Predictoras para Uso na RK

```
#direcionando para diretório de trabalho
setwd("C:/Users/martinkihu/Desktop/regressionk")

#adquirindo os dados de arquivo txt
dados<-read.table("RattcovXX.txt",h=T)
dados
attach(dados)
summary(dados)

#organizando dados e tirando os "missing points"
library(car)
is.na(dados)
cov(dados)
cor(dados)

## critérios de seleção de modelos ##
#Stepwise - carregar a library (MASS)

library(MASS)
#stepAIC(model) é o comando básico.
#quanto menor o valor numerico do AIC, melhor será o modelo ajustado pelo stepwise
#o comando apresenta as variáveis predictoras selecionadas

#exemplo para variável CC
#cria-se o modelo de regressão com o comando abaixo, usando todas as variáveis predictoras
modelCC=lm(CapCa~Elev+DenSi+DenPa+Areia+Silte+Argil+PoTot+Aerac+aacnX+cnblX+demXX
+mrrtf+mrvbf+planX+profi+rspXX+slope+vdn1x+wiXXX)

#roda o comando de stepwise com seleção para o critério AIC
stepAIC(modelCC)

#repete-se o modelo acima usando somente as variáveis selecionadas pelo steAIC
modelCC=lm(CapCa~DenSi+DenPa+PoTot+Aerac+demXX+mrrtf+mrvbf+rspXX+slope)

#testa o modelo para multicolinearidade através do VIF
vif(modelCC)

#repete o comando de modelCC retirando as predictoras que apresentam VIF maior que #5
#retira uma variável de cada vez e vai repetindo o processo
modelCC=lm(CapCa~DenSi+DenPa+PoTot+Aerac+mrrtf+mrvbf+rspXX+slope)
vif(modelCC)
#Após corrigir a multicolinearidade, rodar o modelo novamente
#deixando somente os preditores que resultaram significancia
modelCC=lm(CapCa~DenPa+PoTot+Aerac)

#comando para análise de variância
anova(modelCC)

#depois de completada todas as etapas acima
#esse comando irá mostrar os regressores do modelo para cada variável
summary(modelCC)

#A mesma sequencia é utilizada para as demais variáveis de interesse a serem preditas
#sempre com atenção aos nomes dados aos novos modelos
#por exemplo modelADT, modelAFD, etc.
```