

**UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE CIÊNCIAS RURAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO**

**APLICAÇÃO DE COMPONENTES PRINCIPAIS E  
REGRESSÕES LOGÍSTICAS MÚLTIPLAS EM SISTEMA  
DE INFORMAÇÕES GEOGRÁFICAS PARA A PREDIÇÃO E  
O MAPEAMENTO DIGITAL DE SOLOS**

**DISSERTAÇÃO DE MESTRADO**

**Alexandre ten Caten**

**Santa Maria, RS, Brasil**

**2008**

**APLICAÇÃO DE COMPONENTES PRINCIPAIS E  
REGRESSÕES LOGÍSTICAS MÚLTIPLAS EM SISTEMA  
DE INFORMAÇÕES GEOGRÁFICAS PARA A PREDIÇÃO E  
O MAPEAMENTO DIGITAL DE SOLOS**

**por**

**Alexandre ten Caten**

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Ciência do Solo, Área de Concentração em Processos Físicos e Morfogenéticos do Solo, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de

**Mestre em Ciência do Solo**

Orientador: Prof. Dr. Ricardo Simão Diniz Dalmolin

**Santa Maria, RS, Brasil.**

**2008**

Caten, Alexandre ten, 1976-

C359a

Aplicação de componentes principais e regressões logísticas múltiplas em sistema de informações geográficas para a predição e o mapeamento digital de solos / por Alexandre ten Caten ; orientador Ricardo Simão Diniz Dalmolin. – Santa Maria, 2008.

128 f. ; il.

Dissertação (mestrado) – Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós-Graduação em Ciência do Solo, RS, 2008.

1. Ciência do solo 2. Pedometria 3. Mapeamento preditivo 4. Modelos logísticos politômicos 5. Índice de umidade topográfica 6. Análise estatística multivariada I. Dalmolin, Ricardo Simão Diniz, orient. II. Título

CDU: 631.4:528.7/.9

Ficha catalográfica elaborada por  
Luiz Marchiotti Fernandes – CRB 10/1160  
Biblioteca Setorial do Centro de Ciências Rurais/UFSM

**Universidade Federal de Santa Maria  
Centro de Ciências Rurais  
Programa de Pós-Graduação em Ciência do Solo**

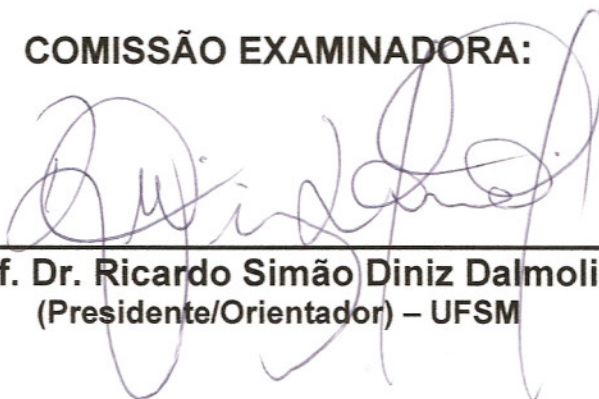
A Comissão Examinadora, abaixo assinada,  
aprova a dissertação de Mestrado

**APLICAÇÃO DE COMPONENTES PRINCIPAIS E  
REGRESSÕES LOGÍSTICAS MÚLTIPLAS EM SISTEMA  
DE INFORMAÇÕES GEOGRÁFICAS PARA A PREDIÇÃO  
E O MAPEAMENTO DIGITAL DE SOLOS**

elaborada por  
**Alexandre ten Caten**

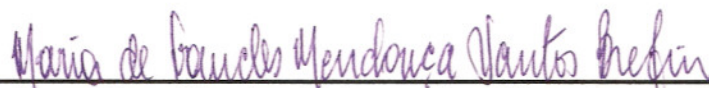
como requisito parcial para obtenção do grau de  
**Mestre em Ciência do Solo**

**COMISSÃO EXAMINADORA:**



---

**Prof. Dr. Ricardo Simão Diniz Dalmolin**  
(Presidente/Orientador) – UFSM



---

**Pesquisadora Dra. Maria de Lourdes Mendonça Santos Brefin**  
EMBRAPA Solos



---

**Prof. Dr. Fabrício de Araújo Pedron**  
UFSM

Santa Maria, outubro de 2008



*A meu pai Marino, minha mãe Vera,  
minha mana Elaine,  
meus irmãos Roberto e Felipe,  
minha querida Leosane,  
dedico.*

## AGRADECIMENTOS

A Universidade Federal de Santa Maria por todo o conhecimento a mim repassado e as oportunidades de crescimento pessoal em todas as etapas de minha vida acadêmica na instituição.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão de bolsa de estudos e disponibilização do portal de periódicos.

Ao Programa de Pós-Graduação em Ciência do Solo da Universidade Federal de Santa Maria pela oportunidade de cursar o mestrado.

Ao professor Ricardo Simão Diniz Dalmolin pela aceitação, amizade e orientação durante o mestrado.

Aos demais membros de meu comitê de orientação Rudiney Soares Pereira e Elódio Sebem pelas conversas e conhecimento, especialmente no assunto geoinformação.

Aos membros da banca examinadora da defesa de dissertação de mestrado Maria de Lourdes Mendonça Santos Brefin e Fabrício de Araújo Pedron pela disponibilidade e contribuições para a melhoria do trabalho.

Aos professores do PPGCS pelo conhecimento e ao funcionário Tarcisio Durgante Uberti pela prestatividade.

Aos professores do departamento de estatística Ivanor Müller, Luis Felipe Dias Lopes e Andreia Zanella pela ajuda no assunto estatística.

Ao estudante do curso Técnico em Geoprocessamento Luis Fernando Chimelo Ruiz pela valorosa ajuda com a produção dos dados.

Ao bibliotecário Luiz Marchiotti Fernandes pelo esforço para aquisição dos livros adequados.

Ao professor Fabrício de Araújo Pedron e o colega de pós-graduação Fabio Pacheco pelo auxílio na coleta de dados a campo.

Aos colegas do setor de Pedologia e do PPGCS pelas conversas e confraternizações;

Ao Colégio Politécnico da Universidade Federal de Santa Maria por possibilitar o uso dos programas.

A minha querida Leosane pelas conversas *científicas*, gargalhadas e carinho.

Obrigado.

*"O Criador modelou o homem com a **argila do solo**,  
soprou-lhe nas narinas um sopro de vida,  
e o homem tornou-se um ser vivente"*

Gênesis 1:7.

## **RESUMO**

Dissertação de Mestrado  
Programa de Pós-Graduação em Ciência do Solo  
Universidade Federal de Santa Maria, RS, Brasil

# **APLICAÇÃO DE COMPONENTES PRINCIPAIS E REGRESSÕES LOGÍSTICAS MÚLTIPLAS EM SISTEMA DE INFORMAÇÕES GEOGRÁFICAS PARA A PREDIÇÃO E O MAPEAMENTO DIGITAL DE SOLOS**

Autor: Alexandre ten Caten  
Orientador: Ricardo Simão Diniz Dalmolin  
Local e data da defesa: Santa Maria, 31 de outubro de 2008.

As demandas da sociedade pela informação solo têm crescido, porém levantamentos pedológicos praticamente não ocorrem mais no país. Técnicas de Mapeamento Digital do Solo podem ser empregadas para inferir a distribuição espacial de classes de solos a partir de mapas existentes e áreas de referência, extrapolando esta informação para áreas não mapeadas. O objetivo deste estudo foi empregar em um Sistema de Informações Geográficas as Regressões Logísticas Múltiplas (RLM) utilizando-se de Componentes Principais (CP) como variáveis explicativas para a predição espacial de classes de solos. A área de estudo foi na região do município de São Pedro do Sul / RS. Para o desenvolvimento dos modelos preditivos foram utilizados um conjunto de nove atributos do terreno. O treinamento dos modelos foi executado em um mapa de solos existente, e em um levantamento realizado em áreas de referência, ambos na escala 1:50.000. As três primeiras CP retidas explicaram 65,57% da variabilidade dos dados. Os modelos preditivos que empregaram CP obtiveram menores valores do índice kappa. O mapa predito mais acurado empregou os nove atributos do terreno e alcançou um valor de kappa de 63,20%. A acurácia do mapeamento é sensível a semelhança entre as classes mapeadas, e o mapeamento em níveis categóricos mais homogêneos reduz a precisão dos mapas preditos. Classes de solos relativamente pouco representativas não são corretamente espacializadas. O emprego de RLM permite espacializar classes de solos para áreas não mapeadas, embora o emprego de CP necessite ser testado com um maior número de covariáveis.

Palavras chaves: pedometria, mapeamento preditivo, análise estatística multivariada, modelos logísticos politômicos, índice de umidade topográfica.

## **ABSTRACT**

Dissertação de Mestrado  
Programa de Pós-Graduação em Ciência do Solo  
Universidade Federal de Santa Maria, RS, Brasil

### **APPLICATION OF PRINCIPAL COMPONENTS AND MULTIPLE LOGISTIC REGRESSION IN A GEOGRAPHICAL INFORMATION SYSTEM FOR PREDICTION AND DIGITAL SOIL MAPPING**

Author: Alexandre ten Caten

Adviser: Ricardo Simão Diniz Dalmolin

Place and date of the defense: Santa Maria, October 31, 2008.

Social demands on soil information have grown dramatically, meanwhile the soil surveys are seldom carried out in the country. Digital soil mapping techniques can be applied to infer the spatial distribution of soil from existing soil maps or from reference areas, extrapolating this information to areas not mapped. The purpose of this study was to apply in a Geographic Information System the Multiple Logistic Regressions (MLR) using Principal Components (PC) as explanatory variables to predict soil classes spatial distribution. The study area was the region of municipality São Pedro do Sul / RS. For the development of predictive models a set of nine terrain attributes were used. Model training was executed on an existing soil map and with a survey carried out in a reference area, both in a 1:50.000 scale. The first three retained PC explained 65.57% of the data variability. The predictive models which used PC had lower values of kappa index. The most accurate predicted map reached a kappa value of 63.20% and was generated by using the nine attributes of land as predictive covariates. The mapping accuracy is sensitive to similarities between the mapped classes, and mapping in a more homogeneous categorical level reduces the accuracy of the predicted maps. Soil classes relatively not representative in the training maps are not properly spatialized. The use of MLR allows spatializing of soil classes to areas not mapped, although the use of PC needs to be tested with a larger number of covariates.

Key words: pedometry, predictive mapping, statistical multivariate analysis, polytomous logistic regression, topographic wetness index.

## LISTAS DE ABREVIATURAS

<b>AC</b>	Área de Contribuição
<b>ACE</b>	Área de Contribuição Específica
<b>ACP</b>	Análise de Componentes Principais
<b>AG</b>	Acurácia Geral
<b>AM</b>	Acurácia do Mapeador
<b>AU</b>	Acurácia do Usuário
<b>C</b>	Cambissolo
<b>CP</b>	Componentes Principais
<b>CP1</b>	Primeira Componente Principal
<b>CP2</b>	Segunda Componente Principal
<b>CP3</b>	Terceira Componente Principal
<b>CPLN</b>	Curvatura Planar
<b>CPRF</b>	Curvatura de Perfil
<b>CTS</b>	Capacidade de Transporte de Sedimento
<b>CX</b>	Cambissolo Háplico
<b>CX_NV</b>	Associação Cambissolo Háplico e Nitossolo Vermelho
<b>DECL</b>	Declividade
<b>DIST</b>	Distância à Drenagem
<b>ELEV</b>	Elevação
<b>F</b>	Plintossolo
<b>FT</b>	Plintossolo Argilúvico
<b>GPS</b>	Sistema de Posicionamento Global
<b>ILUM</b>	Iluminação
<b>IUT</b>	Índice de Umidade Topográfica
<b>K</b>	Kappa

<b>KMO</b>	<i>Kaiser Meyer Olkin</i>
<b>MDE</b>	Modelo Digital de Elevação
<b>MDS</b>	Mapeamento Digital de Solos
<b>MSA</b>	<i>Measure of Sample Adequacy</i>
<b>P</b>	Argissolo
<b>PAC</b>	Argissolo Acinzentado
<b>PBAC</b>	Argissolo Bruno-Acinzentado
<b>PI</b>	Plano de Informação
<b>PV</b>	Argissolo Vermelho
<b>PVA</b>	Argissolo Vermelho-Amarelo
<b>R</b>	Neossolo
<b>RL</b>	Neossolo Litólico
<b>RLB</b>	Regressão Logística Binomial
<b>RLM</b>	Regressão Logística Múltipla
<b>SH</b>	Solos Hidromórficos
<b>SIG</b>	Sistema de Informações Geográficas
<b>SRTM</b>	<i>Shuttle Radar Topography Mission</i>
<b>TAPES</b>	<i>Terrain Analysis Programs for the Environmental Sciences</i>
<b>1°NC</b>	1°Nível Categórico
<b>2°NC</b>	2°Nível Categórico
<b>WRB</b>	<i>World Reference Base</i>

# SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	12
<b>2 REVISÃO BIBLIOGRÁFICA</b> .....	15
<b>2.1 Abordagem quantitativa da distribuição espacial de solos</b> .....	15
2.1.1 Importância da informação solo .....	15
2.1.2 O mapa de solos .....	16
2.1.3 Mapeamento convencional de solos .....	17
2.1.4 A abordagem quantitativa.....	19
2.1.5 Tecnologias envolvidas .....	21
2.1.6 Estudos em pedometria realizados no país.....	23
2.1.7 Futuros estudos e a importância dos dados de campo .....	24
<b>2.2 Análise de componentes principais</b> .....	26
<b>2.3 Atributos do terreno como preditores</b> .....	30
<b>2.4 Áreas de referência</b> .....	31
<b>2.5 Regressões logísticas múltiplas</b> .....	32
<b>2.6 Avaliação dos mapas gerados</b> .....	34
<b>3 MÉTODOS E TÉCNICAS</b> .....	37
<b>3.1 Caracterização da área de estudo</b> .....	37
<b>3.2 Solos do município</b> .....	38
<b>3.3 Levantamento de solos nas áreas de referência</b> .....	39
<b>3.4 Atributos do terreno</b> .....	40
<b>3.5 Geração dos mapas de solos</b> .....	43
3.5.1 Mapa predito a partir do mapa de solos Klamt.....	44
3.5.2 Mapa predito a partir do mapa de solos das áreas de referência.....	46
<b>4. RESULTADOS E DISCUSSÃO</b> .....	50
<b>4.1 Atributos do terreno</b> .....	50
<b>4.2 Análise de Componentes Principais</b> .....	61
<b>4.3 Mapeamento pedométrico a partir do mapa Klamt</b> .....	69
4.3.1 Mapa predito ao nível de ordem.....	69
4.3.1.1 Componentes principais como variáveis explicativas.....	69



4.3.1.2 Atributos do terreno como variáveis explicativas.....	80
4.3.2 Mapa predito ao nível de subordem .....	87
4.3.2.1 Componentes principais como variáveis explicativas.....	87
4.3.2.2 Atributos do terreno como variáveis explicativas.....	93
<b>4.4 Mapeamento pedométrico a partir do mapa referência. ....</b>	<b>97</b>
4.4.1 Mapa predito na área de referência com CP como variáveis explicativas.....	97
4.4.2 Mapa predito com área de referência e atributos do terreno como variáveis explicativas.....	102
<b>4.5 Discussão geral do estudo .....</b>	<b>107</b>
<b>4.6 Considerações finais .....</b>	<b>110</b>
<b>5 CONCLUSÕES .....</b>	<b>112</b>
<b>6 PERSPECTIVAS.....</b>	<b>113</b>
<b>7 REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>114</b>
<b>APÊNDICES .....</b>	<b>121</b>

# 1 INTRODUÇÃO

Levantamentos pedológicos não estão mais ocorrendo com a mesma frequência com que já foram executados no passado, ou que deveriam continuar acontecendo, haja vista sua importância para a formação de novos pedólogos (HUDSON, 1992), e para o conhecimento de relações ainda não descritas entre o solo e a paisagem. A diminuição dos levantamentos pedológicos está ligada aos financiamentos cada vez mais restritos para ciência básica. Descobertas em ciência de base só serão feitas se atreladas a projetos de pesquisas maiores, para os quais a sociedade e os gestores públicos vêem aplicações práticas (BOUMA, 2005). Este é o caminho que deverá tomar o pedólogo que deseja realizar mapeamento de solos.

Projetos em ciência do solo estão voltados para questões como: o uso adequado do solo para a disponibilização de novos recursos energéticos, mensuração e identificação de áreas propensas ao pagamento de serviços ambientais, melhoria e criação de zoneamentos edafoclimáticos, quantificação da contribuição do solo como dreno para o carbono liberado pelas ações antrópicas, e muitos outros títulos poderiam ainda ser citados como justificativa por um melhor conhecimento da distribuição espacial do solo. Os levantamentos pedológicos precisam estar associados a estes projetos como necessidade e oportunidade. Necessidade, pois é uma informação fundamental. Oportunidade, pois os financiamentos para pesquisa estão voltados para questões ditas “maiores”, que envolvam distintas áreas em torno de uma “necessidade palpável” da sociedade.

Um importante papel em informar os gestores públicos da importância de continuar investindo na pesquisa básica e aplicada, em ciência do solo, pode ser cumprido pelo Sistema de Informações Geográficas (SIG). Esta ferramenta possibilita demonstrar os resultados da simulação de cenários de ações antrópicas sobre o ambiente, sendo uma ferramenta importante para a sensibilização da população ou dos tomadores de decisão quanto a ações a serem tomadas, ou, quanto aos efeitos da alocação de recursos e implementação de determinada regulamentação (BOUMA, 2005). É no SIG que se estrutura o Mapeamento Digital de Solos (MDS) definido como *criação e população de sistemas espaciais de informação sobre o solo através do uso de métodos de observação em campo e*

*laboratório aliados a sistemas espaciais e não espaciais de inferência em solos* (LAGACHERIE; MCBRATNEY, 2007).

A disponibilização de tecnologias as quais coletam dados sobre os fatores e processos de formação do solo, aliada ao emprego de técnicas matemáticas e estatísticas, possibilita que o MDS seja empregado para atender esta demanda crescente por informação espacial em solos. Levantamentos de solos através do MDS devem estar associados a projetos maiores, voltados à ciência aplicada, gerando conhecimento básico em pedologia, ao mesmo tempo em que gera as informações necessárias sobre a distribuição espacial de classes, características ou propriedades do solo para as outras fases do projeto.

Neste aspecto a Análise de Componentes Principais (ACP) pode ser empregada para estudar as relações entre os fatores de formação do solo, além de possibilitar a simplificação através da redução do número de variáveis que precisam ser manipuladas. As Regressões Logísticas Múltiplas (RLM) irão gerar a probabilidade de se encontrar cada uma das classes de solos presentes em uma região de referência, a qual será mapeada para servir de área de treinamento para os modelos preditivos.

A justificativa para esse estudo pauta-se na necessidade de se desenvolverem novas metodologias para os levantamentos de solos. Empregando tecnologias que possibilitem a geração das informações necessárias dentro da melhor relação custo benefício, gerando mapas de solos com qualidade associada e compatível com um grande número de aplicações, além de permitir que os mesmos estejam prontamente disponíveis para aplicações de modelagem e simulação de cenários em um SIG.

A hipótese deste estudo é que as Componentes Principais (CP) podem substituir os atributos do terreno em RLM empregadas para descrever matematicamente a relação entre preditores e as classes de solos.

Os objetivos do estudo foram:

Geral: Testar modelos para a espacialização de classes de solos a partir das relações entre solos e atributos do terreno descritas em mapas existentes e levantamentos em áreas de referência.

Específicos:

a) Gerar atributos do terreno os quais tenham relações com os fluxos, principalmente da água na superfície e que sejam fatores de formação do solo.

b) Aplicar ACP ao conjunto de atributos do terreno como ferramenta para a análise das correlações entre os atributos.

c) Aplicar ACP ao conjunto de atributos do terreno para a diminuição do número de covariáveis a serem utilizadas, analisando os efeitos da perda de variabilidade.

e) Utilizar RLM para estabelecer relações matemáticas entre as componentes principais e as classes de solos em diferentes níveis categóricos.

f) Utilizar RLM para estabelecer relações matemáticas entre os atributos do terreno e as classes de solos em diferentes níveis categóricos.

g) Empregar um mapa de solos existente para prever as classes de solos na mesma região e em regiões vizinhas.

h) Gerar um mapa de solos para uma área de referência e extrapolar esta informação para as áreas vizinhas não mapeadas.

## **2 REVISÃO BIBLIOGRÁFICA**

### **2.1 Abordagem quantitativa da distribuição espacial de solos**

#### **2.1.1 Importância da informação solo**

Há cerca de dez mil anos, desde quando o homem aprendeu a cultivar a terra e lhe dar de volta os nutrientes exportados com as colheitas, o conhecimento em solo não parou de se desenvolver e de ter sua importância cada vez mais valorizada para a humanidade (MCNEILL; WINIWARTER, 2004). Várias civilizações tiveram sua prosperidade e declínio relacionados aos eventos ligados ao mau uso do solo (DAVIDSON, 2007). É preciso que experiências do passado sejam convertidas em conhecimento para tratar das questões modernas envolvendo o solo, como a produção de energia e os problemas ambientais (YAALON, 2000). Neste aspecto, a pedologia tem um papel muito importante no estudo da sua distribuição espacial e seus fatores de formação.

As demandas da sociedade pela informação sobre o recurso solo têm se modificado, fazendo emergir novas áreas de interesse como, qualidade do solo e do ambiente, reconhecimento de processos de degradação e arenização e enfoques em poluição do solo. Estes enfoques foram adicionados aos convencionais interesses em fertilidade e manejo do solo para uso agrícola, aumentando os bancos de dados, bem como, atraindo novos usuários para a informação solo. De tal forma que, as propriedades do solo que necessitam serem preditas ou mapeadas não são mais apenas aquelas selecionadas pelo pedólogo, mas sim aquelas demandadas por outros cientistas que necessitam suprir seus modelos (MERMUT; ESWARAN, 2001). Com o aumento das preocupações ambientais, os mapeamentos de solos tem tido um foco em modelagem quantitativa com o acompanhamento de questões envolvendo acurácia e incertezas (MCBRATNEY et al., 2000).

Contudo, o que é de consenso entre os cientistas do solo é a importância do pedólogo e dos levantamentos de campo. Conforme Basher (1997) o pedólogo é o ator principal dos levantamentos de solos sendo o profissional que detém o conhecimento do ambiente solo, e aquele que irá tomar a decisão das técnicas a serem empregadas para a execução de novos levantamentos. Um cientista do solo experiente pode delinear acuradamente a distribuição espacial do solo na paisagem apenas observando sua fração superior (HUDSON, 1992). Em modelos

automatizados de mapeamento, a decisão de quais variáveis ambientais a serem utilizadas são escolhidas a partir do conhecimento de um pedólogo experiente da região (ZHU et al., 1997).

### 2.1.2 O mapa de solos

Um mapa de solos é um documento bidimensional apresentado em papel ou outras formas, e representa uma simplificação da organização espacial do solo nos ambientes naturais. O mapa disponível aos usuários é o resultado de quatro fatores (Figura 1).

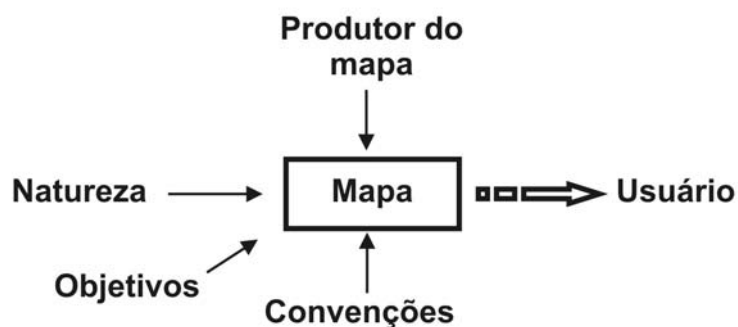


Figura 1 - O mapa é produto da interação de quatro fatores. Fonte: Adaptado de (LEGROS, 2005).

O mapa é uma transposição das entidades naturais relacionadas à geografia dos solos, sendo este um modelo reduzido, com simplificações de acordo com os objetivos. O produtor do mapa (pedólogo) através de seus conhecimentos e experiência utiliza-se das convenções para a representação cartográfica do solo, sendo então o mapa parte científico, cultural, belo, estético, e até mesmo emocional (LEGROS, 2005). Para o autor o mapa de solos é resultado de um procedimento verdadeiramente científico, resultado de um grande número de observações em campo. Além do que, o mapeamento inicia-se no campo e deve retornar a ele na fase de validação e comprovação das hipóteses a respeito da relação solo-paisagem.

O objetivo do levantamento de solos não é o de gerar somente um mapa, pois o mapa não é um fim. O objetivo prático de um levantamento de solos é permitir que inferências mais numerosas, úteis e acuradas para determinado propósito sejam

feitas, e que de outra forma não seriam possíveis (DENT; YOUNG, 1981). O levantamento de solos deve levar em consideração as necessidades do usuário dos produtos a serem gerados. Tão importante quanto a acurácia das propriedades do solo representadas no mapa, é o atendimento às demandas de informações sobre o solo, para que seu uso se dê baseado em decisões acertadas quanto ao gerenciamento dos recursos naturais.

Benefícios gerados pelo uso de informações constantes dos levantamentos de solos, são muito maiores que os custos associados à execução dos levantamentos (GIASSON et al., 2006). Já no primeiro ano após o levantamento, pode haver um aumento de rendimento, pelo uso adequado das terras. Uma analogia inversa poderia ser feita analisado os prejuízos causados pelo desconhecimento e mau uso das terras (erosão, degradação, deslizamentos, etc.). Infelizmente, esta última situação é mais comum do que a primeira.

### 2.1.3 Mapeamento convencional de solos

O mapeamento convencional dos solos é, muitas vezes, a única informação disponível referente ao inventário do solo de uma região (SCULL et al., 2003), sendo este executado em três etapas: (1) observação direta de dados existentes como fotografias aéreas, geologia e vegetação, bem como das características dos perfis de solo; (2) as observações dos atributos do solo são incorporadas em um modelo conceitual implícito o qual é utilizado para prever a variação espacial do solo; e (3) por último, está envolvida a aplicação do modelo conceitual à área levantada para prever a variação do solo em locais não amostrados. O modelo conceitual de variação do solo é então transformado em um modelo cartográfico, o mapa cloroplético (mapa temático em cores), através do desenho dos polígonos uniformes sobre fotografias aéreas.

Através do modelo conceitual discreto os solos no campo são representados pelo delineamento de polígonos, com cada polígono descrevendo a extensão espacial de uma dada classe de solo ou grupo de classes (ZHU, 1997). Esta técnica impõe limitações quanto ao tamanho do “corpo solo” o qual poderá ser delineado como um polígono neste mapa de papel, e quanto a capacidade destes polígonos em representar apenas a distribuição de um conjunto de classes de solos descritas (ZHU et al., 2001). Observações pontuais do solo são extrapoladas para regiões

maiores utilizando-se modelos mentais complexos e qualitativos da relação entre o solo e elementos mais prontamente observáveis da paisagem (MCKENZIE et al., 2000).

Do ponto de vista dos profissionais envolvidos, o mapeamento convencional é uma atividade solitária, executada por grupos de trabalhos menores, com habilidades similares envolvidos em todos os estágios do mapeamento (LAGACHERIE; MCBRATNEY, 2007). Para estes autores o desenvolvimento de mapas detalhados (maior do que 1:20.000) por meio das metodologias convencionais é lento e caro. Esta é a razão pela qual, segundo McBratney et al. (2003), grandes áreas do planeta permaneçam não mapeadas ou com levantamentos de solos apenas em escala pequena.

Os levantamentos de solos estão baseados na aplicação rigorosa de critérios diagnósticos não ambíguos (pelo menos idealmente), tendo sempre envolvido uma abordagem qualitativa quanto à importância da paisagem no levantamento. As bases científicas para o mapeamento de solos sempre foram difíceis de serem estabelecidas e reproduzidas (PENNOCK; VELDKAMP, 2006) sendo os modelos aplicados nos levantamentos vinculados ao conhecimento do pedólogo e à experiência adquirida (HUDSON, 1992; BUI, 2004). Parte da falta de bases quantitativas para pedologia durante o século 20 se deve às limitações das técnicas estatísticas utilizadas. Sendo as análises de variância e pesquisas baseadas em análises de tratamento, clássicas na abordagens agronômicas, excludentes quanto a seu uso em pedologia (PENNOCK; VELDKAMP, 2006).

Quanto à execução de levantamentos de solos pelas metodologias convencionais, as situações parecem ser bastante distintas em várias partes do planeta. Dobos et al. (2006) afirmam que novos levantamentos de solos muito provavelmente não irão ocorrer no território europeu. Sendo que os Estados Unidos está próximo de completar a cobertura de todo o país com levantamentos de solos. Nas áreas ainda não mapeadas os planejadores americanos aguardam ansiosos a disponibilização dos mapas de polígonos de solos (HOWELL et al., 2008). No caso do Brasil, a situação é uma mistura de nostalgia para com os levantamentos do passado (OLIVEIRA, 2007), constatação de que não existem investimentos em levantamentos de escalas maiores (MENDONÇA-SANTOS; SANTOS, 2007) e de que não estão sendo formados novos pedólogos para executar esta tarefa (DALMOLIN, 1999).



#### 2.1.4 A abordagem quantitativa

Historicamente os levantamentos quantitativos do solo vieram a ocorrer por uma demanda fora da área agrônômica. Primeiro, na década de 50, engenheiros civis e militares sentiram a necessidade de prever informações que não existiam sobre o solo. Em seguida nos anos 60, o uso de computadores possibilitou que um volume maior de dados fosse processado, permitindo que os pesquisadores utilizassem técnicas multivariadas para tratarem seus dados. Outra técnica surgida nesta época, muito importante para ciência do solo, foi a geoestatística (WEBSTER, 1994).

No entanto, anos antes, a publicação da obra “Fatores de Formação do Solo” (JENNY, 1941) trouxe uma das questões mais empolgantes para os cientistas do solo. Jenny equacionou o solo como uma função de fatores de formação, sendo estes: clima (cl), organismos (o), relevo (r), material de origem (p), tempo (t) e fatores não determinados (...). Este modelo ficou conhecido como CLORPT:

$$S = f(\text{cl, o, r, p, t, ...})$$

Este princípio-chave para os levantamentos foi estabelecido pelos cientistas russos contemporâneos a Vasily V. Dokuchaev, que verificaram que os perfis do solo tinham uma relação funcional para com os ambientes onde se localizavam estando em equilíbrio (MCKENZIE et al., 2000). A partir da publicação de Jenny muitos pedólogos passaram a propor funções, na sua maioria lineares, demonstrando seus resultados graficamente, buscando associar propriedades do solo aos fatores de formação.

Com a informatização e a aplicação de modelos matemáticos e estatísticos mais complexos consolidou-se a pedometria, que tem origem no neologismo derivado do Grego “pedos” (solo) e “metron” (medida, mensuração), ou seja, o estudo do solo (pedologia) através da aplicação de métodos quantitativos (matemáticos e estatísticos) (WEBSTER, 1994). A pedometria reúne muitos campos científicos do conhecimento, desde a geoestatística até a microbiologia do solo (HENGL, 2003). O domínio da pedometria pode ser melhor definido pela interdisciplinaridade, envolvendo ciência do solo, estatística-matemática aplicada e ciência da geoinformação (Figura 2). A pedometria esta relacionada à abordagem quantitativa, a qualidade da informação solo com respeito a variabilidade, incerteza e confiabilidade (BURROUGHS et al., 1994).



**Figura 2 - Pedometria como uma ciência interdisciplinar onde ciência do solo, estatística aplicada e geoinformação se interceptam. Fonte: Adaptado de (HENGL, 2003).**

Como a base teórica da pedometria está relacionada aos fatores de estado de Jenny, os problemas relacionados a ela estavam ligados a solução matemática da proposição de Jenny, parecendo esta insolúvel e sem dados espacialmente distribuídos a respeito dos fatores de formação. Avanços em teorias matemáticas (lógica nebulosa e redes neurais) e métodos estatísticos, bem como, sistemas remotos de aquisição de dados e informática, têm possibilitado um crescimento dos estudos envolvidos em pedometria (BUI, 2007). Ao ponto de McBratney et al. (2003) proporem uma nova abordagem para a equação de Jenny. Segundo os autores classes e atributos do solo podem ser preditos utilizando-se:

$$Sc = f(s, c, o, r, p, a, n)$$

ou

$$Sa = f(s, c, o, r, p, a, n)$$

Onde, Sc = classe do solo, Sa = atributo do solo, s = próprio solo, c = clima, o = organismos, r = relevo, p = material de origem, a = tempo, n = localização geográfica. Segundo os autores após a equação "f" ser ajustada em um conjunto de dados, esta poderá ser usada em outros locais onde somente as variáveis ambientais são conhecidas. O solo pode ser predito a partir de uma série de

técnicas: a partir de outros atributos do solo na mesma localização, a partir de si mesmo, de atributos do solo e de atributos ambientais em posições vizinhas.

Em pedometria este modelo ficou conhecido como SCORPAN e pode ser visto como uma evolução do modelo CLORPT de Jenny, uma vez que este último não considera que o solo pode ser predito a partir dele próprio como em equações de pedotransferência, nem a partir de coordenadas geográficas. No entanto, a abordagem SCORPAN é uma síntese das tecnologias utilizadas atualmente para a inferência espacial do solo. Na medida em que novas tecnologias de coleta de dados virão a surgir, também novos parâmetros poderão ser incorporados a este modelo. O que neste aspecto, torna o modelo SCORPAN similar ao CLORPT que contava com os fatores desconhecidos representados por “...” (reticências), uma vez que Jenny não desconsiderava que existissem outros preditores para a formação do solo.

No entanto, para Hengl (2003), existem dois grandes desafios para a pedometria ser mais aplicada. Primeiro não existe ainda um conhecimento sistematizado sobre a variabilidade espacial do solo. Segundo, a pedometria ainda não é adequada para mapear solos com características específicas como horizontes revolvidos ou transições horizontais abruptas. Estas condições são mais facilmente mapeadas usando o modelo mental e fotointerpretação do que a geoestatística e variáveis auxiliares.

Contudo, os mapas digitais poderão não ser mais acurados do que os convencionais, devido a dois fatores: variação espacial do solo e incerteza nos dados ambientais (preditores). No entanto, esta abordagem é potencialmente mais barata que os levantamentos convencionais, além de tratar a informação de maneira quantitativa e, portanto, permitir a associação dos erros envolvidos. Há a necessidade de mais pesquisas para planejamento da amostragem, estimativa de qualidade dos mapas e treinamento de pessoal para trabalhar com esta nova forma de mapeamento do solo (MCBRATNEY et al., 2003).

#### 2.1.5 Tecnologias envolvidas

O SIG é uma ferramenta fundamental para o pedólogo, é para onde os dados convergem e as informações são geradas dentro do MDS (BUI, 2007). O papel do SIG na geração de mapas digitais de propriedades e classes de solos é muito

importante, possibilitando a geração de informação mesmo com as limitações impostas pelos elevados custos dos levantamentos de campo e análises de laboratório (MCBRATNEY et al., 2003). A possibilidade de se poder manipular os dados no ambiente SIG em formato digital, permite que através da modelagem utilizando-se de arquivos matriciais (raster), uma área possa ser representada por muitos quadros em uma matriz (pixels). Com a representação raster as generalizações serão reduzidas ao tamanho da resolução espacial, podendo ser esta muito pequena, o que torna o mapeamento mais sensível as variações do ambiente (ZHU et al., 2001). Da mesma forma, esta abordagem raster baseada em uma estrutura de pixels permite a análise do terreno por métodos baseados em grade, como a geração de índices topográficos que se destinam a descrever quantitativamente o relevo (MOORE et al., 1993).

Estes avanços tecnológicos têm permitido aos pedólogos integrar conhecimentos de outras disciplinas à ciência do solo. Técnicas de Sensoriamento Remoto e Fotogrametria Digital disponibilizam dados espaciais digitais da representação da superfície terrestre, os quais podem ser combinados com mapas analógicos digitalizados em um SIG, permitindo uma caracterização e análise eficiente de um grande volume de dados. Embora a digitalização dos mapas analógicos para sua incorporação em pesquisas de modelagem ambiental implantadas em um SIG, passa pelos mesmos problemas que os mapas cloropléticos originais, os quais são assumidos como unidades homogêneas com variabilidade interna desconhecida e bordas entre classes definidas de maneira truncada.

Uma técnica promissora para coleta de dados é a técnica de Sensoriamento Remoto por Radar. O *Shuttle Radar Topography Mission* (SRTM) pela primeira vez disponibilizou um Modelo Digital de Elevação (MDE) no nível de resolução de 30 e 90 metros. O MDE/SRTM cobre o planeta entre as latitudes de 60°N e 57°S tendo sido adquirido com o mesmo sensor em uma única missão e produzido através de uma única técnica, a interferometria por radar de abertura sintética (RABUS et al., 2003). Para Grohmann et al. (2007) a interpolação dos dados SRTM com resolução de 90 para 50 metros, produz um modelo digital de elevação com qualidade comparáveis as tradicionais cartas topográficas na escala 1:50.000.

Outro conjunto de dados com grande potencial para estudos de inferência espacial em solos é o banco de dados da série LANDSAT do Serviço Geológico dos

Estados Unidos (USGS, 2008), recentemente disponibilizados gratuitamente na Internet. Extremamente úteis para gerar informações como mapas temáticos, índices de vegetação e geologia.

O desenvolvimento da análise digital do terreno e tecnologias auxiliares tem criado uma oportunidade para novos métodos científicos serem aplicados nos levantamentos de solos, os quais poderão superar as limitações dos levantamentos convencionais. A análise do terreno tem o potencial de melhorar os levantamentos pedológicos em três áreas. Podendo ser utilizado para: gerar informações ambientais em alta resolução para direcionar o uso da terra (radiação, declividade, etc.), criar estratificações ambientais explícitas, e fornecer preditores espaciais quantitativos para propriedades do solo específicas (MCKENZIE et al., 2000). Programas para a geração e espacialização de preditores têm sido desenvolvidos em estudos relacionados à análise do terreno, um exemplo é o *Terrain Analysis Programs for the Environmental Sciences* (TAPES) (WILSON; GALLANT, 2000). Este software tem sido aplicado com sucesso em mapeamento de solos, os quais relacionam a distribuição espacial do solo a parâmetros do terreno (MCKENZIE et al., 2000).

#### 2.1.6 Estudos em pedometria realizados no país

Utilizando-se redes neurais artificiais, Chagas (2006) afirma que os mapas produzidos empregando-se esta técnica contêm muito mais detalhes espaciais do que o mapa de solos convencional, pelo fato de que a variação ambiental pôde ser quantificada com grande detalhe utilizando-se um software de SIG, que tem a capacidade de manipular muito mais variáveis simultaneamente. A presença de mais de uma classe de solo associada a um determinado pixel somente ocorreu devido à dificuldade ou impossibilidade de estabelecimento de relações solo-paisagem mais precisas. Os mapas inferidos tiveram uma baixa concordância com o mapa de solos convencional. As principais causas da discordância foram: limitações de escala do mapa convencional e maior detalhamento espacial dos mapas inferidos, dificuldades e inexatidão dos limites de unidades de mapeamento do mapa convencional, heterogeneidade geológica da área, profundidade do contato lítico e/ou exposição da rocha, e problemas relacionados com o modelo de correlação ambiental utilizado.

A maioria dos estudos publicados no Brasil tem empregado as geotecnologias como ferramenta de apoio aos levantamentos, ou ao entendimento das relações solo-paisagem. Sousa Junior; Demattê (2008) realizaram uma descrição quantitativa da relação solo-paisagem empregando um MDE e associando a ocorrência de classes de solos à declividade e altitude. O estudo apenas faz uma descrição da ocorrência, nenhum modelo de inferência é empregado. Para os autores a complexidade da paisagem torna necessário o conhecimento e experiência do pedólogo. No estudo conduzido por Ippoliti et al. (2005) técnicas automatizadas foram utilizadas para delinear as geoformas e pedopaisagens com isto obtendo um ganho de tempo e economia de recursos, se comparado a metodologia convencional por fotointerpretação. A utilização de carta topográfica na escala 1:50.000 permitiu que fossem gerados dados morfométricos do terreno, os quais foram utilizados para delinear as unidades de solos. Os autores também não realizaram um levantamento de solos, foi executado uma análise quantitativa das formas da paisagem e sua correlação com levantamentos já existentes, o que pode ser interessante do ponto de vista de auxiliar a formulação quantitativa das, até então, qualitativas relações solo-paisagem.

A utilização de cartas topográficas na escala 1:50.000 em um ambiente SIG, mostra-se como uma excelente estratégia para mapeamentos sistemáticos de solos na abrangência de áreas territoriais maiores (WEBER et al., 2008). Para esses autores, as cartas e a tecnologia SIG foram um excelente material de suporte para os trabalhos de campo; para organização de uma amostragem georreferenciada; delineamento, correção e checagem dos polígonos; manutenção da consistência do banco de dados; e, possibilitaram agilidade e uniformidade na geração do material impresso. Embora os próprios autores reconheçam que seu trabalho não se caracteriza como preditivo, não são feitas menções à acurácia final do mapeamento, nem ao efeito da utilização apenas do parâmetro elevação para determinar os polígonos de solo, mesmo sabendo que o solo está sendo determinado por fatores multivariados (LEGROS, 2005).

#### 2.1.7 Futuros estudos e a importância dos dados de campo

Para Walter et al. (2007) o mais importante não é o melhor modelo, com a melhor acurácia e menor incertezas, mas sim, que o modelo deve ser explicitamente

fundamentado em fatores e processos, podendo ser validado, refinado progressivamente e extrapolado para outras regiões além daquela para o qual foi gerado. Assim, os cientistas do solo podem ter na pedometria uma ferramenta para entender melhor a formação e distribuição do solo na paisagem.

Sendo a relação entre o solo e a paisagem no contexto da formação do solo bem compreendida, relações preditivas entre o solo e os fatores de formação do solo poderão ser elaboradas (GIASSON et al., 2008). No entanto, para Minasny; McBratney (2007) nisso está uma das grandes dificuldades da abordagem quantitativa e há muita pesquisa a ser realizada para a compreensão das relações entre o ambiente e a formação do solo. Deverá ser dada mais atenção às relações que considerem não apenas um fator de estado (topofunções, climofunções, etc.), mas sim modelos multivariados, capazes de considerar mais de um fator do modelo SCORPAN simultaneamente (MCBRATNEY et al., 2003). Estratégias neste sentido, poderiam ser os estudos de processos de formação (MINASNY; MCBRATNEY (2006); MINASNY et al. (2008)).

A necessidade de se continuar coletando dados de campo, especialmente a descrição de perfis é unanimidade entre os pesquisadores, uma vez que estas são fontes de informações seguras para a checagem dos modelos (MCBRATNEY et al. (2000); MCBRATNEY et al. (2003)). Como uma das colocações a favor da abordagem pedométrica tem sido a melhor utilização dos recursos, a coleta de informação sobre o solo deve ser planejada para racionalizar os gastos, incluindo técnicas de predição da localização das amostras, número ótimo de amostras e outros pré-mapeamentos. Quanto às variáveis ambientais a serem utilizadas como preditoras, a decisão de quais serão utilizadas, passará pela sua disponibilidade e ou recursos para que sejam adquiridas (LAGACHERIE; MCBRATNEY, 2007), bem como pela análise de suas correlações.

Técnicas quantitativas têm de ser validadas a campo onde cientistas experientes parecem ser cada vez mais raros, sendo um grande desafio o de treinar pedólogos que unam bons trabalhos de campo às modernas técnicas de análise espacial (WALTER et al., 2007). Possivelmente a tarefa de formar jovens pedólogos com conhecimento em informática será bem menos complexa, para Bui (2007) a presença de jovens pesquisadores com bom conhecimento em Tecnologia da Informação, e com grande vontade de enfrentar desafios é uma oportunidade para

que novas técnicas sejam empregadas, e com isto mais conhecimento sobre os elementos que controlam a distribuição espacial do solo será acumulado.

## 2.2 Análise de componentes principais

A ACP é um método multivariado que permite transformar um conjunto de variáveis iniciais correlacionadas entre si, num outro conjunto de variáveis não correlacionadas, as chamadas CP. É uma técnica poderosa que pode ser utilizada para a redução do número de variáveis e para fornecer uma visão privilegiada do conjunto de dados. Algebricamente, as componentes são combinações lineares ponderadas das  $p$  variáveis aleatórias  $X_1, X_2, \dots, X_p$  (JOHNSON; WICHERN, 1992).

Generalizando para  $m$  componentes e  $p$  variáveis ( $m \leq p$ ):

$$\begin{aligned} CP_1 &= a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ CP_2 &= a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p \\ &\dots \\ CP_m &= a_{1m}X_1 + a_{2m}X_2 + \dots + a_{pm}X_p \end{aligned} \quad (1)$$

Onde,

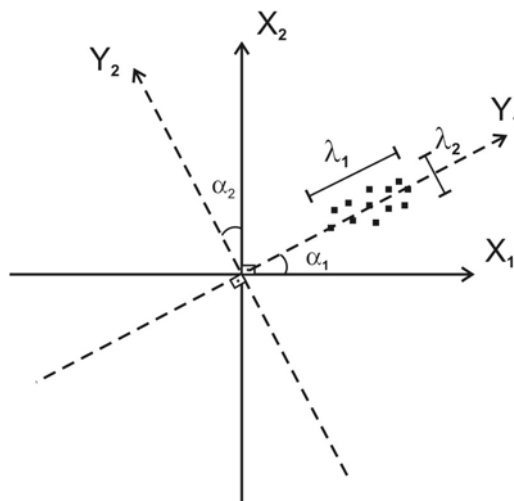
$CP_m$  : m-éssima componente principal;

$a_{pm}$  : autovetor ou peso de cada uma das variáveis para a CP;

$X_p$  : variáveis observadas.

Geometricamente, as componentes principais representam um novo sistema de coordenadas obtido pela rotação dos sistemas originais. Os novos eixos fornecem as direções da máxima variabilidade. Considerando uma amostra com duas variáveis de  $n$  observações, representadas na figura 3. Obtêm-se, por meio da rotação dos eixos, um novo sistema de coordenadas, em que  $Y_1$  representa o eixo principal e  $Y_2$  o eixo secundário, sendo  $\alpha_1$  e  $\alpha_2$  os ângulos formados entre os eixos originais  $X_1$  e  $X_2$  e os eixos  $Y_1$  e  $Y_2$  do novo sistema de coordenadas. A direção dos eixos é determinada pelos autovetores, aplicando-se aos ângulos os cossenos diretores ( $e_{11} = \cos(\alpha_1)$  e  $e_{21} = \cos(\alpha_2)$ ). Os autovalores  $\lambda_1$  e  $\lambda_2$  representam a variabilidade contida em cada um dos novos eixos (JOHNSON; WICHERN, 1992).





**Figura 3 - Sistema de coordenadas  $X_1$  e  $X_2$ , novo sistema  $Y_1$  e  $Y_2$ , ângulos de rotação dos novos eixos e autovalores.**

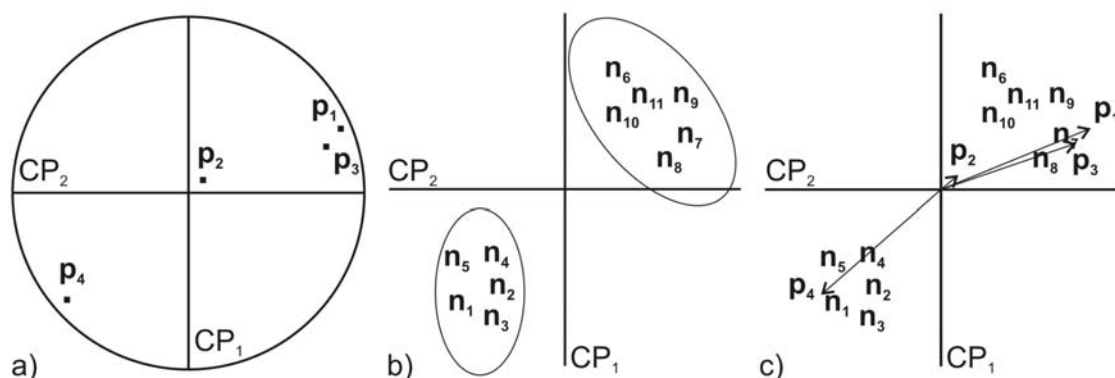
As CP são obtidas por meio da matriz de covariâncias, ou, quando houver necessidade de padronização dos dados, pela matriz de correlações, ambas extraídas da matriz original  $X_1, X_2, \dots, X_p$  (JOHNSON; WICHERN, 1992). As CP são calculadas por ordem decrescente de importância, isto é, a primeira explica o máximo possível da variância dos dados originais, a segunda o máximo possível da variância ainda não explicada, e assim por diante. As correlações das variáveis passam a ser suprimidas uma vez que as CP são ortogonais entre si.

Para Webster (2001) a ACP constitui-se em uma transformação de dados, sendo uma rotação rígida dos dados originais para um novo conjunto de eixos. É um procedimento matemático, não estatístico. Não requer pressuposição da normalidade da distribuição e não leva a teste estatístico de significância, podendo ser usada para reduzir o número de variáveis (VITHARANA et al., 2008), descobrir padrões (GOMES et al., 2004), otimizar amostragem (HENGL, 2003) e gerar variáveis não correlacionadas (HENGL et al., 2007).

Para a obtenção das CP, torna-se necessário verificar se as variáveis observadas estão sendo mensuradas na mesma unidade. Caso isso não aconteça, torna-se imprescindível utilizar como estrutura de variabilidade a matriz de correlação no lugar da matriz de covariância (ODEH et al., 1991). Para tal, faz-se necessário que os dados sejam estandardizados, o que significa que cada variável é transformada de tal forma que tem média igual a zero e variância igual a um. A variância total no conjunto de dados é simplesmente a soma da variância destas

variáveis observadas. Contudo, deve ser considerado que as CP da matriz de correlação não são necessariamente os mesmos que os obtidos da matriz de covariância (REIS, 1997).

Como as CP são ortogonais e, portanto, não correlacionadas, é possível examinar as relações entre as amostras através dos gráficos dos escores, projetados nas primeiras CP, e entre as variáveis, através dos gráficos de correlações. A representação destas correlações é normalmente feita em um gráfico de círculo com dimensão unitária, correlação máxima igual a um (Figura 4a). Quanto mais próximo do círculo estiver uma variável maior sua correlação nesta dimensão ( $p_1$ ), de maneira inversa quanto mais próxima do centro menor sua importância ( $p_2$ ). Assim como, variáveis muito próximas entre si estarão correlacionadas e explicando a mesma informação ( $p_1$  e  $p_3$ ), variáveis em lados opostos do círculo unitário são inversamente correlacionadas ( $p_1$  e  $p_4$ ), e apenas uma delas, em cada uma das situações, necessita ser considerada. Quanto à representação dos escores, sua plotagem em gráficos de dispersão possibilita que possam ser feitas relações entre as amostras em uma menor dimensão (POPPI; SENA, 2000) (Figura 4b). Outra alternativa é a representação simultânea das correlações e dos escores com a utilização dos diagramas de ordenação biplots (Figura 4c). Nestes biplots, ambos, escores e correlações, são normalizados e representados em um mesmo espaço, facilitando, desta maneira, a visualização das relações recíprocas entre objetos, variáveis e CP (ODEH et al., 1991).



**Figura 4 - a) Correlações entre variáveis e as Componentes Principais (CP). b) Escores nas duas primeiras CP. c) Biplot.**

A interpretação de cada componente é feita com base nas correlações entre as variáveis e os eixos, e estaria simplificada se cada variável tivesse uma correlação relativamente mais elevada para apenas uma das componentes, e valores pequenos ou próximos de zero para todas as restantes componentes. Isto pode ser conseguido através da rotação das CP. O método mais popular de rotação das componentes principais é o método VARIMAX. Este é um método ortogonal e pretende que para cada CP, existam apenas alguns pesos significativos e todos os outros sejam próximos de zero, isto é, o objetivo é maximizar a variação entre os pesos de cada CP.

Deve ser observado então que foram aplicadas duas rotações às variáveis iniciais. A primeira para a definição do novo conjunto de eixos (CP), e a segunda para potencializar as correlações entre as variáveis e as componentes retidas.

A aplicação de ACP em ciência do solo é bastante antiga. Na década de 70 Webster; Burrough (1972) utilizaram de ACP como metodologia para selecionar entre um conjunto de variáveis aquelas que explicavam a maior parte da variabilidade dos dados com o objetivo de analisar a coerência e homogeneidade dos delineamentos de classe de solos. Ao longo dos anos predominam estudos como este, que visam a redução do número de variáveis, e posterior aplicação em modelos (OVALLES; COLLINS (1988); BEAUCHEMIN et al. (2002); VITHARANA et al. (2008)). A ACP, por estar baseada em um modelo linear, tem boa aplicabilidade aos estudos que relacionem o solo a preditores ambientais, pois raramente existe entre eles uma relação não linear (Gaussiana) (ODEH et al., 1991). Ainda são comuns no país trabalhos que se utilizam da ACP como um método apenas de ordenação, buscando estabelecer relações entre características químicas, físicas ou biológicas e o ambiente (GOMES et al. (2004); PISSARRA et al. (2004); BENEDUZI et al. (2008)), sem qualquer aplicação de modelagem. Por outro lado, a aplicação de CP pode apresentar desvantagens como a transformação dos dados originais para novas variáveis sem significado físico, além de ser de interpretação mais complexa (HENGL; ROSSITER, 2003).

Sendo a ACP uma rígida rotação dos dados para novos eixos de coordenadas, a existência de algum significado pedológico ou biológico será mais por “obra de sorte” do que fruto de um delineamento experimental (WEBSTER, 2001). Ficando ressaltada a importância do pesquisador, que terá de associar seus

conhecimentos às novas informações geradas, visando dar a elas algum significado para futuras pesquisas.

### **2.3 Atributos do terreno como preditores**

Para Moore et al. (1993) em muitos locais, o desenvolvimento do solo em uma catena ocorre em resposta ao movimento da água através e sob a paisagem. Atributos do terreno podem caracterizar os fluxos e, desta forma, os atributos do solo. O terreno modifica a distribuição de processos hidrológicos, erosivos e a temperatura do solo. Desta forma afeta a distribuição do intemperismo mineral, lixiviação, erosão, sedimentação, decomposição, desenvolvimento do perfil e, finalmente, os atributos do solo.

O parâmetro relevo da equação de Jenny tem sido comumente derivado de MDE. Fontes de dados de elevação podem ser a digitalização de isolinhas, medidas pontuais a partir de levantamentos topográficos convencionais, medidas efetuadas por receptores do Sistema de Posicionamento Global (GPS) e estereopares oriundos do Sensoriamento Remoto. A primeira etapa no uso desta informação é a sua parametrização. Os parâmetros objetivam a caracterização da geomorfometria da superfície ou da classificação do relevo. Diferentes atributos podem ser parametrizados a partir do MDE, como a altitude, declividade, orientação, diferentes curvaturas, área de contribuição e índice topográfico composto (MCBRATNEY et al., 2003).

Walker et al. (1968) relacionaram variações do solo com parâmetros do terreno estudando gradiente de declividade, direção do comprimento do declive, curvatura, distância do divisor de águas, e elevação. Os parâmetros que mostraram melhores resultados como preditores nas regressões foram elevação e declividade. Embora grande parte da variabilidade dos valores observados não foi explicada pelos modelos. Para os autores, isto foi ocasionado pelo fato dos modelos ajustados não considerarem a influência dos organismos e do material de origem na formação do solo do local estudado.

Correlações significativas foram encontradas por Moore et al. (1993) entre atributos do terreno e do solo. Sendo que a declividade e o índice de umidade topográfica, explicaram até 50% da variabilidade da espessura do horizonte A, pH, P extraível, conteúdo de areia, silte e matéria orgânica do solo. Para os autores o uso

da técnica preditiva pode auxiliar no adensamento das informações contidas nos relatórios e mapas de levantamentos de solos. Além ainda, de ser uma ferramenta para guiar o levantamento de áreas até então não mapeadas.

## **2.4 Áreas de referência**

O método baseado na área de referência assume que é possível delimitar pequenas regiões naturais (LAGACHERIE et al., 1995), as quais irão conter um finito número de classes de solos, ocorrendo em associações com outras classes, e formando um padrão reconhecível e repetidamente distribuído. Conseqüentemente, uma representativa área de referência poderia ser suficiente para identificar todas as classes de solos em áreas maiores, e estabelecer suas relações espaciais.

As regiões naturais são os locais onde é possível caracterizar o solo de regiões topográfica ou geologicamente identificáveis. O primeiro estágio consiste de um levantamento detalhado em uma pequena, mas representativa área da região natural, a qual é denominada de área de referência. Este processo define as principais classes de solos de toda a região e estabelece as regras (relações) do mapeamento. Este estágio acelera e facilita o passo seguinte de produzir novos levantamentos de solos na mesma região (VOLTZ et al., 1997).

Assumindo que os mapas de solos contêm as relações entre o solo e o ambiente da área mapeada, estes podem ser vistos como áreas de referência. São o registro do modelo mental desenvolvido pelo pedólogo quando este espacializou o solo na paisagem seguindo relações com os fatores de formação. De tal forma que permanecem como a materialização do conhecimento desenvolvido pelo pedólogo quando este não mais realiza mapeamentos. O conhecimento das relações solo-paisagem contidas nos mapas de solos, quando adequadamente extraído e formulado, poderá servir de guia para novos mapeamentos em áreas similares (QI; ZHU, 2003). Técnicas de mineração de dados podem ser utilizadas para “aprender” com os mapas de solos. A extração das informações contidas em mapas de solos publicados pode possibilitar a geração de novos mapas, em regiões onde as regras geradas pela aplicação do paradigma sejam as mesmas. É importante que sejam pesquisadas então técnicas de “desconstrução” dos mapas de solos existentes para extrair destes as “regras” utilizadas pelo pedólogo durante sua produção (BUI; MORAN, 2001).

Um dos pontos fracos da metodologia de área de referência é o fato de que os relatórios dos levantamentos pedológicos não retratam a plenitude do conhecimento das relações solo-paisagem desenvolvidos pelo pedólogo no levantamento. Assim, as regras desenvolvidas a partir de uma área de referência não serão 100% efetivas em reproduzir o mapeamento em uma nova área. Sendo também uma estratégia de difícil utilização em áreas onde o padrão de distribuição do solo não é claramente identificável, devido a um grande número ou muito sutis fontes de variações. Da mesma forma que falharia em locais onde tendências regionais causam um contínuo gradiente de variação no solo, o qual não pode ser facilmente generalizado (LAGACHERIE et al., 1995).

As aplicações das áreas de referência tem se dado em uma variedade de abordagens. Qi; Zhu (2003) utilizaram de mineração de dados por árvores de decisão para extrair o conhecimento pedológico contido em mapas de solos. Os autores reformularam o modelo de relação solo-paisagem na forma de regras, descrição do solo e na seqüência de uma catena. Estes conhecimentos reelaborados sobre a relação solo-paisagem poderão ser empregados com fins didáticos, ou em levantamentos de solos onde as mesmas relações sejam válidas. Voltz et al. (1997) espacializaram propriedades do solo a partir de classes e propriedades de uma área de referência anteriormente mapeada. Os pesquisadores observaram que a densidade da amostragem para a espacialização das propriedades era fundamental. Também foi observado que a qualidade das propriedades inferidas caía bastante, na medida em que se afastava do local de amostragem.

## **2.5 Regressões logísticas múltiplas**

Em casos onde o resultado de uma inferência pode ser dado sob muitas categorias (politômica), uma alternativa é trabalhar com a probabilidade de ocorrência de cada uma das categorias, para tanto se aplica a RLM. A qual se constitui em uma técnica muito mais flexível do que, por exemplo, a regressão múltipla, uma vez que a RLM não tem qualquer exigência para sua aplicação a respeito da distribuição das variáveis explicativas. Não necessitando ter distribuição normal, correlação linear, medidas em mesma escala ou homogeneidade de

variância. As variáveis explicativas podem ser ainda, uma mistura de dados contínuos, discretos ou binários (CHATTERJEE; HADI, 2006).

A função logarítmica ou logit é a razão entre a probabilidade do evento ocorrer e a probabilidade deste evento não ocorrer (*odds* ou razão de chance). Ou ainda, exemplificando com classes de solo, a logit é a função logarítmica da razão entre a probabilidade ( $P_i$ ) de um pixel ser membro de uma classe de solo “i”, e a probabilidade de que ele não seja ( $1 - P_i$ ). Seu valor pode ser determinado através da regressão:

$$\text{Logit}_i = \ln\left(\frac{P_i}{(1 - P_i)}\right) = a + b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \quad (2)$$

A equação 2 demonstra como calcular o logit de uma categoria predita a partir de um conjunto de variáveis ( $X_{i\dots k}$ ). O logit é o logaritmo natural de uma razão entre probabilidades. O valor de “a” indica o intercepto da regressão e os valores de “b” são os coeficientes de cada variável explicativa.

Como é mais fácil de se perceber a razão entre probabilidades do que seu logaritmo, a equação 2 pode ser reescrita como:

$$\frac{P_i}{(1 - P_i)} = e^{a + b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k} \quad (3)$$

Para os casos em que a variável resposta tem apenas duas alternativas distintas (dicotômica), aplica-se a Regressão Logística Binomial (RLB), da qual a RLM é uma extensão. Na RLB a probabilidade de um evento ocorrer é calculada pelo rearranjo de equação 3 para:

$$P_i = \frac{e^{a + b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k}}{1 + e^{a + b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k}} \quad (4)$$

No entanto, tratando-se de um conjunto de respostas politômicas (várias classes de solo), faz-se necessário estabelecer uma categoria de referência ou base, da qual, em um primeiro momento, não é calculada a probabilidade de ocorrência. Para as demais se aplica:

$$P_i = \frac{e^{a + b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k}}{1 + \sum_1^{m-1} (e^{a + b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k})} \quad (5)$$

A qual é usada para prever a probabilidade  $P_i$  de que a classe de solo “i” esteja presente em determinado pixel, dadas as variáveis preditoras  $X_1, X_2, \dots, X_k$ .

Através da razão entre o logit de “i” e a soma dos logit de todas as outras classes, exceto da classe tomada como referência, mais um.

A classe tomada como referência, não terá seu logit determinado. Sendo sua probabilidade determinada por:

$$P_i = \frac{1}{1 + \sum_1^{m-1} (e^{a+b_0+b_1X_1+b_2X_2+\dots+b_kX_k})} \quad (6)$$

Todos os valores de “a” e “b” das regressões serão determinados para cada classe de solo a partir de dados experimentais (DEBELLA-GILO et al., 2007).

Em contraste com as regressões lineares, as quais selecionam os coeficientes que minimizam o quadrado da distância, regressões logísticas resultam em coeficientes os quais tornam o resultado observado mais provável através de um algoritmo iterativo por verossimilhança. O processo iterativo de busca dos melhores parâmetros para as k equações não retorna um coeficiente de determinação (R<sup>2</sup>), como nas regressões convencionais, mas é realizado um teste de significância para os modelos por teste Chi<sup>2</sup> (OHLMACHER; DAVIS, 2003).

Modelos RLM foram utilizados por Ohlmacher; Davis (2003) para criar um mapa de susceptibilidade do terreno a deslizamento. Paulin; Bursik (2008) desenvolveram um software para o mapeamento de áreas de risco ao deslizamento, empregando RLM aliado a SIG e Redes Neurais. No entanto, a aplicação de RLM ao mapeamento do solo havia recebido pouco enfoque da pesquisa até a publicação da revisão em mapeamento digital do solo por McBratney et al. (2003).

Para Hengl et al. (2007) o uso de RLM é mais trabalhoso, exigindo mais trabalho operacional na geração dos modelos. Outra grande limitação se deve ao fato de que a RLM não considera as autocorrelações entre os dados durante a estimação dos parâmetros dos modelos.

## 2.6 Avaliação dos mapas gerados

O mapeamento do solo, seja por técnicas convencionais ou pedométricas, não estará completo se não for verificada a qualidade da informação contida no mapa. Os usuários dos mapas devem conhecer a qualidade do levantamento para decidir se este é aplicável ao uso pretendido, ou não. Da mesma forma, o responsável pela produção do mapa precisa avaliar se seu levantamento atingiu os



objetivos esperados. Além do que, ambos poderão desejar comparar mapas distintos para avaliar qual é o melhor, ou sua concordância (ROSSITER, 2004).

Uma alternativa para reportar o erro contido em um levantamento de solos é a preparação de uma matriz de erros, também conhecida como matriz de confusão ou matriz de contingência. Esta matriz mostra a tabulação cruzada entre duas informações. Este cruzamento pode ser entre dois mapas gerados por metodologias diferentes, ou ainda, para comparar um mapa com a verdade de campo. Uma série de procedimentos estatísticos poderá ser aplicada aos valores contidos na matriz de erros (BANKO, 1998).

Após amostragem comparativa a matriz de erros é estruturada como demonstra a tabela 1. Nas linhas tem-se o preenchimento com os dados inferidos, e nas colunas são marcados os valores correspondentes aos dados de referência. Os valores na diagonal corresponderão aos dados corretamente preditos.

Um dos primeiros estimadores da qualidade do levantamento será a Acurácia Geral (AG), correspondendo ao total de dados corretamente preditos dividido pelo total de amostras. Outro estimador é a Acurácia do Mapeador (AM), a qual irá determinar os acertos da área mapeada. Nesta estarão incluídos os erros por omissão, que se referem à proporção de dados verificados a campo e que não constam no mapa. Quanto mais omissões existirem, menor a qualidade do mapa. Há ainda, a Acurácia do Usuário (AU) que dá uma idéia de confiabilidade do mapa, indicando se as classes estão verdadeiramente presentes no campo. Erros neste valor são ditos como de inclusão, pois o mapa indica uma classe de solo a qual não é verificada no mesmo local no campo. Da mesma forma, inclusões irreais no mapa diminuem sua acurácia para o usuário (BANKO, 1998).

Tabela 1: Arranjo da matriz de erros para caso com três categorias (A, B e C). Acurácia do mapeador (AM), acurácia do usuário (AU) e acurácia geral (AG).

		Dados de referência				AU (%)
		A	B	C	$\Sigma$	
Predito	A	$x_{11}$	$x_{12}$	$x_{13}$	$\Sigma x_{i1} (i=1,2,3)$	$(x_{11} * 100) / \Sigma x_{i1} (i=1,2,3)$
	B	$x_{21}$	$x_{22}$	$x_{23}$	$\Sigma x_{i2} (i=1,2,3)$	$(x_{22} * 100) / \Sigma x_{i2} (i=1,2,3)$
	C	$x_{31}$	$x_{32}$	$x_{33}$	$\Sigma x_{i3} (i=1,2,3)$	$(x_{33} * 100) / \Sigma x_{i3} (i=1,2,3)$
	$\Sigma$	$\Sigma x_{1j} (j=1,2,3)$	$\Sigma x_{2j} (j=1,2,3)$	$\Sigma x_{3j} (j=1,2,3)$	$\Sigma x_{ij} (i,j=1,2,3)$	
AM		$(x_{11} * 100) / \Sigma x_{1j} (j=1,2,3)$	$(x_{22} * 100) / \Sigma x_{2j} (j=1,2,3)$	$(x_{33} * 100) / \Sigma x_{3j} (j=1,2,3)$		
						$\Sigma x_{ij} (i=j) * 100 / \Sigma x_{ij} (i,j=1,2,3)$
						AG (%)

O índice kappa (K) mede a proporção de concordância entre os dados após as concordâncias devidas somente ao acaso serem removidas (ROSSITER, 2004).

$$K = (n * \Sigma x_{ii} - \Sigma x_{i+}x_{+i}) / (n^2 - \Sigma x_{i+}x_{+i}) \quad (7)$$

onde,

$n$  = número de observações;

$x_{ii}$  = observações na diagonal;

$x_{i+}$  = totais em linha;

$x_{+i}$  = totais em coluna.

O índice K é um indicador bastante utilizado para atestar a qualidade dos mapeamentos preditivos (HENGL et al., 2007). Sendo que valores da ordem de 0,3 a 0,6 são relatados pela literatura em testes de modelos preditivos aplicados ao mapeamento de solos. Valores maiores, acima de 0,6, são normalmente alcançados com simplificação das legendas de mapeamento (ZEILHOFER, 2006). Para Uebersax (2008) as tabelas que ordenam os valores do índice K em “bom”, “moderado”, “ruim”, etc. deveriam ser evitadas, e trabalhos realizados em uma mesma linha de pesquisa serem utilizados para se avaliar a real qualidade dos valores de K encontrados.

### 3 MÉTODOS E TÉCNICAS

#### 3.1 Caracterização da área de estudo

A área de estudo localiza-se no município de São Pedro do Sul, situado na região central do estado do Rio Grande do Sul, na microrregião de Santa Maria. Possui uma superfície de 873,59 km<sup>2</sup>, e está compreendido entre as coordenadas 29°46' a 29°26' de latitude sul e 54°30' a 53°56' de longitude oeste. Abrange uma região transitória entre as regiões fisiográficas do planalto médio e depressão central conforme ilustra a figura 5. O clima da região é temperado chuvoso e quente, tipo Cfa, com temperaturas médias mínimas de 14° e máxima de 25°C. A precipitação média anual chega a 1500 mm (CABRAL; SILVA, 1998).

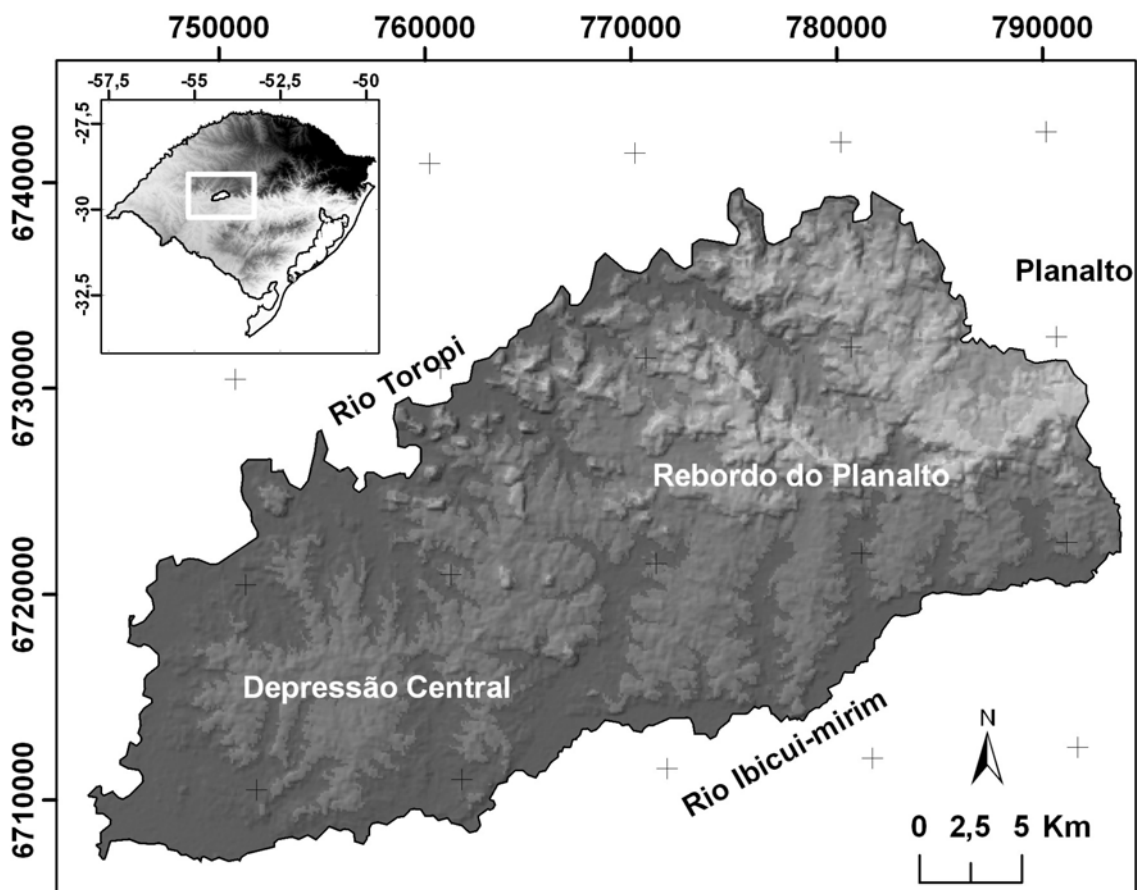


Figura 5 – Regiões fisiográficas e relevo do município. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Imagem interna localiza São Pedro do Sul no estado do Rio Grande do Sul.

No município ocorrem várias unidades geológicas (Figura 6). Sendo estas: Sedimentos Quaternários e Cenozóicos, Formação Serra Geral, Formação Botucatu, Formação Caturrita, Formação Santa Maria e Formação Sanga do Cabral. Quanto a geomorfologia, o município apresenta formas de relevo que se diferenciam em função de estar situado entre o compartimento geomorfológico do planalto e da depressão central, sendo identificadas unidades de relevo como topo do planalto, rebordo do planalto e depressão central (KLAMT et al., 2001).

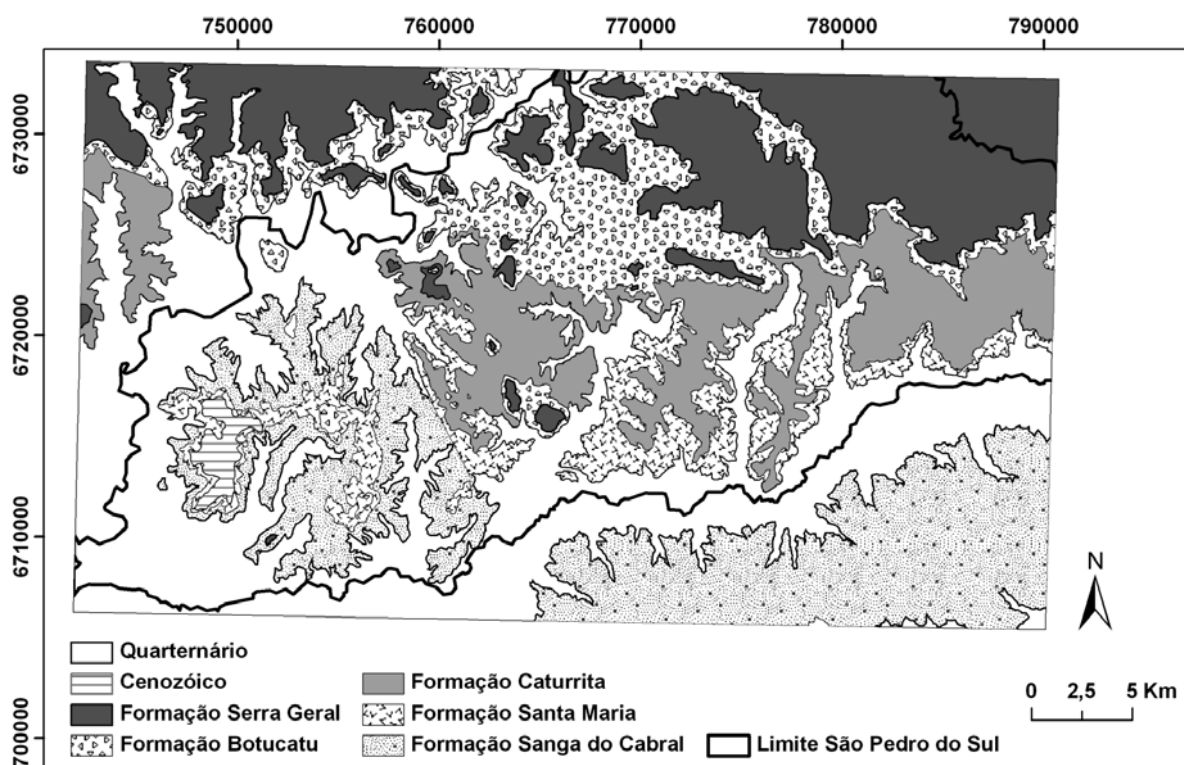


Figura 6 – Unidades geológicas em São Pedro do Sul. Adaptado de Santos e Moreira (1987) e não disponível para todo o município. Coordenadas E e N em metros, UTM/SAD69, Fuso 21.

### 3.2 Solos do município

Esta área foi escolhida por apresentar um levantamento semidetalhado de solos na escala 1:50.000 (KLAMT et al., 2001) (Apêndice A). A legenda de solos foi atualizada conforme EMBRAPA (2006) (Tabela 2).

**Tabela 2 – Solos do município de São Pedro do Sul (KLAMT et al., 2001) atualizados de acordo com EMBRAPA (2006) e respectivas áreas.**

Legenda	Classe	Área	
		(km <sup>2</sup> )	(%)
PVd1	ARGISSOLO VERMELHO Distrófico arênico	2,70	0,31
PVd2	ARGISSOLO VERMELHO Distrófico típico	27,76	3,19
PVAd1	ARGISSOLO VERMELHO-AMARELO Distrófico espessarênico	18,53	2,13
PVAd2	ARGISSOLO VERMELHO-AMARELO Distrófico arênico	37,07	4,26
PVAd3	ARGISSOLO VERMELHO-AMARELO Distrófico abrupto	116,34	13,37
PVAal	ARGISSOLO VERMELHO-AMARELO Alítico abrupto	25,32	2,91
PACe	ARGISSOLO ACINZENTADO Eutrófico abrupto	83,62	9,61
PBACal	ARGISSOLO BRUNO-ACINZENTADO Alítico úmbrico	119,65	13,75
NVd	NITOSSOLO VERMELHO Distrófico argissólico	6,70	0,77
FTd	PLINTOSSOLO ARGILÚVICO Distrófico arênico	19,75	2,27
CXbe	CAMBISSOLO HÁPLICO Ta Eutrófico léptico	168,73	19,39
RLe1	NEOSSOLO LITÓLICO Eutrófico típico	40,90	4,70
RLe2	NEOSSOLO LITÓLICO Eutrófico chernossólico	4,44	0,51
SH	Solos Hidromórficos ( Associação NEOSSOLO QUARTZARÊNICO Órtico e GLEISSOLO MELÂNICO Eutrófico)	198,66	22,83

As unidades de solos do mapa foram digitalizadas no programa ArcGIS 9.2, através do aplicativo Editor com auxílio do procedimento Auto-completar, o que possibilitou um delineamento sem conflitos entre polígonos vizinhos.

### 3.3 Levantamento de solos nas áreas de referência

Visando a produção de mapas de solos a partir de áreas de referência, foi realizado um levantamento de solos convencional em duas regiões do município de São Pedro do Sul. As áreas de referência encontram-se na figura 7. Para a definição das classes de solos foram utilizadas informações extraídas de Klamt et al. (2001). Visando a verificação e checagem das classes de solos das áreas de referência, foram amostrados 56 pontos no campo de acordo com a metodologia convencional de levantamentos (Apêndice B). As classes de solos foram descritas até o nível de subordem. As informações de campo também foram utilizadas para testar a acurácia do mapa gerado.

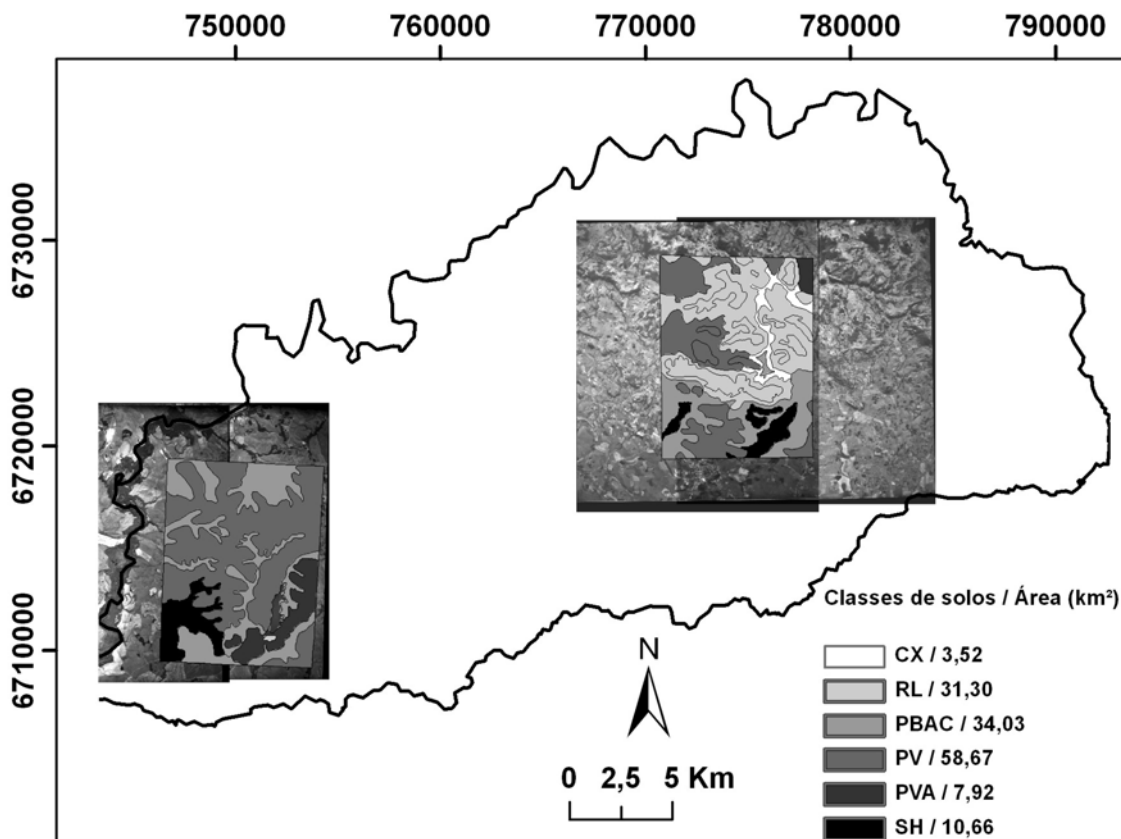


Figura 7 – Pares estereoscópicos e classes de solos descritas nas áreas de referência no município de São Pedro do Sul. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA) e Argissolo Bruno-Acinzentado (PBAC).

### 3.4 Atributos do terreno

Para a derivação dos atributos do terreno a serem utilizadas no estudo foi utilizado o MDE oriundo da técnica de Sensoriamento Remoto por Radar. As informações do SRTM foram baixadas do sítio na internet [http://www.ecologia.ufrgs.br/labgeo/SRTM\\_BR.php](http://www.ecologia.ufrgs.br/labgeo/SRTM_BR.php). Estes dados haviam sido tratados para o preenchimento de falhas existentes nos dados originais através de interpolação, conversão para valores inteiros e transformação para o Datum SAD 69. A resolução original do MDE de 90 m foi interpolada para uma nova resolução de 50 m, utilizando-se o procedimento *Spline* do aplicativo Análise Espacial do programa ArcGIS. O Quadro 1 traz os nove atributos do terreno gerados a partir do SRTM para este estudo, sua abreviatura, significado físico e unidade do atributo.

Atributos do terreno (abreviatura)	Significado Físico (unidade)
Elevação (ELEV)	Tem importante papel na definição do clima local (metros)
Declividade (DECL)	Afeta a velocidade de fluxos superficiais e subsuperficiais, e influencia no conteúdo de água no solo, no potencial de erosão e na formação do solo (graus).
Curvatura Planar (CPLN)	Indica a concentração ou dispersão da água sob o relevo, assim como, as posições de topo e depressão na paisagem (metros <sup>-1</sup> ).
Curvatura de Perfil (CPRF)	Importante para a velocidade da água, e processos de transporte de sedimento. Irá indicar as superfícies côncavas e convexas na paisagem (metros <sup>-1</sup> ).
Distância à Drenagem (DIST)	Está relacionada com a densidade da rede de drenagem, e com a presença e exposição do material de origem e do solo ao lençol freático (metros).
Iluminação (ILUM)	Representa diferentes exposições da paisagem a energia solar, refletindo em condições ambientais distintas (adimensional).
Logaritmo Natural da Área de Contribuição (LNAC)	Relaciona-se com o fluxo da água sob a paisagem, além da disponibilidade de água para os processos ambientais (logaritmo natural de metro <sup>2</sup> ).
Índice de Umidade Topográfica (IUT)	Representa o controle da topografia sobre a umidade do solo (adimensional).
Capacidade de Transporte de Sedimento (CTS)	Representa o potencial de perda de solo caracterizando processos de erosão e deposição na paisagem (adimensional).

**Quadro 1: Atributos do terreno e seu significado físico. Fonte: Wilson; Gallant (2000)**

A partir do MDE foi derivada a Área de Contribuição (AC) acima do pixel utilizando o aplicativo TAPES, com o algoritmo FD8 sendo executado em ambiente ArcGIS (WILSON; GALLANT, 2000). O valor da AC representa a área acima de determinado ponto da paisagem (pixel) que está contribuindo para o fluxo de água que passa por este ponto. Foi extraído o logaritmo natural do valor da AC, para reduzir a amplitude desta variável, passando a mesma a ser denominada de LNAC neste estudo.

O IUT também denominado de índice topográfico, índice de umidade, e ainda, índice topográfico composto, foi obtido pelo logaritmo natural da razão entre a Área de Contribuição Específica ( $AC_E$ ) e a tangente da declividade do terreno (Equação 8) (WILSON; GALLANT, 2000).

$$IUT = \ln\left(\frac{AC_E}{tg(\beta)}\right) \quad (8)$$

Onde,  $AC_E$  foi obtida pela razão entre a AC e a resolução do pixel (Equação 9).

$$AC_E = \frac{AC}{resolução\_pixel} \quad (9)$$

Os cálculos de IUT e  $AC_E$  foram efetuados no aplicativo Análise Espacial (ArcGIS) pelo procedimento Calculadora Raster, o qual possibilita realizar operações com Plano de Informação (PI) do tipo raster.

Para o cálculo da CTS foi utilizada a equação 10, (WILSON; GALLANT, 2000):

$$CTS = \left(\frac{AC_E}{22.13}\right)^{0.6} \left(\frac{\text{sen}(\beta)}{0.0896}\right)^{1.3} \quad (10)$$

Onde,  $\text{sen}(\beta)$  representa o seno da declividade do terreno. A função foi processada na função Calculadora Raster (ArcGIS).

A variável representando a ELEV foi obtida diretamente do MDE/SRTM. Tendo sido realizado o procedimento *FillSinks* (ArcGIS), para correção de informações espúrias (picos anômalos e pontos ou áreas com ausência de dados). As informações de DECL, CPLN, CPRF foram derivadas do MDE/SRTM utilizando-se do aplicativo TAPES (WILSON; GALLANT, 2000) em ambiente ArcGIS. A variável ILUM foi obtida através da função *Hillshade* do programa ArcGIS, com um azimute de 0° e inclinação solar de 60°. Para determinação da DIST, foi utilizada a função



*Cost Weighted* no programa ArcGIS, o qual calcula a distância até a drenagem mais próxima utilizando de um ponderador, neste caso, o ponderador foi definido como sendo a elevação do terreno.

### 3.5 Geração dos mapas de solos.

Um fluxograma geral dos procedimentos executados para a geração dos mapas preditos de solos pode ser visualizado na figura 8. Como foram utilizadas duas fontes distintas de informações sobre a distribuição espacial das classes de solos no município, esta sessão de materiais e métodos abordará cada uma separadamente.

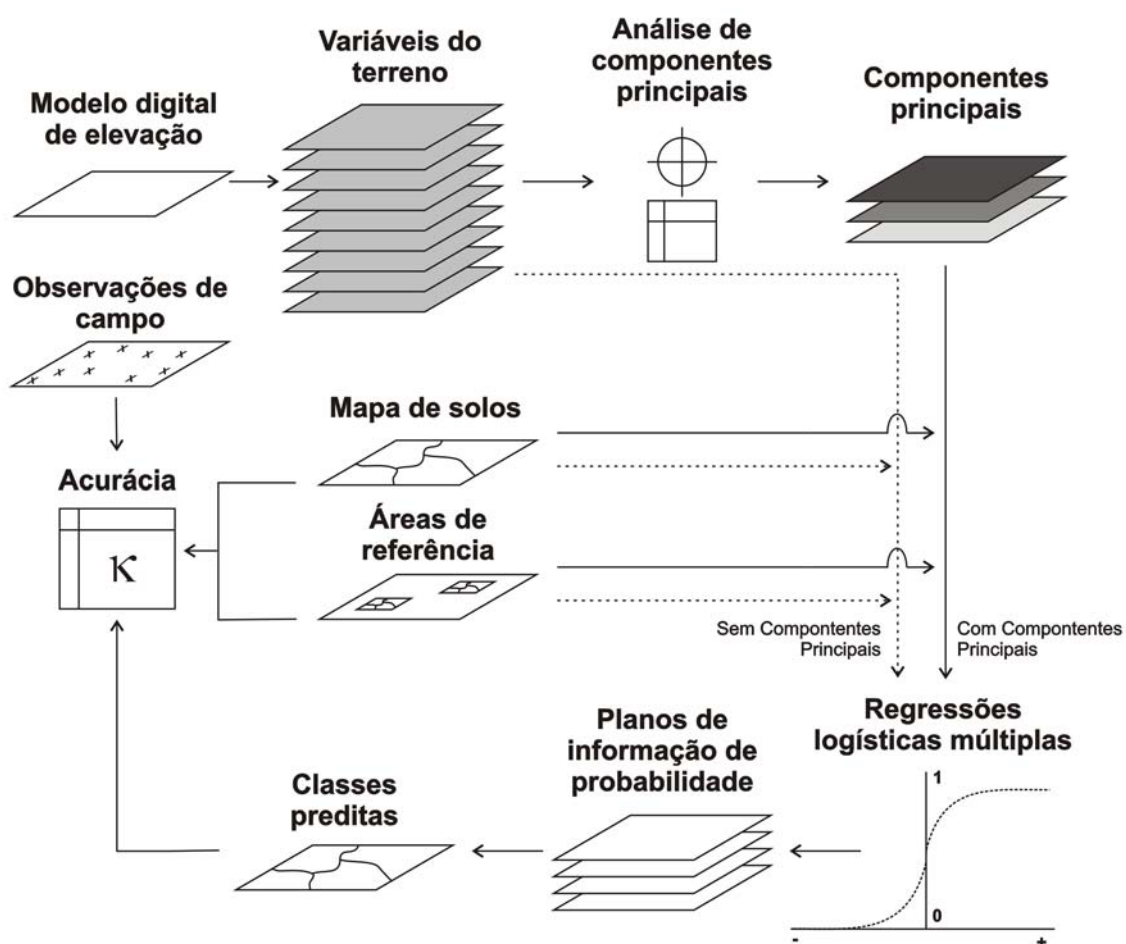


Figura 8 – Fluxograma geral das etapas de processamento do estudo.

### 3.5.1 Mapa predito a partir do mapa de solos Klamt.

Os nove PI de atributos do terreno, bem como, o arquivo de classes de solos do mapa de solos de São Pedro do Sul (KLAMT et al., 2001), doravante denominado mapa Klamt, foi convertido para o formato raster e passou a fazer parte de um projeto do programa ArcGIS. Um total de 70.000 pontos, representando aproximadamente 20% da área total do município, foram aleatoriamente gerados para a tabulação dos dados de variáveis de terreno e das classes de solos. A amostragem foi executada com a função *Sample* no programa ArcGIS, tendo sido gerada uma matriz de dados de 11 colunas e 70.000 linhas (Quadro 2).

ID	SOLO	ELEV	DECL	DIST	CPLN	CPRF	ILUM	LNAC	IUT	CTS
1	3	115,18	8,05	791,05	0,02	0,04	199,00	10,05	8,10	11,12
2	3	138,01	3,61	615,90	0,01	0,00	168,00	9,46	8,31	2,76
3	1	403,13	0,78	71,53	0,01	-0,01	177,00	8,90	9,28	0,27
4	6	146,11	9,00	275,81	0,00	-0,35	156,00	9,12	7,05	7,36
5	3	119,30	4,61	213,48	0,02	-0,04	181,00	9,30	7,91	3,45
6	1	372,67	6,83	2776,10	-0,03	0,05	190,00	9,57	7,78	6,73
7	3	143,07	1,44	240,61	0,03	0,07	177,00	8,15	7,92	0,38
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
70000	6	393,45	9,66	2585,56	0,13	0,16	169,00	8,61	6,47	5,94

**Quadro 2 – Parte da matriz de pontos amostrados e tabulados para a análise de componentes principais. Identificador (ID), Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Iluminação (ILUM), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).**

A matriz em formato texto (ASCII) foi processada para a ACP, tendo sido verificada a adequabilidade das amostras pelo teste individual *Measure of Sample Adequacy* (MSA) de acordo com a equação 11:

$$MSA = \frac{\sum_{j \neq i} r_{ij}^2}{\left( \sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} a_{ij}^2 \right)} \quad (11)$$

e geral *Kaiser Meyer Olkin* (KMO) pela equação 12:

$$KMO = \frac{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2}{\left( \sum_{i \neq j} \sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} \sum_{i \neq j} a_{ij}^2 \right)} \quad (12)$$

Onde, em ambas as equações:

$r_{ij}$  = é o coeficiente de correlação entre as variáveis  $i$  e  $j$ ;

$a_{ij}$  = é o coeficiente de correlação parcial entre as variáveis  $i$  e  $j$ .

Estes testes foram aplicados visando à verificação do grau de correlação entre as variáveis e a justificativa da ACP. As CP foram rotacionadas com o procedimento VARIMAX, o que possibilita que as componentes retidas sejam denominadas de acordo com sua relação com as variáveis originais. O número de autovalores retidos foi condicionado para aqueles que tivessem valor mínimo igual a um. Os autovetores rotacionados, resultantes da ACP, foram utilizados para o cálculo dos valores das novas variáveis.

Na seqüência, os valores das CP foram utilizadas como variáveis explicativas em RLM. Como variável resposta utilizou-se as classes de solos ao nível de ordem, 1º Nível Categórico (1ºNC), e subordem, 2º Nível Categórico (2ºNC), do mapa Klamt. Também foram gerados modelos por RLM tomando-se como variáveis explicativas as variáveis de terreno, sem serem consideradas as CP. Neste caso, novamente a variável resposta foi obtida do levantamento de solo, em 1º e 2º NC. Em todos os modelos testados, foram considerados apenas os coeficientes acima de 5% de significância.

No programa ArcGIS, utilizando-se da função *Raster Calculator*, foram processadas todas as etapas de produção dos PI das CP. Bem como, os PI das RLM os quais geraram arquivos rasters de probabilidade de ocorrência de cada classe de solo. E finalmente, a reunião dos pixels de maior valor entre todos os PI de probabilidade para um único PI, este por fim correspondendo à posição de cada classe de solo na paisagem.

Para o teste da acurácia do mapa predito, dois novos grupos de pontos foram aleatoriamente gerados. A verificação da acurácia do mapa predito na área de onde a informação foi extraída para a geração do modelo (região dos 70.000 pontos - área externa) serviu para testar a capacidade do modelo reproduzir o mapa original. Por outro lado, a comparação de uma área onde não haviam sido utilizadas informações

para a geração dos modelos, permitiu verificar a capacidade do modelo prever as classes de solo (área interna) (Figura 9). Os pontos da área de onde as informações foram utilizadas para a geração dos modelos, corresponderam a aproximadamente 2% de toda área do município, cerca de 7.000 pontos. A capacidade de o modelo prever as classes de solos foi verificada com um conjunto de 1.000 pontos.

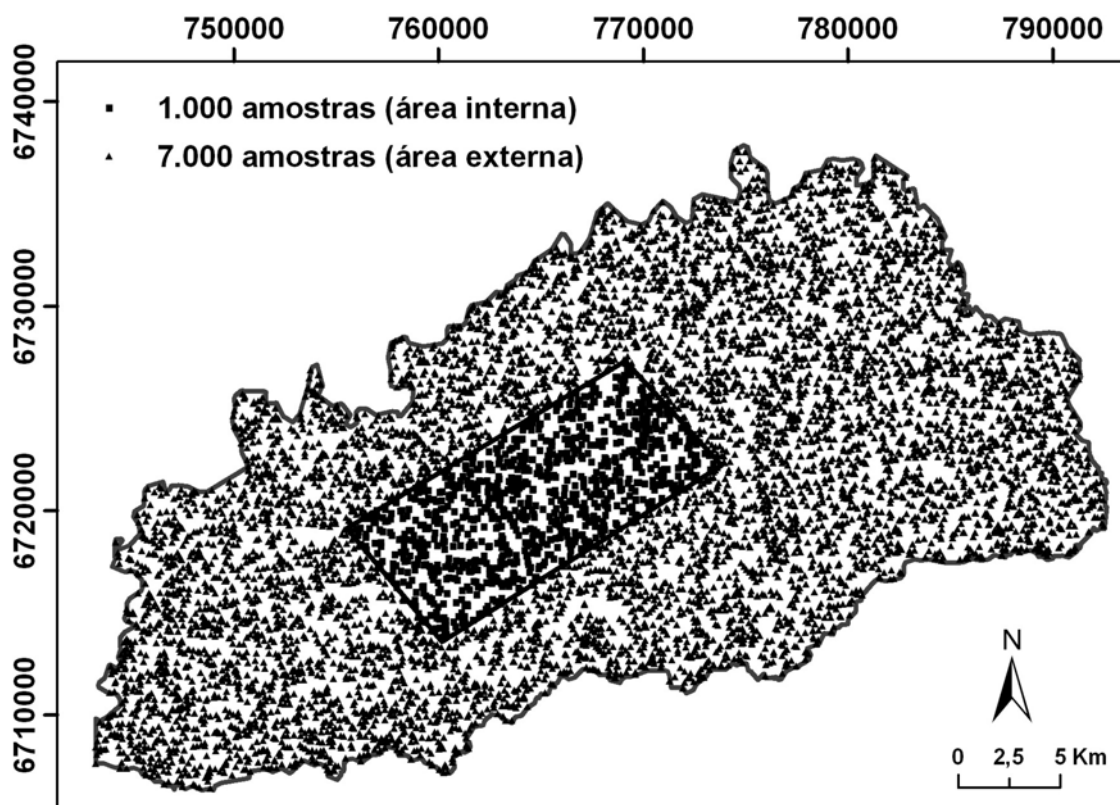


Figura 9 – Espacialização dos pontos utilizados para verificar a acurácia dos modelos que utilizaram o mapa Klamt. Coordenadas E e N em metros, UTM/SAD69, Fuso 21.

Ambos os conjuntos de pontos foram utilizados para gerar tabelas com dados oriundos do mapa de referência e do mapa predito, as tabelas no formato ASCII foram processadas para a geração do índice kappa ponderado.

### 3.5.2 Mapa predito a partir do mapa de solos das áreas de referência.

O arquivo de classes de solos das áreas de referência, doravante denominado mapa referência, foi convertido para o formato raster e passou a fazer parte de um projeto do programa ArcGIS. Para a tabulação dos dados de terreno e classes de solos, foram aleatoriamente gerados 6.000 pontos em cada uma das

áreas de referência, correspondendo a aproximadamente 20% do total das áreas. A amostragem e tabulação seguiram o mesmo procedimento já descrito para o mapa Klamt. O procedimento para a geração dos dados por CP não foi novamente executado, por se tratar da mesma área, foram utilizados os mesmos dados gerados na sessão anterior (3.5.1).

As RLM foram geradas utilizando-se dos dados de CP e de atributos do terreno como variáveis explicativas, e das classes de solos no mapa referência ao nível de subordem (2ºNC) como variáveis respostas. Para os modelos RLM também só foram consideradas os parâmetros significativos ao nível de 5%. A seqüência de passos para a implementação das funções de RLM, geração dos mapas de probabilidades e dos mapas de solos em ambiente SIG foi a mesma descrita na sessão 3.5.1.

A verificação da acurácia dos mapas gerados foi executada por uma comparação entre o mapa referência e o predito, e também com dados de campo. Para tal, um conjunto de 1.000 pontos foi aleatoriamente gerado em cada uma das áreas de referência. Nestes pontos foi verificada a capacidade dos modelos reproduzirem o mapa de referência. Outro teste de acurácia foi procedido utilizando-se os 56 pontos amostrados no campo, nestes foram realizadas tradagens e descrições de perfis durante a fase de levantamento de solos das áreas de referência (Figura 10). Este conjunto de pontos não se constitui de um conjunto totalmente independente de dados, uma vez que não foi gerado em áreas fora daquelas utilizadas para a geração dos modelos. O teste de acurácia realizado utilizou-se do índice K.

A verificação da acurácia dos mapas preditos utilizando-se dos mapas existente e do mapa produzido para as áreas de referência, não tem a intenção de afirmar que estes mapas são iguais. Estes mapas foram produzidos por metodologias distintas e possivelmente não sejam até mesmo comparáveis. No entanto, este estudo buscou determinar a acurácia dos mapas gerados pelos modelos logísticos comparando-os com os convencionais partindo-se do princípio de que estes representam a “verdade” existente. Uma alternativa mais adequada para a checagem da qualidade dos mapas preditos seria a sua conferencia pontual em relação a perfis de solos, o que demandará um número adequado de perfis descritos, e conseqüentemente de recursos.

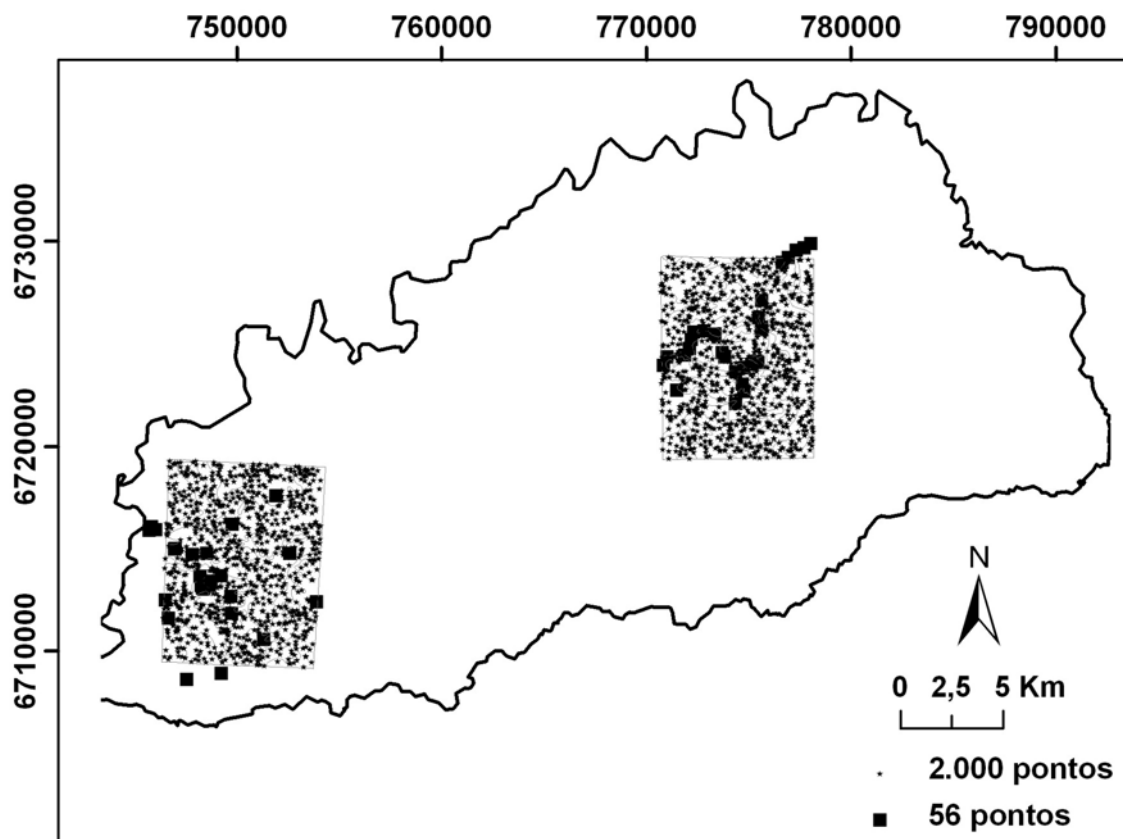


Figura 10 – Espacialização dos pontos utilizados para verificar a acurácia dos modelos nas áreas de referência. Coordenadas E e N em metros, UTM/SAD69, Fuso 21.

Um resumo das etapas, variáveis utilizadas, número de amostras geradas e avaliações realizadas neste estudo pode ser observado no quadro 3.

Variáveis Explicativas		Variáveis Resposta	Amostras	Avaliação			
Componente Principal	Atributos de Terreno	Levantamentos de Solos		Repetibilidade	Predição	Veracidade a Campo	Visual
Primeira Componente - CP1	Elevação - ELEV	<b>mapa Klamt - 1°NC</b> Cambissolo - C Plintossolo - F	70.000	7.000 pontos aleatoriamente gerados em área ao longo de todo o município exceto na área central - Área Externa	1.000 pontos aleatoriamente gerados na área central do município - Área Interna	Não executado	Comparação entre os mapas
	Declividade - DECL	Nitossolo - N					
Segunda Componente - CP2	Distância à Drenagem - DIST	Argissolo - P Neossolo - R Solos Hidromórficos - SH	70.000	7.000 pontos aleatoriamente gerados em área ao longo de todo o município exceto na área central - Área Externa	1.000 pontos aleatoriamente gerados na área central do município - Área Interna	Não executado	
	Curvatura Planar - CPLN	<b>mapa Klamt - 2°NC</b> Assoc. Cambissolo Háplico e Nitossolo Vermelho - CX_NV Plintossolo Argilúvico - FT Solos Hidromórficos - SH					
	Curvatura de Perfil - CPRF	Argissolo Acinzentado - PAC					
	Exposição a Radiação - ILUM	Argissolo Bruno-Acinzentado - PBAC Argissolo Vermelho - PV					
Terceira Componente - CP3	Logaritmo Natural da Área de Contribuição - LNAC	Argissolo Vermelho - PVA Neossolo Litólico - RL	12.000	1.000 pontos aleatoriamente gerados em cada uma das áreas de referência.	Não executado	56 pontos visitados durante a fase de checagem de campo da fotointerpretação da área de referência.	
	Índice de Umidade Topográfica - IUT	<b>mapa Referência - 2°NC</b> Solos Hidromórficos - SH Argissolo Bruno-Acinzentado - PBAC Argissolo Vermelho - PV Argissolo Vermelho-Amarelo - PVA					
	Capacidade de Transporte de Sedimento - CTS	Cambissolo Háplico - CX Neossolo Litólico - RL					
Modelo utilizado: Regressão Logística Múltipla (RLM)			Modelo utilizado: Kappa (K)				

**Quadro 3 – Quadro resumo das variáveis explicativas e resposta, número de amostras e avaliações da qualidade dos mapas executadas neste estudo.**

## 4. RESULTADOS E DISCUSSÃO

### 4.1 Atributos do terreno

O município de São Pedro do Sul possui uma diferença de ELEV de aproximadamente 400 m entre as regiões ao sul (depressão central) e norte (rebordo do planalto) do município (Figura 11). Esta diferença em ELEV pode estar condicionando variáveis meteorológicas como precipitação e temperatura, as quais por sua vez irão se refletir na alteração diferenciada do material de origem, e por conseqüência, na formação do solo. Embora fatores climáticos não tenham sido considerados diretamente neste estudo, estes são contemplados de maneira indireta pelo atributo ELEV. O histograma da imagem indica um predomínio de valores em torno de 100 m, apesar de ocorrer freqüências de valores em praticamente todas as classes de ELEV.

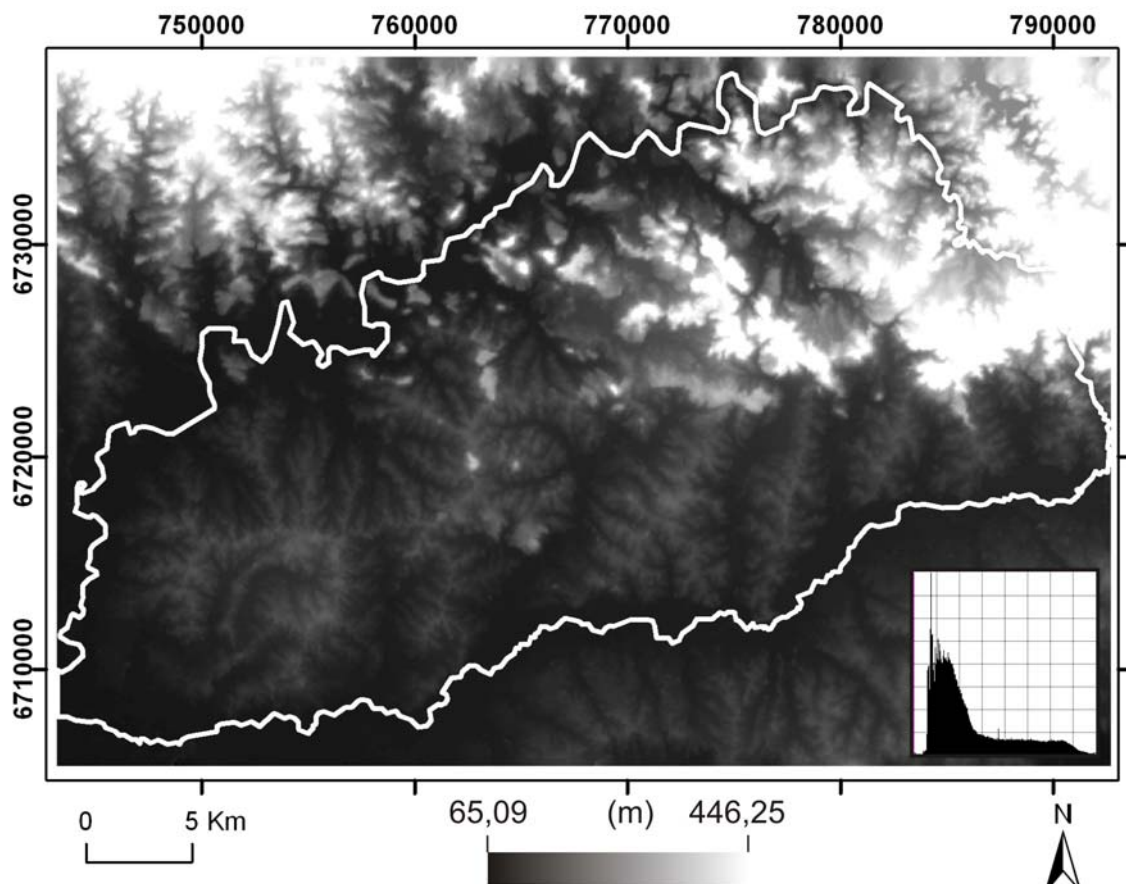


Figura 11 – Distribuição espacial da elevação na região de estudo. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Na imagem interna o histograma de freqüências das elevações.



Os maiores valores de DECL na região (Figura 12) ocorrem na presença dos chamados morros testemunhos e no rebordo do planalto. Os morros testemunhos apresentam mais resistência ao processo de intemperismo geológico, e na medida em que a encosta retrocede em direção ao norte, destacam-se na paisagem como os locais mais elevados, suas encostas tornam-se íngremes, o que resulta em maiores ângulos de DECL.

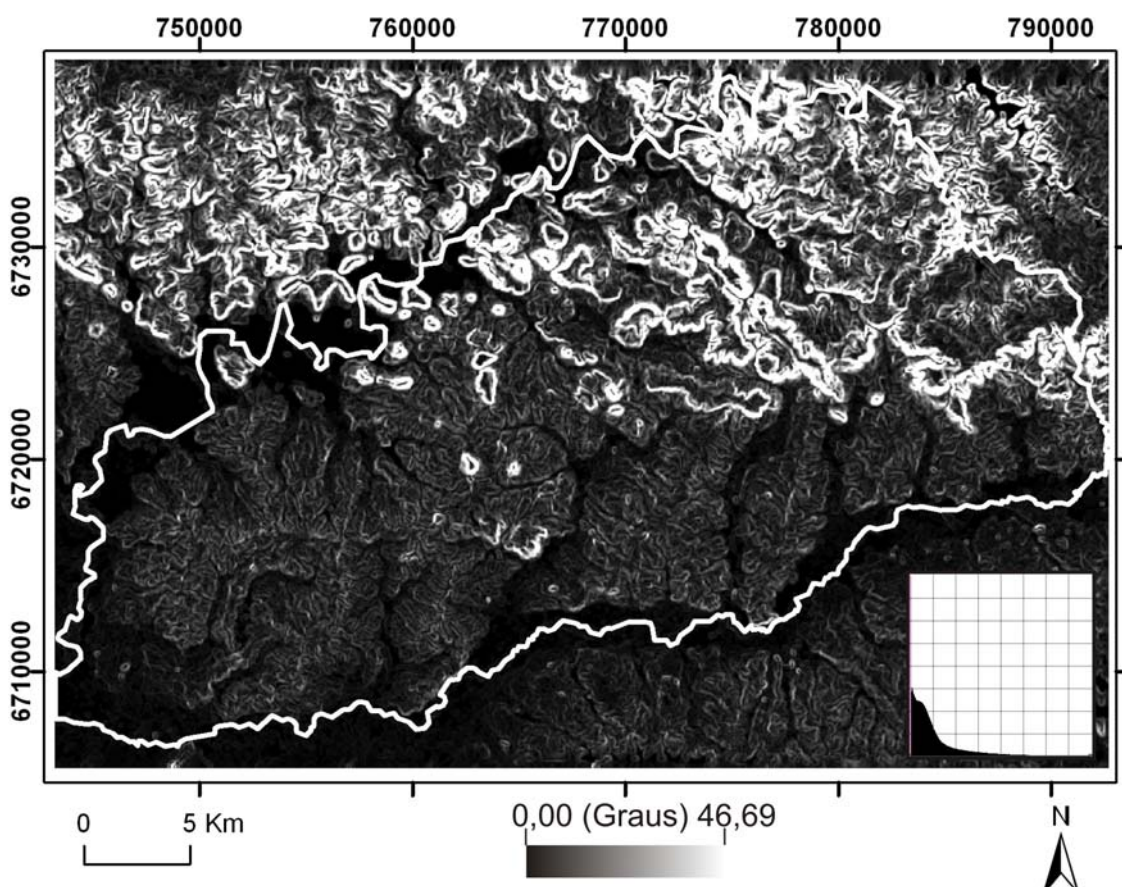
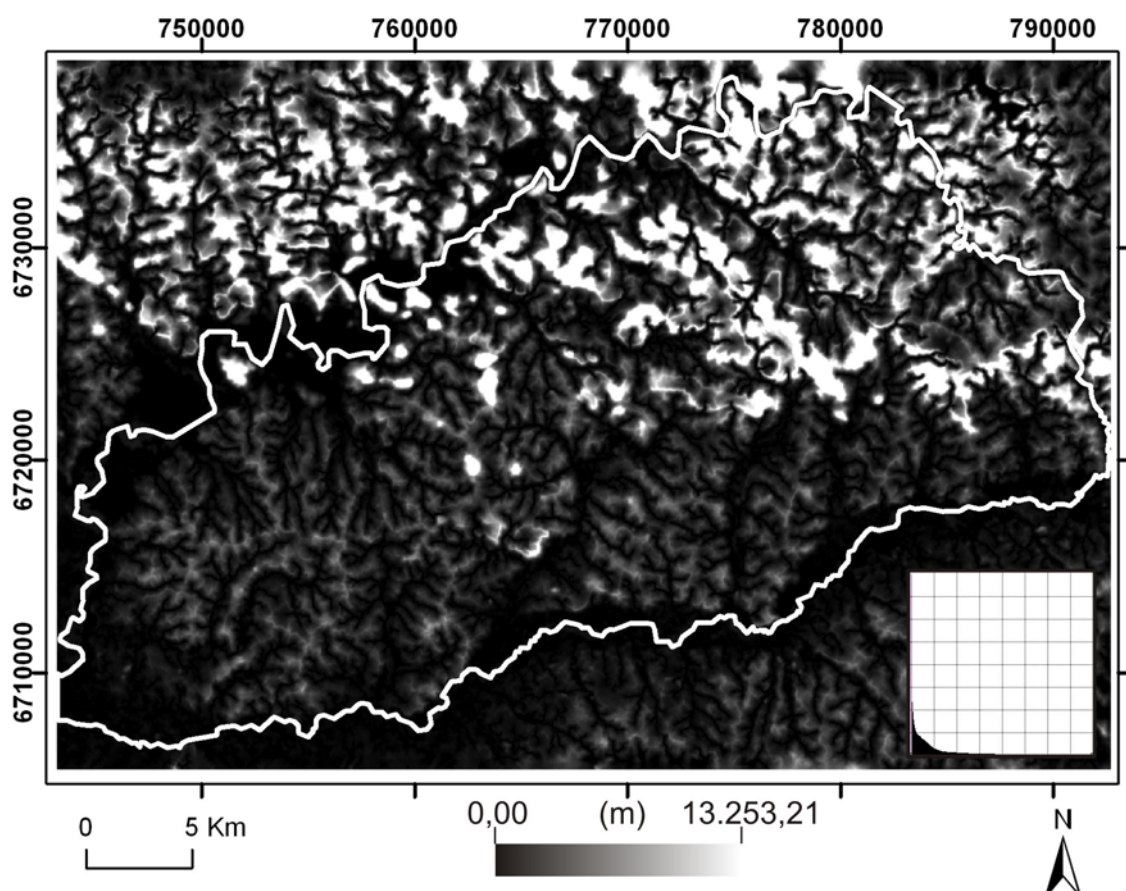


Figura 12 – Distribuição espacial da declividade na região de estudo. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Na imagem interna o histograma de frequências das declividades.

A figura 13 representa a DIST. Este atributo foi gerado utilizando-se a elevação como ponderador, buscando evitar com isso um padrão semelhante ao de um *buffer* em ambas as margens da drenagem, o que caracterizaria uma distância uniforme até a drenagem, independente da elevação do terreno. Contudo, este padrão de *buffer* pode ser verificado a sudoeste nas áreas de menores elevações,

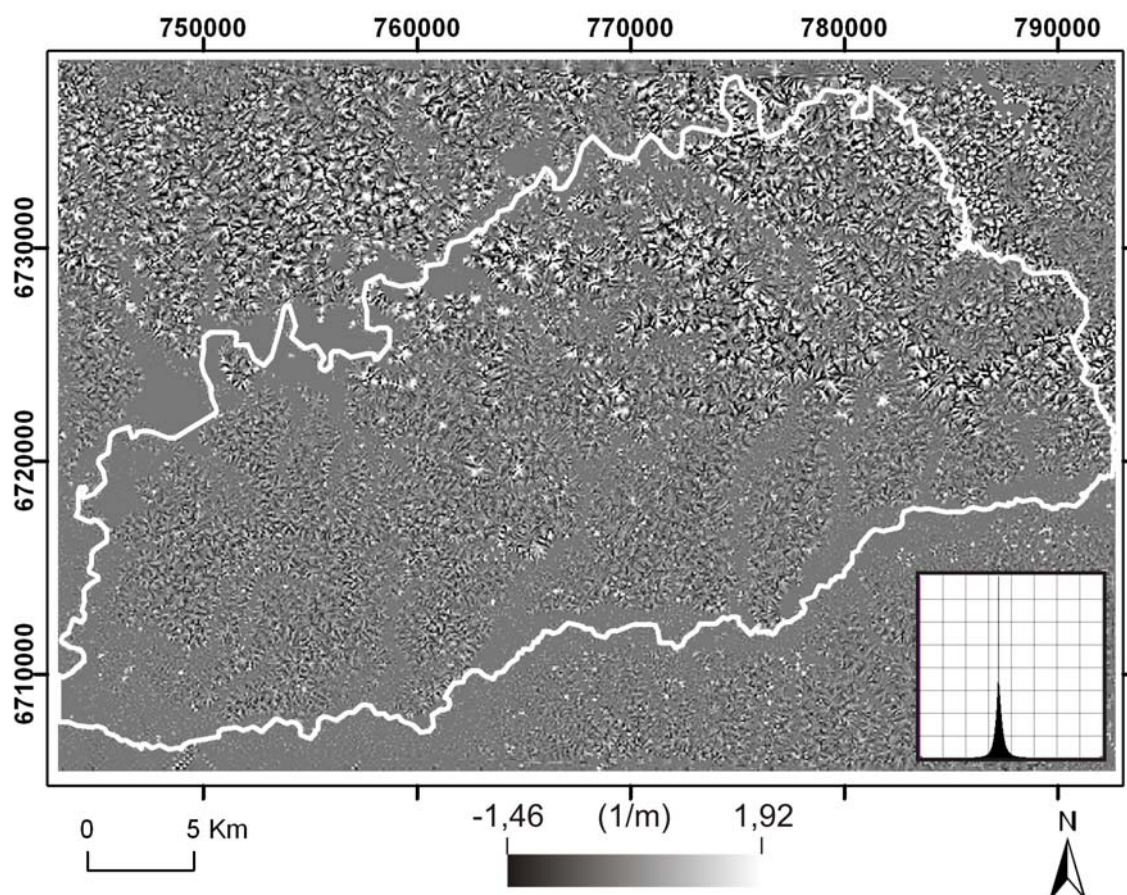
onde, devido à elevação praticamente constante do relevo, a DIST se mantém semelhante em pontos vizinhos paralelamente à rede de drenagem. Embora não tenham sido consideradas as diferentes larguras da drenagem para o cálculo deste atributo, verifica-se pelo histograma que predominam as menores DIST. Isto é um indicativo de uma região rica em canais responsáveis pelo carreamento de sedimentos e rebaixamento da paisagem, possibilitando a formação de planícies e de solos hidromórficos.



**Figura 13 – Distribuição espacial da distância a drenagem mais próxima na área de estudo. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Na imagem interna o histograma de freqüências da drenagem.**

O atributo do terreno CPLN é formado a partir da intersecção de um plano horizontal ao terreno (Figura 14), seu valor será um indicativo da divergência (curvatura positiva) ou convergência (curvatura negativa) do relevo, indicando a presença de vales e de topo de morros. Na imagem predominam valores

intermediários (tom cinza), o que também é demonstrado pela forma e posição do histograma.



**Figura 14 – Distribuição espacial da curvatura planar na área de estudo. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Na imagem interna o histograma de frequências da curvatura planar.**

A CPRF (Figura 15) é formada a partir da intersecção de um plano vertical com o relevo, seu valor indica a taxa de mudança do relevo ao longo do declive. Posições convexas na paisagem (curvatura positiva) poderão estar localizadas em locais de perda de material na paisagem. Por sua vez, as posições côncavas da paisagem (curvatura negativa), indicam locais onde o material carreado de outros locais pode estar se acumulando. Nas posições côncavas da paisagem pode haver também uma disponibilidade mais prolongada de umidade para os processos de intemperização do material de origem. A imagem da CPLN ressalta muito bem os



talvegues da paisagem, onde localiza-se a rede de drenagem e a curvatura a jusante é próxima de zero (tom cinza).

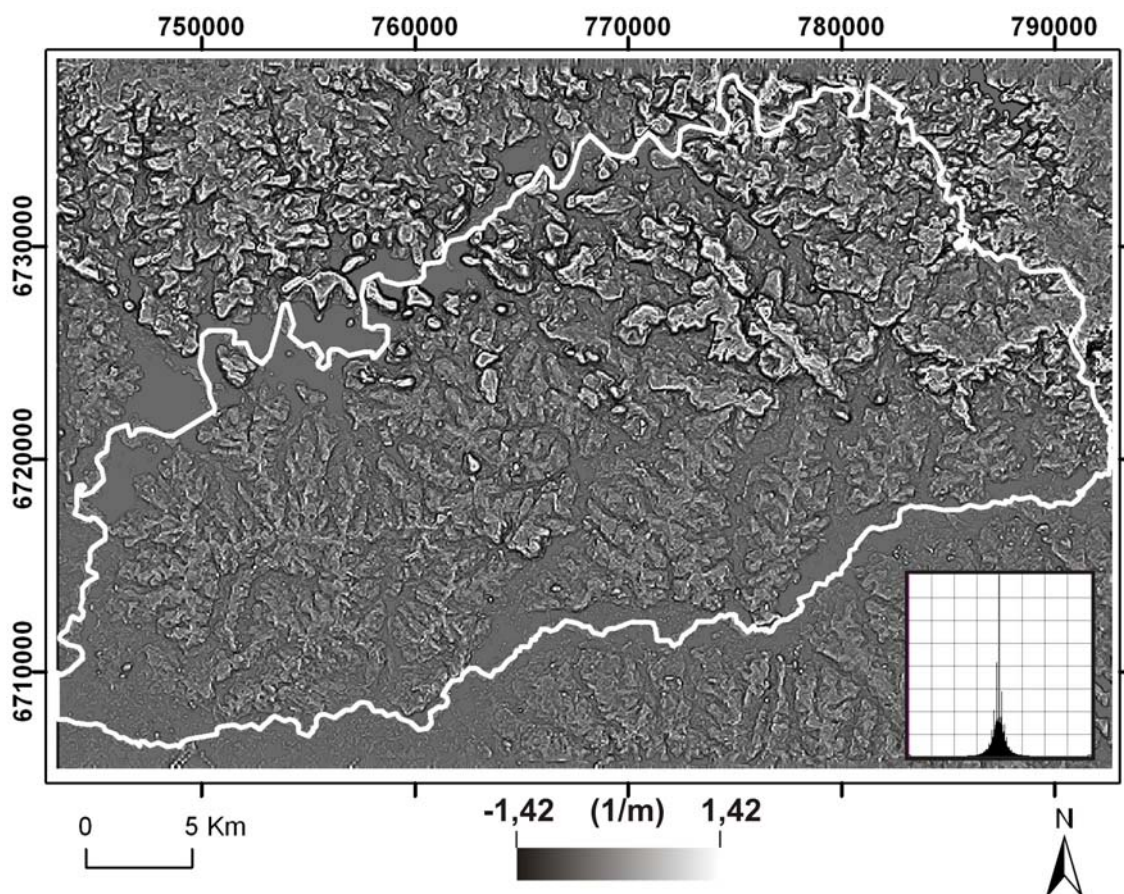


Figura 15 – Distribuição espacial da curvatura de perfil na área de estudo. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Na imagem interna o histograma de freqüências da curvatura de perfil.

O valor de ILUM representa uma estimativa relativa da disponibilidade de energia (calor/luminosidade) para os processos de intemperismo (Figura 16). Pode estar também ligada à distribuição da vegetação sob a paisagem, bem como, ao uso antrópico da terra, ambos fatores de formação do solo. Na região norte do município, ao longo do rebordo do planalto, as faces do relevo voltadas para o sul (tom preto), podem estar associadas a solos mais rasos, podendo ser uma variável com bom poder de explicação destes solos na paisagem, apesar de não ter uma relação direta tão forte com os processos de formação do solo quanto outros atributos.

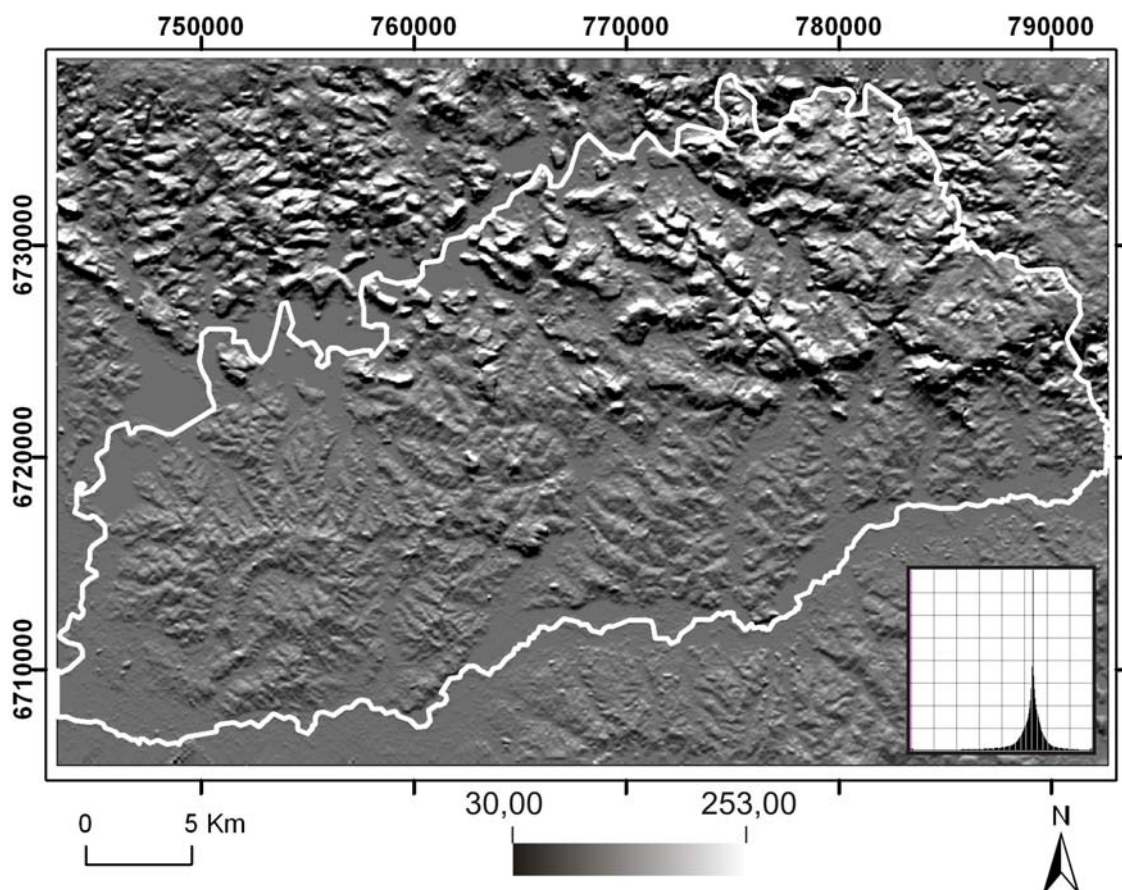


Figura 16 – Distribuição espacial da exposição da paisagem a radiação na área de estudo. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Na imagem interna o histograma de freqüências da exposição da paisagem a radiação.

A figura 17 traz a espacialização dos valores do LNAC. Este atributo representa uma área acima de cada ponto da paisagem a qual acumula material (água/sedimento) que converge para o dado ponto. Na imagem estão muito bem delineados os pontos mais elevados da paisagem, pontos mais escuros com menores valores do LNAC. Estes pontos localizam-se em topos de morros e coxilhas, podendo estar associados a solos típicos de ambientes com menor disponibilidade hídrica.



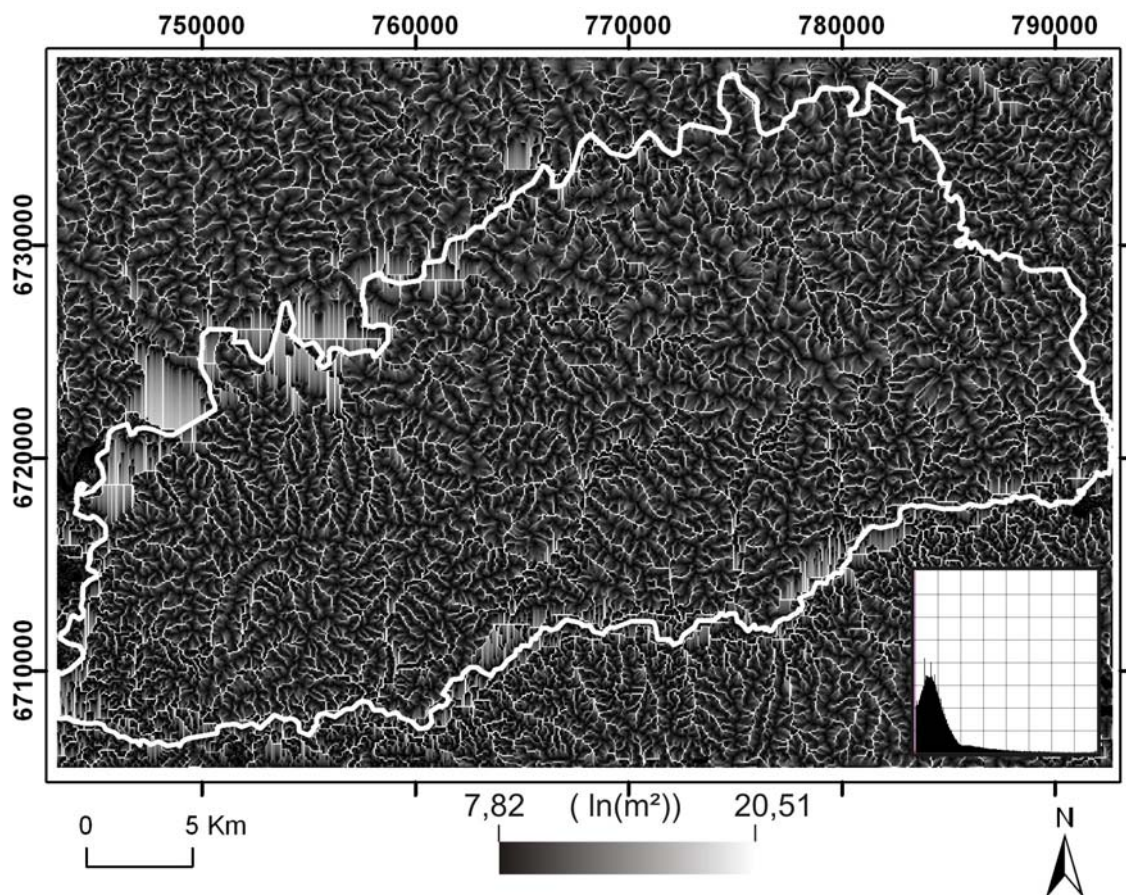


Figura 17 – Distribuição espacial do logaritmo natural da área de contribuição na área de estudo. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Na imagem interna o histograma de freqüências do logaritmo natural da área de contribuição.

A umidade disponível para os processos de formação do solo está associada ao IUT (Figura 18), este por sua vez, relacionado com os pontos da paisagem com menor declividade e com maior área de contribuição a montante. Sendo um importante indicador de locais favoráveis a formação de ambientes hidromórficos. Além de visualmente delinear a rede de drenagem com bastante fidelidade sob a paisagem, como ocorreu nas várzeas dos rios Ibicui-mirim e Toropi.

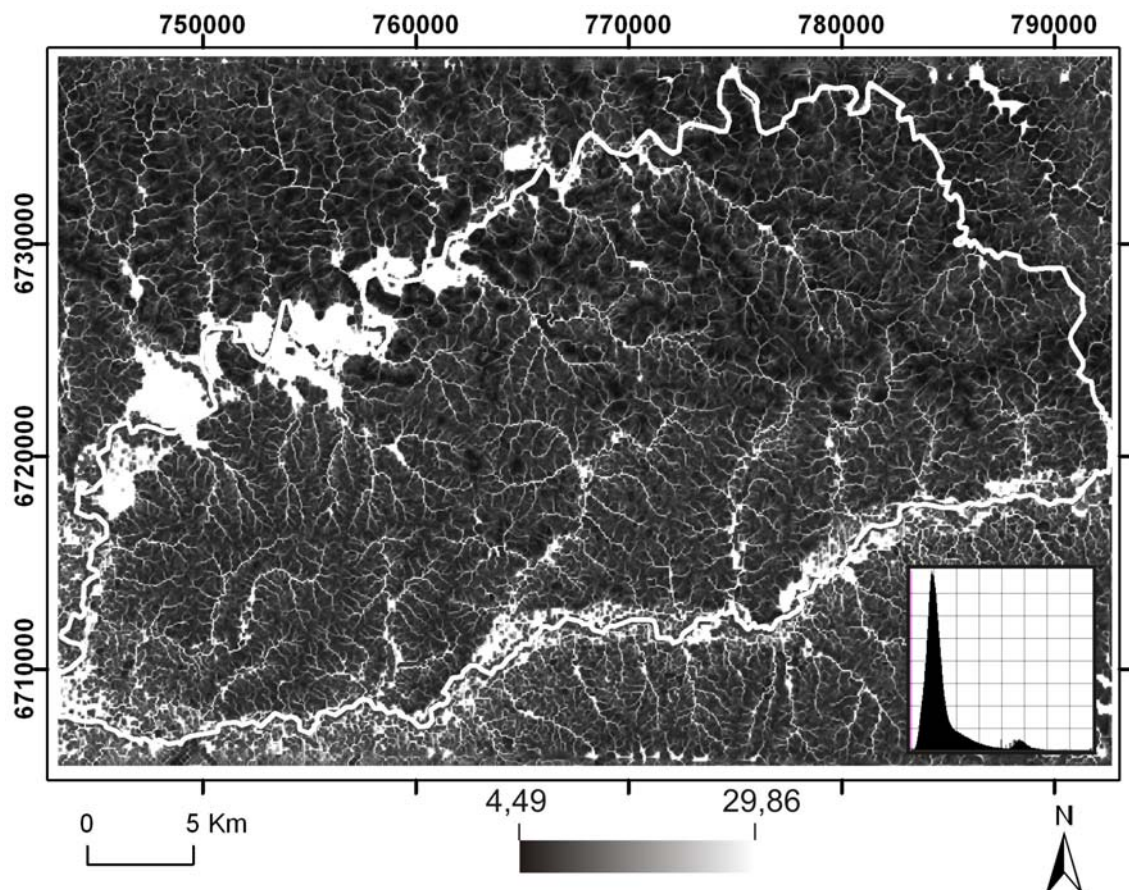
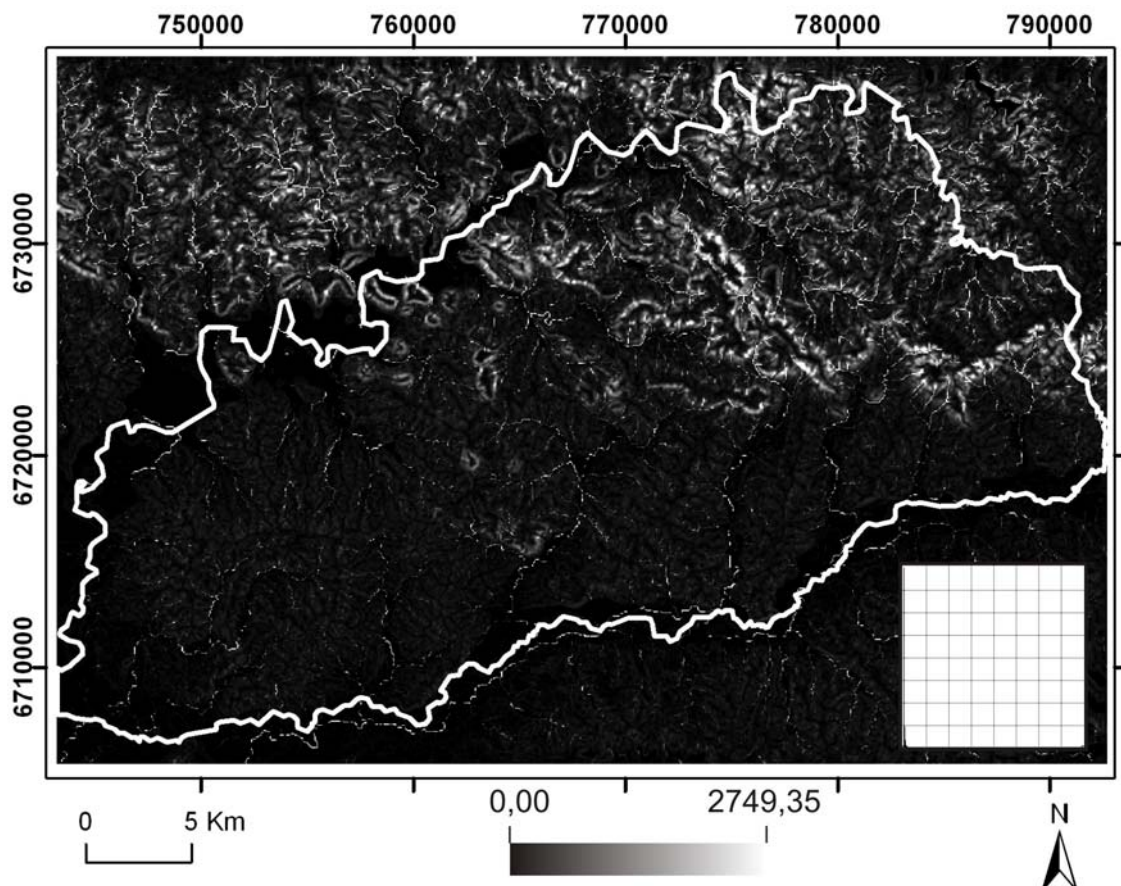


Figura 18 – Distribuição espacial do índice de umidade topográfica na área de estudo. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Na imagem interna o histograma de freqüências do índice de umidade topográfica.

A CTS (Figura 19) possui os maiores valores na região do rebordo do planalto, associado, principalmente, aos maiores valores de DECL do terreno. Maiores valores de CTS estão associados à disponibilidade de energia para o carreamento de sedimento e ao não aprofundamento da frente de formação do solo, sendo verificada a presença de solos rasos em pontos da paisagem com elevada CTS. Embora no município de São Pedro do Sul predominem valores mais baixos como indicado pelo histograma.



**Figura 19 – Distribuição espacial da capacidade de transporte de sedimento na área de estudo. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Na imagem interna o histograma de freqüências da capacidade de transporte de sedimento.**

Giasson et al. (2008) verificaram que os atributos do terreno mais recorrentes quando da utilização de regressões logísticas foram ELEV, DIST, IUT, curvaturas e DECL, os quais são atributos relacionados com acúmulo e dinâmica da água. Para Debella-Gilo et al. (2007) os atributos do terreno que mais influenciaram em modelos utilizando-se de RLM foram ELEV, DIST, tempo de exposição à radiação, DECL e IUT.

A relação existente entre as classes de solos de um fragmento do mapa Klamt, e os atributos do terreno ELEV e DECL pode ser visualizada nas figuras 20 e 21. Na figura 20 foi feita uma sobreposição das classes de solos delineadas no mapa Klamt e o atributo ELEV, a qual também foi utilizada pelos autores do levantamento de solos para delinear as unidades sob a paisagem. Devido ao fato do delineamento das classes de solos por Klamt et al. (2001) ter sido realizado em carta topográfica, pode ser verificada uma boa relação entre pontos de mesma cota e as



classes no mapa de solos. A posição dos solos hidromórficos no ponto mais baixo da paisagem, e os solos rasos nos topos dos morros tem boa correlação visual.

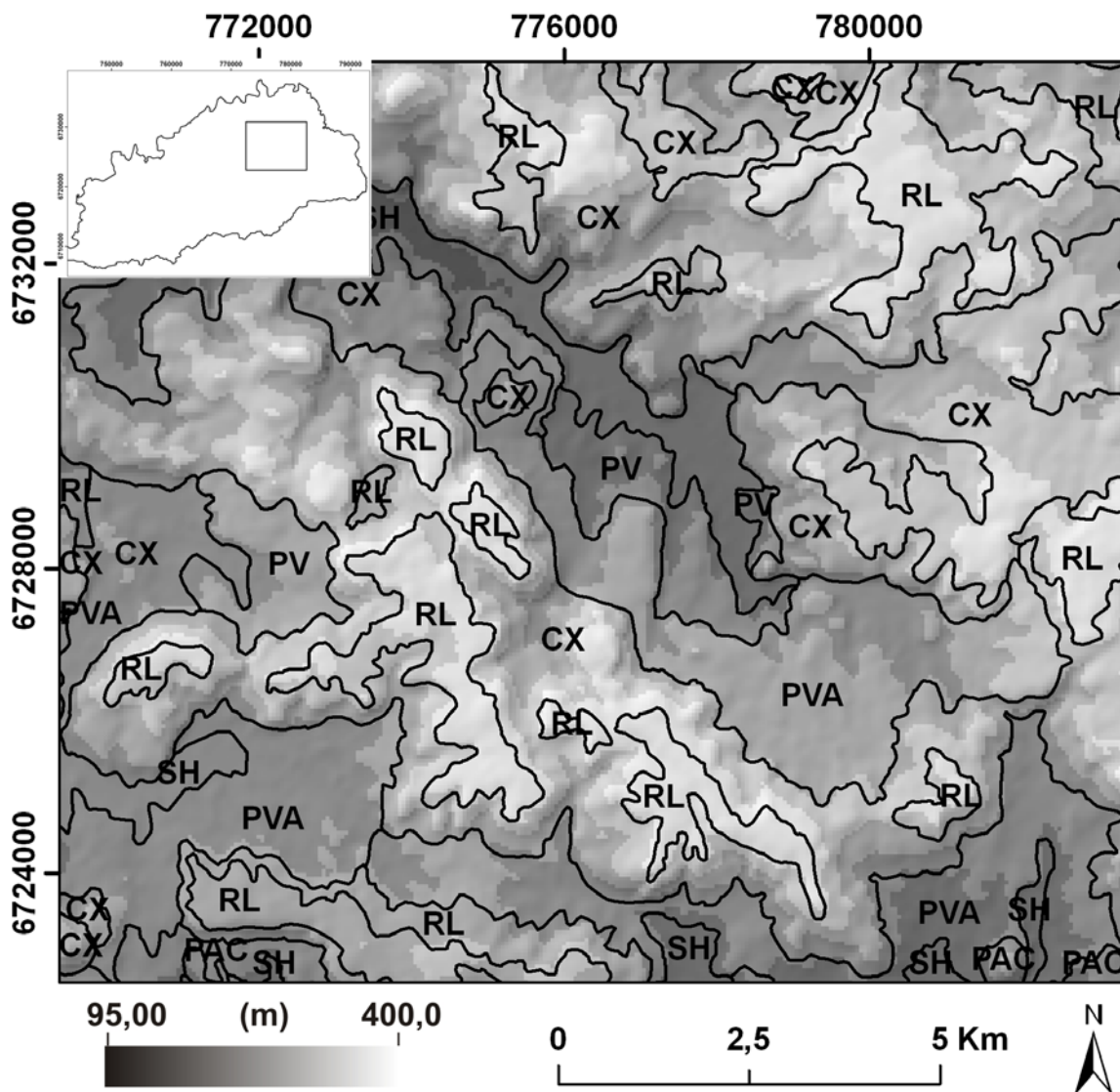


Figura 20 – Sobreposição entre as unidades de solos do mapa Klamt e o atributo ELEV em uma sessão da região de estudo. Imagem interna representa a localização da área no município. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA) e Argissolo Acinzentado (PAC).

A figura 21 traz como plano de fundo ao mapa de solos a distribuição espacial da DECL do terreno na mesma região. Nesta imagem são perceptíveis algumas incompatibilidades quanto ao delineamento da classe de solo e a presença dentro

deste mesmo polígono de diferentes declividades de terreno. Sendo esta situação atribuída ao fato de que os polígonos no levantamento foram traçados apenas tendo como referência as curvas de nível das cartas topográficas, o que pode significar a inclusão de fatores distintos em uma mesma unidade fisiográfica quanto aos fatores e processos de formação do solo. Acredita-se que o MDS, por tratar as informações no formato matricial raster e com múltiplos PI simultaneamente, possa conduzir a melhores relações entre atributos do terreno e classes de solo. Uma vez que em um SIG estas relações são feitas pixel a pixel em cada um dos PI utilizados.

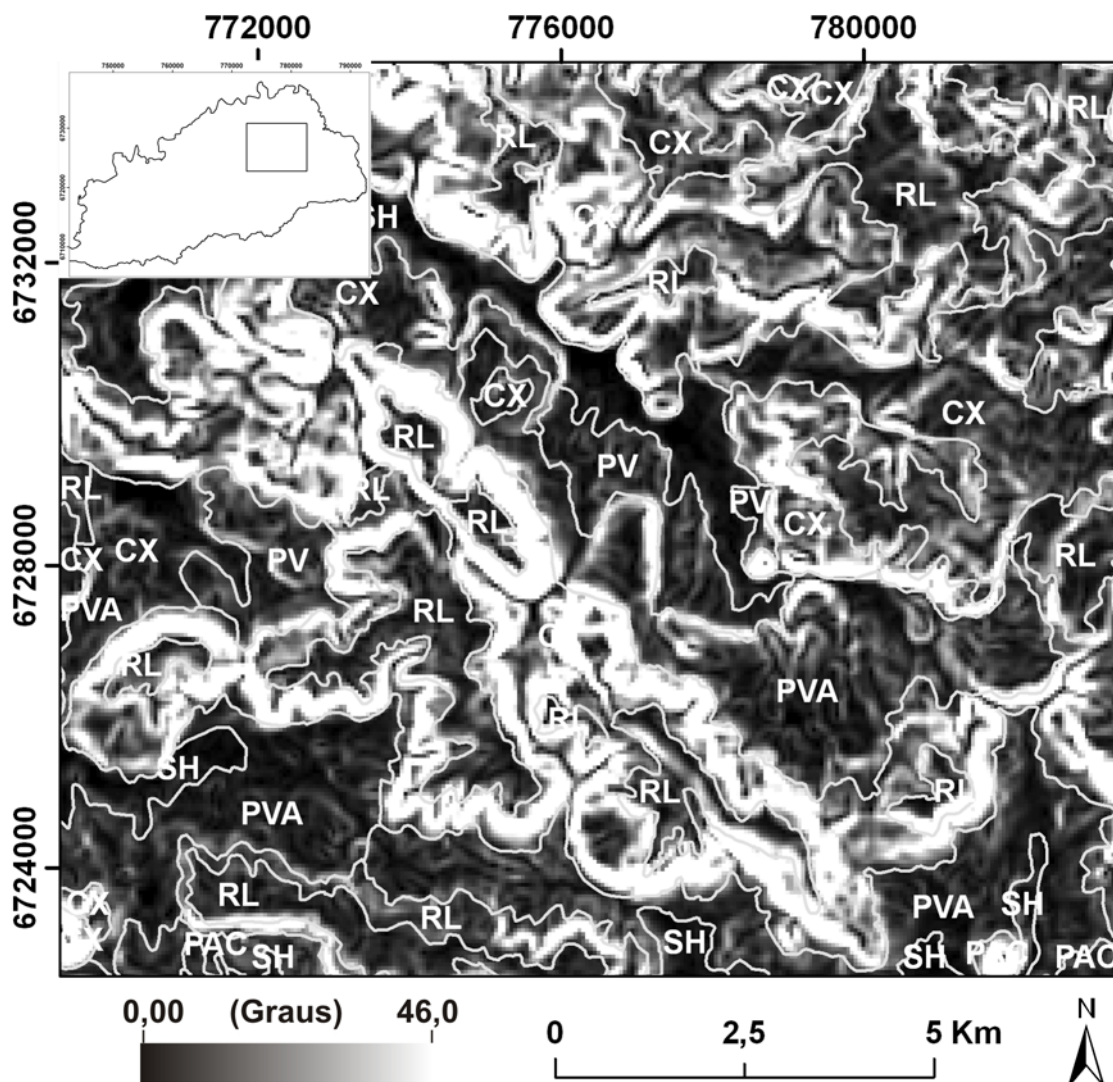


Figura 21 – Sobreposição entre as unidades de solos do mapa Klamt e o atributo DECL em uma sessão da região de estudo. Imagem interna representa a localização da área no município. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA) e Argissolo Acinzentado (PAC).

A utilização de cartas topográficas em levantamentos sistemáticos de solos, onde o delineamento dos polígonos de solo sobre a paisagem é feito a partir das curvas de nível, vem se tornando uma metodologia cada vez mais comum. Exemplos desta prática são os levantamentos executados na Serra Gaúcha (Weber et al., 2008), e vários outros levantamentos pedológicos no estado do Rio Grande do Sul (Streck et al., 2008).

#### **4.2 Análise de Componentes Principais**

Os dados de atributos do terreno não possuem uma distribuição normal, o que é preconizado pela maioria das análises estatísticas que se utilizam dos parâmetros de tendência central. Isto pode ser visualizado no baixo contraste das figuras 11 a 19 (predomínio de tons muito escuros ou muito claros), ou na forma assimétrica da distribuição dos histogramas dos atributos do terreno, demonstrando o predomínio das formas mais planas na região. Outra característica dos atributos do terreno gerados é o de possuírem correlação entre seus valores. O que de certa forma é indesejável para variáveis explicativas que serão utilizadas em modelos. Uma vez que poderão estar explicando a mesma informação na predição dos valores da variável resposta. Por isso deve-se evitar o termo “variável independente”, uma vez que poderá não haver a “independência” completa entre as variáveis explicativas. No caso dos atributos do terreno ELEV, DECL e DIST estão correlacionadas, como pode ser confirmado pelas imagens dos atributos. Assim como maiores valores do LNAC estão associados a maiores valores do IUT.

A análise das correlações entre os atributos do terreno nos 70.000 pontos utilizados para a geração dos modelos preditivos pode ser visualizada na figura 22. A diagonal principal da matriz mostra os histogramas de distribuição dos atributos do terreno nos pontos aleatoriamente gerados, os quais têm a forma muito similar aos demonstrados nas imagens das figuras 11 a 19. A diagonal inferior traz os diagramas de dispersão dos atributos analisados dois a dois. Estes gráficos retratam na forma de imagem o que está sendo reportado pelos valores e sinais das correlações na diagonal superior da matriz. As maiores correlações ocorrem entre os atributos do relevo ELEV e DECL com 0,53; ELEV e DIST com 0,64; e DECL e DIST de 0,69. Igualmente elevada é a correlação entre LNAC e IUT com 0,70. Correlações negativas também podem ser verificadas entre o IUT e ELEV, DECL e DIST, de -0,37, -0,47 e -0,40, respectivamente. As CPLN e CPRF também esboçam

uma correlação de 0,36. Praticamente inexistente correlação entre os atributos ILUM e LNAC, bem como, entre os atributos IUT e CTS. Embora todas as correlações na matriz sejam significativas ao nível de 5% de significância testado.

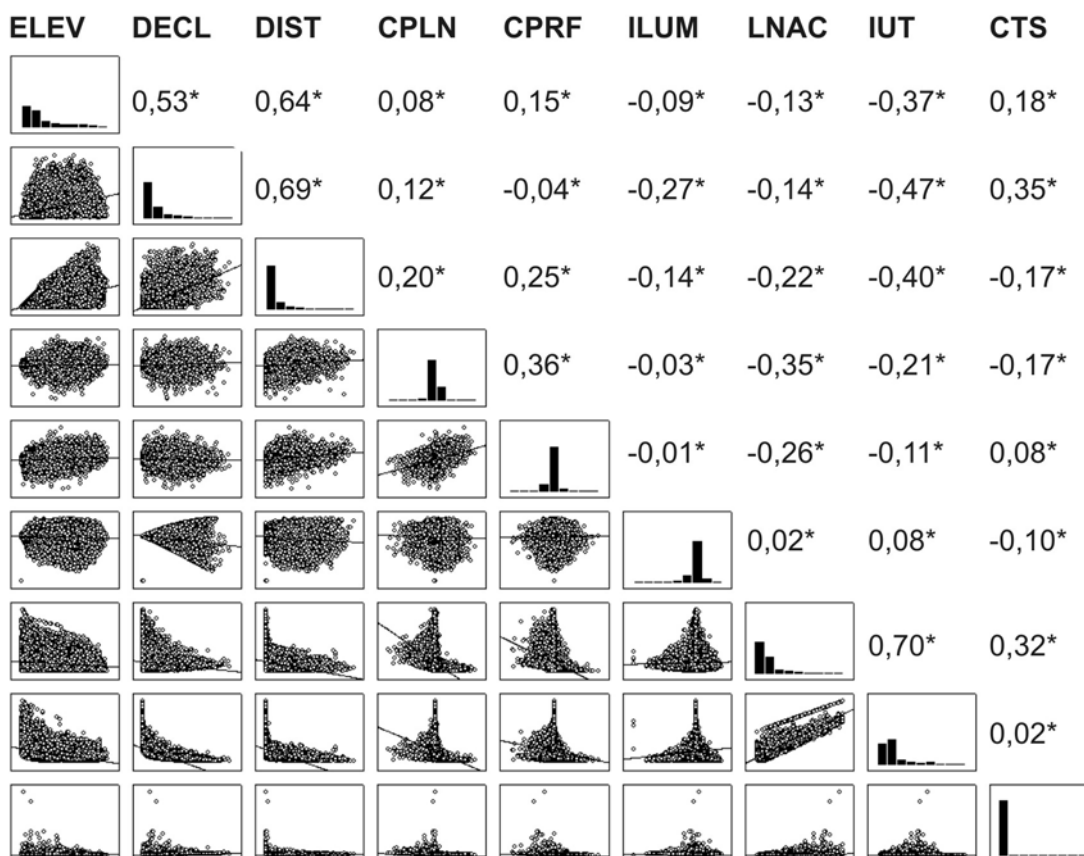


Figura 22 - Matriz de correlação entre os atributos do terreno. Diagonal inferior: diagramas de correlação. Diagonal principal: histograma de freqüências. Diagonal superior: correlações entre os atributos do terreno, nestes os valores marcados com asterisco são significativos ao nível 5%. Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Iluminação (ILUM), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).

Para a aplicação da análise de CP na matriz de correlação dos atributos, primeiramente foi verificada a adequabilidade dos dados pelo testes individual de MSA e geral KMO (Tabela 3).

**Tabela 3 – Valores de medida de adequabilidade dos atributos do terreno para análise de componentes principais. Individual *Measure of Sample Adequacy* (MSA) e geral *Kaiser Meyer Olkin* (KMO).**

MSA									KMO
ELEV	DECL	DIST	CPLN	CPRF	ILUM	LNAC	IUT	CTS	
0,80	0,67	0,70	0,73	0,58	0,67	0,56	0,62	0,67	0,66

Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Iluminação (ILUM), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).

Como preconizado pela literatura, valores de MSA abaixo de 0,5 indicam que a variável não é apropriada para a aplicação da ACP. Entre os atributos do terreno CPRF e LNAC obtiveram valores de 0,58 e 0,56, respectivamente, os quais podem ser considerados valores bastante baixos para a aplicação destas variáveis em ACP. No entanto, como o número de atributos é de apenas nove, optou-se por manter todas as variáveis. O valor de KMO do conjunto de todos os atributos foi de apenas 0,66, também podendo ser considerado baixo para aplicação da ACP, contudo ainda praticável. Vitharana et al. (2008) utilizaram CP para identificar atributos do terreno que se relacionassem com zonas de manejo em agricultura de precisão, e reportaram um valor de KMO de 0,67, muito similar ao encontrado neste estudo. Apesar disto, os autores relatam que a metodologia empregando CP auxiliou o adequado delineamento das zonas de manejo a partir de atributos do terreno.

Após a aplicação da ACP aos nove atributos do terreno foram gerados nove CP, cada uma concentrando um percentual decrescente da variabilidade dos dados iniciais (Tabela 4).

Tabela 4 – Autovalores, variância explicada, acumulada nas nove componentes principais.

<b>Componentes principais</b>	<b>Autovalor</b>	<b>Variância explicada (%)</b>	<b>Acumulado (%)</b>
<b>1</b>	<b>2,97</b>	<b>33,00</b>	<b>33,00</b>
<b>2</b>	<b>1,87</b>	<b>20,84</b>	<b>53,84</b>
<b>3</b>	<b>1,05</b>	<b>11,74</b>	<b>65,57</b>
4	0,96	10,69	76,26
5	0,71	7,88	84,14
6	0,60	6,66	90,80
7	0,40	4,46	95,26
8	0,25	2,75	98,01
9	0,18	1,99	100,00

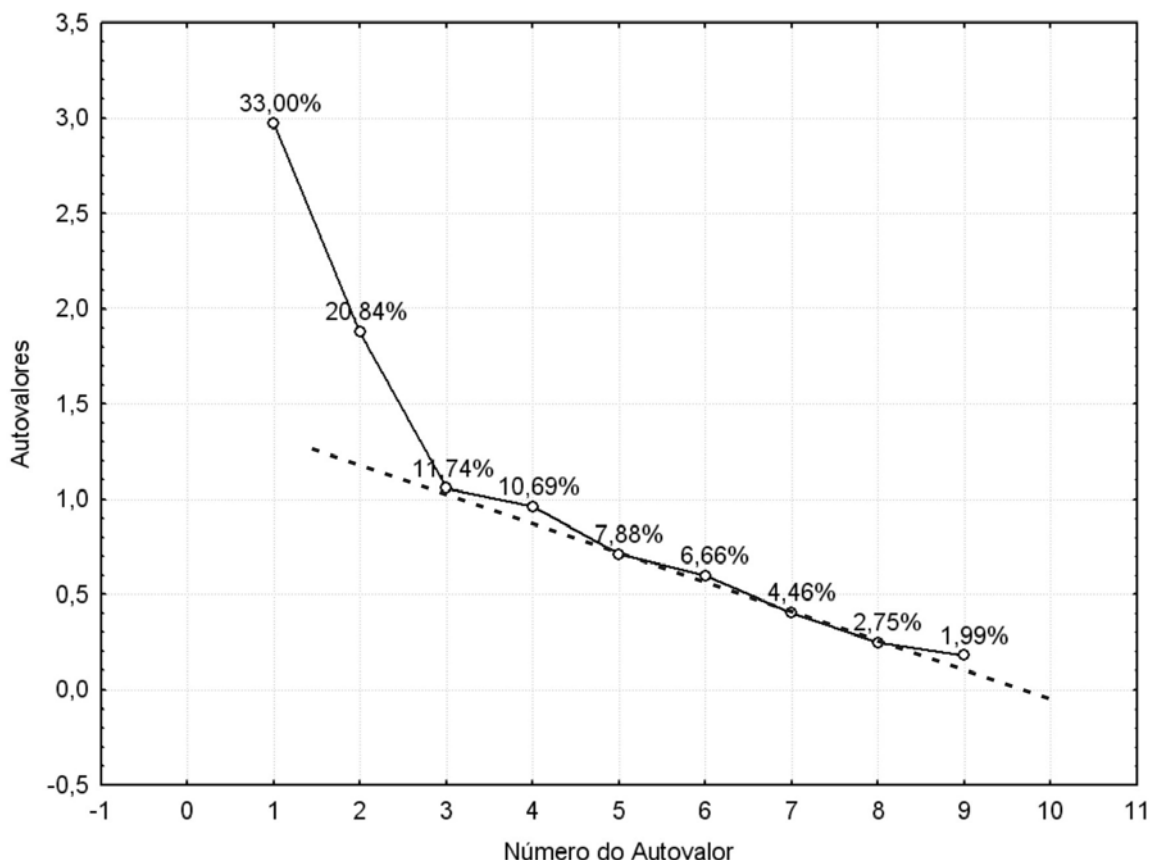
Os autovalores possuem valores superiores a um para as três primeiras CP (valores em negrito). Como a ACP foi realizada a partir de dados standardizados (média igual a zero e variância igual a um), apenas serão retidas as CP com valores de autovalor superior a um. Isto justificaria o uso da ACP para a redução do número de variáveis e aumento do poder de explicação de cada uma das novas variáveis. Quanto ao percentual de variância explicada pelas novas variáveis, verifica-se que a primeira componente é responsável por um terço da variância dos dados originais. A segunda e a terceira componentes têm juntas um poder de explicação equivalente ao da primeira componente, em torno de um terço da variabilidade total. A retenção apenas das três primeiras componentes, significa uma perda de um terço da variabilidade dos dados acumulada nas novas variáveis da quarta a nona componente. Ganha-se com a simplificação do número de variáveis, embora havendo uma perda significativa do padrão de variabilidade dos dados. Estas novas variáveis não possuem nenhum significado físico na paisagem. Não sendo possíveis relações como haviam sido feitas com os atributos do terreno, em transporte de sedimento, exposição à radiação ou direcionamento dos fluxos sob a paisagem.

Hengl (2003) utilizou ACP para mitigar o efeito de multicolinearidade entre atributos do terreno. O autor identificou que as cinco primeiras componentes foram responsáveis por cerca de 80% da variabilidade de um conjunto de nove atributos do terreno. Este resultado é muito semelhante ao reportado pela tabela 4, onde as cinco primeiras componentes acumulam cerca de 84,14% da variância. O autor aplicou também a ACP aos atributos do terreno dividindo os dados em duas subregiões, relevo de planície e de montanha. As proporções da variância explicada

pelas CP praticamente não se alteraram, sendo que até a terceira componente, em ambas as áreas, por volta de 70% da variabilidade era retida.

A proporção considerável de variabilidade distribuída entre as CP de maior valor (Tabela 4) indica que os atributos do terreno possuem um grau elevado de independência. A presença de pouca correlação entre os dados poderá tornar o uso da ACP pouco eficiente ao reduzir o número de variáveis para os modelos logísticos. Hengl; Rossiter (2003) descartaram o uso de CP devido a baixa redundância (alta proporção de variabilidade nas CP mais altas) entre os atributos daquele estudo, bem como, pelo fato de que as CP são de mais difícil interpretação e não tem nenhum significado físico, químico ou biológico. Scull et al. (2005) utilizando-se de ACP para reduzir o número de 25 variáveis iniciais, relatou que as novas variáveis (CP) não foram uma saída interessante para a utilização em árvores de decisão, as quais por si só já produzem modelos de complexa interpretação. Além disso, a concentração da variabilidade em um menor número de variáveis (sete primeiras CP), praticamente não modificou o número de ramificações das árvores geradas pela técnica, comparativamente ao emprego das 25 variáveis originais.

A figura 23 traz a imagem do *scree plot* o qual também pode ser usado para a verificação da importância das componentes mais elevadas para a explicação da variância. Este gráfico também é utilizado para a decisão de quantas componentes devem ser retidas após a ACP. Para uma aplicação mais efetiva da ACP esperava-se que a reta tracejada que intercepta estas componentes fosse mais próxima e paralela ao eixo horizontal.



**Figura 23 – Gráfico *scree plot* dos autovalores e o percentual decrescente de variância explicada.**

Os pesos dos nove atributos do terreno nas três primeiras componentes retidas podem ser visualizados nos diagramas de ordenação unitários da figura 24. Os atributos mais significativos para os 33,0% da variabilidade explicada pelo primeiro eixo são DIST, DECL e ELEV, os quais são correlacionados entre si como indica a sua proximidade dentro do círculo unitário. Da mesma forma IUT tem correlação positiva com a primeira CP. Devido à posição praticamente simétrica destas variáveis no diagrama de ordenação unitária, pode-se dizer que as mesmas são inversamente correlacionadas, o que já havia sido indicado pelos valores em torno de -0,40 na matriz de correlação da figura 22. Os demais atributos do terreno possuem pequena contribuição na variabilidade dos dados desta primeira componente.



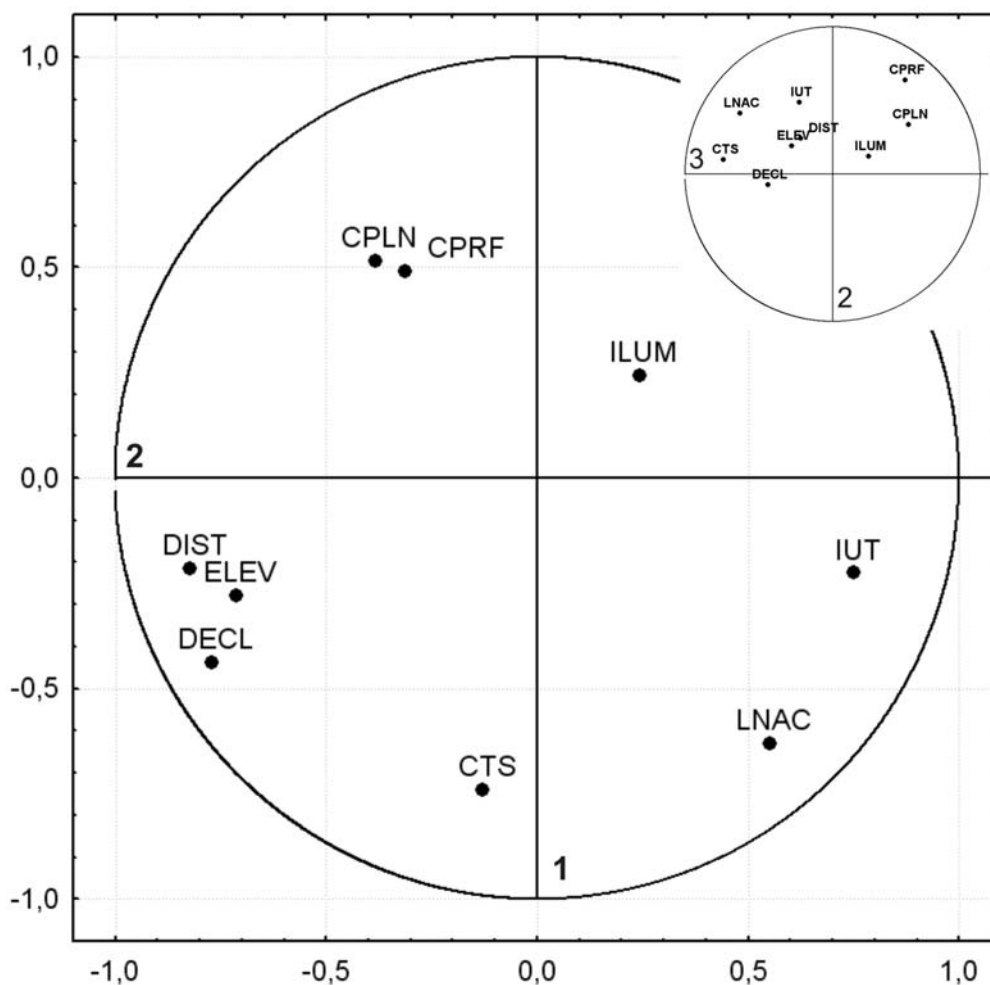


Figura 24 – Diagrama de ordenação unitário da primeira (1) e segunda (2) componente principal. Diagrama de ordenação unitária da segunda (2) e terceira (3) componente principal (círculo interno). Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Iluminação (ILUM), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).

As contribuições dos atributos do terreno na segunda componente, com uma variância acumulada de 22,84%, podem ser visualizadas nas projeções destes atributos para a linha vertical do círculo unitário. Os atributos mais significativos são a CTS e a LNAC, embora ainda com correlações bastante baixas com esta componente. Os valores de CPLN e CPRF mostraram estas variáveis bastante correlacionadas entre si nestas duas primeiras componentes. O círculo menor da figura 24 mostra as correlações dos atributos do terreno com a segunda e terceira componentes. Todos os atributos encontram-se menos dispersos e mais próximos ao centro do círculo unitário, o que era de se esperar visto que o poder de

explicação das variáveis diminui na medida em que é menor a variância ainda retida pelas componentes mais elevadas. Em todas as três primeiras componentes o atributo ILUM encontra-se próximo ao centro do círculo unitário, o que é um indicativo de sua pouca relação com as novas variáveis geradas.

Visando a potencialização das relações entre as componentes retidas e os atributos do terreno, foi aplicado o método de rotação ortogonal VARIMÁX aos eixos das componentes. Os resultados das novas correlações entre atributos do terreno e as componentes constam na tabela 5. Os valores mais significativos em cada componente estão em negrito.

**Tabela 5 – Pesos dos nove atributos do terreno nas três primeiras componentes principais (CP) retidas após a rotação por VARIMÁX. A comunalidade indica o total de variabilidade de cada variável explicada pelas três componentes retidas.**

Atributo de terreno	CP			Comunalidade
	1	2	3	
<b>ELEV</b>	<b>0,756</b>	-0,074	0,218	0,63
<b>DECL</b>	<b>0,870</b>	-0,184	-0,053	0,79
<b>DIST</b>	<b>0,810</b>	-0,131	0,327	0,78
<b>CPLN</b>	0,038	-0,264	<b>0,673</b>	0,52
<b>CPRF</b>	0,027	-0,019	<b>0,865</b>	0,75
<b>ILUM</b>	-0,325	0,062	0,151	0,13
<b>LNAC</b>	-0,029	<b>0,904</b>	-0,231	0,87
<b>IUT</b>	-0,419	<b>0,822</b>	-0,010	0,85
<b>CTS</b>	0,549	0,420	-0,309	0,57

Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Iluminação (ILUM), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).

A primeira componente poderia ser denominada de **componente de relevo** por ter uma maior relação com ELEV, DECL e DIST. A segunda poderia ser denominada de **componente hidrológica**, devido aos valores mais significativos de correlações com as variáveis ligadas ao movimento da água na paisagem como LNAC e IUT. A terceira componente pode ser denominada de **componente forma do relevo**, devido aos maiores pesos dos atributos CPLN e CPRF. A proporção da variância, de cada atributo, explicada pelas três primeiras componentes pode ser

visualizada na coluna das comunalidades. Valores próximos a um indicam que o atributo foi bem contemplado pelas componentes retidas. Valores baixos como 0,52 (CPLN), e até mesmo 0,18 (ILUM) indicam que estas variáveis foram menos contempladas pelo novo conjunto de variáveis. Os baixos valores de comunalidades estão relacionados com o fato de que as três primeiras componentes explicam juntas apenas 65,57% da variabilidade dos dados originais. Sendo que mais de um terço da variância dos atributos originais será perdida com a nova estrutura de dados. É questionável o fato da rotação VARIMÁX aplicada ter realmente simplificado a estrutura das componentes. Uma vez que neste estudo estão sendo utilizadas apenas nove variáveis, que por si só já caracterizam uma estrutura simplificada. Embora, como se observa na tabela 5, para cada CP temos maiores pesos (em módulo) associados a algumas variáveis, e valores próximos a zero associados as demais variáveis.

Estudos como os realizados por Debella-Gilo et al. (2007) e Giasson et al. (2008) apontaram que os atributos do terreno como ELEV e DECL tiveram grande importância para os modelos logísticos, no entanto, os autores não fazem nenhuma menção quanto a multicolinearidade das variáveis. Os dados gerados neste estudo e analisados por CP demonstram que estas variáveis foram correlacionadas. Utilizando-se das informações da figura 24 e da tabela 5, poder-se-ia optar por utilizar apenas as informações da variável DECL, LNAC e CPRF. Embora modelos apenas com estas três variáveis não tenham sido executados neste estudo.

### **4.3 Mapeamento pedométrico a partir do mapa Klamt**

#### **4.3.1 Mapa predito ao nível de ordem**

##### **4.3.1.1 Componentes principais como variáveis explicativas**

Estudos científicos que se utilizam das RLM aplicam a regra prática de uma proporção mínima de 10 amostras para cada categoria (classe) a ser predita (10:1) para a geração dos modelos. Esta regra prática foi cumprida em todos os conjuntos de dados utilizados para formulação das funções logit neste estudo. Algumas publicações também utilizam a regra de um conjunto mínimo de 100 amostras para cada classe a ser predita quando a estimativa dos parâmetros da função logit é feita por máxima verossimilhança (PENG et al., 2002). A quantidade de observações usadas neste estudo também superou este limite mínimo.

Tomando-se as CP como variáveis explicativas e as classes de solos ao nível de ordem como variáveis resposta, foram aplicados os modelos logísticos múltiplos para a espacialização da probabilidade de encontrar as classes de solos na paisagem. A tabela 6 traz os valores de intercepto e coeficientes dos modelos logísticos.

**Tabela 6 – Valores dos interceptos e dos coeficientes das componentes principais nos modelos logísticos desenvolvidos ao nível de ordem.**

	C <sup>1</sup>				
	R	SH	P	N	F
<b>Intercepto</b>	-1,168	-7,805	-0,562	-2,612	-8,104
<b>CP1</b>	-0,002	-0,115	-0,058	-0,013	-0,127
<b>CP2</b>	-0,005	0,042	-0,032	0,005	-0,056
<b>CP3</b>	0,002	0,288	0,127	0,032	0,287

<sup>1</sup>Classe de referência. Todos os valores da tabela são significativos ao nível de 5%. Todos os modelos apresentaram diferença significativa para o teste dos modelos apenas com o intercepto e os mesmos com as variáveis explicativas. Todas as variáveis explicativas têm relação significativa com a variável resposta (*Likelihood Ratio Test*). Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N), Plintossolo (F), Primeira Componente Principal (CP1), Segunda Componente Principal (CP2) e Terceira Componente Principal (CP3).

Utilizando-se o intercepto e os coeficientes da tabela acima definem-se os parâmetros da função logit para a classe Neossolos como sendo (Equação 13):

$$Logit_R = e^{-1,168-0,002*CP1-0,005*CP2+0,002CP3} \quad (13)$$

De forma similar procedeu-se os cálculos dos logit para o restante das classes de solos da tabela 6, exceto para a classe referência. Para o cálculo da probabilidade de encontrar a classe Neossolos sob a paisagem utilizou-se a equação 14.

$$Pr ob_R = \frac{Logit_R}{(1+Logit_{SH} + Logit_P + Logit_N + Logit_F)} \quad (14)$$

Onde cada função logit utiliza os valores da tabela 6 como parâmetros para os modelos. O cálculo da probabilidade de encontrar a classe de referência (Cambissolos) na paisagem foi determinado pela equação 15.

$$Pr ob_C = 1 - (Logit_R + Logit_{SH} + Logit_P + Logit_N + Logit_F) \quad (15)$$

A escolha da classe dos Cambissolos como referência não teve qualquer motivação especial. A influência de cada uma das classes de solos como referência na fase de determinação dos parâmetros das equações logit pode inclusive ser estudada em futuros trabalhos utilizando-se das RLM.

Como resultado deste procedimento, foram gerados PI para as probabilidades da presença das classes de solos na paisagem. Estes planos foram reunidos em um único plano, o qual contém os maiores valores de probabilidades gerados espacializados na paisagem (Apêndice C). Na seqüência foi associada a cada valor de probabilidade a classe de solo da função logit utilizada, gerando-se o mapa de solos propriamente dito. A tabela 7 indica em uma amostra de 15 pontos como este procedimento foi executado em todos os pixels da área de estudo.

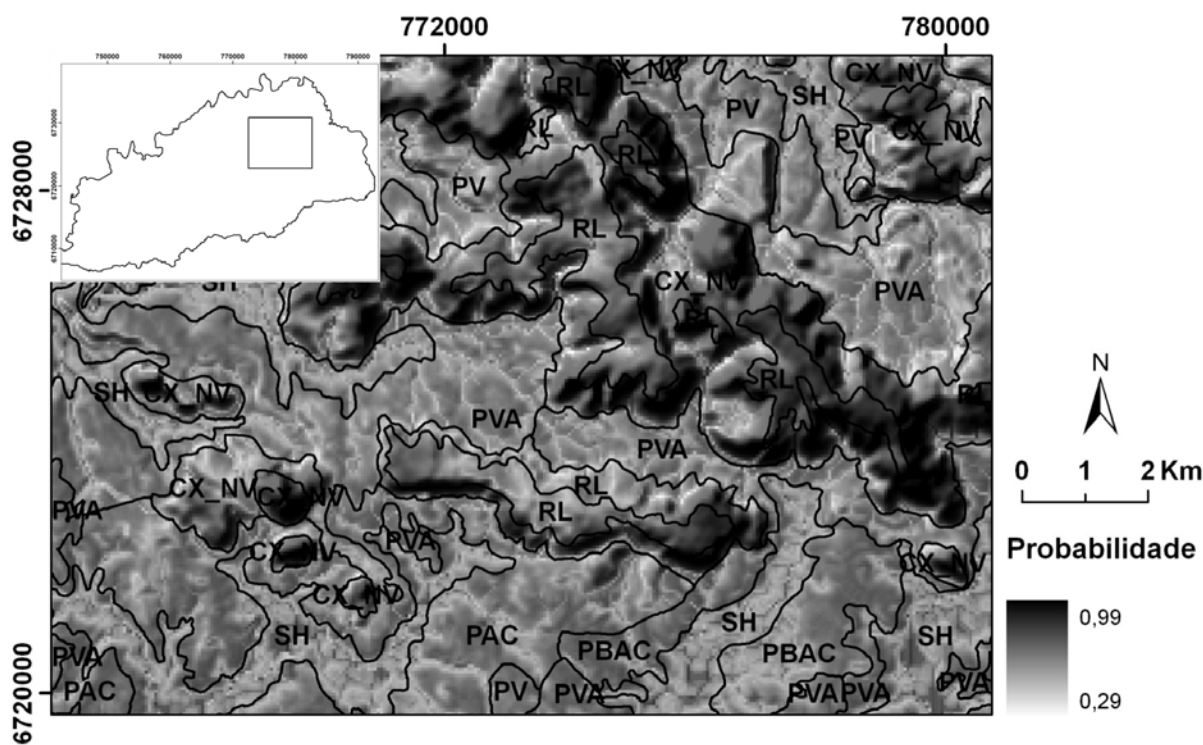
**Tabela 7 – Cada classe de solo corresponde a um Plano de Informação (PI) com um valor de probabilidade. O maior valor encontra-se hachurado e define a classe de solo no PI da Classe Predita (Cl.Pr.) com a respectiva probabilidade (Probab.) em outro PI.**

Pixel	Coordenadas (m)		Probabilidades						Cl.Pr.	Probab.
	X	Y	R	SH	P	N	F	C		
1	776645,070	6713567,840	0,040	0,004	0,680	0,010	0,020	0,220	P	0,680
2	779918,930	6732757,760	0,090	0,000	0,010	0,010	0,000	0,880	C	0,880
3	746078,190	6715250,380	0,003	0,560	0,350	0,003	0,060	0,010	SH	0,560
4	772881,930	6729357,200	0,040	0,000	0,001	0,003	0,000	0,950	C	0,950
5	755116,820	6715046,600	0,010	0,100	0,730	0,010	0,040	0,080	P	0,730
6	758234,950	6714347,110	0,004	0,500	0,420	0,004	0,040	0,010	SH	0,500
7	776908,660	6725414,680	0,110	0,000	0,070	0,020	0,000	0,770	C	0,770
8	769785,230	6721120,670	0,003	0,570	0,350	0,003	0,050	0,009	SH	0,570
9	766800,290	6721511,170	0,030	0,020	0,720	0,010	0,010	0,160	P	0,720
10	746752,680	6719858,060	0,002	0,610	0,300	0,002	0,070	0,006	SH	0,610
11	747692,830	6713448,370	0,010	0,060	0,760	0,010	0,060	0,070	P	0,760
12	778837,320	6726794,270	0,140	0,000	0,002	0,040	0,000	0,800	C	0,800
13	774080,650	6717485,130	0,040	0,005	0,660	0,020	0,020	0,240	P	0,660
14	762842,510	6713845,180	0,007	0,510	0,410	0,006	0,020	0,020	SH	0,510
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
656355	761770,000	6719507,150	0,010	0,500	0,410	0,009	0,010	0,030	SH	0,500

Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N) e Plintossolo (F).

No mapa de probabilidades não são encontrados valores iguais a zero (0%) ou um (100%) (Apêndice C). Disso se infere que os modelos não afirmam com absoluta certeza que em um determinado ponto da paisagem será, ou não será encontrada determinada classe de solo. Os maiores valores de probabilidade estão ligados aos solos encontrados em regiões bem distintas, quanto aos atributos do terreno utilizados nos modelos, as quais são a região de encosta do rebordo e a de

várzea ao longo das drenagens do Rio Toropi e Ibicui-mirim. A região de encosta também coincide com a região da classe dos Cambissolos, que haviam sido usados como classe de referência para as regressões logísticas. A figura 25 confirma a relação entre os delineamentos dos Cambissolos e as maiores probabilidades.



**Figura 25 – Polígonos das unidades de solos no mapa Klamt sobreposto ao mapa de probabilidade. Imagem interna representa a localização da área no município. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA), Argissolo Acinzentado (PAC) e Argissolo Bruno-Acinzentado (PBAC).**

Os PI representando a probabilidade de encontrar cada uma das classes de solos têm uma boa relação visual com o que se esperava encontrar pela relação solo-paisagem para a região (Apêndice D). As maiores probabilidades de encontrar os solos de várzea se deu nos locais mais baixos, onde a presença de água é um importante fator no processo de formação de solos como Gleissolos, Planossolos e Plintossolos. Nos topos dos morros e na encosta do planalto observa-se as maiores probabilidades de encontrar as classes dos Neossolos Litólicos e Cambissolos, onde, especialmente para o primeiro, o processo de remoção constante do material intemperizado não permite que ocorra a formação de um perfil mais desenvolvido.

Nas áreas de coxilha, intermediárias entre várzeas e encostas de morros, encontram-se as maiores probabilidades para os Argissolos. Solos com horizontes mais espessos e bem drenados. Uma relação semelhante entre as probabilidades espacializadas e cada uma das classes de solo na paisagem foi descrita por Debella-Gilo et al. (2007) em solos Noruegueses. Os autores não geraram um mapa de solos, mas consideraram bastante positivos os resultados visuais entre a relação estabelecida visualmente para as classes de solos em mapas existentes e os valores das probabilidades gerados.

Sugere-se que estes mapas de probabilidades associados às classes de solos possam ser utilizados como indicadores de qualidade da informação. Este dado poderia ser utilizado como peso das respectivas classes de solos em modelagens nas quais características ou propriedades do solo estejam sendo preditas, como carbono orgânico no caso de um inventário de gases de efeito estufa, ou na modelagem de perdas de solo por erosão. O uso dos mapas de probabilidade precisa ser investigada em futuros estudos.

O mapa de solos gerado não espacializou as classes Neossolos, Plintossolos e Nitossolos (Figura 26). As classes Cambissolos, Argissolos e os Solos Hidromórficos foram espacializadas visualmente de maneira muito semelhante ao que se encontra no mapa Klamt, e na relação solo-paisagem da área de estudo. A figura 27 traz as unidades delineadas do mapa Klamt sobrepostas a uma sessão do mapa predito. Nesta imagem pode-se perceber melhor o que foi dito acima. Os Solos Hidromórficos foram espacializados nas partes mais baixas da paisagem. Os Cambissolos nas encostas e topos de morros. Ao passo que os Argissolos foram distribuídos nas coxilhas, e a classe dos Cambissolos foi atribuída a regiões onde encontram-se os Neossolos. Supõe-se que a motivação para a não espacialização das classes Neossolos, Plintossolos e Nitossolos tenha sido devido a pouca representatividade destas classes no total das amostras utilizadas nos modelos logísticos. Estas classes correspondem somente a cinco, três e um % do total das 70.000 amostras randomicamente selecionadas para a geração dos modelos.

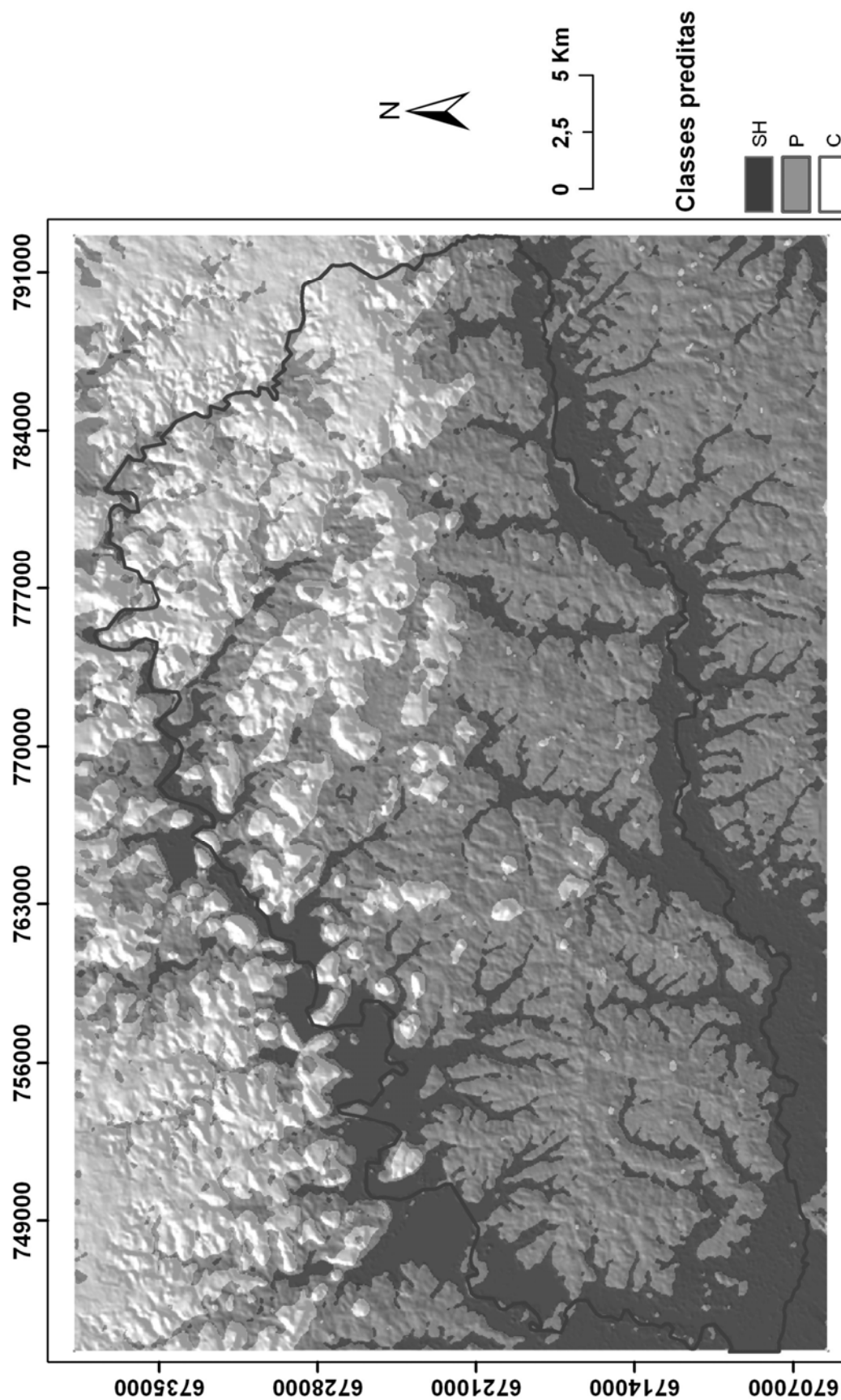


Figura 26 – Mapa de solos predito utilizando-se de componentes principais e o mapa Klamt ao nível de ordem. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Solos Hidromórficos (SH), Argissolo (P) e Cambissolo (C).



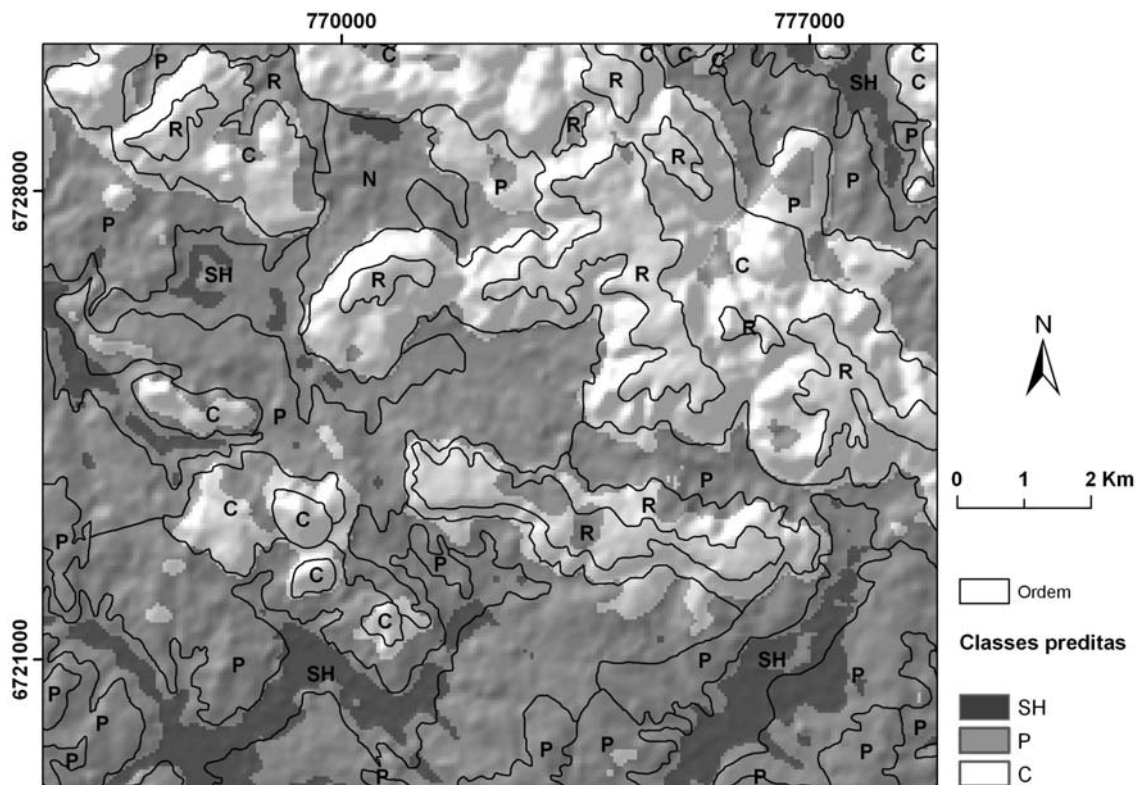


Figura 27 – Unidades de solos sobrepostas ao mapa predito utilizando-se de componentes principais e o mapa Klamt ao nível de ordem. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Solos Hidromórficos (SH), Argissolo (P) e Cambissolo (C).

Para melhorar a capacidade dos modelos em predizer os solos ao nível de série, Bailey et al. (2003) desconsideraram as classes com área menor do que 5% da área total. Real et al. (2006) afirmam que as regressões logísticas geram os valores de probabilidades baseadas não somente nos valores dos preditores, mas também na proporção relativa entre as classes das variáveis resposta. O que tem levado pesquisadores a propor metodologias de ponderamento entre as frequências de ocorrência de cada uma das variáveis preditas. Neste estudo nenhuma transformação dos dados foi realizada, assumindo-se o risco de que os resultados possam assumir valores tendenciosos em favor das classes com as maiores frequências, como foi reportado por Debella-Gilo et al. (2007).

Hengl et al. (2007) estudaram a potencialidade de emprego de RLM para a interpolação de classes de solos em todo o território do Irã a partir de 5.200 perfis descritos. Os pesquisadores reportaram que os modelos logísticos foram bastante dependentes de uma forte correlação entre os preditores e as classes de solos. Ao ponto de que se esta relação não fosse expressiva, os modelos não espacializavam

adequadamente as classes de solos. Analisando os resultados do estudo daqueles autores, verifica-se que as classes de solos com o menor número de perfis descritos, não foram corretamente preditas ( $K = 0$ ) pelos modelos logísticos. Confirmando a idéia de que todas as classes devem estar mapeadas em uma proporção relativa mínima, para que seus dados tabulados sejam submetidas à RLM sem problemas de tendência dos resultados em favor das classes proporcionalmente mais representativas.

Devido ao mapa ter sido gerado em um ambiente de sistema de informações geográficas trabalhando com arquivos matriciais, a menor unidade de mapeamento utilizada é o pixel, neste caso, com 50 m de resolução. Verifica-se no mapa que existem pixels isolados, definindo classes de solos em meio a outras classes de solos. Estes, devido à resolução de trabalho, e a escala em que os mapas venham a serem utilizados, podem ser considerados como inclusões, sendo dissolvidos nos polígonos vizinhos até o tamanho mínimo da área mínima mapeável.

A análise da tabela 8 permite verificar que, do ponto de vista do mapeador, apesar de três classes não terem sido acuradamente espacializadas, as demais classes obtiveram valores bastante satisfatórios. Chegando a determinar que o mapa predito espacializa corretamente 85% dos pontos amostrados para a classe Argissolos. Do ponto de vista do usuário esta mesma classe de solos poderá ser encontrada em mais de 76% dos locais. A acurácia geral do mapa predito foi de 74,3%, o que pode ser considerado bom, embora não considere os pontos corretamente mapeados devido ao acaso. Para contornar isso, considera-se o valor do índice K de 58,20% como uma medida mais realista da qualidade do mapa predito. Podendo ser afirmado então, que ao nível de ordem, o modelo preditivo corretamente reproduz o mapa Klamt em aproximadamente 60% de seu padrão original.

Tabela 8 – Matriz de erros da área externa, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de ordem produzido a partir das componentes principais como variáveis dos modelos logísticos.

Mapa predito	Mapa Klamt						Total	AU (%)	
	R	SH	P	N	F	C			
R	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
SH	0,03	17,92	4,69	0,01	1,55	0,12	24,33	73,65	
P	0,86	4,93	39,72	0,57	1,15	4,42	51,66	76,89	
N	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
F	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
C	4,68	0,12	2,28	0,19	0,09	16,66	24,02	69,36	
<b>Total</b>	5,57	22,97	46,69	0,77	2,79	21,20	100,00		
<b>AM (%)</b>	0,00	78,01	85,07	0,00	0,00	78,58			
<b>AG (%)</b>	74,30		K <sup>1</sup> (%)			58,20			

<sup>1</sup>erro padrão assintótico =  $\pm 0,009$  e intervalo de confiança 95% = 0,56...0,60. Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N) e Plintossolo (F).

Os erros mais comuns na matriz foram entre classes espacialmente próximas quanto ao delineamento no mapa (Figura 28). Os solos de várzea foram confundidos com os Argissolos, estes por sua vez foram erroneamente espacializados nas posições dos Cambissolos, os quais confundiram-se com os Neossolos. Estes erros da matriz podem ter origem nas bordas de cada classe de solos. A inferência da verdadeira classe, sob o ponto de vista do mapa Klamt, pode ser bastante difícil para os modelos, devido a problemas no próprio delineamento que serviu de treinamento. Ou ainda, devido a diferenças muito tênues entre os atributos do terreno, os quais podem não apresentar nenhum tipo de gradiente na borda dos polígonos.

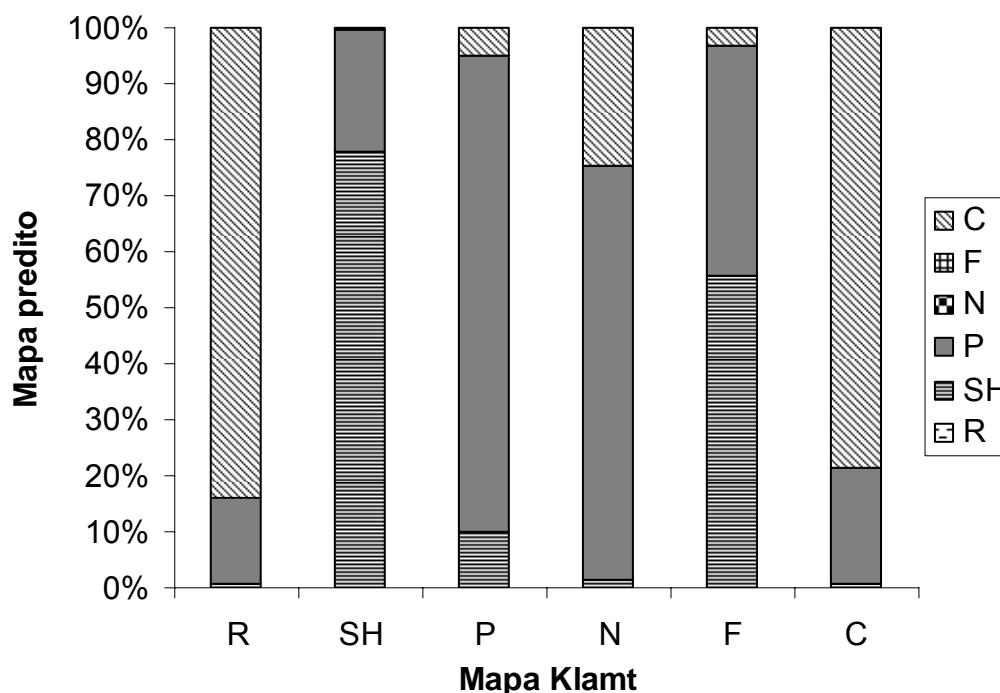


Figura 28 – Proporção relativa dos erros de inferência das classes de solos no mapa obtido por componentes principais no primeiro nível categórico (área externa). Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N), Plintossolo (F).

Boruvka; Penizek (2007) utilizaram redes neurais para a predição de classes de solos e verificaram que classes muito semelhantes sob o ponto de vista dos processos de formação tendem a ser confundidas pelos modelos. A semelhança entre solos como os *Luvisols* e *Albeluvisols* (WRB) ocasionou um incremento de solos incorretamente classificados. Os solos com a acurácia mais elevada foram os *Cambisols haplic* (WRB) em 85% dos casos. Classes devem ser bem definidas e distintas entre elas para uma boa espacialização. O emprego de qualquer metodologia, segundo os autores, deverá considerar o nível categórico a ser predito em função da heterogeneidade local, além da disponibilidade de informações para a geração dos modelos (número de perfis ou área).

Para a verificação da capacidade preditiva do modelo utilizado, foi realizado um teste de acurácia em uma área da qual não foram utilizados dados para a geração do modelo, doravante denominada área interna (Figura 9). Este procedimento valida o modelo utilizado, ou de outra maneira, testa a sua real capacidade de inferência ou capacidade preditiva. Os resultados estão na tabela 9. Nesta área existem Neossolos os quais não foram preditos pelo modelo. No entanto,

a classe dos Argissolos corresponde ao mapa Klamt em aproximadamente 90% dos pontos testados. E do ponto de vista do usuário, esta mesma classe encontra-se corretamente posicionada em 86,47% dos locais amostrados. A AG chegou a ser até mesmo maior do que a AG da área de onde foram usados pontos para o modelo, alcançando um valor de 79,4%. No entanto, uma medida mais realista da qualidade do mapa nesta região, mostra que a acurácia do mapa foi menor, o índice K de 48,53% ficou dez pontos percentuais abaixo da área de onde foram utilizados pontos para o modelo.

**Tabela 9 – Matriz de erros da área interna, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de ordem produzido a partir das componentes principais como variáveis dos modelos logísticos.**

Mapa predito	Mapa Klamt				Total	AU (%)
	R	SH	P	C		
R	0,00	0,00	0,00	0,00	0,00	0,00
SH	0,00	6,50	4,30	0,00	10,80	60,19
P	0,20	8,40	67,10	1,90	77,60	86,47
C	2,40	0,00	3,40	5,80	11,60	50,00
<b>Total</b>	2,60	14,90	74,80	7,70	100,00	
<b>AM (%)</b>	0,00	43,62	89,71	75,32		
<b>AG (%)</b>	79,40		<b>K<sup>1</sup> (%)</b>	48,53		

<sup>1</sup>erro padrão assintótico =  $\pm 0,035$  e intervalo de confiança 95% = 0,41...0,55. Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH) e Argissolo (P).

A espacialização de séries de solos por Bailey et al. (2003) atingiu uma acurácia geral de 65% quando o mapa predito foi verificado em relação ao mapa utilizado para treinar os modelos, o que segundo esses autores não é um indicativo preciso de que esta informação corresponda ao que se encontra no campo. Os autores não fazem menção ao índice K. Buscando avaliar a acurácia do mapa em relação a uma área da qual não foram utilizadas informações para o modelo, os autores utilizaram informações de 293 observações em uma área vizinha. A acurácia geral do mapa predito para a região foi de 47,4%. Os autores também verificaram a que distância de um pixel erroneamente classificado encontrava-se a classe predita. Para 68,8% dos 154 pontos testados, a informação adequada para aquele pixel encontrava-se a uma distância de até 300m. E para 42,9% dos pontos testados a

correta informação estava a uma distância de 100m. Para os autores, como os modelos logísticos utilizam-se de regressões, a qualidade dos dados é ponto chave para a qualidade da informação predita. Tal como nas regressões “convencionais”.

#### 4.3.1.2 Atributos do terreno como variáveis explicativas

Buscando avaliar o real ganho com as CP como alternativa para diminuir o número de variáveis explicativas nos modelos logísticos, foi realizado um novo conjunto de regressões utilizando-se desta vez, todos os nove atributos do terreno como variáveis explicativas. A tabela 10 traz os valores dos interceptos e dos coeficientes dos modelos para cada atributo do terreno. Diferentemente do que havia ocorrido para as CP, encontrou-se valores não significativos para os modelos.

**Tabela 10 – Valores dos interceptos e dos coeficientes das variáveis do terreno nos modelos logísticos para o 1º Nível Categórico.**

	C <sup>1</sup>				
	R	SH	P	N	F
<b>Intercepto</b>	1,087	15,344	10,963	<b>-1,551</b>	20,006
<b>ELEV</b>	0,011	-0,087	-0,031	0,006	-0,111
<b>DECL</b>	-0,123	-0,209	-0,221	-0,218	-0,488
<b>DIST</b>	<b>0,001</b>	-0,003	-0,003	-0,001	0,005
<b>CPLN</b>	-1,608	3,716	-0,449	<b>-0,287</b>	1,598
<b>CPRF</b>	0,901	2,756	1,869	1,697	3,495
<b>ILUM</b>	-0,011	0,017	-0,002	0,008	-0,008
<b>LNAC</b>	-0,419	-0,408	-0,092	-0,368	<b>-0,007</b>
<b>IUT</b>	0,200	0,080	-0,153	0,187	-0,276
<b>CTS</b>	<b>-0,004</b>	0,003	-0,010	<b>-0,011</b>	<b>-0,007</b>

<sup>1</sup>Classe de referência. Valores em negrito não são significativos ao nível de 5%. Todos os modelos apresentaram diferença significativa para o teste dos modelos apenas com o intercepto e os mesmos com as variáveis explicativas. Todas as variáveis explicativas têm relação significativa com a variável resposta (*Likelihood Ratio Test*). Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N), Plintossolo (F), Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Iluminação (ILUM), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).

O cálculo da probabilidade de se encontrar a classe de Neossolos na paisagem inicia-se pela função logit (Equação 16).

$$\text{Logit}_R = e^{1,087+0,011*\text{ELEV}-0,123*\text{DECL}-1,608*\text{CPLN}+0,901*\text{CPRF}-0,011*\text{ILUM}-0,419*\text{LNAC}+0,200*\text{IUT}} \quad (16)$$

A equação 16 tem uma complexidade, pelo menos em forma, maior do que a equação que se utilizou das CP (Equação 13). Apesar de, nesta última, as variáveis não terem qualquer significado físico, diferente do que ocorre na equação 16.

A seqüência das operações para o cálculo das probabilidades e a definição das classes de solos sob a paisagem é a mesma já executada na sessão 4.3.1 deste texto.

O mapa dos maiores valores de probabilidades encontram-se no apêndice E. É interessante verificar que existe um padrão na imagem, bastante associado às formas da paisagem. A análise visual da espacialização das probabilidades permite visualizar as principais drenagens, regiões de coxilhas ao longo de todo o centro e sul do município, as encostas do rebordo, bem como o topo do planalto. Supõe-se que a correta associação entre estes valores de probabilidade e as classes de solos, pode levar a mapas de solos de qualidade compatível com uma série de aplicações.

O mapa de solos predito com modelos utilizando-se atributos do terreno como variáveis explicativas e as classes de solo ao nível de ordem como variáveis resposta pode ser visualizado na figura 29. A classe Plintossolos não foi espacializada com estes modelos. Contudo, visualmente parece ter havido um aumento da qualidade da espacialização das demais classes de solos. Com o diferencial da alocação da classe dos Neossolos nos topos dos morros, bem como, da espacialização dos Nitossolos nas regiões mais elevadas. Quanto ao padrão geral do mapa gerado, este parece ter um aspecto bastante relacionado com a relação solo-paisagem para o município, como havia sido reportado a cima, quando da visualização do mapa gerado por CP (Figura 26).

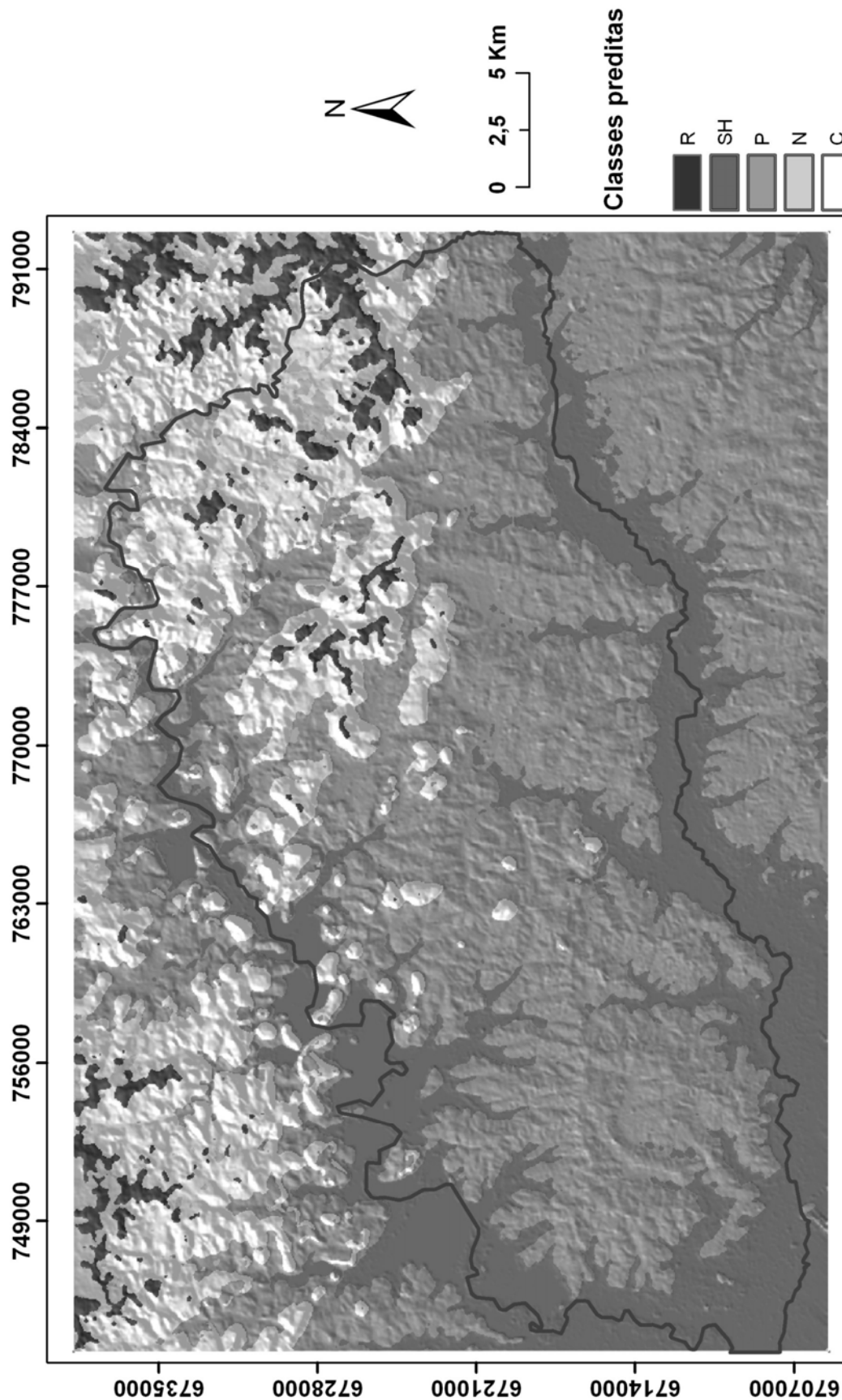


Figura 29 – Mapa de solos predito utilizando-se de atributos do terreno e o mapa Klamt ao nível de ordem. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P) e Nitossolo (N).



Na figura 30 pode-se visualizar as classes delineadas no mapa Klamt sobrepostas ao mapa predito. Na região visualizada os Solos Hidromórficos estão alocados com menor qualidade comparativamente a figura 27, que havia sido realizado com as CP. No entanto, visualmente há um ganho na espacialização da classe dos Neossolos nos topos dos morros. A classe dos Nitossolos, apesar de ter sido predita sob a paisagem, reproduz visualmente muito pouco do que estava localizado no mapa Klamt. Como já foi dito, isto pode estar relacionado ao baixo número de amostras utilizadas desta classe na geração dos modelos. O que é uma consequência da área ocupada por esta classe nos quase 874 km<sup>2</sup> do município, de apenas 0,77% (6,7 km<sup>2</sup>).

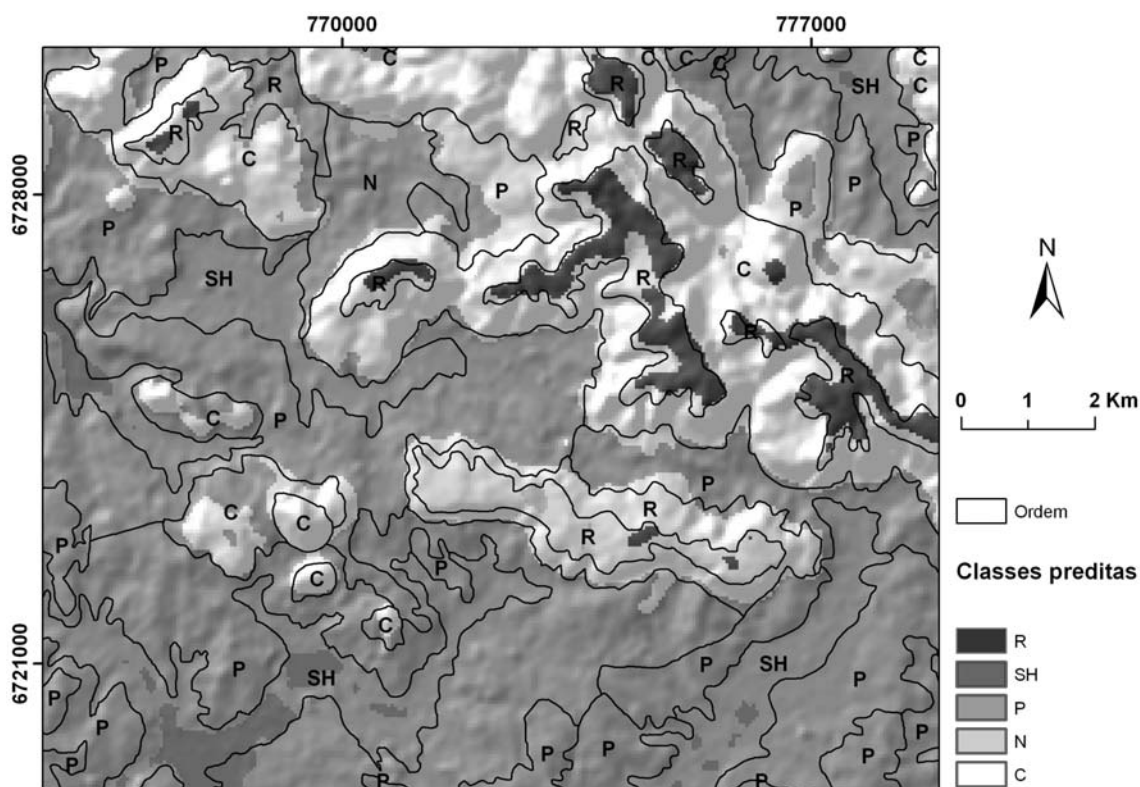


Figura 30 – Unidades de solos sobrepostas ao mapa predito utilizando-se de atributos do terreno e o mapa Klamt ao nível de ordem. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P) e Nitossolo (N).

A capacidade dos modelos logísticos em reproduzir o mapa Klamt utilizando-se dos atributos do terreno está reportada na tabela 11. Apesar das classes Neossolos e Nitossolos terem sido preditas, tanto a AM quanto a AU destas duas

classes possuem valores bastante baixos, contudo os valores AG e K tiveram incremento. A acurácia, considerando as classes mapeadas devido ao acaso, atingiu o valor de 63,2%. Este valor é um pouco superior aos 58,02% encontrados utilizando-se as três CP. Possivelmente devido ao fato de que com as CP, na busca de uma simplicidade dos modelos em detrimento do poder de explicação (variância total) das novas variáveis, ocorre uma perda de variância com a diminuição de nove para apenas três variáveis.

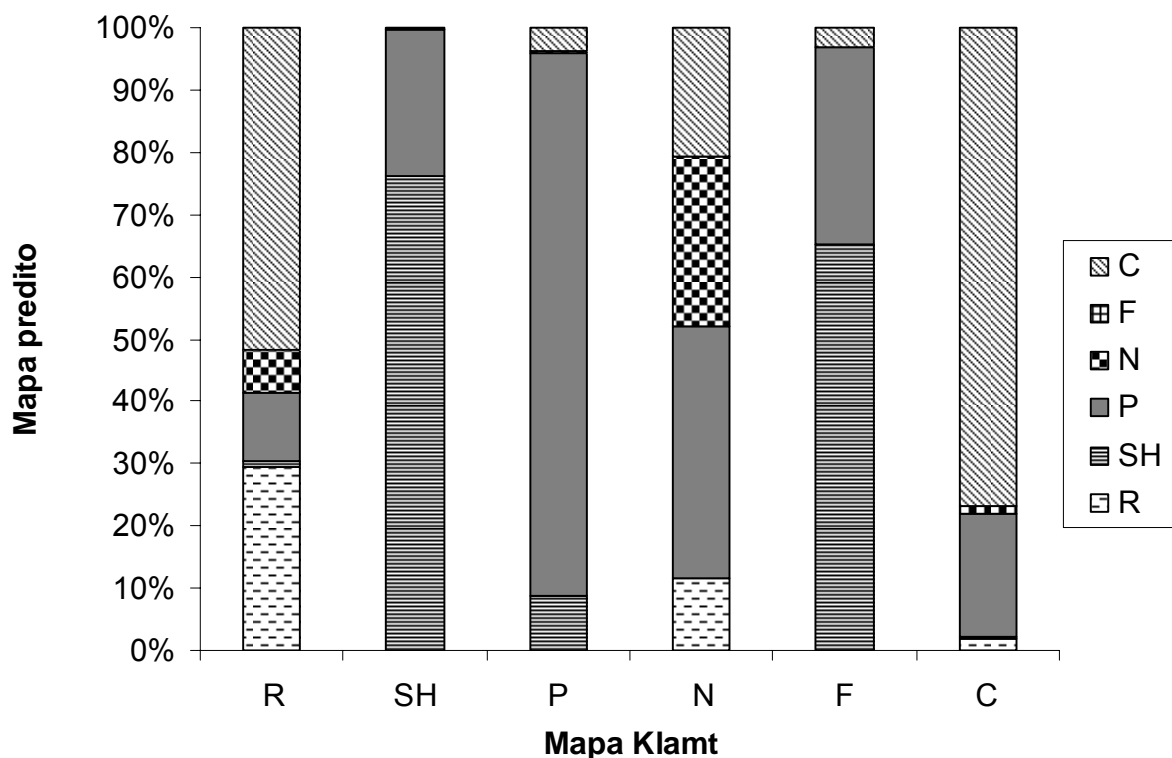
**Tabela 11 – Matriz de erros da área externa, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de ordem produzido a partir dos atributos do terreno como variáveis dos modelos logísticos.**

Mapa predito	Mapa Klamt						Total	AU (%)
	R	SH	P	N	F	C		
R	1,64	0,00	0,00	0,09	0,00	0,43	2,16	75,93
SH	0,06	17,49	4,17	0,00	1,82	0,03	23,57	74,20
P	0,61	5,41	40,64	0,31	0,88	4,22	52,07	78,05
N	0,39	0,00	0,13	0,21	0,00	0,27	1,00	21,00
F	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
C	2,88	0,07	1,74	0,16	0,09	16,25	21,20	76,65
<b>Total</b>	<b>5,57</b>	<b>22,97</b>	<b>46,69</b>	<b>0,77</b>	<b>2,79</b>	<b>21,20</b>	<b>100,00</b>	
<b>AM (%)</b>	<b>29,44</b>	<b>76,14</b>	<b>87,04</b>	<b>27,27</b>	<b>0,00</b>	<b>76,65</b>		
<b>AG (%)</b>	<b>76,24</b>		<b>K<sup>1</sup> (%)</b>		<b>63,20</b>			

<sup>1</sup>erro padrão assintótico =  $\pm 0,009$  e intervalo de confiança 95% = 0,61...0,65. Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N) e Plintossolo (F).

Os erros de mapeamento ocorreram nas classes de solos próximas quanto ao delineamento na paisagem (Figura 31). Havendo áreas de Cambissolos classificadas como Neossolos (2,88) maior do que a própria área de Neossolos (1,64). Os Solos Hidromórficos foram erroneamente espacializados em torno de 24% dos locais amostrados, sendo confundidos com os Argissolos na maior parte destes locais. Os Nitossolos foram confundidos com uma diversidade de outras classes. Observando-se sua posição no mapa Klamt, pode ser verificado que esta classe ocupa uma posição intermediária entre os solos de coxilha e do planalto. Tornando a sua associação com os atributos do terreno mais difícil para os modelos. Além de ser uma área muito pouco representativa no total mapeado, o que torna sua

previsibilidade crítica para os modelos logísticos, por serem estes últimos sensíveis ao tamanho relativo das amostras.



**Figura 31 – Proporção relativa dos erros de inferência das classes de solos no mapa obtido por atributos do terreno no primeiro nível categórico (área externa). Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N) e Plintossolo (F).**

A capacidade preditiva dos modelos foi analisada na área de onde não haviam sido utilizados dados para a geração dos mesmos (Tabela 12). O que se verifica é uma capacidade preditiva menor do que os modelos utilizados na sessão 4.3.1 deste texto. Apesar de que em ambas os valores de K na área de validação não atingiram 50%.

**Tabela 12 – Matriz de erros da área interna, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de ordem produzido a partir dos atributos do terreno como variáveis dos modelos logísticos.**

Mapa predito	Mapa Klamt				Total	AU (%)
	R	SH	P	C		
R	0,00	0,00	0,00	0,00	0,00	0,00
SH	0,00	3,70	1,10	0,00	4,80	77,08
P	0,30	11,20	72,80	3,30	87,60	83,11
C	2,30	0,00	0,90	4,40	7,60	57,89
<b>Total</b>	2,60	14,90	74,80	7,70	100,00	
<b>AM (%)</b>	0,00	24,83	97,33	57,14		
<b>AG (%)</b>	80,90			<b>K<sup>1</sup> (%)</b>	43,61	

<sup>1</sup>erro padrão assintótico =  $\pm 0,039$  e intervalo de confiança 95% = 0,35..0,51. Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH) e Argissolo (P).

Ambas as matrizes de erros, tabelas 11 e 12, repetem o padrão de confusão entre as classes vizinhas de solos, sob o ponto de vista do delineamento na paisagem, percebido nas tabelas 8 e 9.

Em estudo realizado para a interpolação de 15 classes de solos ao nível de ordem, Hengl et al. (2007) obtiveram através de RLM uma acurácia pelo índice K de 36,7%. Outros modelos não logísticos de interpolação utilizados no estudo obtiveram valores para K de 33,7%, 42,4% e 54,2%. Segundo os autores, nenhuma das classes preditas alcançou um K superior a 60%, o que segundo eles, foi devido a problemas na amostragem dos dados (não representatividade dos perfis descritos), e grande variabilidade dos solos. Para Zeilhofer (2006) a acurácia dos mapas gerados através de RLM estará limitada a acurácia das variáveis utilizadas. O autor relata que a predição dos solos do Pantanal, originalmente mapeados em 21 classes e reagrupadas para seis associações, rendeu uma acurácia de 65%.

Também foram realizados testes utilizando-se as classes de solos ao nível de subordem, para verificar a empregabilidade das CP e regressões logísticas múltiplas no mapeamento digital de solos. As sessões 4.3.2.1 e 4.3.2.2, a seguir, tratam dos resultados destes testes.

#### 4.3.2 Mapa predito ao nível de subordem

##### 4.3.2.1 Componentes principais como variáveis explicativas.

Os valores de interceptos e coeficientes dos modelos utilizando-se de CP estão na tabela 13. A função logit a qual espacializa a probabilidade de se encontrar os Plintossolos sobre a paisagem, não utilizou o intercepto e a terceira CP como parâmetros do modelo.

**Tabela 13 – Valores dos interceptos e dos coeficientes das componentes principais nos modelos logísticos desenvolvidos ao nível de subordem.**

	SH <sup>1</sup>						
	FT	PAC	PBAC	PV	PVA	RL	CX_NV
<b>Intercepto</b>	<b>-0,384</b>	5,497	4,375	6,379	7,249	7,047	8,258
<b>CP1</b>	-0,013	0,055	0,029	0,080	0,073	0,121	0,122
<b>CP2</b>	-0,097	-0,080	-0,090	-0,083	-0,063	-0,044	-0,038
<b>CP3</b>	<b>0,001</b>	-0,157	-0,097	-0,218	-0,196	-0,304	-0,305

<sup>1</sup>Classe de referência para os modelos logísticos. Valores em negrito não são significativos ao nível de 5%. Todos os modelos apresentaram diferença significativa para o teste dos modelos apenas com o intercepto e os mesmos com as variáveis explicativas. Todas as variáveis explicativas têm relação significativa com a variável resposta (*Likelihood Ratio Test*). Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA), Argissolo Acinzentado (PAC), Argissolo Bruno-Acinzentado (PBAC), Plintossolo Argilúvico (FT), Primeira Componente Principal (CP1), Segunda Componente Principal (CP2) e Terceira Componente Principal (CP3).

O desenvolvimento das funções logit, cálculo das probabilidades e espacialização das classes de solos segue o mesmo procedimento já detalhado na sessão 4.3.1 deste texto.

A distribuição espacial das maiores probabilidades encontradas na paisagem teve o mesmo padrão e relações com a geomorfologia da região já descritos para os apêndices C e E. Embora neste caso, tenha sido utilizada para o cálculo das probabilidades a classe dos Solos Hidromórficos como referência nos modelos logísticos.

A espacialização das classes de solos ao nível de subordem pode ser visualizada na figura 32. As classes dos Plintossolos Argilúvicos e dos Argissolos Acinzentados não foram espacializadas pelos modelos.

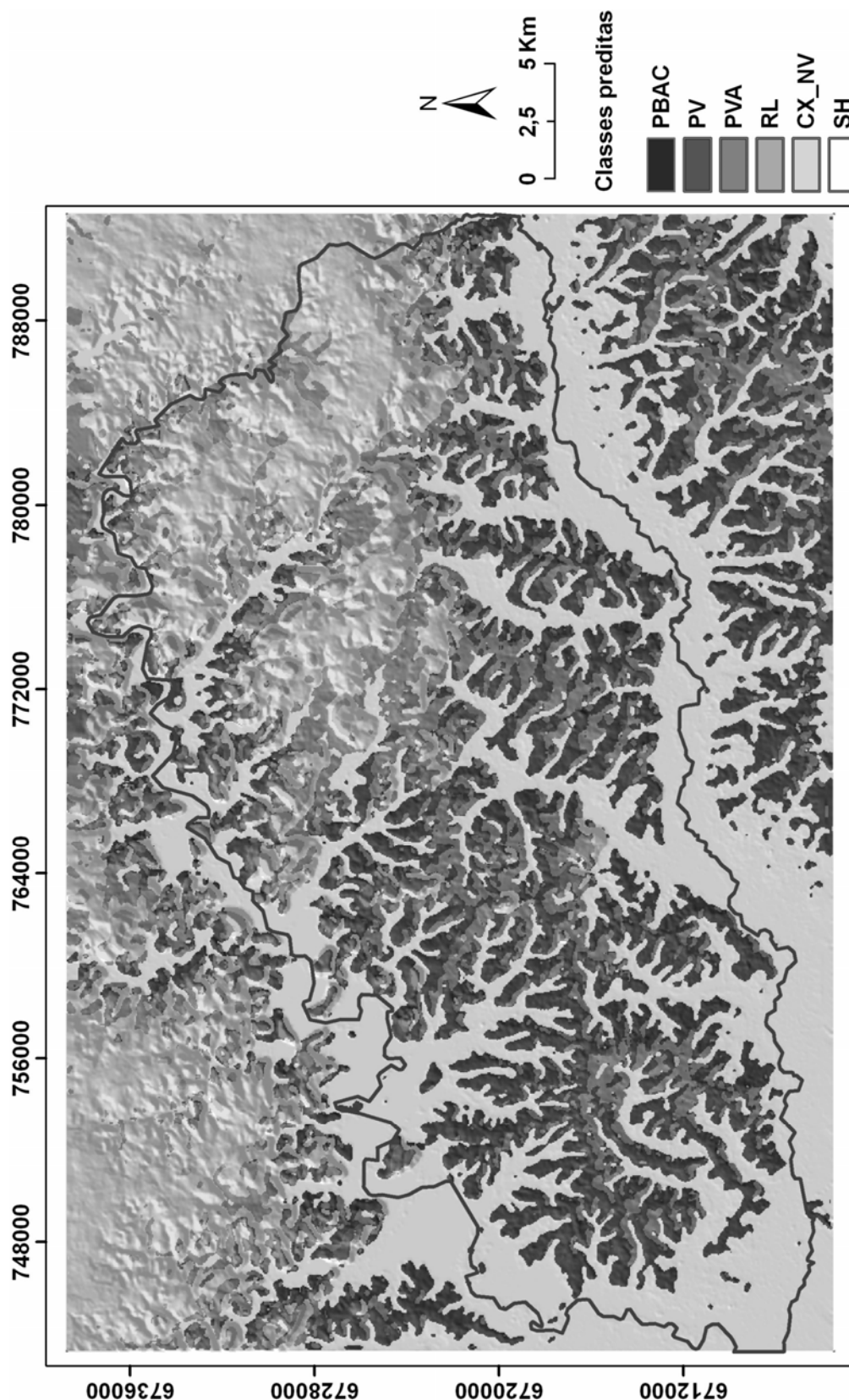


Figura 32 – Mapa de solos predito utilizando-se de componentes principais e o mapa Klamt ao nível de subordem. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA) e Argissolo Bruno-Acinzentado (PBAC).

As classes não preditas dos Plintossolos Argilúvicos e dos Argissolos Acinzentados corresponderam a 2,5 e 7,6%, respectivamente, do total de amostras utilizadas nos modelos. Corroborando com os resultados verificados anteriormente e informações da literatura, de que deve haver uma proporção relativa mínima entre as classes a serem preditas pelas RLM. Supõe-se que uma solução alternativa para este problema poderia ser a amostragem que considera alguma relação inversa entre as áreas. A viabilidade deste procedimento deveria ser adequadamente estudada na literatura especializada de regressões logísticas.

Entretanto as classes dos Argissolos Vermelhos e dos Neossolos Litólicos, mesmo correspondendo a apenas 3,5% e 5,65%, respectivamente, dos locais amostrados foram espacializadas pelos modelos. Uma comparação com os solos de várzea das figuras 26 e 29, mostra que houve uma espacialização muito mais intensa destes solos na figura 32. A classe de solos Argissolos que agora foi dividida em quatro outras classes mostra-se, aparentemente, muito intrincada. Com muitos pixels pertencentes à determinada subordem isolados dentro de outra classe. Com este padrão de distribuição muitos pixels deveriam ser reeditados e considerados inclusões por estarem abaixo da área mínima mapeável.

Uma comparação entre o mapa Klamt e as classes preditas pode ser visualizada na figura 33. Na imagem pode ser verificada a presença dos Solos Hidromórficos nas áreas mais baixas do relevo, ao longo da rede de drenagem. Os Neossolos Litólicos e os Cambissolos Háplicos ocuparam posições mais altas do terreno sobre a região de rebordo e topo dos morros. Quanto aos Argissolos, estes distribuíram-se em um padrão bastante variado, predominantemente sobre a região de coxilhas.

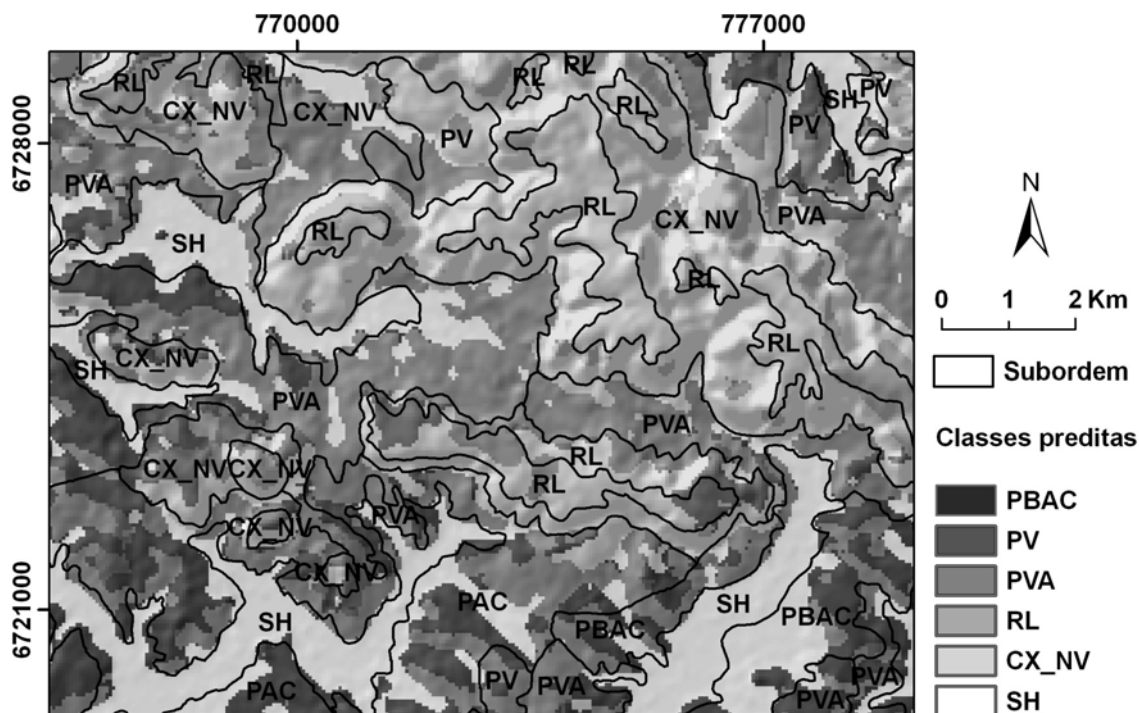


Figura 33 – Unidades de solos sobrepostas ao mapa predito utilizando-se de componentes principais e o mapa Klamt ao nível de subordem. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA) e Argissolo Bruno-Acinzentado (PBAC).

A matriz de erros entre o mapa Klamt e o mapa predito por CP ao nível de subordem pode ser visualizada na tabela 14. As classes com os resultados mais inferiores foram aquelas que corresponderam, proporcionalmente, aos menores percentuais de amostras entre as 70.000 aleatoriamente geradas para a RLM. As classes Argissolo Vermelho-Amarelo, Solos Hidromórficos, Cambissolos Háplicos e Nitossolos Vermelhos, corresponderam, cada uma delas, em torno de 22% do total dos locais amostrados para a geração dos modelos. Estas foram também as classes com os melhores resultados e AM e de AU. Os Solos Hidromórficos obtiveram uma AM de mais de 92%, enquanto que sua AU ficou abaixo de 60%, o que corrobora com o que havia sido verificado anteriormente (Figura 32) a respeito destes solos terem sido demasiadamente espacializados sobre a paisagem. Ou seja, do ponto de vista do mapeador os delineamentos preditos cobrem bem as áreas desta classe no mapa Klamt. No entanto, sob o ponto de vista do usuário esta classe, em 40% dos casos, não condiz com a realidade de campo. O índice K ficou em torno de 40%,



sendo assim, o mapeamento ao nível de subordem foi menos acurado do que a nível de ordem utilizando-se as CP como variáveis explicativas nos modelos.

**Tabela 14 – Matriz de erros da área externa, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de subordem produzido a partir de componente principais como variáveis dos modelos logísticos.**

Mapa predito	Mapa Klamt								Total	AU (%)
	FT	PAC	PBAC	PV	PVA	RL	CX_NV	SH		
FT	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PAC	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PBAC	0,64	2,82	6,71	1,25	9,73	0,25	1,53	1,35	24,28	27,64
PV	0,00	0,13	0,00	0,12	0,59	0,39	1,69	0,00	2,93	4,10
PVA	0,01	2,12	1,40	1,37	7,20	0,52	2,32	0,36	15,30	47,06
RL	0,00	0,00	0,00	0,00	0,01	0,48	1,93	0,00	2,42	19,83
CX_NV	0,09	0,18	0,01	0,18	1,31	3,80	13,53	0,09	19,19	70,51
SH	2,03	2,18	4,96	0,30	3,96	0,12	0,91	21,43	35,89	59,71
<b>Total</b>	<b>2,78</b>	<b>7,44</b>	<b>13,08</b>	<b>3,21</b>	<b>22,81</b>	<b>5,55</b>	<b>21,90</b>	<b>23,23</b>	<b>100,00</b>	
<b>AM (%)</b>	<b>0,00</b>	<b>0,00</b>	<b>51,30</b>	<b>3,74</b>	<b>31,57</b>	<b>8,65</b>	<b>61,78</b>	<b>92,25</b>		
<b>AG (%)</b>	<b>49,47</b>			<b>K' (%)</b>					<b>40,02</b>	

<sup>1</sup>erro padrão assintótico =  $\pm 0,008$  e intervalo de confiança 95% = 0,38...0,41. Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA), Argissolo Acinzentado (PAC), Argissolo Bruno-Acinzentado (PBAC) e Plintossolo Argilúvico (FT).

Valores similares de acurácia foram encontrados por Giasson et al. (2008) em estudo empregando mapas de solos existentes nas escalas 1:30.000 a 1:50.000. Os autores relataram um valor de K igual a 36% quando verificando a potencialidade dos modelos em reproduzir os mapas utilizados para treinamento. Após uma simplificação da legenda este valor foi para 54%. De acordo com os autores os baixos valores da acurácia ocorrem devido a complexa distribuição espacial das classes e a grande variabilidade do material de origem da região, o qual não havia sido utilizado nos modelos por causa da inexistência de informações de qualidade.

Os resultados da validação do modelo ao nível de subordem estão na tabela 15. Estes resultados são muito inferiores aos anteriores utilizados para a verificação da capacidade do modelo em repetir o mapa Klamt. Uma das possíveis causas é de que na região de validação, predominam os Argissolos, classe esta que obteve os piores resultados devido ao diluído número de amostras de cada uma das

subordens utilizadas no modelo. O valor K indica uma acurácia de apenas 29,41% do modelo utilizado quando predizendo as classes de solos na paisagem.

**Tabela 15 – Matriz de erros da área interna, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de subordem produzido a partir de componente principais como variáveis dos modelos logísticos.**

Mapa predito	Mapa Klamt							Total	AU (%)
	PAC	PBAC	PV	PVA	RL	CX_NV	SH		
<b>PAC</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>PBAC</b>	11,90	8,10	1,40	10,10	0,10	0,30	1,80	33,70	24,04
<b>PV</b>	0,40	0,30	0,10	1,50	0,80	2,50	0,00	5,60	1,79
<b>PVA</b>	6,60	4,50	1,90	12,50	0,10	0,90	1,10	27,60	45,29
<b>RL</b>	0,00	0,00	0,00	0,00	0,10	0,70	0,00	0,80	12,50
<b>CX_NV</b>	0,30	0,10	0,20	1,10	1,50	3,00	0,00	6,20	48,39
<b>SH</b>	6,20	4,30	0,20	3,10	0,00	0,30	12,00	26,10	45,98
<b>Total</b>	25,40	17,30	3,80	28,30	2,60	7,70	14,90	100,00	
<b>AM (%)</b>	0,00	46,82	2,63	44,17	3,85	38,96	80,54		
<b>AG (%)</b>	35,80			<b>K<sup>1</sup> (%)</b>		29,41			

<sup>1</sup>erro padrão assintótico =  $\pm 0,022$  e intervalo de confiança 95% = 0,24...0,33. Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA), Argissolo Acinzentado (PAC) e Argissolo Bruno-Acinzentado (PBAC).

A matriz de erros demonstra a dificuldade dos modelos em distinguir entre as classes de solos na medida em que a classificação torna-se mais específica. A classe dos Argissolos foi dividida em quatro novas classes o que passou a representar uma dificuldade maior para a modelagem espacial desta classe. Em vários pontos da paisagem ocorrem erros de classificação entre os Argissolos ao nível de subordem. Interessante de serem analisados os erros entre Solos Hidromórficos e as diferentes subordens dos Argissolos. A medida em que se vai de um ponto mais baixo da paisagem (Argissolo Acinzentado) para locais mais elevados, distantes das várzeas (Argissolos Vermelhos) os erros entre solos de várzea e os Argissolos diminuem. A dificuldade do modelo em espacializar as diferentes classes de solos, devido a estes serem muito semelhantes quanto a relação com os atributos do terreno utilizados, também foi reportada por Figueiredo (2006). Na tentativa de melhorar a capacidade preditiva de um modelo por RLM o

autor simplificou a legenda agrupando algumas classes de solos. No entanto, como entre as variáveis a serem preditas estavam dois Latossolos, os quais apenas se diferenciavam pela fase relevo, os novos modelos gerados melhoraram a acurácia em menos de 2%, de 36,6% para 38,3%.

#### 4.3.2.2 Atributos do terreno como variáveis explicativas.

Os parâmetros utilizados nas regressões logísticas múltiplas estão na tabela 16.

**Tabela 16 – Valores dos interceptos e dos coeficientes dos atributos do terreno nos modelos logísticos desenvolvidos ao nível de subordem.**

	SH <sup>1</sup>						
	FT	PAC	PBAC	PV	PVA	RL	CX_NV
<b>Intercepto</b>	4,742	-8,555	-2,838	-8,173	-7,732	-16,304	-17,080
<b>ELEV</b>	-0,024	0,067	0,035	0,083	0,073	0,114	0,103
<b>DECL</b>	-0,284	0,029	-0,101	-0,062	0,071	0,134	0,264
<b>DIST</b>	0,003	0,002	0,003	0,003	0,003	0,003	0,003
<b>CPLN</b>	-2,134	-3,960	-4,221	-4,887	-4,141	-5,429	-3,850
<b>CPRF</b>	<b>0,763</b>	-0,871	<b>-0,058</b>	-1,963	-1,539	-2,369	-3,230
<b>ILUM</b>	-0,026	-0,016	-0,017	-0,033	-0,022	-0,032	-0,020
<b>LNAC</b>	0,403	0,357	0,443	0,386	0,286	<b>-0,030</b>	0,261
<b>IUT</b>	-0,355	-0,214	-0,346	-0,353	-0,160	0,160	0,072
<b>CTS</b>	-0,012	-0,048	-0,039	<b>-0,004</b>	-0,009	-0,006	-0,002

<sup>1</sup>Classe de referência para os modelos logísticos. Valores em negrito não são significativos ao nível de 5%. Todos os modelos apresentaram diferença significativa para o teste dos modelos apenas com o intercepto e os mesmos com as variáveis explicativas. Todas as variáveis explicativas têm relação significativa com a variável resposta (*Likelihood Ratio Test*). Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA), Argissolo Acinzentado (PAC), Argissolo Bruno-Acinzentado (PBAC), Plintossolo Argilúvico (FT), Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Iluminação (ILUM), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).

As classes Plintossolo Argilúvico, Argissolo Acinzentado e Argissolo Vermelho não foram preditas (Figura 34).

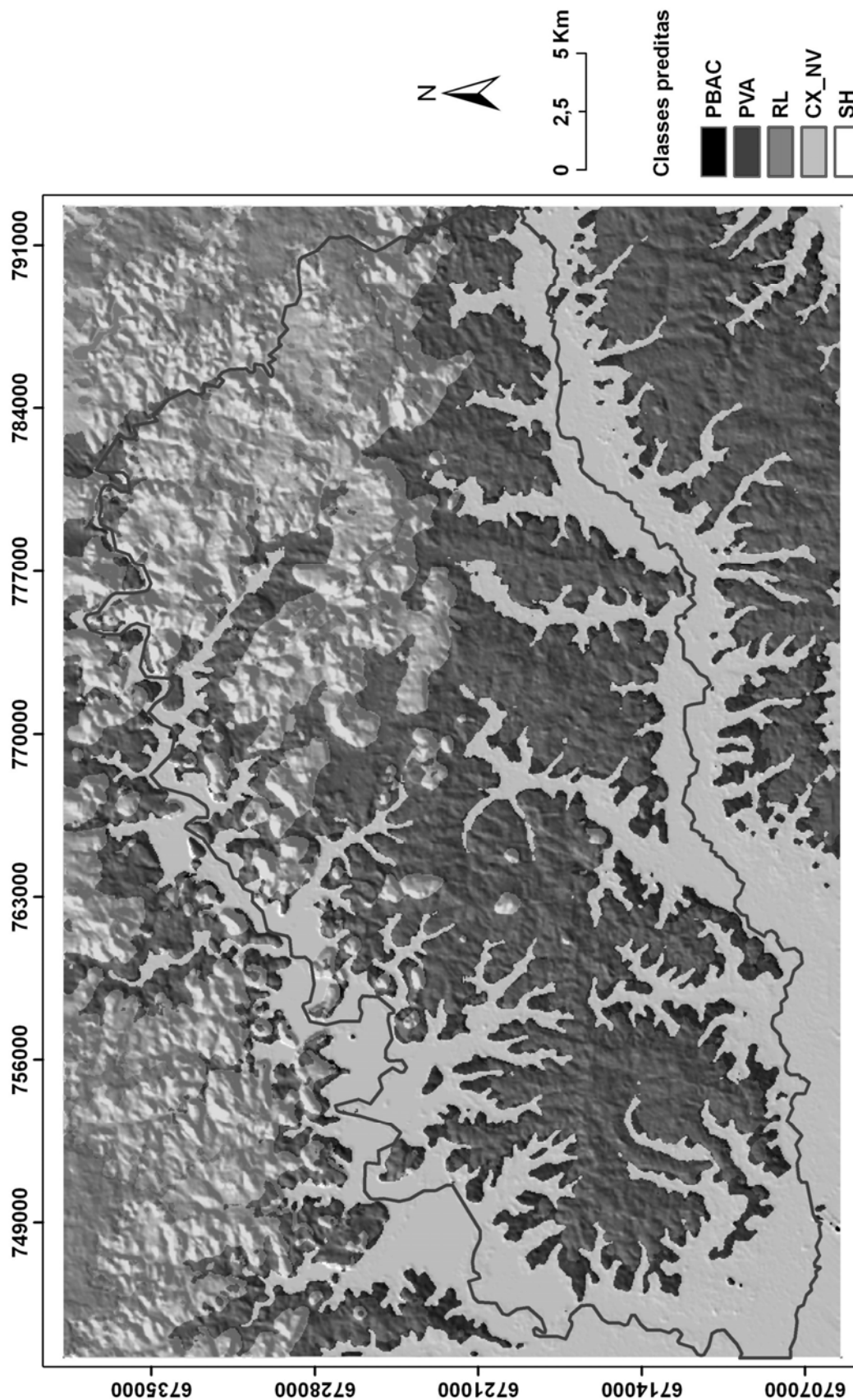


Figura 34 – Mapa de solos predito utilizando-se de atributos do terreno e o mapa Klamt ao nível de subordem. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA) e Argissolo Bruno-Acinzentado (PBAC).

A classe Argissolo Vermelho correspondeu a apenas 3,4% dos pontos amostrados. No entanto, a análise visual da distribuição espacial das classes de solos preditas, corresponde bastante a distribuição explicada pela relação solo-paisagem para a região. Com um detalhe interessante sendo a posição dos Argissolos Bruno-Acinzentados ocupando uma posição intermediária entre Solos Hidromórficos e os Argissolos Vermelho-Amarelos localizados nas posições mais altas das coxilhas. Os Cambissolos Háplicos e os Neossolos Litólicos foram espacializados na região de encosta e de topos de morros, bem como, em alguns morros testemunhas. Sendo que os Neossolos Litólicos ocuparam os pontos mais altos do relevo.

A visualização da sobreposição entre os mapas Klamt e predito está demonstrada na figura 35. Chama atenção a posição dos solos de várzea, dos solos de coxilha e dos solos de encosta e topos de morros. Contudo uma reprodução do mapa Klamt ficou comprometida, especialmente nas subordens dos Argissolos.

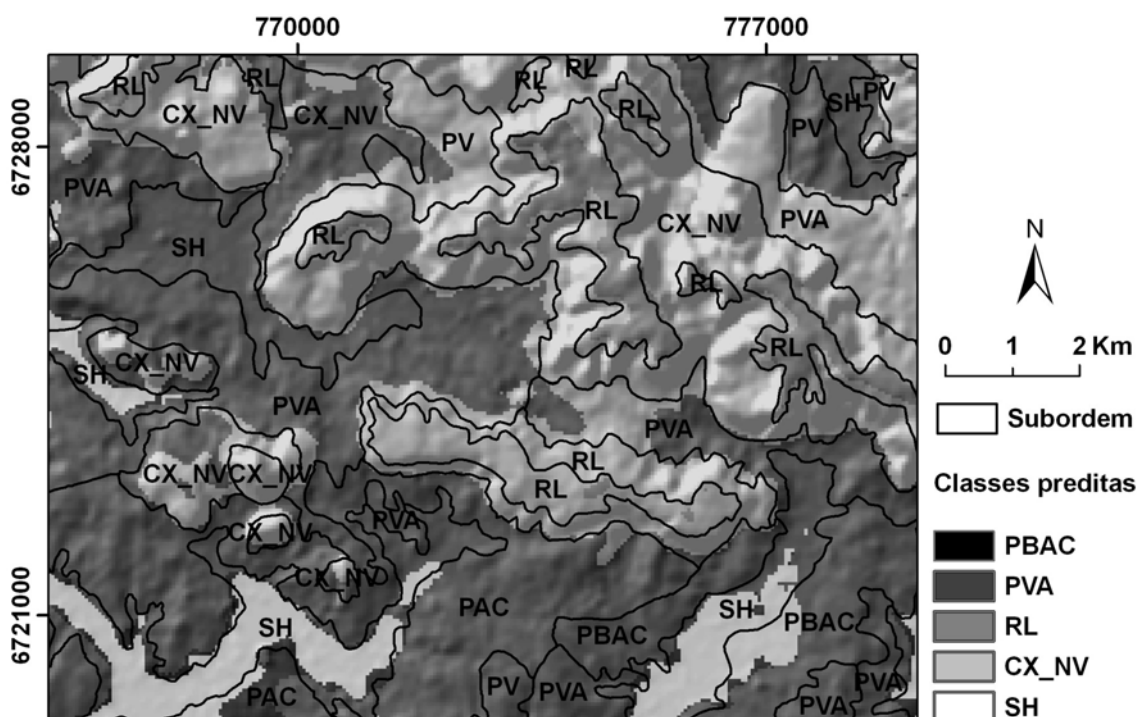


Figura 35 – Unidades de solos sobrepostas ao mapa predito utilizando-se de atributos do terreno e o mapa Klamt ao nível de subordem. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA) e Argissolo Bruno-Acinzentado (PBAC).

A matriz de erros para o mapa das subordens previstas pelos atributos do terreno consta na tabela 17. Corroborando com o que foi verificado visualmente, os índices de acurácia foram melhores quando se utilizou os atributos do terreno nos modelos. O valor da AG chegou a aproximadamente 60%. E o valor de K ficou em 45,88%, confirmando que este último modelo produz melhores resultados do ponto de vista de reprodução do mapa original.

**Tabela 17 – Matriz de erros da área externa, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de subordem produzido a partir de atributos do terreno como variáveis explicativas nos modelos logísticos.**

Mapa predito	Mapa Klamt								Total	AU (%)
	F	PAC	PBAC	PV	PVA	RL	CX_NV	SH		
F	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PAC	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PBAC	0,52	0,73	3,82	0,15	2,48	0,13	0,18	0,94	8,94	42,73
PV	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PVA	0,07	5,24	5,29	2,72	16,11	0,39	3,61	1,75	35,18	45,79
RL	0,00	0,00	0,00	0,00	0,00	2,05	0,65	0,00	2,70	75,93
CX_NV	0,09	0,16	0,00	0,27	1,95	2,91	17,90	0,06	22,53	79,45
SH	2,09	1,31	3,98	0,07	2,27	0,07	0,37	20,48	30,65	66,82
<b>Total</b>	<b>2,78</b>	<b>7,44</b>	<b>13,08</b>	<b>3,21</b>	<b>22,81</b>	<b>5,55</b>	<b>21,90</b>	<b>23,23</b>	<b>100,00</b>	
<b>AM</b>	<b>0,00</b>	<b>0,00</b>	<b>29,20</b>	<b>0,00</b>	<b>70,63</b>	<b>36,94</b>	<b>81,74</b>	<b>88,16</b>		
<b>AG (%)</b>	<b>59,55</b>		<b>K' (%)</b>		<b>45,88</b>					

erro padrão assintótico =  $\pm 0,008$  e intervalo de confiança 95% = 0,44...0,47. Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA), Argissolo Acinzentado (PAC), Argissolo Bruno-Acinzentado (PBAC) e Plintossolo Argilúvico (FT).

No entanto, na área de validação do modelo os valores de acurácia foram tão ou mais ruins do que o modelo anterior que se utilizou de CP (Tabela 18). Possivelmente o causador desta baixa qualidade preditiva, tenha sido também a predominância na região de testes das subordens dos Argissolos, os quais têm proporcionalmente um número muito baixo de amostras, especialmente os Argissolos Vermelhos e os Argissolos Acinzentados, 3,48 e 7,68%, respectivamente. Ficando clara a dificuldade dos modelos em distinguir entre classes muito próximas taxonomicamente.

Tabela 18 – Matriz de erros da área interna, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de subordem produzido a partir de atributos do terreno como variáveis explicativas nos modelos logísticos.

Mapa predito	Mapa Klamt							Total	AU (%)
	PAC	PBAC	PV	PVA	RL	CX_NV	SH		
PAC	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PBAC	3,30	2,20	0,00	0,30	0,00	0,00	0,80	6,60	33,33
PV	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PVA	19,40	12,70	3,60	25,60	0,20	3,00	6,60	71,10	36,01
RL	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
CX_NV	0,10	0,00	0,20	0,80	2,40	4,70	0,00	8,20	57,32
SH	2,60	2,40	0,00	1,60	0,00	0,00	7,50	14,10	53,19
<b>Total</b>	25,40	17,30	3,80	28,30	2,60	7,70	14,90	100,00	
<b>AM (%)</b>	0,00	12,72	0,00	90,46	0,00	61,04	50,34		
<b>AG (%)</b>	40,00		<b>K<sup>1</sup> (%)</b>		25,16				

<sup>1</sup>erro padrão assintótico =  $\pm 0,019$  e intervalo de confiança 95% = 0,21...0,28. Associação Cambissolo Háplico e Nitossolo Vermelho (CX\_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA), Argissolo Acinzentado (PAC) e Argissolo Bruno-Acinzentado (PBAC).

#### 4.4 Mapeamento pedométrico a partir do mapa referência.

##### 4.4.1 Mapa predito na área de referência com CP como variáveis explicativas.

Os parâmetros dos modelos logísticos para as áreas de referência utilizando-se das CP constam na tabela 19. Estes valores definiram as relações entre os PI das CP no ambiente SIG, e foram utilizados para a geração dos PI de probabilidade de cada uma das funções logit para as classes de solos a partir do mapa referência.

**Tabela 19 – Valores dos interceptos e dos coeficientes das componentes principais nos modelos logísticos das áreas de referência.**

	CX <sup>1</sup>				
	SH	PBAC	PV	RL	PVA
<b>Intercepto</b>	-1,948	1,967	2,963	-1,660	<b>-0,118</b>
<b>CP1</b>	-0,096	-0,062	-0,041	<b>0,003</b>	-0,074
<b>CP2</b>	-0,068	-0,074	-0,068	0,040	-0,111
<b>CP3</b>	0,202	0,119	0,073	0,021	0,139

<sup>1</sup>Classe de referência para os modelos logísticos. Valores em negrito não são significativos ao nível de 5%. Todos os modelos apresentaram diferença significativa para o teste dos modelos apenas com o intercepto e os mesmos com as variáveis explicativas. Todas as variáveis explicativas têm relação significativa com a variável resposta (*Likelihood Ratio Test*). Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-Amarelo (PVA), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Bruno-Acinzentado (PBAC), Primeira Componente Principal (CP1), Segunda Componente Principal (CP2) e Terceira Componente Principal (CP3).

O apêndice F traz a espacialização dos maiores valores de probabilidades para as seis classes de solos preditas a partir do mapa referência. Pode ser verificado que na região de rebordo do planalto predominam os valores próximos a um. Este valor ainda não havia sido registrado nos mapas de probabilidade anteriores. Na região de depressão predominam valores baixos de probabilidade para a determinação das classes de solos.

Utilizando-se de CP na área de referência não foi possível espacializar os Solos Hidromórficos e os Argissolos Vermelho-Amarelos. A área espacializada dos Cambissolos Háplicos também foi muito pequena, apesar deste último solo constar na legenda da figura 36. Estas três classes correspondem justamente as menores áreas delineadas nos mapas referência. O que também havia ocorrido quando utilizou-se o mapa existente (mapa Klamt) para a produção das regressões logísticas. Das 12.000 amostras aleatoriamente geradas para a produção das RLM, os Solos Hidromórficos corresponderam a 7%, os Argissolos Vermelho-Amarelos a 5% e os Cambissolos Háplicos a apenas 2%. Por outro lado as classes Argissolo Bruno-Acinzentado, Argissolo Vermelho e Neossolo Litólico, corresponderam a 24, 40 e 21%, respectivamente, do total de 12.000 pontos gerados.



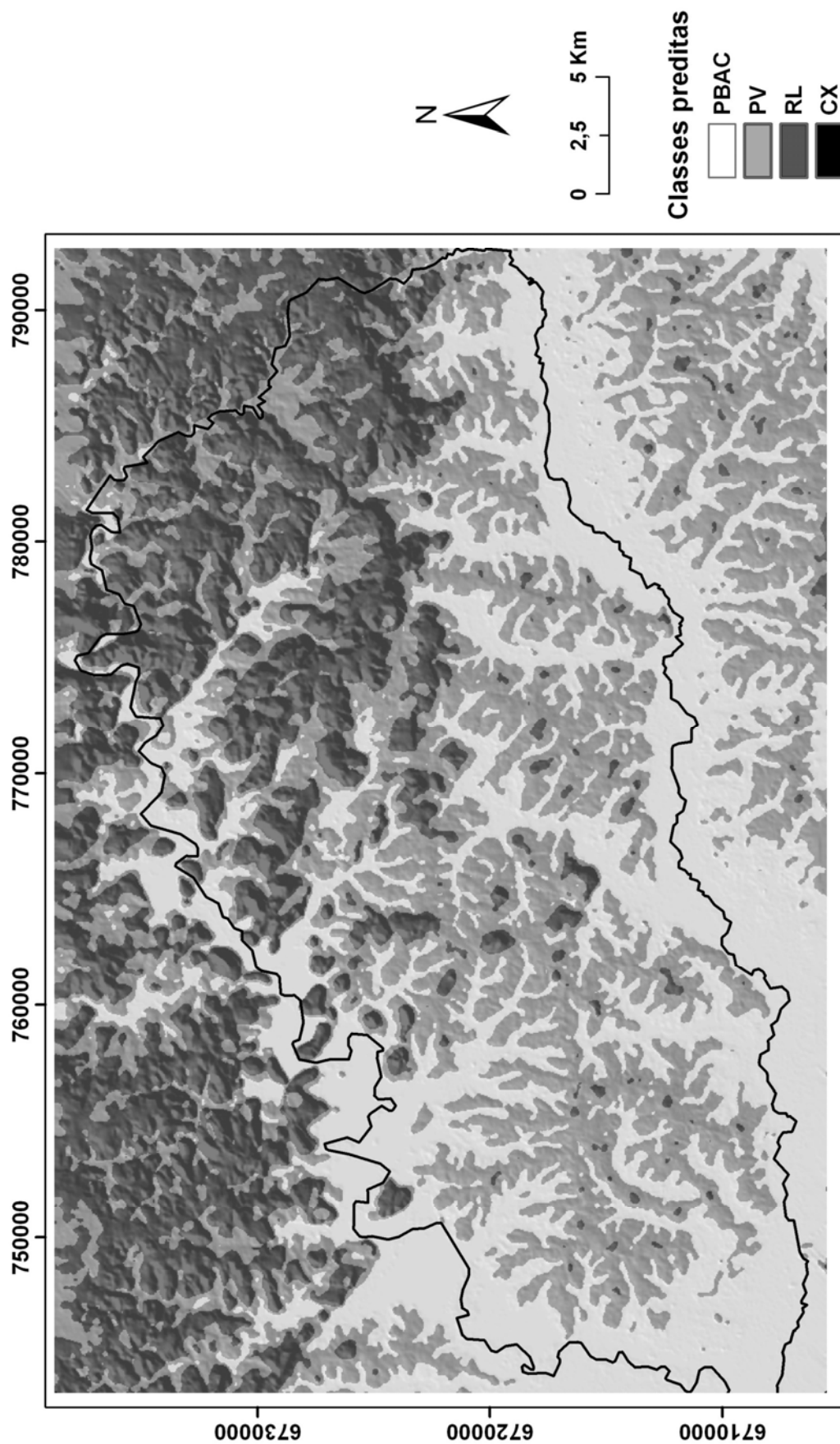


Figura 36 – Mapa de solos predito utilizando-se de componentes principais e o mapa referência. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Neossolo Litólico (RL) e Argissolo Bruno-Acinzentado (PBAC).

Visualizando as unidades delineadas por fotointerpretação sobrepostas ao mapa predito (Figura 37) é possível verificar que o modelo espacializou incorretamente os Argissolos Bruno-Acinzentados nos locais de Solos Hidromórficos. No entanto, a espacialização dos Neossolos Litólicos e dos Argissolos Vermelhos, pelo menos visualmente, ocorreu de maneira bastante razoável.

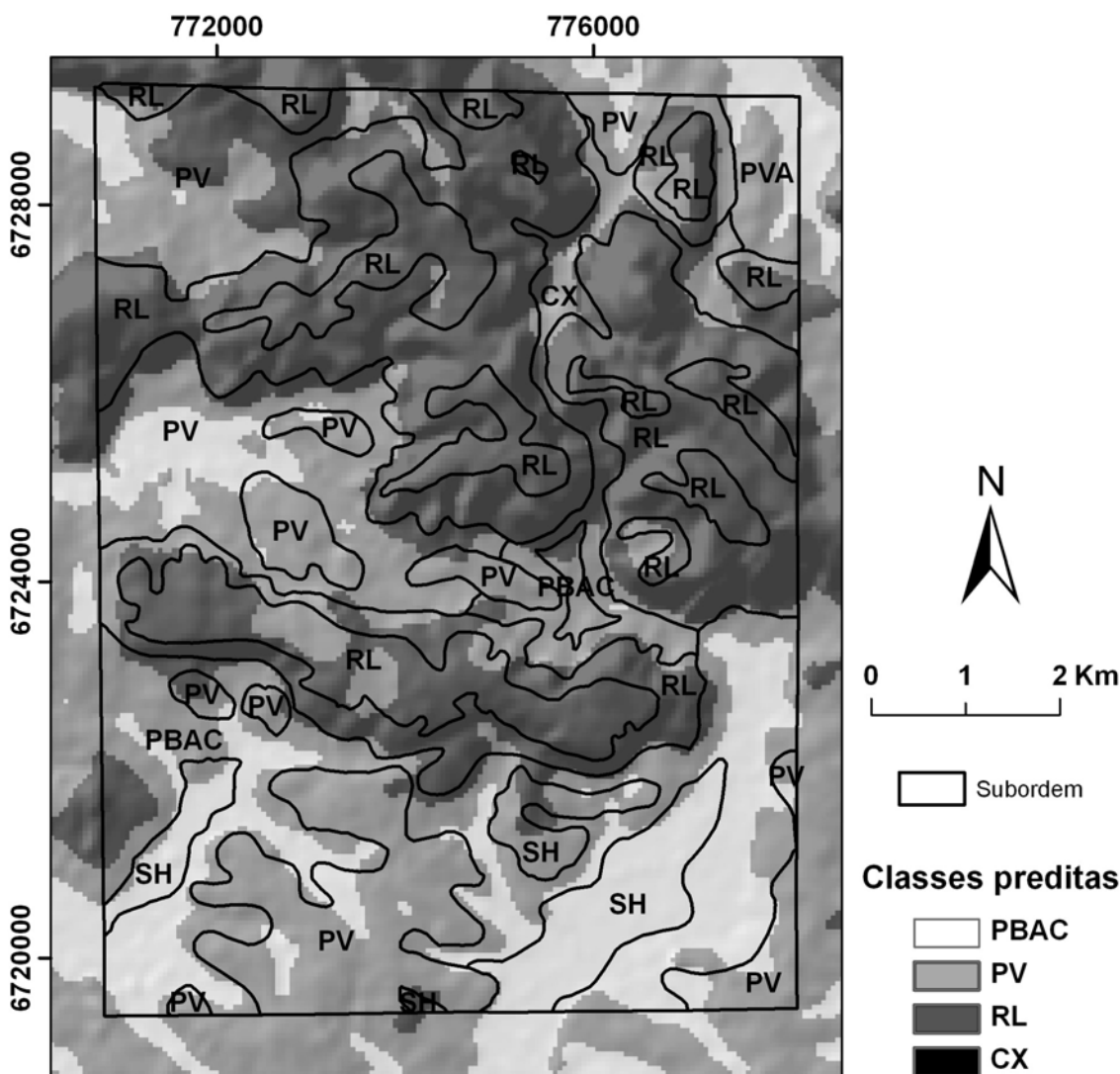


Figura 37 – Unidades de solos oriundas da fotointerpretação sobrepostas ao mapa predito utilizando-se de componentes principais. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo Háptico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-Amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno-Acinzentado (PBAC) e Solo Hidromórfico (SH).

A acurácia do mapa predito confirma o que havia sido visualmente confirmado a respeito dos Neossolos (Tabela 20). A AM destes solos foi de mais de 86%, já a AU ficou em 80%, confirmando que esta classe foi adequadamente espacializada,

pelo menos do ponto de vista de reproduzir o que estava mapeado no mapa referência. A AG do mapa foi de 60,53%, e seu índice K de 46,46%. Estes resultados foram bastante similares aos valores encontrados quando da espacialização das classes de solos a partir do mapa Klamt ao nível de subordem, tabelas 14 e 17.

**Tabela 20 – Matriz de erros entre o mapa da área de referência e o predito produzido a partir das componentes principais como variáveis dos modelos logísticos. Acurácia do mapeador (AM), acurácia do usuário (AU) e acurácia geral (AG) e índice kappa (K).**

Mapa predito	Mapa referência						Total	AU(%)
	SH	PBAC	PV	RL	PVA	CX		
SH	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
PBAC	5,92	14,27	8,76	0,00	2,02	0,00	30,97	46,08
PV	1,62	8,30	27,38	2,83	3,74	1,62	45,50	60,18
RL	0,10	0,71	2,73	18,88	0,00	1,11	23,53	80,24
PVA	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
CX	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>Total</b>	7,64	23,28	38,87	21,71	5,77	2,73	100,00	
<b>AM(%)</b>	0,00	61,30	70,44	86,96	0,00	0,00		
<b>AG(%)</b>	60,53		<b>K'(%)</b>		46,46			

<sup>1</sup>erro padrão assintótico =  $\pm 0,014$  e intervalo de confiança 95% = 0,43...0,49. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-Amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno-Acinzentado (PBAC) e Solo Hidromórfico (SH).

A verificação do mapa predito com a verdade de campo mostra que em apenas 1/3 (33,68%) dos locais visitados os modelos inferiram corretamente a classe de solo (Tabela 21). Pode-se prever que possivelmente este valor seria ainda menor, caso houvessem sido utilizados pontos de campo fora das áreas onde foi realizada a fotointerpretação e o delineamento das classes de solos.

**Tabela 21 – Matriz de erros entre os pontos de campo e o mapa predito produzido a partir das componentes principais como variáveis dos modelos logísticos. Acurácia do mapeador (AM), acurácia do usuário (AU) e acurácia geral e índice kappa (K).**

Mapa predito	Observações a campo						Total	AU(%)
	SH	PBAC	PV	RL	PVA	CX		
<b>SH</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>PBAC</b>	16,07	7,14	5,36	0,00	0,00	0,00	28,57	24,99
<b>PV</b>	1,79	3,57	33,93	3,57	12,50	3,57	58,39	58,11
<b>RL</b>	0,00	0,00	3,57	3,57	1,79	3,57	12,50	28,56
<b>PVA</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>CX</b>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>Total</b>	17,86	10,71	42,86	7,14	14,29	7,14	100,00	
<b>AM(%)</b>	0,00	66,67	79,16	50,00	0,00	0,00		
<b>AG(%)</b>	44,64		<b>K1(%)</b>	33,68				

erro padrão assintótico =  $\pm 0,056$  e intervalo de confiança 95% = 0,22...0,44. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-Amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno-Acinzentado (PBAC) e Solo Hidromórfico (SH).

#### 4.4.2 Mapa predito com área de referência e atributos do terreno como variáveis explicativas.

Os parâmetros dos modelos utilizados para a geração das probabilidades de se encontrar as classes de solos na paisagem a partir de atributos do terreno constam na tabela 22. Uma série de coeficientes foram não significativos para os modelos, ao ponto de a CTS não ter sido utilizada em nenhuma das funções logit.

Novamente entre os maiores valores de probabilidades sobre a paisagem foi verificada a presença de valores iguais a um. Supõe-se que a possível causa poderia ter sido a pouca área ocupada pelos Cambissolos Háplicos, que no caso do mapa Klamt, dividiam a região de encosta com os Neossolos. De tal forma que os modelos logísticos gerados a partir do mapa referência tiveram menos dificuldade em separar os Neossolos das demais classes, resultando com isso em maiores valores de probabilidade nesta região.

**Tabela 22 – Valores dos interceptos e dos coeficientes dos atributos do terreno nos modelos logísticos das áreas de referência.**

	CX <sup>1</sup>				
	SH	PBAC	PV	RL	PVA
<b>Intercepto</b>	36,384	26,510	17,125	<b>-0,278</b>	20,460
<b>ELEV</b>	-0,077	-0,053	-0,027	0,014	-0,045
<b>DECL</b>	-0,238	<b>-0,031</b>	<b>-0,027</b>	<b>-0,050</b>	-0,198
<b>DIST</b>	-0,001	-0,001	0,0003	0,001	0,0005
<b>CPLN</b>	<b>1,293</b>	<b>-0,070</b>	-2,605	-2,018	<b>-1,806</b>
<b>CPRF</b>	4,221	4,147	3,331	1,554	3,000
<b>ILUM</b>	-0,090	-0,053	-0,023	<b>0,001</b>	-0,026
<b>LNAC</b>	-1,572	-1,472	-1,383	<b>-0,311</b>	-1,421
<b>IUT</b>	1,096	1,077	0,989	<b>0,084</b>	0,835
<b>CTS</b>	<b>0,007</b>	<b>-0,006</b>	<b>-0,009</b>	<b>0,010</b>	<b>0,016</b>

<sup>1</sup>Classe de referência para os modelos logísticos. Valores em negrito não são significativos ao nível de 5%. Todos os modelos apresentaram diferença significativa para o teste dos modelos apenas com o intercepto e os mesmos com as variáveis explicativas. Todas as variáveis explicativas têm relação significativa com a variável resposta (*Likelihood Ratio Test*). Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-Amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno-Acinzentado (PBAC), Solo Hidromórfico (SH), Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Iluminação (ILUM), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).

A separação da área de estudo em áreas de várzea, encosta e coxilha, e a associação com os mapas de probabilidade, permite verificar que os solos situados em áreas de coxilha tiveram os menores valores de probabilidade nos apêndices C, E e F. A origem disto pode estar na homogeneidade (menor contraste) dos atributos na região, ou até mesmo, no fato de que as classes desta área não foram utilizadas como classes de referência em nenhuma das etapas deste estudo. A aplicação dos PI de probabilidade como variáveis para outros estudos e modelos deve ser melhor avaliada. Uma vez que um mesmo solo pode ter sido adequadamente alocado na paisagem, no entanto, com valores diferentes de probabilidade.

A utilização dos atributos do terreno como variáveis explicativas possibilitou a espacialização dos Solos Hidromórficos, estes foram posicionados ao longo das drenagens dos rios Ibicuí-mirim e Toropi, especialmente ao sudoeste (Figura 38).

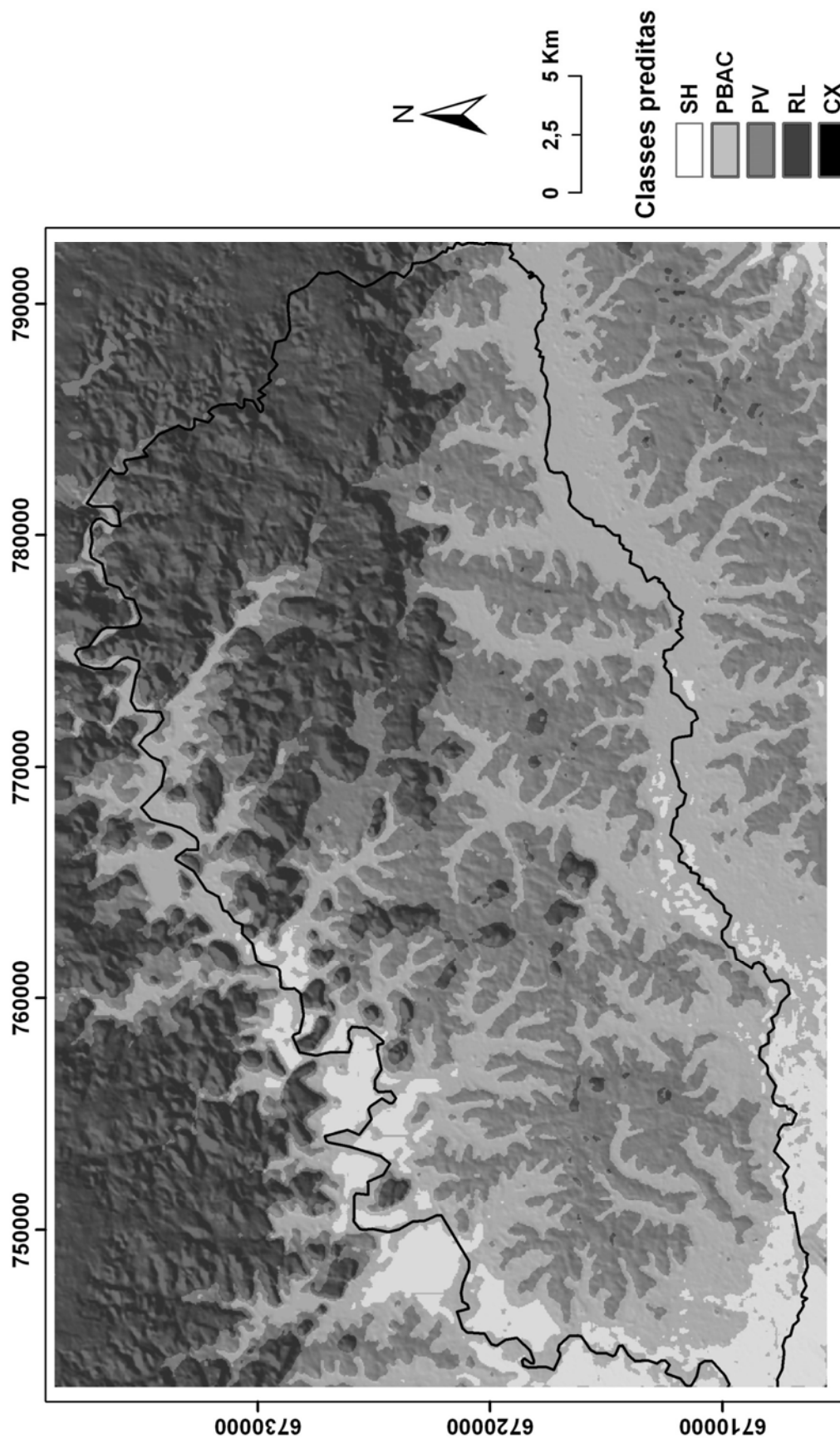


Figura 38 – Mapa de solos predito utilizando-se de atributos do terreno e o mapa referência. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Neossolo Litólico (RL), Argissolo Bruno-Acinzentado (PBAC) e Solo Hidromórfico (SH).

Este resultado pode estar associado ao fato de que a utilização de todos os nove atributos do terreno nos modelos permite a manutenção do padrão de variabilidade dos dados. Diferente do que ocorre quando da utilização somente das CP, as quais descartaram 1/3 da variabilidade dos dados originais. No entanto, ainda ocorreu uma grande confusão na espacialização dos Argissolos Bruno-Acinzentados nas demais áreas de várzea, quando utilizou-se os atributos do terreno como preditores. A posição dos Neossolos na paisagem corresponde visualmente ao que estava delineado no mapa referência (Figura 39). Reproduzindo um resultado muito similar ao visualizado na figura 37.

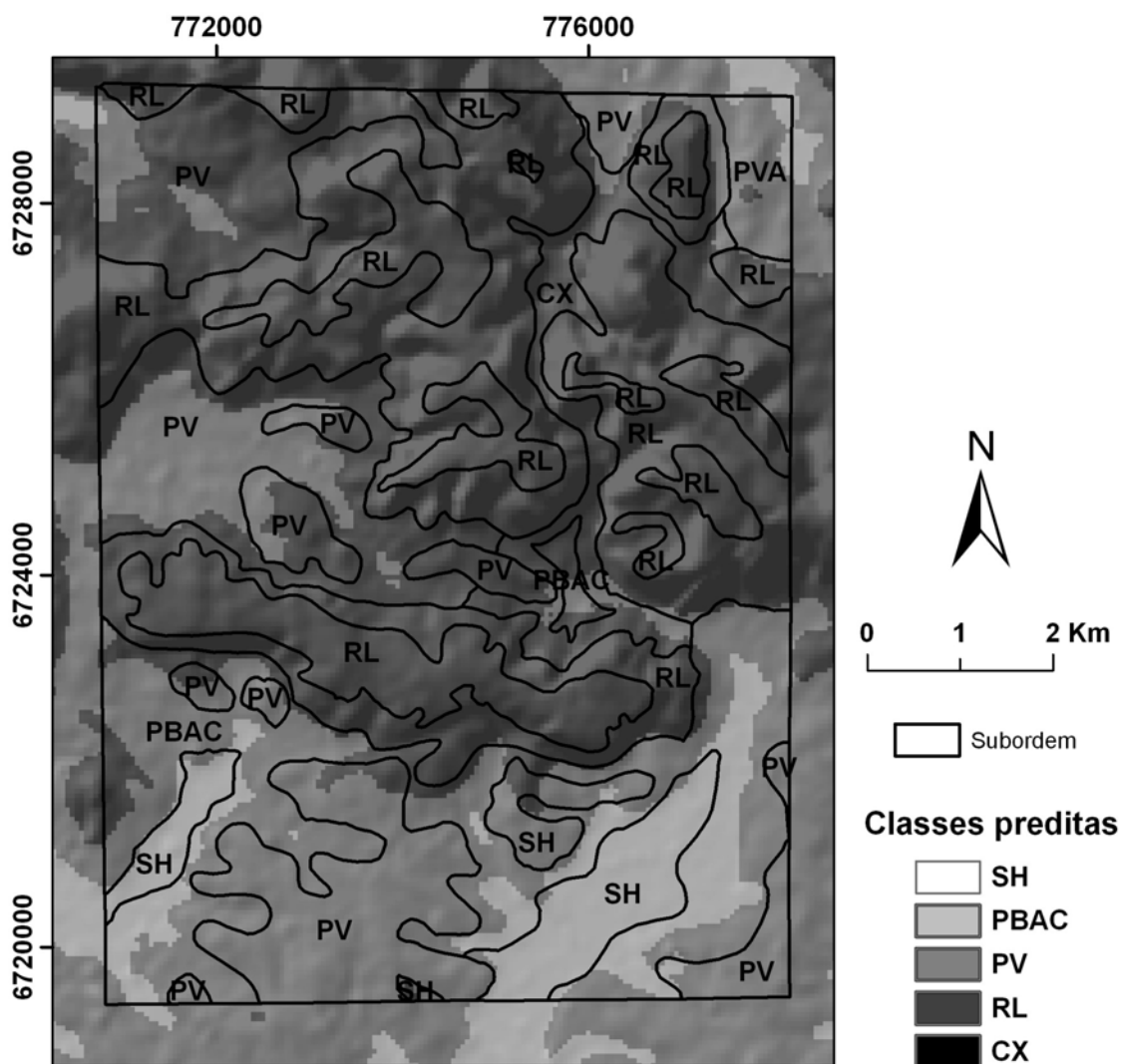


Figura 39 – Unidades de solos oriundas da fotointerpretação sobrepostas ao mapa predito utilizando-se de atributos do terreno. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Neossolo Litólico (RL), Argissolo Bruno-Acinzentado (PBAC) e Solo Hidromórfico (SH).

A similaridade entre os Argissolos Vermelhos em ambos os mapas demonstra que esta classe foi visualmente reproduzida. Contudo em vários pontos mais baixos da paisagem, onde haviam sido localizados no mapa referência os Solos Hidromórficos, o modelo espacializou os Argissolos Bruno-Acinzentados.

A acurácia do mapa predito por atributos do terreno foi sensivelmente superior a aquela gerada pelos modelos com CP (Tabela 23). Sendo que o mapa predito reproduziu em 50% as informações que constavam no mapa referência. A AM para os Solos Hidromórficos demonstra que estes foram confundidos com os Argissolos Bruno-Acinzentados em grande parte dos locais amostrados para o teste da acurácia.

**Tabela 23 – Matriz de erros entre o mapa da área de referência e o predito produzido a partir dos atributos do terreno como variáveis dos modelos logísticos. Acurácia do mapeador (AM), acurácia do usuário (AU) e acurácia geral e índice kappa (K).**

Mapa predito	Mapa referência						Total	AU(%)
	SH	PBAC	PV	RL	PVA	CX		
SH	0,46	0,30	0,00	0,00	0,00	0,00	0,76	60,53
PBAC	6,02	13,61	7,29	0,00	1,42	0,00	28,34	48,02
PV	1,01	8,60	26,32	0,30	4,35	0,40	40,99	64,21
RL	0,15	0,76	5,26	21,41	0,00	2,33	29,91	71,58
PVA	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
CX	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>Total</b>	7,64	23,28	38,87	21,71	5,77	2,73	100,00	
<b>AM(%)</b>	6,02	58,46	67,71	98,62	0,00	0,00		
<b>AG(%)</b>	61,79		<b>K<sup>1</sup>(%)</b>	51,15				

erro padrão assintótico =  $\pm 0,013$  e intervalo de confiança 95% = 0,48...0,53. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-Amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno-Acinzentado (PBAC) e Solo Hidromórfico (SH).

Em relação a verdade de campo os Solos Hidromórficos foram adequadamente espacializados (Tabela 24). No entanto, é importante frisar que esta é a realidade do conjunto de 56 pontos visitados. Não sendo este número nem muito elevado, e nem correspondente a pontos fora dos locais utilizados como mapa referência. A acurácia obtida pelo índice K confirma resultados similares aos encontrados para o nível de subordem utilizando o mapa Klamt.



**Tabela 24 – Matriz de erros entre os pontos de campo e o mapa predito produzido a partir dos atributos do terreno como variáveis dos modelos logísticos. Acurácia do mapeador (AM), acurácia do usuário (AU) e acurácia geral e índice kappa (K).**

Mapa predito	Observações a campo						Total	AU(%)
	SH	PBAC	PV	RL	PVA	CX		
SH	5,36	0,00	0,00	0,00	0,00	0,00	5,36	100,00
PBAC	10,71	5,36	3,57	0,00	0,00	0,00	19,64	27,29
PV	1,79	3,57	23,21	1,79	10,71	0,00	41,07	56,51
RL	0,15	1,79	16,07	5,36	3,57	7,14	33,93	15,80
PVA	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
CX	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
<b>Total</b>	17,86	10,71	42,86	7,14	14,29	7,14	100,00	
<b>AM(%)</b>	30,01	50,05	54,15	75,07	0,00	0,00		
<b>AG(%)</b>	39,29		<b>K¹(%)</b>	37,90				

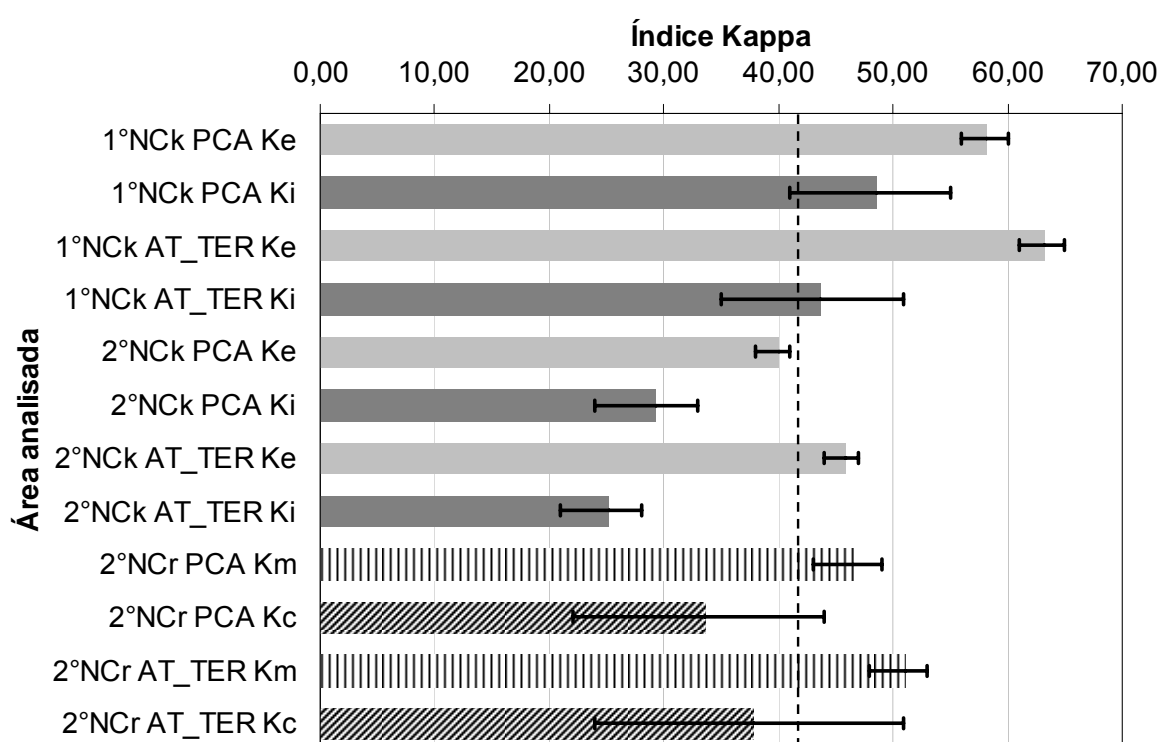
<sup>1</sup>erro padrão assintótico =  $\pm 0,068$  e intervalo de confiança 95% = 0,24...0,51. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-Amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno-Acinzentado (PBAC) e Solo Hidromórfico (SH).

#### 4.5 Discussão geral do estudo

A aplicação da ACP possibilitou visualizar as relações entre os atributos do terreno, assim como, o grau de multicolinearidade entre os dados. As CP utilizadas permitiram uma simplificação dos modelos, embora com duas importantes implicações. A primeira foi uma sensível menor qualidade dos modelos preditivos por CP (Figura 40). Os modelos preditivos que empregaram CP obtiveram menores valores do índice K na área de treinamento para os modelos. Uma possível explicação para este fato seria a de que a retenção de apenas três componentes causou uma perda de 1/3 do poder de explicação das variáveis, conseqüentemente os modelos desempenharam com qualidade inferior a aqueles nos quais foram empregados todos os nove atributos do terreno. Embora uma análise do índice K médio geral entre os modelos que utilizaram CP, e aqueles que utilizaram atributos do terreno como variáveis explicativas, indique um valor de K apenas 4% inferior nos modelos que utilizaram as CP.

A segunda implicação, e talvez a mais grave sob o ponto de vista pedológico, é a de que os modelos gerados por CP não expressam nenhuma relação clara entre os atributos do terreno e as classes de solos a serem preditas. Especialmente se considerarmos que as novas variáveis ortogonalizadas são utilizadas em funções

logit, que por sua vez irão produzir um valor de probabilidade. Observa-se que o uso desta metodologia torna as relações entre o solo e variáveis preditivas, e a possibilidade de se estabelecer relações mais claras entre o solo e a paisagem mais obscuras e complexas. Essa é uma grande desvantagem desta metodologia se comparada a outras técnicas de predição de classes de solos como as árvores de decisão, as quais produzem claras relações entre os fatores de formação e as classes de solos.



**Figura 40 – Valores do índice kapa em todos os modelos testados. Linhas sobre as barras indicam Intervalo de Confiança de 95%. Linha vertical tracejada indica o kapa médio de todos os valores obtidos. Variável resposta oriunda de mapa Klamt ao nível de ordem (1°NCK), variável resposta oriunda de mapa Klamt ao nível de subordem (2°NCK), variável resposta oriunda de mapa referência ao nível de subordem (2°NCKr), componente principal como variável explicativa (PCA), atributos do terreno como variável explicativa (AT\_TER), índice kapa utilizando mesma área de onde foram extraídas informações para os modelos (Ke), índice kapa utilizando área de onde não foram extraídas informações para os modelos (Ki), índice kapa utilizando mapa da área de referência (Km), índice kapa gerado a partir de comparação com dados de campo (Kc).**

Contudo, o uso da ACP pode ser importante em estudos que utilizem um grande número de covariáveis para predição de classes, características ou propriedades do solo. Vários estudos em MDS têm reportado que a qualidade dos mapas preditos melhoraria caso houvessem sido utilizados mapas de geologia, vegetação, uso da terra, ou até mesmo, uma variedade maior de atributos do terreno. Neste estudo os valores de K encontrados foram similares aos de outros estudos já realizados. Embora o real ganho com a aplicação de CP não tenha sido alcançado, a teoria envolvida nesta técnica permite afirmar que ela tem grande aplicabilidade no sentido de diminuir o número de preditores, além de permitir visualizar as correlações entre as variáveis explicativas.

À medida que a metodologia foi empregada em um nível do sistema de classificação cada vez mais baixo, a acurácia dos mapas preditos diminuiu (Figura 40). Pelo menos duas podem ter sido as causas. Uma delas seria a redução da área de treinamento dos modelos para cada uma das classes de solos. Isto tem implicações na medida em que as RLM são sensíveis a frequência relativa das variáveis a serem preditas. Ou seja, haveria um limite mínimo (escala ou nível categórico) para que uma dada categoria pudesse ser predita, a partir deste limite ela teria que ser considerada apenas como uma inclusão ou por associações.

Um segundo elemento que pode ter contribuído para a menor acurácia dos modelos, seria a maior homogeneidade entre as classes à medida que níveis categóricos mais baixos são empregados. Especialmente utilizando-se um baixo número de preditores como neste estudo. A capacidade dos modelos em predizer adequadamente as classes de solos à medida que homogeneízam-se as classes, poderia ser melhorada com o uso de mais preditores como a geologia, uso da terra, ou no aumento da resolução dos arquivos rasters utilizados. Esta última opção representaria uma maior amostragem das variações do relevo, que poderia ser crucial na tentativa de se distinguir entre classes de solos muito próximas taxonomicamente, ou sob o ponto de vista dos processos que as formam.

Os modelos têm uma capacidade preditiva menor do que a reprodução do padrão das áreas de treinamento (Figura 40). Em todos os níveis categóricos, áreas de treinamento e variáveis preditivas empregadas os valores do índice K foram menores quando testou-se a validação dos modelos do que quando se verificou a reprodutibilidade dos mapas de treinamento. Este resultado já era esperado, uma vez que o emprego de um modelo de regressão em uma área distinta daquela onde

tiveram origem seus dados, estará condicionada a uma similaridade entre os locais (ou fenômenos modelados). De tal forma que a utilização de mapas de solos existentes para o treinamento de modelos de predição para áreas não mapeadas, passará primeiro por um estudo da real similaridade do ponto de vista de classes, processos e fatores de formação entre as áreas. Da mesma maneira, no emprego da técnica de áreas de referência, precisa-se primeiramente definir a dimensão (raio de abrangência), até a qual a região pode ser considerada homogênea. Nela a área de referência ainda poderá ser considerada válida, mesmo afastada da área onde foi originalmente concebida.

A técnica de mapeamento por área de referência produziu resultados superiores àqueles obtidos pelo mapa Klamt em um mesmo nível categórico (Figura 40), embora os modelos empregados no mapa Klamt tenham tido uma desvantagem, na medida em que havia neste mapa duas classes a mais de solos a serem preditas (Plintossolos Argilúvicos e Argissolos Acinzentados). Contudo, assumindo que ambas as técnicas gerassem (em reais condições comparativas) resultados similares, pode-se afirmar que o emprego da metodologia baseada em área de referência apresenta-se como uma alternativa para a produção de mapas de solos com qualidade compatível a uma série de aplicações, principalmente se considerarmos os custos e rapidez para a geração das informações. Quanto às áreas de referência não foi determinada sua região de abrangência ou até que distância sua aplicação ainda seria válida.

A acurácia do mapeamento empregando o índice K tende a ter uma qualidade menor quando um menor número de observações é utilizado para seu cálculo (Figura 40). Como foram utilizadas apenas 56 amostras de campo para a verificação da acurácia do mapa predito por área de referência, o Intervalo de Confiança desta medida tem uma abrangência muito maior. Embora os intervalos mais baixos do K para a área referência ainda sejam comparáveis aos valores de acurácia encontrados para o mapa Klamt no mesmo nível categórico.

#### **4.6 Considerações finais**

Com o emprego de regressões logísticas para a predição de classes de solos obteve-se neste estudo um índice K médio de 43,58% (mínimo = 25,16% e máximo = 63,20%) (Figura 40). Uma série de elementos podem estar colaborando para que este valor não seja mais elevado, tais como: a) a qualidade e a resolução do modelo

digital de elevação utilizado; b) problemas na derivação dos atributos do terreno utilizados os quais são muito sensíveis aos algoritmos utilizados; c) seleção não adequada ou em número insuficiente dos preditores para a região considerando os solos nela existentes; d) a qualidade dos mapas de solos utilizados para treinar os modelos, principalmente quanto aos seus delineamentos e à acurácia; e) análise estatística empregada; e f) a distribuição espacial do solo, que é de difícil modelagem.

A metodologia foi capaz de gerar informações sobre os solos para regiões onde estas informações não estão disponíveis (Apêndice Ga). Cabe aos pesquisadores continuar investigando as potencialidades desta ferramenta para o mapeamento pedométrico de solos. Pontos importantes a serem considerados em futuros estudos são: 1) covariáveis a serem utilizadas no mapeamento preditivo; 2) emprego de CP em um número consideravelmente maior de covariáveis; 3) utilização das RLM como ferramenta de mapeamento prévio e planejamento; 4) estudo da aplicabilidade dos mapas de probabilidades gerados pelas RLM (Apêndice Gb); e 5) influência das proporções relativas entre as categorias a serem preditas por RLM.

Muito ainda deve ser melhorado no sentido de tornar a metodologia “empregável” em levantamentos sistemáticos. Uma proposta neste sentido poderia ser a geração de rotinas automatizadas em programas bem difundidos de SIG, através da programação de metodologias como a empregada neste estudo. Figueiredo (2006) acredita que o emprego de RLM é de fácil compreensão e operacionalização. No entanto, para Webster (2001) os pesquisadores em ciência do solo têm dificuldades em empregar adequadamente matemática e estatística. O próprio estudo das probabilidades envolvidas nas regressões logísticas poderá representar, para muitos pedólogos, uma barreira. Aliado a isso, se for considerada sua associação com as CP, as quais não têm nenhuma relação com o fenômeno natural, poderá significar o não emprego definitivo desta metodologia, a qual precisa ser ainda muito melhor estudada quanto ao seu potencial para a disponibilização de informações sobre os solos.

## 5 CONCLUSÕES

As relações entre classes de solos e atributos do terreno em mapas existentes e de áreas de referência podem ser descritas por modelos matemáticos, os quais possibilitam a predição e a espacialização das classes de solos na paisagem.

A Análise de Componentes Principais permite uma melhor visualização das correlações entre os atributos do terreno, além de possibilitar uma diminuição do número de covariáveis nos modelos logísticos.

As Regressões Logísticas Múltiplas têm potencial para a predição e o mapeamento de classes de solos, embora classes relativamente não representativas nos mapas de treinamento não sejam acuradamente mapeadas pela metodologia.

O emprego das Componentes Principais como variáveis explicativas nos modelos logísticos implica em perda da acurácia dos mapas preditos, devido a um menor poder de explicação das componentes retidas comparativamente as variáveis originais, embora ocorra um ganho quando um menor número de variáveis é utilizado nos modelos.

A utilização de áreas de referência para o treinamento de modelos logísticos produz resultados sensivelmente superiores àqueles que utilizam mapas de solos existentes. Embora esta conclusão esteja baseada em um único teste com um mapa da área de referência simplificado em relação ao mapa existente.

## 6 PERSPECTIVAS

A predição de classes de solos pode ser realizada utilizando-se da técnica de árvores de decisão. Empregada em uma área previamente mapeada, como área de referência ou em um mapa existente, esta técnica de mineração dos dados pode estabelecer as “regras” do modelo mental utilizada pelo pedólogo para espacializar as classes de solo na paisagem. Em estudos futuros de MDS com árvores de decisão pretender-se-á o estabelecimento de relações mais claras entre o solo e a paisagem a partir do conhecimento contido nos mapas pedológicos.

O uso de preditores, que não apenas a variável relevo ( $r$ ) da função de Jenny, poderá representar um ganho na capacidade preditiva dos modelos. Mapas de uso da terra podem ser gerados para todas as regiões em uma grande variedade de resoluções espaciais. Muitas áreas contam ainda com mapas geológicos, que uma vez usados como preditores nos modelos potencializarão a predição de classes e propriedades, haja vista a importância do elemento geologia para a formação dos solos.

Por tudo que está envolvido no MDS seu emprego representa a oportunidade e motivação para o aprendizado em matemática/estatística, ciência do solo e geoinformação. Muitas são as possibilidades para a melhor função ( $f$ ) que descreva as relações entre o solo e a paisagem ou os preditores. A busca da função mais adequada para cada situação, e a possibilidade de se conhecer os erros associados à cada escolha, permite ao cientista do solo envolvido no MDS potencializar seus conhecimentos em matemática e estatística. O estudo dos fatores que determinam a distribuição espacial de classes e propriedades do solo possibilita aprendizado e aprimoramento constante dos conhecimentos em ciência do solo. Por fim, a utilização de ferramentas de geoinformação na geração dos preditores ou na implementação das funções em um SIG, torna o MDS uma fascinante área para o desenvolvimento de atividades de pesquisa.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

BAILEY, N. et al. Modelling soil series data to facilitate targeted habitat restoration: a polytomous logistic regression approach. **Journal of Environmental Management**, v. 67, n. 4, p. 395-407, 2003.

BANKO, G. A review of assessing the accuracy of classifications of remotely sensed data and of methods including remote sensing data in forest inventory. International Institute for Applied Systems Analysis. Laxenburg, p.36. 1998.

BASHER, L. R. Is pedology dead and buried? **Australian Journal of Soil Research**, v. 35, n. 5, p. 979-994, 1997.

BEAUCHEMIN, S.; HESTERBERG, D.; BEAUCHEMIN, M. Principal component analysis approach for modeling sulfur K-XANES spectra of humic acids. **Soil Science Society American Journal**, Madison, v. 66, p. 83-91, 2002.

BENEDUZI, A. et al. Evaluation of genetic diversity and plant growth promoting activities of nitrogen-fixing bacilli isolated from rice fields in South Brazil. **Applied Soil Ecology**, v. 39, n. 3, p. 311-320, 2008.

BORUVKA, L.; V., P. A test of an artificial neural network allocation procedure using the Czech Soil Survey of Agricultural Land data. . In: LAGACHERIE, P.; MCBRATNEY, A.; VOLTZ, M. (Ed.). **Digital soil mapping: an introductory perspective**. Amsterdam: Elsevier, 2007, p. 415-424.

BOUMA, J. Soil Scientist in a changing world. **Advances in Agronomy**, v. 88, p. 67-96, 2005.

BUI, E. A review of digital soil mapping in Australia. In: LAGACHERIE, P.; MCBRATNEY, A.; VOLTZ, M. (Ed.). **Digital soil mapping: an introductory perspective**. Amsterdam: Elsevier, 2007, p. 25-37.

BUI, E. N. Soil survey as a knowledge system. **Geoderma**, Amsterdam, v. 120, n. 1-2, p. 17-26, 2004.

BUI, E. N; MORAN, C. J. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. **Geoderma**, Amsterdam, v. 103, p. 79-74, 2001.

BURROUGHS, P. A.; BOUMAB, J.; YATES, S. R. The state of the art in pedometrics. **Geoderma**, Amsterdam, v. 62, n. 3, p. 311-326, 1994.

CABRAL, I. L. L.; SILVA, R. P. **Relatório de consultoria: Mapeamento geomorfológico e uso da terra do município de São Pedro do Sul – RS**. In: MACIEL FILHO, C. L. (Coord.) – UFSM, 1998, 26 p.

CHAGAS, C. D. S. **Mapeamento digital do solo por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de Mar dos Morros**. 2006. 238 f.



Tese (Doutorado em Solos e Nutrição de Plantas) - Universidade Federal de Viçosa, Viçosa.

CHATTERJEE, S.; HADI, A. S. **Regression analysis by example**. John Willey & Sons, 4. ed., 2006. 375 p.

DALMOLIN, R. S. D. Faltam pedólogos no Brasil. **Boletim Informativo da Sociedade Brasileira de Ciência do Solo**, n. 24, p. 13-15, 1999.

DAVIDSON, E. A. Dirt cheap soil. **Nature**, v. 447, p. 777-778, 2007.

DEBELLA-GILO, M.; ETZELMULLER, B.; KLAKEGG, O. Digital soil mapping using digital terrain analysis and statistical modeling integrated into GIS: Examples from Vestfold county of Norway. ScanGIS'2007. In: SCANDINAVIAN RESEARCH CONFERENCE ON GEOGRAPHICAL INFORMATION SCIENCES, 11., 2007, Norway. Proceedings... Norway, 2007, 274 p.

DENT, D.; YOUNG, A. **Soil Survey and Land Evaluation**. E & FN Spon, 1, 1981. 278 p.

DOBOS, E. et al. Digital soil mapping as a support to production of functional maps European Communities. Luxemburg, 2006, 68 p.

EMBRAPA. **Sistema brasileiro de classificação de solos**. 2. ed., Rio de Janeiro: Embrapa Solos, 2006, 306 p.

FIGUEIREDO, S. R. **Mapeamento supervisionado de solos através do uso de regressões logísticas múltiplas e sistema de informações geográficas**. 2006. 96f. Dissertação (Mestrado em Ciência do Solo), Universidade Federal do Rio Grande do Sul, Porto Alegre.

GIASSON, E. et al. Digital soil mapping using logistic regression on terrain parameters for several ecological regions in southern Brazil. In: HARTEMINK, A. E.; MCBRATNEY, A.; MENDONÇA SANTOS, M. DE L. (Ed.). **Digital soil mapping with limited data**. New York: Springer-Verlag, 2008, p. 225-232.

GIASSON, E.; INDA JUNIOR, A. V.; NASCIMENTO, P. C. D. Estimativa do benefício econômico potencial de dois levantamentos de solos no Estado do Rio Grande do Sul. **Ciência Rural**, Santa Maria, v. 36, p. 478-486, 2006.

GOMES, J. B. V. et al. Análise de componentes principais de atributos físicos, químicos e mineralógicos de solos do bioma cerrado. **Revista Brasileira de Ciência do Solo**, v. 28, n. 1, 2004.

GROHMANN, C. H.; RICCOMINI, C.; ALVES, F. M. SRTM-based morphotectonic analysis of the Poços de Caldas Alkaline Massif, southeastern Brazil. **Computers & geosciences**, v. 33, n. 1, p. 10-19, 2007.

HENGL, T. Pedometric mapping: bridging the gaps between conventional and pedometric approaches, 2003, 236 f. Tese (Doutorado). Wageningen University, Enschede.

HENGL, T.; ROSSITER, D. G. Supervised Landform Classification to Enhance and Replace Photo-Interpretation in Semi-Detailed Soil Survey. **Soil Science Society of America Journal**, Madison, v. 67, n. 6, p. 1810, 2003.

HENGL, T. et al. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. **Geoderma**, Amsterdam, v. 140, n. 4, p. 417-427, 2007.

HOWELL, D.; KIM, Y. G.; HAYDU-HOUDESHELL, C. A. Development and application of digital soil mapping within traditional soil survey: What will it grow into? In: HARTEMINK, A. E.; MCBRATNEY, A.; MENDONÇA SANTOS, M. DE L. (Ed.). **Digital soil mapping with limited data**. New York: Springer, 2008, p. 43-51.

HUDSON, B. D. The Soil Survey as Paradigm-based Science. **Soil Science Society of American Journal**, Madison, v. 56, p. 836-841, 1992.

IPPOLITI R, G. A. et al. Análise digital do terreno: ferramenta na identificação de pedoformas em microbacia na região de "Mar de Morros" (MG). **Revista Brasileira de Ciência do Solo**, v. 29, n. 269-276, 2005.

JENNY, H. **Factors of soil formation**: a system of quantitative pedology. New York, London: McGraw-Hill, 1941, 324 p.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice-Hall, 1992. 642 p.

KLAMT, E.; FLORES, C. A.; CABRAL, D. R. **Solos do Município de São Pedro do Sul**. Departamento. de Solos/CCR/UFMS. Santa Maria, 2001, 96 p.

LAGACHERIE, P.; LEGROS, J. P.; BURFOUGH, P. A. A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. **Geoderma**, Amsterdam, v. 65, n. 3-4, p. 283-301, 1995.

LAGACHERIE, P.; MCBRATNEY, A. B. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P.; MCBRATNEY, A.; VOLTZ, M. (Ed.). **Digital soil mapping**: an introductory perspective. Amsterdam: Elsevier, 2007, p. 3-22.

LEGROS, J. P. **Mapping of the soil**. Enfield: Science Publisher, v. 1, 2005, 411 p.

MCBRATNEY, A. B.; MENDONÇA SANTOS, M. L.; MINASNY, B. On digital soil mapping. **Geoderma**, Amsterdam, v. 117, n. 1-2, p. 3-52, 2003.

MCBRATNEY, A. et al. An overview of pedometric techniques for use in soil survey. **Geoderma**, Amsterdam, v. 97, n. 3-4, p. 293-327, 2000.

MCKENZIE, N. J. et al. The role of terrain analysis in soil mapping. In: WILSON, J. P.; GALLANT, J. C. (Ed.). **Terrain analysis**: principles and applications. New York: Wiley & Sons, 2000, p. 245-265

MCNEILL, J. R.; WINIWARTER, V. Breaking the Sod: Humankind, History and Soil. **Science**, v. 304, p. 1627-1629, 2004.

MENDONÇA-SANTOS, M. L.; SANTOS, H. G. D. The state of the art of Brazilian soil mapping and prospects for digital soil mapping In: LAGACHERIE, P.; MCBRATNEY, A.; VOLTZ, M. (Ed.). **Digital soil mapping: an introductory perspective**. Amsterdam: Elsevier, 2007, p. 39-54.

MERMUT, A. R.; ESWARAN, H. Some major developments in soil science since the mid-1960s. **Geoderma**, Amsterdam, v. 100, n. 3-4, p. 403-426, 2001.

MINASNY, B.; MCBRATNEY, A. B. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. **Geoderma**, v. 142, n. 3-4, p. 285-293, 2007.

MINASNY, B.; MCBRATNEY, A. B. Mechanistic soil-landscape modelling as an approach to developing pedogenetic classifications. **Geoderma**, Amsterdam, v. 133, n. 1-2, p. 138-149, 2006.

MINASNY, B.; MCBRATNEY, A. B.; SALVADOR-BLANES, S. Quantitative models for pedogenesis - A review. **Geoderma**, Amsterdam, v. 144, n. 1-2, p. 140-157, 2008.

MOORE, I. D. et al. Soil attribute prediction using terrain analysis. **Soil Science Society American Journal**, Madison, v. 57, p. 443-452, 1993.

ODEH, I. O. A.; CHITTLEBOROUGH, D. J.; MCBRATNEY, A. B. Elucidation of soil-landform interrelationships by canonical ordination analysis. **Geoderma**, Amsterdam, v. 49, p. 1-32, 1991.

OHLMACHER, G. C.; DAVIS, J. C. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. **Engineering Geology**, v. 69, n. 3-4, p. 331-343, 2003.

OLIVEIRA, V. Á. D. As implicações da informatização nos levantamentos pedológicos. **Boletim Informativo da Sociedade Brasileira de Ciência do Solo**. n. 32, p. 46, 2007.

OVALLES, F. A.; COLLINS, M. E. Variability of northwest Florida soils by principal component analysis. **Soil Science Society of American Journal**, Madison, v. 52, p. 1430-1435, 1988.

PAULIN, G. L.; BURSİK, M. Logisnet: A tool for multimethod, multiple soil layers slope stability analysis. **Computers & geosciences**, v. In Press, Corrected Proof, 2008.

PENG, C.-Y. J. et al. The Use and Interpretation of Logistic Regression in Higher Education Journals: 1988–1999. **Research in Higher Education**, v. 43, n. 3, p. 259-293, 2002.

PENNOCK, D. J.; VELDKAMP, A. Advances in landscape-scale soil research. **Geoderma**, Amsterdam, v.133, n.1-2, p.1-5, 2006.

PISSARRA, T. C. T.; POLITANO, W.; FERRAUDO, A. S. Avaliação de características morfométricas na relação solo-superfície da Bacia Hidrográfica do Córrego Rico, Jaboticabal (SP). **Revista Brasileira de Ciência do Solo**, v. 28, n. 2, 2004.

POPPI, R. J.; SENA, M. M. D. Métodos quimiométricos na análise integrada de dados. In: R. T. S. Frighetto P. J. Valarini (Ed.). **Indicadores biológicos e bioquímicos da qualidade do solo**: manual técnico. Jaguariúna: EMBRAPA Meio Ambiente, 2000, p. 198.

QI, F.; ZHU, A. X. Knowledge discovery from soil maps using inductive learning. **International Journal of Geographical Information Science**, v. 17, n. 8, p. 771-795, 2003.

RABUS, B. et al. The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. **Journal of Photogrammetry & Remote Sensing**, v. 57, p. 241- 262, 2003.

REAL, R.; BARBOSA, A.; VARGAS, J. Obtaining Environmental Favourability Functions from Logistic Regression. **Environmental and Ecological Statistics**, v. 13, n. 2, p. 237-245, 2006.

REIS, E. **Estatística Multivariada Aplicada**. Lisboa: Sílabo, 1997. 342 p.

ROSSITER, D. G. Technical Note: Statistical methods for accuracy assesment of classified thematic maps. International Institute for Geo-Information Science and Earth Observation. Enschede, 2004, 46 p.

SANTOS, E. L.; MOREIRA, J. M. **Projeto Sítios Paleontológicos do Estado do Rio Grande do Sul - Fase 1. - Seção de Geologia e Mineralogia**. Ministério das Minas e Energia/Departamento Nacional de Produção Mineral. Porto Alegre, 1987, 15 p.

SCULL, P.; FRANKLIN, J.; CHADWICK, O. A. The application of classification tree analysis to soil type prediction in a desert landscape. **Ecological Modelling**, v. 181, n. 1, p. 1-15, 2005.

SCULL, P. et al. Predictive soil mapping: a review. **Progress in Physical Geography**, v. 27, p. 171-197, 2003.

SOUSA JUNIOR, J. G. D. A.; DEMATTÊ, J. A. M. Modelo digital de elevação na caracterização de solos desenvolvidos de basalto e material arenítico. **Revista Brasileira de Ciência do Solo**, v. 32, p. 449-456, 2008.

STRECK, E. V. et al. **Solos do Rio Grande do Sul**. 2 ed. Porto Alegre: EMATER/RS, 2008. 222 p., v. 1

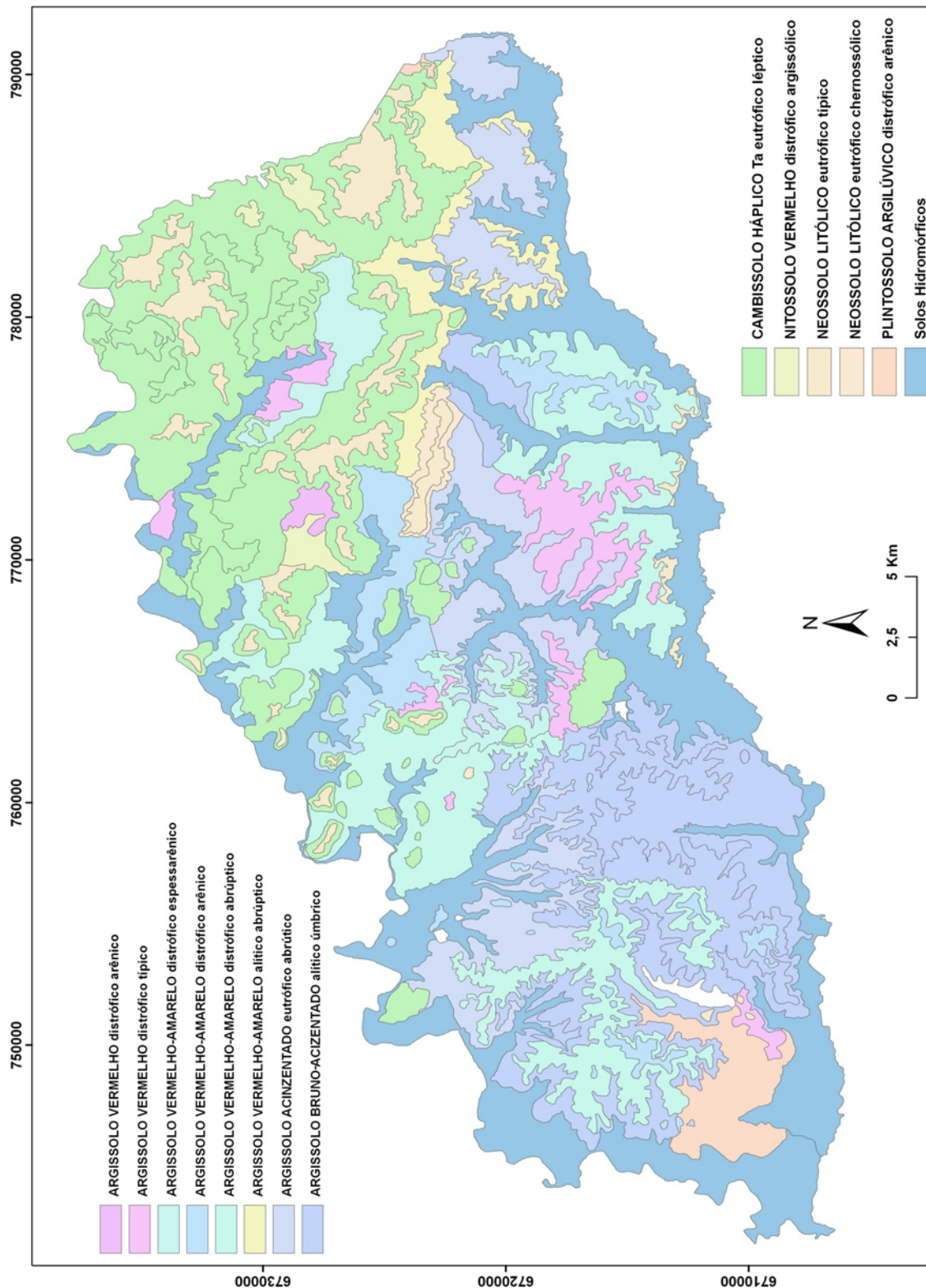
- UEBERSAX, J. Kappa Coefficients. Disponível em: <<http://ourworld.compuserve.com/homepages/jsuebersax/kappa.htm>>. Acesso em 20/09/2008.
- USGS, S. G. D. E. U. Orthorectified Landsat Digital Data Now Available From USGS. Disponível em: <http://www.usgs.gov/newsroom/article.asp?ID=1425>. Acesso em: 2008
- VITHARANA, U. W. A. et al. Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. **Geoderma**, Amsterdam, v. 143, n. 1-2, p. 206-215, 2008.
- VOLTZ, M.; LAGACHERIE, P.; LOUCHARTE, X. Predicting soil properties over a region using sample information from a mapped reference area. **European Journal of Soil Science**, v. 48, n. 1, p. 19-30, 1997.
- WALKER, P. H.; HALL, G. F.; PROTZ, R. Relation between landform parameters and soil properties. **Soil Science Society of American Journal**, v. 32, p. 101-104, 1968.
- WALTER, C.; LAGACHERIE, P.; FOLLAIN, S. Integrating pedological knowledge into digital soil mapping In: LAGACHERIE, P.; MCBRATNEY, A.; VOLTZ, M. (Ed.). **Digital soil mapping: an introductory perspective**. Amsterdam Elsevier, 2007, p. 281-300.
- WEBER, E. et al. GIS as a support to soil mapping in southern Brazil. In: HARTEMINK, A. E.; MCBRATNEY, A.; MENDONÇA SANTOS, M. DE L. (Ed.). **Digital soil mapping with limited data**. New York: Springer, 2008, p. 103-112.
- WEBSTER, R. The development of pedometrics. **Geoderma**, Amsterdam, v. 62, n. 1-3, p. 1-15, 1994.
- WEBSTER, R. Statistics to support soil research and their presentation. **European Journal of Soil Science**, v. 52, p. 331-340, 2001.
- WEBSTER, R.; BURROUGH, P. A.; Computer-based soil mapping of small areas from sample data. **Journal of Soil Science**, v. 23, n. 2, 1972.
- WILSON, J. P.; GALLANT, J. C. Digital terrain analysis. In: \_\_\_\_\_ (Ed.). **Terrain analysis: principles and applications**. New York: Wiley & Sons, 2000, p. 1-27
- YAALON, D. H. Down to earth. **Nature**, v. 407, p. 301-301, 2000.
- ZEILHOFER, P. Soil Mapping in the Pantanal of Mato Grosso, Brazil, using Multitemporal Landsat TM data. **Wetlands Ecology and Management**, v. 14, n. 5, p. 445-461, 2006.
- ZHU, A. X. A similarity model for representing soil spatial information. **Geoderma**, Amsterdam, v. 77, n. 2-4, p. 217-242, 1997.

ZHU, A. X. et al. Derivation of soil properties using a Soil Land Inference Model (SoLIM). **Soil Science Society of American Journal**, Madison, v. 61, n. 2, p. 523-533, 1997.

ZHU, A. X. et al. Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. **Soil Science Society of American Journal**, Madison, v. 65, n. 5, p. 1463-1472, 2001.

## **APÊNDICES**

Apêndice A – Solos de São Pedro do Sul (adaptado de Klamt et al., 2001).  
 Coordenadas E e N em metros, UTM/SAD69, Fuso 21.

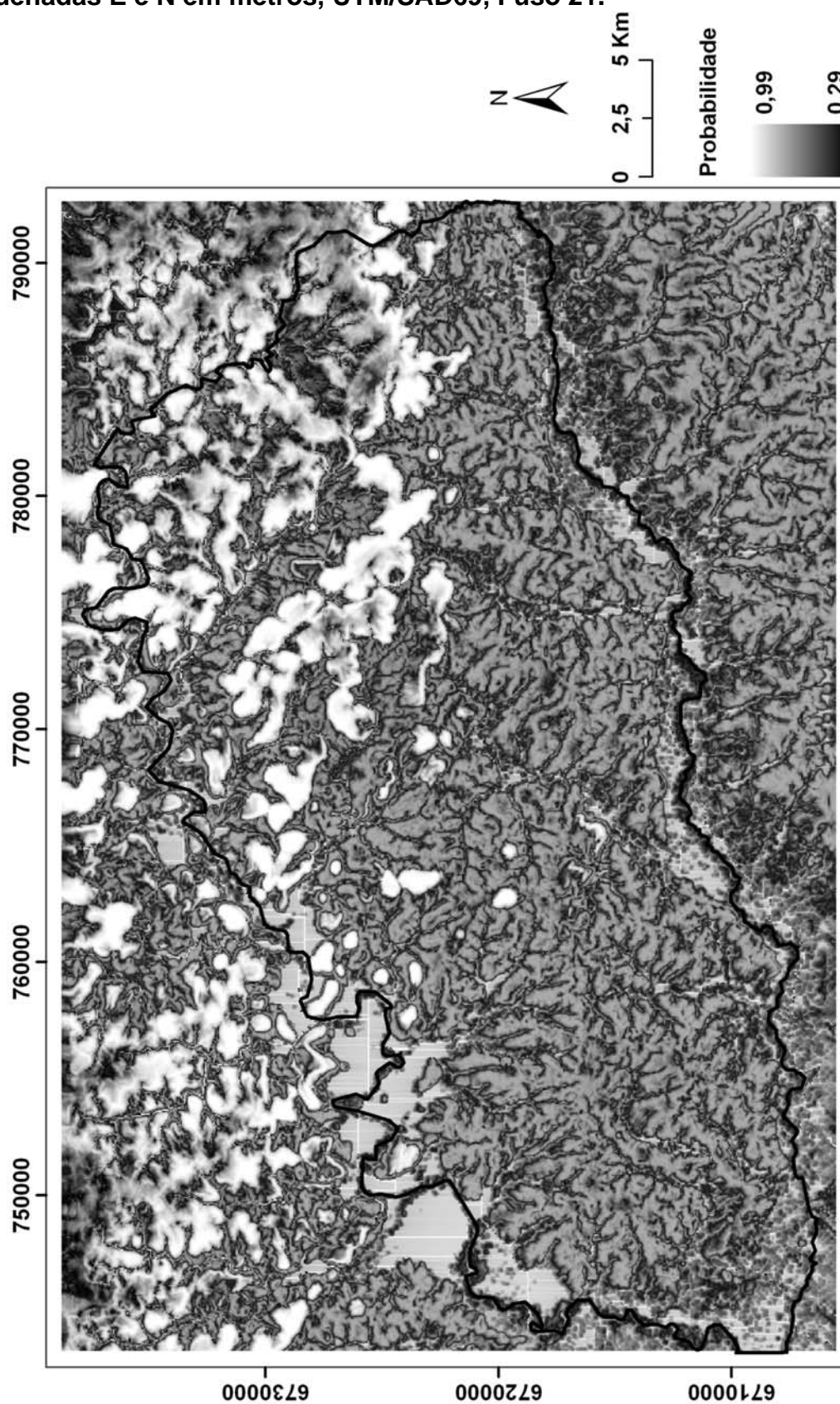




**Apêndice B – Execução do levantamento nas áreas de referência. Imagem superior descrição de um perfil. Imagem inferior checagem da fotointerpretação.**

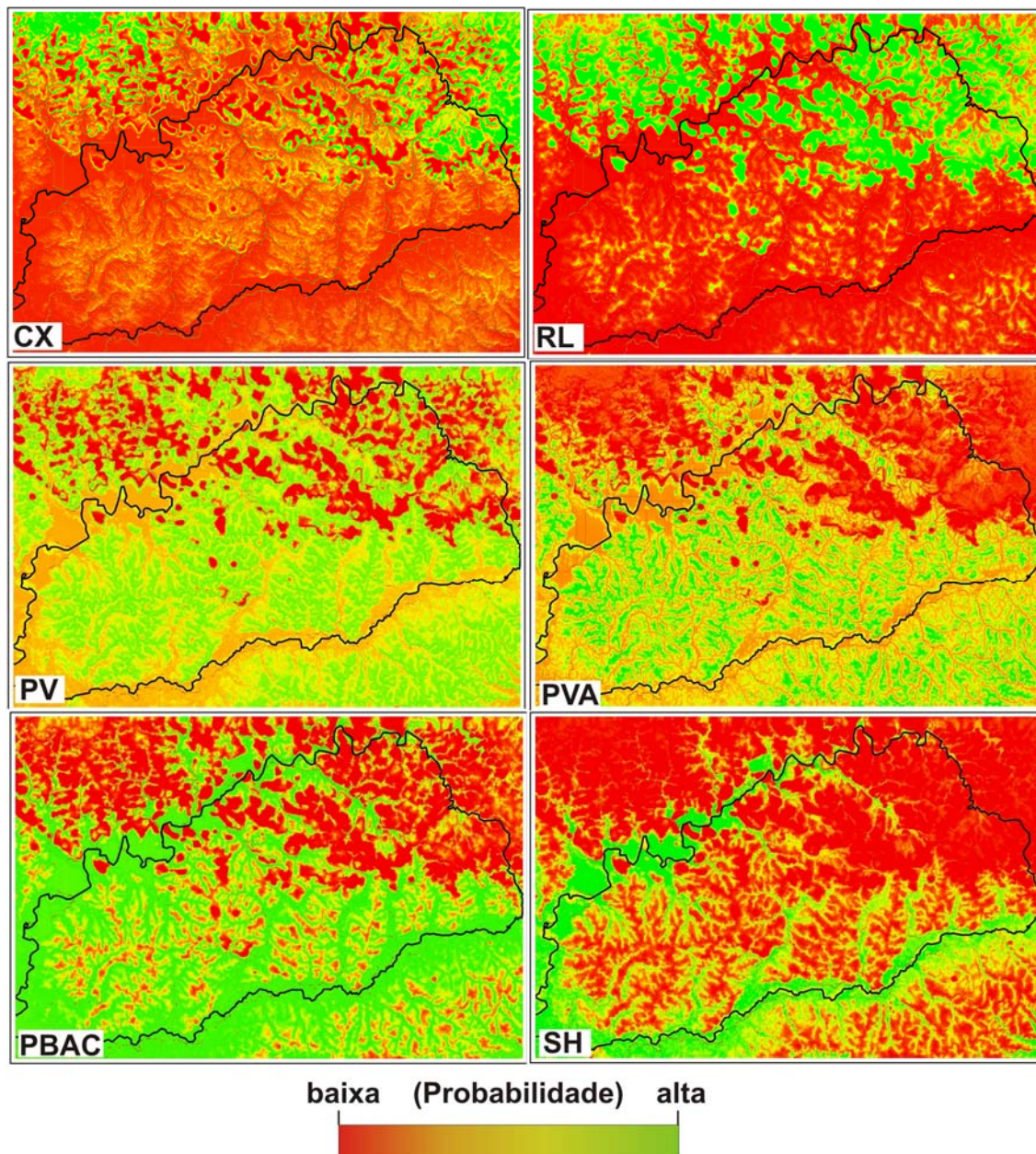


Apêndice C – Distribuição espacial dos maiores valores de probabilidade das funções logit calculadas com componentes principais e o mapa Klamt. Coordenadas E e N em metros, UTM/SAD69, Fuso 21.

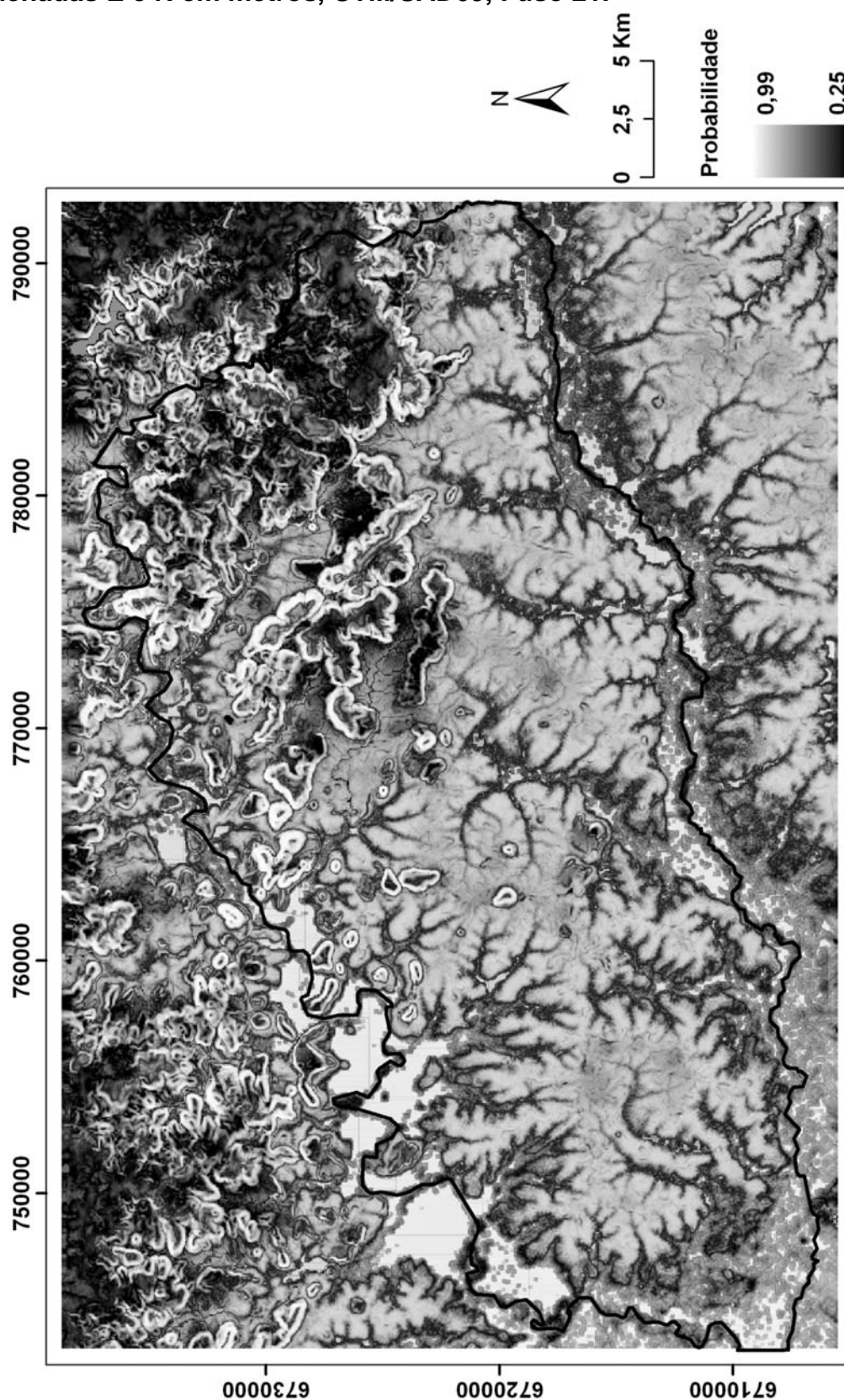




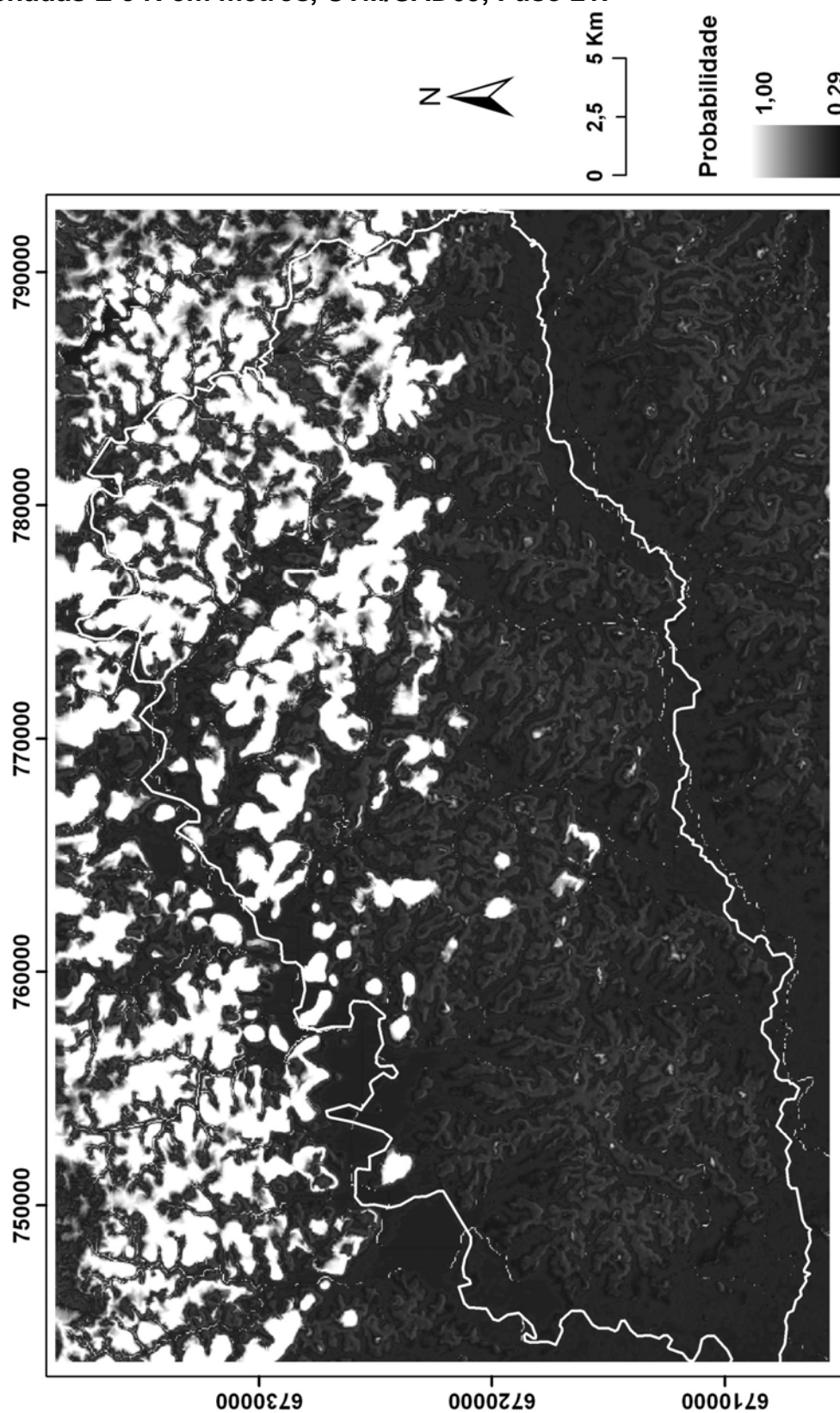
Apêndice D – Distribuição espacial das probabilidades de se encontrar cada uma das classes de solos testadas utilizando-se componentes principais e a área de referência para treinamento dos modelos. Tons em vermelho indicam baixa probabilidade, tons em verde indicam alta probabilidade.



Apêndice E – Distribuição espacial dos maiores valores de probabilidade das funções logit calculadas com atributos do terreno e o mapa Klamt. Coordenadas E e N em metros, UTM/SAD69, Fuso 21.



Apêndice F – Distribuição espacial dos maiores valores de probabilidade das funções logit calculadas com componentes principais nas áreas de referência. Coordenadas E e N em metros, UTM/SAD69, Fuso 21.





Apêndice G – a) Mapa de solos predito, curvas de nível (300m) e rede de drenagem (azul). b) Espacialização da probabilidade para a mesma região. Coordenadas E e N em metros, UTM/SAD69, Fuso 21.

