

CLASSIFICAÇÃO DE PAISAGENS COM PROPRIEDADES FÍSICO-HÍDRICAS E TERMODINÂMICAS DISTINTAS PELO USO DE ATRIBUTOS TOPOGRÁFICOS

Kenny Delmonte Oliveira

Monografia apresentada na disciplina de Introdução ao Geoprocessamento do programa de Pós Graduação em Sensoriamento Remoto, como parte das exigências para conclusão dos créditos.

São José dos Campos 2015

RESUMO

O objetivo do presente estudo foi realizar a classificação temática de classes pedológicas com propriedades físico-hídricas e termodinâmicas distintas para a bacia hidrográfica do rio Jucu. A primeira etapa do trabalho consistiu na criteriosa obtenção dos dados de entrada oriundos de derivações do Modelo Digital de Elevação MDE, tanto por meio de revisão bibliográfica como pelo uso de geotecnologias. A segunda etapa consistiu na classificação temática por classificação com técnicas de iso data e máxima verossimilhança. Foram utilizadas variáveis climáticas e pedológicas das quais realizou-se a eliminação das variáveis mais redundantes por critérios de Análise de Agrupamento e Análise de Componentes Principais.

Ressalta-se que a intensão na elaboração das classes de pedopaisagens do solo não foi a definição de classes pedológicas como usualmente tem sido feito, mas sim o delineamento de áreas inominadas com características físico-hídricas e termodinâmicas semelhantes, o que foi certamente alcançado pela intrínseca correlação da seleção final das variáveis com tais características, com base nos seguintes processos físicos: gradiente de fluxo lateral e aceleração em direção aos corpos d'água, gradiente climático local, convergência e divergência de fluxo d'água, conteúdo de água no solo, potencial energético, distribuição espacial da umidade, previsão de zonas de saturação, e aceleração dos fluxos laterais pelo gradiente natural do relevo, das quais obteve-se 7 classes de paisagens representativas das propriedades de interesse.

Palavras-chave: pedologia, geotecnologias, classificação não supervisionada.

1. INTRODUÇÃO

A topografia está fortemente relacionada com outras variáveis pedogenéticas, como material de origem, clima, atividade biótica, idade dos solos (JENNY, 1980 citado por ROMANO; CHIRICO, 2004), fluxo lateral de superfície e subsuperfície, regime de temperatura e radiação solar incidente, dinâmica de vegetação, dinâmica temporal de umidade, entre outras. Esta se apresenta eficiente na avaliação da distribuição espacial e organização das propriedades pedológicas, já que a hidrologia e geomorfologia são processos motores na formação do solo. Portanto, as características do terreno são as variáveis ambientais mais comumente empregadas na complementação dos dados de solo, pois representam o comportamento hidráulico e as condições ambientais locais (ROMANO; CHIRICO, 2004).

Tricart (1968), após analise dos fatores morfogenéticos e pedogenéticos, afirma que apenas os estudos litológicos, de forma isolada, são insuficientes para explicar o processo de formação dos solos, sendo necessária a consideração dos elementos climáticos. Ab'Sáber (1966) ressalta, ainda, que as paisagens tropicais só poderão ser compreendidas a partir de estudos climáticos, sendo este o fator exógeno que define a dinâmica das bases litológicas nesses ambientes.

Böhner e Selige (2006) utilizaram médias mensais e anuais de diferentes variáveis climáticas em um período de 30 anos (1961-1990) da *German Meteorological Network (DWD*) para predição espacial de atributos do solo adotando tanto atributos topográficos como regionalização climática. Estes autores argumentam que, partindo de um modelo teórico, o impacto sofrido pelas camadas ambientais exige uma integração de variáveis climáticas, o que permite, de forma mais casual, a obtenção de estimativas espaciais de parâmetros e padrões do solo.

Demattê et al. (2004) comparou o método tradicional e técnicas espectrais por sensoriamento remoto para obtenção de mapas de solo, concluindo que dados espectrais podem ser utilizados como instrumento neste tipo de classificação. Oliveira; Lobo Ferreira (2002) e Abreu e Pena (2013); propuseram metodologias para definição de classes de solo com uso de classes de infiltração máxima e permeabilidade. PINEDA, (2008) utilizou o descritor de terreno HAND (*Height Above the Nearest Drainage* – desnível acima da drenagem mais próxima), que se baseia apenas na direção de fluxo (área de contribuição) e na rede de drenagem (distância euclidiana), para gerar mapas detalhados de classes de solo. Carvalho Junior et al.

(2011) utilizando algoritmos de classificação por máxima verossimilhança e redes neurais artificiais, realizou a classificação de unidades pedológicas baseando-se em um conjunto de variáveis geomorfológicas.

Muitas são as fontes de erros na quantificação das variáveis edáficas, sendo uma das principais, a baixa densidade de amostragem associada à alta heterogeneidade e anisotropia do solo, adquirindo magnitudes ainda maiores pela distribuição não sistemática da variável medida. Em função da complexidade e do elevado custo de aquisição das variáveis edáficas, o que tem sido comumente adotado é a atribuição do valor médio de um pequeno número de observações, muitas vezes de levantamento exploratório, como sendo representativo de grandes áreas. Diante disso, se entente que são necessárias novas propostas metodológicas, ou continuidades das já existentes, para suprir a carência dessas informações de tão elevada demanda, enquanto tais dificuldades, associadas aos complexos métodos tradicionais, não sejam superadas.

Considerando a ausência de variáveis físico-hídricas nos relatórios técnicos de levantamento dos perfis dos solos, pela classificação tradicional (OTTONI, 2005), proposta pelo Sistema Brasileiro de Classificação de Solos (SiBCS) (EMBRAPA, 1999), e devido às numerosas restrições impostas à obtenção de informações edáficas, optou-se pela elaboração de um mapa temático representativo das propriedades físico-hídricas e termodinâmicas dos solos para uma bacia hidrográfica de grande importância para o estado do Espírito Santo, a bacia do rio Jucu.

2. MATERIAIS E MÉTODOS

2.1. CARACTERIZAÇÃO DA ÁREA EM ESTUDO

Tomando como ponto de partida a hierarquização descendente da divisão político-administrativa do Brasil, a área de estudo, definida pela bacia hidrográfica do rio Jucú (BHRJ) (Figura 1), encontrar-se inserida na região Sudeste do país, no estado do Espírito Santo e Macrorregião de Planejamento Metropolitana.

Componente da Região Hidrográfica Atlântico Sudeste, o rio Jucú drena a Bacia de nível 4 e código 7.714, classificada pelo sistema de Otto Pfafstetter, composta de nove sub-bacias de nível 5, a saber: rio Barcelos; ribeirão Tijuco Preto; rio Galo; rio Jucú Braço Sul; ribeirão Peixe Verde, rio Jacarandá; rio Camboapina; rio Marinho; e Foz do Rio Jucú. Estes por sua vez recebem 65 tributários de nível 6 segundo esse mesmo sistema de classificação (IJSN, 2014).



Figura 1 – Localização da unidade de estudo.

A BHRJ localiza-se entre as coordenadas UTM mín. E: 276.759 e máx. E: 364.059 m, e mín. N: 7.770.580 e máx. N: 7.724.730 m, compreendendo uma área de drenagem, projetada em plano horizontal, de 2.125 km², área real (com base no declive) de 2.239 km², perímetro de 447 km, altitude máxima de 1.888 m e mínima de -10 m.

Dois tipos de clima são coincidentes à região, o clima quente e úmido e o tropical de altitude (mesotérmico), os quais se enquadram nas categorias Tropical Úmido com Estação Chuvosa no Verão e Seca no Inverno (Aw), Tropical Úmido sem Estação Seca Pronunciada (Am), e subtipos Brando Úmido (Cfa) e Brando de Inverno Seco (Cwa) (HABTEC, 1997; PEEL; FINLAYSON; MCMAHON, 2007).

O tipo de solo predominante na área é o Latossolo Vermelho-Amarelo, ocupando quase que a totalidade da bacia. Ocorrem ainda solos do tipo Cambissolo Háplico na região norte e sudoeste da bacia, Neossolo Fluvico nas regiões mais baixas do relevo, e manchas de Neossolo Litólico e Argissolo Vermelho Amarelo.

3. MAPA DE CLASSES PEDOLÓGICAS

O mais recente mapeamento do solo do estado do ES (ESPÍRITO SANTO, 1978), e com melhor detalhamento, encontra-se em uma escala grosseira (1:400.000) para estudos de micro e mesoescalas, assim, optou-se pela elaboração de um melhor detalhamento de regiões da paisagem com características pedológicas semelhantes.

Para melhor compreensão, as etapas de preparação da representação matricial de paisagens pedológicas, definidas aqui como pedopaisagens, foram subdivididas da seguinte forma:

- Aquisição dos dados para obtenção dos atributos topográficos primários, secundários e climáticos pedogenéticos;
- Elaboração dos parâmetros topográficos e climáticos pedogenéticos;
- ✓ Análise de componentes principais; e
- ✓ Classificação das pedopaisagem por isodata e máxima verossimilhança.

3.1. DADOS NECESSÁRIOS PARA A ELABORAÇÃO DO MAPA DE CLASSES PEDOLÓGICAS

3.1.1. Modelo Digital De Elevação (MDE)

Após uma análise prévia das alternativas de MDEs passíveis de uso, adotouse o *Global Digital Elevation Model (GDEM2*), oriundo do *Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)*, disponível gratuitamente na internet, desenvolvido em conjunto pela *U.S. National Aeronautics and Space Administration (NASA)* e *Japan's Ministry of Economy, Trade, and Industry (METI)*. O ASTER GDEM2 é distribuído em formato GeoTIFF, com coordenadas geográficas no sistema *WGS-84* e unidade de medida vertical em metros.

Esse MDE trata-se da segunda versão do GDEM, com substanciais melhorias, apresentando uma precisão vertical de 17 metros (com 95% de confiança). O deslocamento horizontal foi reduzido para 0,23 pixels, com melhorias na resolução horizontal entre 71 e 82 metros. Apesar do consequente aumento de ruídos, o número de vazios e artefatos foi fortemente reduzido, e praticamente eliminado em algumas áreas (LP DAAC, 2011).

3.1.2. Atributos Topográficos

As informações básicas para análise topográfica do terreno foram obtidas por meio do MDE, através de derivações que apresentam forte significado na descrição de propriedades ambientais e processos ativos da paisagem. Aqueles atributos que podem ser obtidos imediatamente dessa fonte são classificados como atributos primários; porém, se associados a algum índice ou por combinação de dois ou mais atributos primários, serão denominados de atributos secundários (ROMANO; CHIRICO, 2004).

Os atributos primários e secundários para a classificação das pedopaisagens foram obtidos com base em Romano e Chirico, (2004), Florinsky et al. (2002) e Wilson e Gallant, (2000) conforme descrito na Tabela 1 (Apêndices A a N).

Atributo	Definição	Processo físico ou variável ao qual se encontra correlacionado
Altitude	Elevação acima o nível do mar.	Gradiente climático local, padrões vegetacionais, potencial energético.
Declividade	Ângulo entre o plano horizontal e um plano tangencial para a superfície do terreno.	Velocidade do fluxo superficial, subsuperficial e lateral, taxa de escoamento, precipitação, vegetação, geomorfologia, conteúdo de água no solo, classe de aptidão do solo.
Declive de Captação	Inclinação média ao longo da bacia hidrográfica.	Tempo de concentração.
Área de Captação	Área de drenagem para saída de captação.	Volume de escoamento superficial.
Comprimento do Caminho de Fluxo	Distância máxima de fluxo de água para um ponto na área de captação.	Taxa de erosão, produção de sedimento, tempo de concentração.
Curvatura do Perfil	Curvatura da superfície do terreno na direção identificada pelo ângulo do aspecto.	Aceleração dos fluxos laterais a jusante, taxa de erosão e deposição, geomorfologia.
Curvatura Plana	Contorno da curvatura.	Convergência/divergência do fluxo, conteúdo de água no solo, características do solo.
Curvatura Tangencial	Curvatura da superfície do terreno na direção ortogonal que é identificada pelo ângulo do aspecto	Fornece medida alternativa de convergência e divergência do fluxo local lateral, taxa de acumulação lateral.
Aspecto	Ângulo horário do norte da	Direção de fluxos laterais, intensidade

Tabela 1 - Atributos topográficos primários e secundários do solo

	projeção do vetor normal à superfície do terreno no plano horizontal.	relativa da radiação solar.
Índice Topográfico de Umidade	$\operatorname{Ln}\left(\frac{\alpha}{\tan\beta}\right)$	Descrição da distribuição espacial do índice de umidade e extensão das zonas de saturação para geração de escoamento como uma função do declive da área de contribuição. Previsão de zonas de saturação tipicamente em segmentos convergentes da paisagem.
Radiação Solar	Proporção da incidência de radiação solar extraterrestre sobre uma superfície local inclinada em relação a uma superfície horizontal de uma determinada latitude e longitude.	Variabilidade relativa da taxa de evapotranspiração, exposição do terreno, disponibilidade de energia.
Índice de Energia do Fluxo	$\alpha \tan(\beta)$	Medida do poder erosivo da água que flui com base na suposição de que a descarga (q) é um índice proporcional à área de captação específica (α). Prevê erosão líquida em áreas de perfil que apresentam convexidade e concavidade tangencial (aceleração do fluxo e zonas de convergência) e deposição em áreas de perfil côncavo (zonas de diminuição da velocidade do fluxo).
Fator LS	$LS = 0,00984L^{0.63}D^{1.18}$ em que: L = Comprimento de rampa D = Declividade	Este índice de capacidade de transporte de sedimentos foi derivado da teoria de unidade "stream power" e é equivalente ao fator da EUPSR em determinadas circunstâncias. Utilizado também para prever os locais de erosão e áreas de deposição.
Relevo Relativo	Característica de drenagem da paisagem.	A razão entre a diferença de altitudes em um determinado ponto na superfície do solo e o ponto mais baixo de um divisor de águas para a diferença de altitudes entre o mais alto e o mais baixo ponto de uma bacia hidrográfica.
Distância Euclidiana	Descreve a relação de cada célula do raster com uma fonte específica, a qual foi definida como a rede de drenagem.	Associado com outros atributos permite inferir a influência da proximidade desses à fonte selecionada. Como exemplo cita-se a ocorrência de regiões com maior deficiência de drenagem quando próximas a fonte (rede de drenagem) ou menor disponibilidade hídrica, quando distante da fonte.

3.1.3. Variáveis Climáticas Pedogenéticas

Optou-se por suplementar os fatores de formação do solo com as variáveis climáticas pedogenéticas precipitação e temperatura. Porém ressalta-se que como explicado por Böhner e Selige (2006), são parâmetros de uso casual, portanto busca-se por sua inclusão, considerar a possibilidade da promoção de um efeito positivo na identificação de padrões espaciais de parâmetros do solo quando submetidos a um critério de classificação, mesmo que sob pena de se ter uma representatividade apenas superficial do paleoclima regional.

Os dados meteorológicos foram obtidos do trabalho desenvolvido por Castro (2008), que utilizou séries históricas de um período de 30 anos (1977 – 2006) de variáveis climáticas em escala anual de tempo. Além daquelas estações inseridas no estado do Espírito Santo, o autor adicionou 16 postos pluviométricos, pertencentes à Agência Nacional de Águas (ANA), na região circunvizinha do estado, visando minimizar efeitos de borda em processos de espacialização dos dados, obtendo assim um total de 110 estações.

3.1.3.1. Parâmetros Topográficos e Climáticos Pedogenéticos

3.1.3.1.1. Cálculo dos Atributos Topográficos

O GDEM2 é disponibilizado em cenas individuais em um sistema de projeção global, o que implica na necessidade de um processamento de mosaicagem quando da necessidade do uso de mais de um produto, e transformação geográfica para o sistema de projeção local. Para tanto adotou-se a seguinte sequência metodológica:

1. Inserção de 16 cenas do *GDEM2*, que cobrem o estado do ES, em uma armação de dados e importação para um geodatabase, sem especificações dos parâmetros de armazenamento;

 Criação de um conjunto de dados de mosaico vazio para posterior gerenciamento, visualização e consulta a coleção dos rasters, sem alteração do sistema de projeção, tipo de pixel ou número de bandas;

3. Preenchimento do conjunto de dados de mosaico vazio pela adição da coleção de rasters, com definição do tipo de raster como *Digital Terrain Elevation Data (DTED*), permitindo a identificação de metadados,

georreferenciamento, data de aquisição, tipo de sensor, comprimento de onda da banda, armazenamento dos pixels, formato do raster e outros parâmetros específicos. Cálculo do tamanho do intervalo das células de cada conjunto de dados e geração do polígono limite do conjunto de dados mosaico;

4. Construção de pirâmides, por técnica de reamostragem do tipo bilinear, e cálculos estatísticos para aprimoramento do desempenho de exibição e correto simbolismo do conjunto de dados raster além do conjunto de dados mosaico. Uso de compressão do tipo LZ77 visando preservar os valores das células do raster, sem qualquer perda de informação ou alterações;

Construção de "Seamlines" para definição da linha ao longo da qual os 5. rasters, no conjunto de dado de mosaico, foram mosaicados, e com vista a garantir a melhor construção possível do mesmo. O método de construção de "Seamlines" adotado foi do tipo radiometria, nos quais são examinados os valores e padrões das áreas de interseção das cenas e realizado o cálculo de percursos ao longo de pontos de interseção. Devido à correção de cores ao conjunto de dados mosaico não ter sido realizada, não houve necessidade de priorização desse procedimento sobre a construção de "Seamlines" quando do uso do método radiometria. O tamanho de célula do raster para a construção dessa feição matricial foi definida com o mesmo tamanho de célula dos dados de entrada. O método de classificação que define a ordem em que os rasters serão fundidos foi do tipo noroeste. A ordem de classificação dos rasters foi do tipo ascendente. A matização (harmonização) ao longo das "Seamlines" na região onde ocorreu sobreposição das cenas foi definida como "interna", impedindo assim, o uso de valores externos à área sobreposta para tal procedimento, e teve-se como unidade de medição o próprio pixel;

6. Construção de *"Footprints"* (feição de dimensionamento dos dados válidos dos rasters) adotando o método "radiometria", com vista à exclusão de áreas de fronteiras com dados inválidos. Os valores mínimos e máximos determinados pela profundidade dos bits, como valores válido, foram definidos com 0 e 65.535 (2¹⁶) respectivamente. Ressalta-se que esse intervalo de valores não excluiu os valores negativos de altitude dos dados de entrada. De forma a não haver suavização do polígono delimitador,

adotou-se o valor de -1 (valor de configuração) para que o número de vértices, da feição de saída, fosse correspondente ao total dos vértices presentes nas regiões marginais no conjunto do mosaico. Optou-se pela atualização da feição de limite (*Boundary*), mesclando as feições "*Footprints*" para criação de um único polígono delimitador da extensão dos pixels válidos;

7. Mosaicagem das 16 cenas obedecendo à condição de que os rasters possuem o mesmo número de bandas e a mesma profundidade de bits. O sistema de referência espacial e tipo de pixels foram mantidos inalterados. Como a área de sobreposição dos rasters apresentou mesmos valores, o operado de mosaico foi determinado para manter os mesmos valores correspondentes aos dos rasters de imediata sobreposição. O modo de cores de mapa do mosaico foi determinado tomando como referência a primeira cena de entrada.

 Extração do raster mosaico gerado no procedimento anterior com uso do limite (*Boundary*) desenvolvido no procedimento 6, obtendo assim uma MDE apenas com valores válidos e ausência das áreas de fronteira, que correspondem especificamente, a massa de água marítima;

9. Transformação do sistema de coordenadas do produto obtido no procedimento anterior para sistema de coordenadas projetadas Sirgas 2000 UTM Zona número 24 designador K hemisfério Sul, com transformação geográfica do tipo WGS 1984 para SIRGAS 2000 com técnica de reamostragem dos valores do tipo bilinear e conservação do grid de espaçamento (tamanho de célula) com mesma resolução.

Erros sistemáticos presentes em MDEs, podem induzir a sérias interpretações errôneas. Este problema tende a ser agravado em produtos obtidos das derivadas de primeira e segunda ordem, como por exemplo, a declividade, convexidade, aspecto, curvaturas plana e tangencial do perfil. Casos ainda mais críticos podem ocorrer pela derivação de atributos topográficos secundários, como Índice Topográfico de Umidade e Capacidade de Transporte de Sedimentos (MOORE et al. (1993) citado por WILSON e GALLANT, 2000).

A precisão dos parâmetros derivados da elevação do terreno depende não só da precisão dos dados iniciais, ou seja, do próprio MDE, mas também da precisão

da técnica de cálculo adotada (FLORINSKY, 1998), portanto, maior atenção foi despendida nesses dois fatores de geração de erros.

Embora a correção ideal de anomalias seja pela reposição das cotas, é possível o tratamento da imagem por simples remoção dos erros e uso dos valores da vizinhança para o preenchimento dos dados ausentes por interpolação (VALERIANO; PICINI, 2003).

A metodologia adotada para remoção das anomalias presentes no GDEM2 foi realizada com as seguintes diretrizes (VALERIANO; PICINI, 2003).

- Identificação dos locais afetados;
- Eliminação das áreas de influência;
- Exportação das cotas altimétricas válidas; e
- Interpolação dos valores de cota.

Foram executados procedimentos básicos de geoprocessamento para a correção de possíveis valores atípicos presentes no MDE, por detecção de padrões de variação espacial com uso de filtro para destacar mudanças bruscas de altitudes como descrito a seguir:

1. Aplicação do *Laplacian Filter* (*Laplacian of Gaussian*). O método kernel foi definido com ajuste visual pela ponderação entre áreas selecionadas como de referência com presença e ausência de artefatos no MDE. Assim foi observado se o aumento da frequência local era razoável em ambas as extremidades. Os parâmetros ajustados, foram definidos como 100 para o desvio padrão (sigma), 8 para o raio de vizinhança e definição do modo de pesquisa como do tipo circular;

 Reclassificação do raster, originário da filtragem, em duas classes, com quebra de valores relativos em 7%, definindo assim as áreas com mudanças bruscas de altitude;

3. Remoção dos dados de elevação, do GDEM2, nas áreas correspondentes à classe de valores máximos do raster obtido no passo anterior, anulando os valores dessas áreas;

 Conversão da classe de valores atípicos em feição poligonal sem suavização de vértices;

5. Definição de uma área tamponante ao redor da feição poligonal dos valores atípicos, adotada como uma máscara de extração da área circunvizinha às áreas com células nulas no raster de elevação. O

afastamento foi definido em 1 km (aproximadamente 33 células), esse valor correspondeu a um limite suficientemente representativo para uso em interpolação do algoritmo ANUDEM 5.3, como definido no procedimento 7, (HUTCHINSON, et al., 2011);

6. Conversão da área extraída no procedimento 5 em feição de ponto contendo os z-valores a serem interpolados em uma superfície raster;

7. Interpolação com o uso do algoritmo ANUDEM 5.3. O tamanho da célula foi definido para o mesmo tamanho do grid de espaçamento do raster de elevação (30m), já que este produto seria utilizado para reposição das regiões onde foram extraídos os valores atípicos no MDE. Para obter os melhores resultados de interpolação ao longo das bordas da quadrícula de saída, os limites de x e y foram menores do que a extensão dos dados de entrada em pelo menos 20 células em cada um dos lados. Optou-se apenas pela restrição do valor mínimo de altitude para interpolação (com base nos dados de entrada) uma vez que o valor máximo real era desconhecido. Como o objetivo da interpolação era independente de criar uma superfície sem descontinuidade no fluxo do terreno, foi estabelecida a não remoção de depressões no terreno. Com vista a otimizar o método de pesquisa, utilizado durante a geração de fluxos e sulcos, especificou-se a feição dominante de entrada (ponto). O valor máximo definido como limite de interações foi igual a 20, já que o mesmo apresentou melhores resultados na espacialização da superfície quando comparados ao número de depressões espúrias restantes em cada valor de interações testado por variações desse limite. O fator de espacialização do erro - utilizado para o ajuste da suavização durante a conversão dos dados de entrada para uma superfície raster - foi de 1 em um intervalo de 0.25 a 4, em que menores valores resultam em menor suavização. Por padrão, pela feição do tipo ponto ser o dado de entrada dominante, definiu-se como 0 o valor de tolerância que reflete a precisão e a densidade dos pontos de elevação em relação à superfície de drenagem, e 200 como a tolerância de folga na drenagem através de barreiras irrealísticas;

8. Extração dos novos valores (obtidos por interpolação no procedimento anterior) pelo uso de uma feição com afastamento de 300m das regiões de valores atípicos definidas no procedimento 2;

9. Correção do MDE e sua área de influência, por substituição dos valores atípicos e sua área de influência, com o produto gerado no procedimento anterior.

Muitos são os softwares de análise de informações espaciais que apresentam pacotes de ferramentas para análises do terreno. Devido ao elevado número dos atributos que pode ser calculado pela topografia, nenhum desses softwares contempla a totalidade desses parâmetros topográficos. Logo, além do ArcGIS 10.2.19[®], optou-se pelo uso do SAGA 2.1.0[®] e Whitebox 3.0.7[®].

Visando conservar o MDE com suas características originais, quando possível, evitou-se o uso do preenchimento de depressões do terreno. Assim, apenas os atributos dependentes da continuidade espacial do fluxo de superfície foram estimados com o MDE livre de depressões, sendo estes: área de captação modificada, índice topográfico de umidade, índice de energia do fluxo, Comprimento do Caminho de Fluxo e Relevo Relativo. Os procedimentos para o cálculo dos atributos topográficos do terreno encontram-se descritos a seguir:

1. Preenchimento das depressões naturais do terreno e depressões espúrias com o método proposto por Planchon e Darboux (2001). O ângulo mínimo do declive preservado de uma célula para o outra foi de 0,01 graus;

2. Cálculo da Área de Captação realizado com base em uma Área de *Captação Modificada* (BOEHNER et al., 2002), diferindo da primeira por não trata o fluxo como uma fina película, mas como resultado, prevê que as células situadas em vales, com uma pequena distância vertical, apresentem um canal de uma forma mais realística, pela aplicação de um efeito de sucção, o qual, baseando-se em uma feição poligonal de massa d'água, foi definido igual a 10;

3. Cálculo da Declive de Captação (BOEHNER et al., 2002) definindo a mínima declividade como 0, o fator de ponderação para o cálculo do índice de inclinação como 1, e o fator de compensação igual a 0,1;

4. Cálculo do Índice Topográfico de Umidade (BOEHNER et al., 2002) obtido em função da Área de Captação Modificada e Declive de Captação, sendo o tipo de área definido como a raiz quadrada da área de captação;

5. Cálculo do Índice de Energia do Fluxo (MOORE; GRAYSON; LADSON, 1991) com efeito realístico da área de inundação promovido no Índice

Topográfico de Umidade por uso da Área de Captação Modificada e Declive de Captação para o calculo desse índice;

 Cálculo do Comprimento do Caminho de Fluxo (QUINN et al. 1991) adotando o método Determinístico 8 (D8) como algoritmo de roteamento (O'CALLAGHAN; MARK, 1984);

7. Cálculo da Curvatura Tangencial (WILSON; GALLANT, 2000), Curvatura do Perfil e Curvatura Plana (MOORE; GRAYSON; LADSON, 1991; ZEVERBERGEN; THORNE, 1987) com fator de conversão definido para 1, por serem idênticas as unidades verticais e horizontais do MDE;

8. Cálculo do Aspecto (BURROUGH; MCDONELL, 1998) pela identificação da direção da descida da taxa máxima de mudança do valor de cada célula e seus vizinhos;

 Cálculo da Declividade (BURROUGH; MCDONELL, 1998) para obtenção do gradiente de inclinação de cada célula na superfície raster. A unidade de medida foi definida para percentagem com o intervalo de valores variando de 0 a infinito;

10. Cálculo do Fator LS (BERTONI; LONBARD NETO, 2012) utilizado pela *Universal Soil Loss Equation (USLE)* para obtenção da combinação de comprimento de rampa e grau de declive;

11. Cálculo da Radiação Solar (MCCUNE; DYLAN, 2002) obtendo o potencial de radiação incidente direta anual e carga de calor para estimar o potencial de radiação solar. Para este cálculo foi utilizada uma superfície raster com as mesmas dimensões da área de estudo, contendo valores de latitude em graus decimais;

12. Cálculo do Relevo Relativo (FLORINSKY et al., 2002) para caracterização da drenagem da paisagem. Calculou-se inicialmente a direção de fluxo do terreno (GREENLEE, 1987) seguido pela acumulação do mesmo (JENSON; DOMINGUE, 1988) no intuito de estabelecer a hidrografia numérica para uma determinada área de captação. Na definição da área de captação mínima do fluxo acumulado, constituinte de um segmento hidrográfico, tomou-se como referência uma hidrografia vetorizada na escala de 1:50.000. A identificação dos cursos d'água das bacias de cabeceira foi realizada pela conversão do ponto mais a montante, dos cursos d'água de primeira ordem, em feição de ponto (nascentes). Adotando-se o método

espacial de seleção do tipo interseção, isolou-se o conjunto de feição linear, correspondente aos cursos d'água de primeira ordem, para o cálculo do valor médio de comprimento. Esse valor foi utilizado como diâmetro de uma circunferência média das menores áreas de captação dos cursos d'água de primeira ordem. Com o uso do raster de acumulação de fluxo, calculou-se a rede hidrográfica com limiar definido por um número de células referente à área deste círculo, o que correspondeu a 600 células (54 ha). Atribuíram-se valores exclusivos para as seções da rede linear rasterizada localizada entre as interseções. Criou-se uma grade em que cada célula, pertencente a uma determinada área de captação (sub-bacia), possuísse 0 valor correspondente da seção da rede linear rasterizada. Estas por sua vez, foram convertidas em feição poligonal, e acrescidas áreas de captação Muitos pequenos polígonos são formados durante estes costeiras. procedimentos, assim optou-se por mesclá-los com os polígonos vizinhos com a mais longa fronteira em comum. Este procedimento foi repetido até que não houvesse mais polígonos com áreas inferiores a pré-estabelecida. Por fim, o Relevo Relativo foi obtido calculando-se o intervalo entre os valores extremos do MDE dentro de cada uma das zonas correspondentes as áreas de captação mínima;

13. Cálculo da Distância Euclidiana tomando como a fonte mais próxima a rede de drenagem em formato matricial, definida pelo número de células referentes à área mínima de captação obtida no procedimento anterior.

3.1.3.1.2. Cálculo dos Atributos Climáticos Pedogenéticos

Neste tópico aborda-se a obtenção das variáveis climáticas temperatura e precipitação para o estado do ES, adotando-se para tal procedimento as médias anuais, oriundas de dados diários, das séries de 30 anos calculadas para 110 estações existentes no estado e áreas circunvizinhas, utilizadas no trabalho desenvolvido por Castro (2008).

3.1.3.1.2.1. Temperatura

Por não se dispor de estações meteorológicas suficientemente representativas na região de estudo, o valor médio anual de temperatura diária foi estimado a partir do ajuste da Equação 1, em função da altitude (efeito da variação de pressão), latitude (relação com irradiância global), e por julgar-se necessário, incluiu-se também a variável longitude (efeito da maritimidade e continentalidade) (PEREIRA; ANGELOCCI; SENTELHAS, 2002), adotando-se o Mínimo Quadrado Ordinário (MQO) como a técnica de análise de regressão.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \text{Altitude} + \hat{\beta}_2 \text{Latitude} + \hat{\beta}_3 \text{Longitude}$$
(1)

Para espacialização da temperatura, pelo modelo de regressão linear múltipla, foi utilizados o MDE, consistido nas etapas anteriores, e camadas matriciais de latitude e longitude com valores em graus decimais.

Uma eficiente forma para validação da equação de regressão é avaliar a capacidade desse modelo em prever a variável dependente com uso de observação não adotada no cálculo dos coeficientes. Como em grande parte dos casos as observações são insuficientes para estimar o modelo proposto, este procedimento torna-se inviável (NAGHETTINI; PINTO, 2007).

Conclui-se unicamente que o modelo de regressão linear, pelo método dos MQO, é eficaz e de confiança, quando todos os pressupostos exigidos inerentemente a esse método forem satisfeitos. Os problemas avaliados, com base nos pressupostos estatísticos, foram os seguintes: variáveis explicativas omitidas expecificação), relações não lineares, discrepância de dados. não (má estacionariedade, heterocedasticidade, multicolinearidade, variância inconsistente dos resíduos, autocorrelação espacial dos resíduos e distribuição normal dos resíduos enviesada. Assim, para validação do modelo proposto, adotaram-se os seguintes índices e testes: Critério de informação Akaike (AIC), Critério de informação Akaike corrigido (AICc), Coeficiente de determinação (r²), Coeficiente de determinação ajustado (r²Ajustado), F-estatística conjunta, Wald estatística conjunta, Estatística estudentizada Breusch-Pagan de Koenker, Estatística Jarque-Bera, Sigma ao quadrado dos erros (variância), Índice global de Moran, Índice esperado de l' Moran, Variância (do teste de Moran, Z-score, Erro Padrão, Erro Padrão Robusto, Teste t, Teste t Robusto e Fator de Inflação da Variância (VIF).

Pela conversão dos coeficientes do modelo para desvio padrão, estes passaram a serem denominados coeficientes padronizados, tais desvios em valor absoluto (módulo) foram adotados para avaliar o efeito que as variáveis explicativas tiveram sobre a variável dependente, de forma que os maiores valores foram indicativos de maior influência da variável independente no processo de explicação da variável dependente. Esta análise foi feita a luz dos respectivos erros padrões para constatação da real magnitude deste efeito, já que um erro padrão grande para um coeficiente denota que uma reamostragem resultaria numa vasta gama de possíveis valores de coeficientes, sendo o contrário indicativo de coeficientes com suficiente consistência.

Um dos testes adotados para avaliar a significância das variáveis foi o teste "t", ao nível de 5% de probabilidade, sob a hipótese nula H_0 de que o coeficiente é, para todos os efeitos, igual a 0 e, portanto a variável poderá ser removida do modelo.

O teste do Fator de Inflação da Variância (VIF) avalia a correlação entre variáveis independentes e suas combinações lineares, e a menos que existam fortes razões teóricas para manter as variáveis, estas deverão ser removidas do modelo uma a uma até obter-se um valor dentro do intervalo admissível deste teste.

A análise da estacionariedade e heterocedasticidade do modelo foi realizada com uso do teste Koenker (BP), visando determinar a presença de relação sistemática das variáveis independentes com a variável dependente, tanto no (estarionariedade) de dados espaço geográfico quanto no espaço (homocedasticidade), idealizando-se que o processo espacial representado pelas variáveis explicativas proporcione o mesmo comportamento em todos os lugares da área de estudo e que não haja alterações dos valores previstos pela alteração em magnitudes das variáveis explicativas. Este teste foi aplicado ao nível de 5% de probabilidade sob a hipótese nula H₀ de que há estacionariedade е homocedasticidade nas previsões do modelo apresentando. A rejeição de H₀ induzirá na análise do teste estatístico de Wald e na análise do teste "t" robusto.

A análise com teste "t" permite considerar apenas a significância das variáveis de forma individual. Para uma análise de significância das variáveis, de forma conjunta, ou seja, para um exame global do modelo adotou-se os testes Wald

estatística conjunta e a F-estatística conjunta, ao nível de 5% de probabilidade, sob a hipótese nula H₀ que os coeficientes são simultaneamente iguais a 0, de forma que a não rejeição desta hipótese sugere que a remoção das variáveis do modelo não irá prejudicar substancialmente seu ajuste. Estes testes irão indicar quão longe os parâmetros estimados estão de zero, em erros padrão e de forma simultânea. Semelhantemente ao teste "t", a estatística F conjunta é confiável apenas quando o teste Koenker (BP) não for estatisticamente significativo, ao contrário esta análise é desconsiderada, e pelo mesmo motivo, a análise dos coeficientes padronizados deverá ser submetida ao rigor do coeficiente padronizado robusto.

O teste sob o rigor da análise robusta permite emular o método adotado de forma que não seja indevidamente afetado por valores discrepantes ou desvios pressupostos do modelo, apresentando bom desempenho na análise com dados provindos de diferentes distribuições de probabilidade, pois é resistente a erros nos resultados produzidos por desvios de hipóteses, como a de normalidade dos dados.

Em suma, se esclarece que o teste definitivo para indicação da significância ou não de uma variável explicativa é o teste "t" robusto quando da presença de mais de um tipo de distribuição dos dados entre as variáveis explicativas, porém sua análise deve ser precedida daquelas anteriormente realizadas e alusivas ao mesmo, para constatação da possível ocorrência de heterocedasticidade, não estacionariedade e para análise conjunta da significância das variáveis.

A análise da distribuição normal dos resíduos padronizados foi realizada primariamente pela estatística Jarque-Bera, ao nível de 5% de probabilidade sob a hipótese nula H_0 que os resíduos são normalmente distribuídos, de forma que a presença de um valor-p pequeno (menor que 0,05) é indicativa da predição tendenciosa dos dados, tendo-se também como possíveis consequências a presença de relações não lineares, valores atípicos, e/ou forte heterocedasticidade.

A avaliação da auto-correlação espacial residual foi realizada mediante a estatística Global de Moran, visando identificar se o padrão expresso dos dados é agrupado, disperso, ou aleatório, pois agrupamento ou dispersão dos dados, estatisticamente significativos, pode ser indicativo de má especificação das variáveis independentes. Este diagnóstico foi realizado mediante cinco valores: índice Moran, índice esperado de Moran, variância, z-score e valor-p. Quando o z-score ou o valor-p são estatisticamente significativos, o valor do índice de Moran positivo refletirá tendência de agrupamento dos dados, enquanto que valores negativos serão

indicativos de tendência à dispersão. Assim, este índice foi avaliado ao nível de 10% de probabilidade sob a hipótese nula H₀ que os resíduos padronizados apresentam distribuição aleatória em toda a área de estudo, ou seja, os processos espaciais que promovem o padrão observado são ao acaso, e sendo o valor-p significativo, rejeitase a hipótese nula.

Para o cálculo da estatística Global Moran, utilizou-se os resíduos padronizados, distribuídos espacialmente, nos pontos referentes às estações meteorológicas. A conceituação das relações espaciais entre os pontos amostrais foi realizada pelo método do Inverso da Distância, de forma que maior influência terão os pontos mais próximos para as características alvo do que os mais distantes. Para especificar como as distâncias são calculadas entre estações, adotou-se a Distância de Manhattan, que calcula a distância entre dois pontos ao longo de eixos perpendiculares, pela soma da diferença (absoluta) entre as coordenadas X e Y.

Resguardando-se da possível inadequação do modelo proposto, bibliograficamente fundamentado, e visando identificar o maior número de combinações possíveis com diferentes alternativas de variáveis explicativas para construção de um ou mais modelos de regressão válidos, submeteu-se a um método de regressão exploratória um total de 23 variáveis, com propriedades hidráulicas do solo, hidrológicas, climáticas, térmicas e locacionais, sendo estas: altitude, aspecto do terreno, curvatura do perfil, curvatura plana, curvatura tangencial, declividade, deficiência hídrica, disponibilidade hídrica, distância euclidiana, direção da distância euclidiana, evapotranspiração potencial, evapotranspiração real, excedente hídrico, fator LS, comprimento do caminho de fluxo, latitude, longitude, área de captação modificada, declive de captação modificado, precipitação, radiação solar, índice de energia do fluxo e índice topográfico de umidade.

Além do expressivo número de variáveis, essas foram, ainda, submetidas a seis diferentes tipos de transformações, a saber: boxcox, potência ao expoente 2 e 3, raíz cúbica, logaritmo na base 10 e exponencial. Outras transformações não foram adotadas devido à inadequação dos tipos de variáveis à normalização dos dados.

Mediante a adequação das transformações para os diferentes tipos de dados, obteve-se 124 variáveis explicativas, que testadas com a variável dependente, também sob efeito de transformações, passou a haver um total de 744 variáveis. O número mínimo de termos para construção de um modelo de regressão linear múltipla foi determinado por tentativas ascendentes, iniciando em uma variável explicativa para estimação da variável dependente não transformada e sob transformações, de forma que a combinação apropriada deveria obedecer a todos os critérios pré-definidos de validação do modelo. Não houve nenhuma combinação válida com um número menor ou igual a três de variáveis explicativas, porém identificou-se 18 combinações possíveis com combinações de quatro variáveis independentes, que corresponderam aos seguintes crivos:

1. Teste t ao nível de 5% de probabilidade para significância da variável explicativa;

2. Teste Jarque Bera ao nível de 5% de probabilidade para análise de normalidade da distribuição dos erros padronizados;

3. Coeficiente r² ajustado igual ou maior a 0,5;

4. Teste VIF com fator máximo igual a 7,5;

 Teste Global Moran's I ao nível de 10% de probabilidade para o teste de ausência de autocorrelação espacial dos erros.

Por se ter avaliado mais de um modelo consecutivamente, pôde-se adotar o Critério de Informação Akaike corrigido (AICc) como auxiliar no julgamento do melhor modelo dentro de cada categoria de análise, que neste caso trata-se dos diferentes tipos de transformações da variável dependente. O modelo com o menor valor de AICc pode ser considerado o melhor modelo, isto é, tendo em conta a complexidade dos modelos, aquele com menor AICc fornece um melhor ajuste dos dados observados.

3.1.3.1.2.2. Precipitação

O processo de reprodução espacial das características do fenômeno precipitação, baseado nos pontos amostrais (correspondentes às estações pluviométricas), foi realizado com uso do método de interpolação Krigagem Bayesiana Empírica (EBK). A decisão da escolha desse interpolador foi tomada com base em uma "Árvore de Classificação de Diferentes Métodos de Interpolação" (ESRI, 2013). O método selecionado é um método geoestatístico que automatiza os aspectos de maior complexidade na construção de um modelo de krigagem válido.

Ao contrário dos demais métodos de krigagem que utilizam os mínimos quadrados ponderados para estimar os parâmetros do semivariograma, a EBK é um método de interpolação geoestatística que estima os parâmetros do semivariograma pela máxima verossimilhança restrita. (CHILÈS; DELFINER, 1999).

Através de processos de subdivisões e simulações, a EBK realiza a construção de um modelo de krigagem válido, automatizando os parâmetros mais complexos de ajuste do modelo de krigagem, além disso, realiza a escolha do algoritmo de mapeamento adequado e é capaz de lidar com as incertezas sobre a distribuição das observações e suas dependências espaciais (PILZ; SPÖCK, 2007).

As principais vantagens no uso da EBK são: requer mínima modelagem interativa, os erros padrão de previsão são mais precisos do que outros métodos de krigagem, permite previsões precisas de dados moderadamente não estacionários e é mais preciso do que outros métodos de krigagem para pequenos conjuntos de dados.

Os seguintes parâmetros de calibração manual mínimos, das propriedades gerais e de vizinhança de pesquisa foram ajustados pelo critério de tentativa e erro (CHILÈS; DELFINER, 1999):

✓ Tamanho do subconjunto: com referência no tamanho do conjunto de dados, estes são inicialmente subdivididos em subconjuntos de sobreposição de tamanho padrão igual a 100 pontos por subconjunto.

✓ Fator de sobreposição: representa o grau de sobreposição entre os modelos locais (subconjuntos), no qual cada ponto de entrada poderá cair em vários subgrupos, sendo o fator de sobreposição que define o número médio de subgrupos aos quais os pontos irão cair. Os valores típicos variam entre 0.01 e 5, de forma que maiores fatores promovem superfícies mais suaves.

✓ Número de simulações: especifica o número de semivariogramas que serão simulados para cada subconjunto.

✓ Ângulo: ângulo do círculo de pesquisa de um ponto, com propriedade de influenciar nos dados de vizinhança quando esta é subdividida em quatro ou mais setores ou por ajuste do tamanho dos semieixos do círculo. O tipo de setor com oito subdivisões apresenta amplitude de calibração correspondente à máxima 22,5°.

 Raio: distância em unidade de mapa, referente ao comprimento do raio do círculo de pesquisa.

Os seguintes parâmetros foram definidos com base nas características naturais dos dados de entrada e com a finalidade de obter o *layer* geoestatístico de maior rigor qualitativo:

✓ Tipo de superfície de saída: quatro tipos de superfícies estão disponíveis nesta análise de técnica geoestatística, a predição, quantil, probabilidade e predição do erro padrão. Pelo objetivo do estudo escolheu-se a "predição" como tipo de superfície de saída.

Transformação dos dados: em KBE uma das transformações disponível é a transformação "multiplicative skewing normal score" com escolha dos tipos de distribuição empírica ou log-empírica. A transformação com distribuição logempírica é particularmente sensível a outliers, podendo criar superfície de previsão com valores de ordens de magnitude maior ou menor que os valores de entrada. Os pontos amostrais (estações pluviométricas) são representativos de uma distribuição espacial, de uma variável não estacionária, em uma região de grandes amplitudes de variações climáticas, pois se encontra sob efeito de maritimidade, continentalidade, relevo fortemente descontínuo topograficamente, restrito número de pontos amostrais, alocação aleatória e localização em amplo espaço territorial. Apesar da distribuição log-empírica ser adequada unicamente para dados positivos, como chuvas, a possível implicação resultante do uso desse tipo de transformação sobre as restrições impostas pelas características dos dados de entrada, não justificou a escolha da mesma, portando por obviedade escolheu-se a distribuição empírica para a transformação dos dados.

✓ Tipo de semivariograma: ao optar pela transformação dos dados, tornam-se disponíveis os modelos de semivariograma *exponential*, *exponential detrended*, *whittle*, *whittle detrended*, *k-bessel e k-bessel detrended*. Ao avaliar as vantagens e desvantagens dos diferentes métodos, concluiu ser o método *k-bessel detrended* o ideal para o presente caso, pela maior flexibilidade, precisão e capacidade de remoção das tendências de primeira ordem.

✓ Tipo de vizinhança: define quais pontos circundantes serão utilizados para controlar os resultados de saída, sendo elas a circular padrão e suavização circular. No intuito de não adicionar tendências aos dados por influência na variação dos valores de pontos amostrais de vizinhança, e pela ausência do fator de suavização (*smoothing factor*) na lista de parâmetros disponíveis para calibração – que permite avaliar a amplitude de variação sobre os dados de vizinhança – decidiuse por não usar as técnicas de suavização para na estimativa dos dados.

✓ Tipo de setor: trata-se de uma feição circular que apresenta propriedades definidas pelos demais parâmetros de pesquisa de vizinhança (raio do círculo, número de setores, máximo e mínimo de vizinhos, etc.). Esta feição pode ser subdividida em quatro setores com quadrantes de mesmas dimensões, em quatro setores permitindo 45° de deslocamento ou ainda em oito setores. Por comparação às demais, a feição subdividida em oito setores apresentou maior capacidade de ajuste dos dados de vizinhança na previsão do layer geoestatístico.

✓ Número máximo e mínimo de vizinhos: corresponde ao número máximo e mínimo de pontos amostrais que serão compreendidos na área abrangida pelo setor de pesquisa na interpolação de um ponto qualquer. Por padrão, não são permitidas mudanças no número máximo de vizinhos quando o tipo de setor apresenta oito subdivisões, sendo este limite definido como igualmente correspondente ao limite máximo de setores, e pela característica do tipo de setor escolhido não se justificou a redução do número de vizinhos em função da perda na capacidade de previsão dos resultados por sub-amostragem de vizinhança.

A análise do ajuste do método foi realizada com as seguintes predições de erros, calculados por validação cruzada: quadrado médio do erro (RMS); média padronizada; quadrado médio do erro padronizado; e erro padrão médio. Devido ao pequeno número de amostras, provenientes de dados médios de uma série histórica climática com prévia consistência, optou-se pela não remoção dos valores que apresentassem elevado erro na estimação do modelo geoestatístico.

3.2. ANÁLISE DE COMPONENTES PRINCIPAIS

O elevado número de parâmetros envolvidos na predição das pedopaisagens, e as derivações oriundas, em sua grande maioria, diretamente do MDE, foram os fatores motivacionais para aplicação da Análise de Agrupamento (AA) e Análise de Componentes Principais (ACP), objetivando a remoção dos atributos topográficos redundantes.

Considerou-se conveniente o controle das escalas pela padronização dos dados para média 0 e desvio padrão igual a 1, sendo dessa forma utilizada para a ACP a matriz de correlação (R).

A remoção das variáveis menos explicativas foi realizada utilizando AA, seguida pela ACP, para remoção da variável menos explicativa dentro de cada grupo formado pela primeira técnica. O número de variáveis a serem retidas corresponde ao número de grupos definidos pela AA, e a seleção das variáveis

dentro de cada grupo é feita retendo-se a ultima variável a ser excluída pela ACP dentro daquele grupo, resultando, portanto, na permanência de variáveis não redundantes e mais explicativas da variabilidade espacial das informações de caráter pedológico.

Dos métodos propostos por Jolliffe (1972, 1973), para a redução do número de variáveis na ACP, decidiu-se pela escolha do método dinâmico baseado em componentes principais. Nesse método, define-se previamente um número de variáveis a serem retidas; remove-se o coeficiente de ponderação (autovetor) associado ao último componente principal (variável menos explicativa), e o procedimento é novamente refeito até se atingir o número total, pré-estabelecido, de variáveis a serem retidas.

A definição do número de grupos ideal na AA foi realizada com auxílio de um gráfico de distâncias das ligações da matriz de dissimilaridade, que identifica descontinuidades naturais em termos de distância (VICINI; SOUZA, 2005).

3.3. CLASSIFICAÇÃO DAS CLASSES DE PEDOPAISAGEM POR ISODATA E MÁXIMA VEROSSIMILHANÇA

Para a classificação das pedopaisagens, adotou-se o método de classificação não supervisionada, com uso de ferramenta baseada em classificação ISODATA e Máxima Verossimilhança (BALL; HALL, 1965; RICHARDS, 2013), tal método executa classificação não-supervisionada em uma série de bandas raster de entrada combinando as funcionalidades do Iso Cluster e máxima verossimilhança. Apenas os dados resultantes da ACP foram submetidos a essa classificação. O valor considerado como tamanho mínimo de uma classe foi de 10 vezes o número de camadas nas bandas raster de entrada. O valor inserido para o intervalo de amostragem foi definido correspondente a 20 células, o qual foi considerado como suficiente para representar as menores categorias desejáveis existentes nos dados de entrada. A definição dos grupos foi realizada por comparação ocasional do mapa de solos do estado e daquele gerado pelo classificador. Para adequação da conformação e do total de classes, realizaram-se simulações variando-se o número de classes e a inclusão e/ou exclusão de variáveis, com base no significado intrínseco e importância dos dados considerados.

Para aprimoramento da classificação, todas as bandas tiveram seus valores

transformados por padronização dos valores em um intervalo de 0 a 100. O grid de espaçamento adotado para a modelagem hidrológica foi definido como 50 m. Assim, para a mudança do tamanho de pixel de 30 m para 50 m, e para obtenção de um melhor refinamento e homogeneização dos grupos, aplicou-se uma filtragem ao mapa de pedopaisagens, por reamostragem majoritária, o qual determina o valor de cada célula baseando-se nos valores mais frequentes das células vizinhas.

É importante advertir que, sob um olhar atento, há recursos para criação de classes especializadas do solo, as quais não devem ser apenas voltadas para simular as classes já existentes, mas sim características desejadas, e quando submetidos a pesquisas mais criteriosas, estes recursos poderão indicar resultados em um nível de confiança com erro admissível. E justamente pela escassez de estudos nesse direcionamento que deixa-se claro que os critérios adotados, tratamse apenas de uma proposta para a realidade de áreas com ausência de levantamentos detalhados, e não uma metodologia de substituição das técnicas já consagradas de classificação do solo.

4. RESULTADOS E DISCUSSÃO

4.1. MAPA DE PRECIPITAÇÃO

Na Tabela 2 são apresentados os parâmetros do modelo geoestatístico utilizado para a interpolação da precipitação, resultando na espacialização apresentada na Figura 2.

Tabela 2 – Parâmetros do modelo geoestatístico (KBE) utilizado para interpolação da precipitação.

Propriedades Gerais					
Tamanho do subconjunto:	90				
Fator de Sobreposição:	3				
Número de Simulações:	100				
Tipo de Superfície de Saída:	Predição				
Transformação dos dados:	Empírica	Empírica			
Tipo de semivariograma:	K-Bessel Detrended	K-Bessel Detrended			
Pesquis	a de Vizinhança				
Tipo de vizinhança:	Circular padrão				
Máximo de vizinhos:	8				
Mínimo de vizinhos:	8				
Tipo de setor:	8 setores				
Ângulo (º):	0				
Raio (m): 47.554,22					





A validação do modelo adotado foi fundamentada em estatísticas de erro de previsão por validação cruzada. As previsões centradas nos valores verdadeiros são indicativas de imparcialidade, dessa forma o erro médio de previsão deve ser próximo de zero, porém, por este depender da escala dos dados, os erros de previsão são submetidos à padronização, dividindo-os por seus erros padrão de previsão, da qual a média também deve ser próxima de zero. O erro médio e o Erro Médio Padronizado foram respectivamente 3,373 mm e 0,008. Baseando-se em uma média livre de influências escalares, o erro médio apresentou resultado dentro do esperado e com imparcialidade das previsões.

A validação do erro padrão de previsão foi obtida por analise da variabilidade da previsão dos dados verdadeiros. Para tanto, uma adequada estimação requer que o Erro Médio Padrão deva apresentar valores próximos à Raiz do Quadrado Médio do Erro. Apesar da imprecisão da estimativa, com valores na ordem de 122,95 para a Raiz do Quadrado Médio do Erro e 123,94 para o Erro Médio Padrão, a semelhança entre estes valores é indicativa da avaliação correta da variabilidade e exatidão das predições. Dividindo-se o erro de previsão pelo erro padrão de previsão, para obtenção da Raiz do Quadrado Médio do Erro Padronizada, obtevese o valor de 1,0064, próximo de 1. Tal resultado corrobora com a validação do modelo ajustado e, ainda, por apresentar-se ligeiramente superior a 1, é indicativo de subestimativa da previsão. Esta subestimativa também pode ser observada ao considerar o intercepto e o coeficiente angular, com respectivos valores de 457,15 e 0,64, e r² de 0,62, portanto com capacidade explicativa da variabilidade dos dados de entrada acima de 60%.

4.2. MAPA DE TEMPERATURA

O modelo de regressão linear múltipla obtido para estimação da temperatura é apresentado na Equação 2, composto da altitude, latitude e longitude como variáveis independentes (explicativas).

Temperatura =
$$19,7170 - 0,0065$$
 Altitude + $0,1763$ Latitude - $0,2246$ Longitude (2)

O r² apresentado pela Equação 2 foi de 0,9314, indicativo de excelente capacidade preditiva do modelo. Por ser composto de mais de uma variável independente, adotou-se o r² ajustado para refletir a complexidade do modelo quanto ao número total de variáveis utilizadas em sua elaboração. Esse índice apresentou um valor não muito diferente daquele não ajustado, com valor de 0,9307, de forma que o modelo ainda mantem-se com capacidade explicativa de mais de 93% da variabilidade da variável dependente.

A análise gráfica dos resultados medidos contra os estimados (Figura 3) ilustra maior precisão e exatidão para os valores máximos de temperatura e pequena superestimativa dos valores mínimos e intermediários, evidenciados pela representação linear da tendência dos dados, ajustada pelos respectivos valores do intersépto e do coeficiente angular.



Figura 3 – Confronto entre a temperatura média anual histórica medida e simulada para o estado do ES.

Os coeficientes (Equação 2 e Tabela 3) são representativos do grau de associação e do tipo de relação com cada variável do modelo à qual o mesmo é associado. As variáveis altitude e longitude apresentaram correlação linear negativa com a variável dependente e correlação positiva para variável latitude, portanto com base no comportamento natural esperado. Os coeficientes do modelo apresentam-se nas mesmas unidades que suas variáveis independentes associadas, de forma que mantendo-se todas as outras variáveis constantes, um coeficiente refletirá a variação esperada na variável dependente para cada 1 unidade de variação na variável independente, assim sendo, a exemplo da altitude, a temperatura é reduzida em -0,00649 °C para cada metro adicional desta variável explicativa, o que corrobora com o valor do "*lapse rate*" de temperatura (taxa de queda da temperatura com a altitude) apresentado por Hartmann (1994), de 6,5°C km⁻¹, para o perfil de temperatura médio global.

A variável altitude apresentou um coeficiente padronizado de 0,06364 e erro padrão de 0,000186, portanto considerada a variável mais explicativa do modelo, o que também se tem indício pelo elevado valor (em módulo) do coeficiente de correlação de Pearson com -0,97. As variáveis latitude e longitude apesaram pequena diferença entre seus coeficientes padronizados, com valor inferior a 0,03, porém, pela análise dos erros padrão, a variável longitude apresentou menor

consistência e, portanto, desempenha uma função menos significativa no processo explicativo do modelo, com respectivos valores de 0,05327 para a primeira e 0,11153 para a última variável.

Em referência ao valor-p do teste "t", rejeitou-se a hipótese nula, pois não se pode afirmar, a um nível de confiança superior a 5% de certeza, que os coeficientes das variáveis independentes são iguais a 0, no entanto, por ter sido o índice estatístico Koenker (BP) (*Koenker's studentized Bruesch-Pagan statistic/Godfrey*) significativo, com valor-p de 0,001 (ausência de estacionariedade e presença de heterocedasticidade), adotou-se o teste de probabilidade "t" robusto para uma nova avaliação da significância estatística das variáveis explicativas, pois quando as relações são não estacionárias, só se pode dizer que uma variável explicativa é estatisticamente significativa mediante este tipo de probabilidade.

Constatou-se a ausência de multicolinearidade entre as variáveis explicativas que se mantiveram com valores do teste VIF, variando entre 1,33 e 1,71, portanto, bem abaixo do limite máximo aceitável de 7,5, sendo a extrapolação deste valor é indicativa de viés, conduzindo a instabilidade na capacidade de estimação do modelo.

Tabela 3 – Coeficientes e	respectivas	estatísticas	para o	o modelo	de regressão	linear
múltipla proposto bibliogra	aficamente.					

Variável	Coeficiente	Erro	Teste t	Valor-p	EP	Teste t	Valor-p	Coeficiente	VIF
		Padrao		•	Robusto	Robusto	Robusto	Padronizado	
Intercepto	19,716033	4,1433	4,7586	0,0000*	5,3638	3,6758	0,0004*	0,0000	-
Altitude	-0,006491	0,0002	-34,9602	0,0000*	0,0002	-33,1785	0,0000*	-0,9786	1,34
Longitude	-0,224559	0,1115	-2,0134	0,0466*	0,1309	-1,7161	0,0891	-0,0636	1,70
Latitude	0,176320	0,0533	3,3099	0,0013*	0,0529	3,3334	0,0012*	0,0925	1,33
*0:			مام (از ما مرما مرمر						

*Significativo ao nível de 5% de probabilidade

O teste Koenker (BP) resultou na não aceitação de H₀, com índice de 16,2765 e um valor-p de 0,000995, portanto, conclui-se que as predições do modelo de regressão ajustado são definitivamente heterocedásticas e não estacionárias.

Pelo teste estatístico *Joint Wald* não rejeitou-se a hipótese nula H_0 , pois obteve-se um valor-p igual a 0, e um alto valor para esta estatística, com índice de 2.126,93, excessivamente superior ao valor chi-quadrado, que apresentou-se igual a 7,81.

Em correspondência a recomendação pelos resultados do teste "t" e Koenker (BP) realizou-se a análise do teste "t" robusto, que resultou na não significância da variável longitude ao nível de 5% de probabilidade, com um valor-p de 0,089064, portanto passível de remoção.

O valor obtido para a estatística Global Moran's I foi de 0,1068 de um intervalo que pode variar de -1,0 a 1,0. Apesar de ser próximo ao valor nulo, porém positivo, o conjunto apresenta uma suave tendência de se agrupar espacialmente, ou seja, valores elevados dos resíduos padronizados aglomeram-se próximos de outros valores elevados e valores baixos próximos a outros valores baixos. Após o cálculo do índice de Moran, obteve-se o valor do índice esperado de -0,009174 e da variância entre eles de 0,005272 (com base no número de amostras) a qual determina se esta diferença é estatisticamente significativa ou não. Assim, pelos critérios adotados, obteve-se um valor-p de 0,110132 e um valor positivo para zscore, portanto não se rejeita a hipótese nula, sendo bem possível que a distribuição espacial dos valores seja resultante de processos espaciais aleatórios, com tendência que o padrão de distribuição dos ruídos ocorra ao acaso, dado também a um z-score de 1,5976. Na Figura 4 podem-se observar os diferentes tipos de disposição dos resíduos mediante os resultados da estatística Global Moran's I.



Figura 4 – Ilustração da distribuição e respectivos índices da autocorrelação espacial dos resíduos.

A análise da distribuição de cada variável do modelo e dos resíduos, e seus respectivos gráficos de dispersão, mostra a relação entre a variável dependente e cada uma das variáveis explicativas. Por essa análise foi possível determinar que o viés – indicado por um valor-p estatisticamente significativo para Jarque Bera, resultante da ausência de normalidade dos resíduos – é decorrente de heterogeneidade ente os histogramas, com ocorrência de mais de um tipo de distribuição, o que resulta na necessidade de transformação dos dados para melhorar o desempenho do modelo de regressão.

A presença de heterocedasticidade indicada pela estatística de Koenker (BP) é fortemente expressa pela relação da temperatura com as variáveis de localização espacial (latitude e longitude), e ilustrada na Figura 5, de forma a apresentar disposição em forma de cone, pela tendência ao aumento na difusão dos erros partindo daqueles de menor magnitude para o outro extremo ao longo do gráfico de dispersão. Esta propriedade advém da relação sistemática entre as variáveis que estão sendo modelada, consequência da variabilidade espacial dos dados e/ou magnitude dos mesmos, resultando, de forma inerente, na precisão de previsão. Foi observada também, forte relação linear da temperatura com a variável explicativa altitude, consequente das inversas formas assintóticas da distribuição. O resumo dos índices e testes estatísticos diagnósticos para validação do modelo de regressão linear múltipla é apresentado na Tabela 4.



Figura 5 – Dispersão dos resíduos padronizados no processo de estimativa dos dados de temperatura.

Tabela 4 – Resumo estatístico dos parâmetros diagnósticos adotados para validação do modelo de regressão linear múltipla.

Diagnóstico	Símbolo/ Abreviação	Valor
Coeficiente de determinação	r ²	0,9379
Coeficiente de determinação ajustado	r ² Ajustado	0,9362
F-estatística conjunta	F-Estatística	534,0525
Valor-p do teste F-estatística conjunta	F-Probabilidade	0,0000
Wald estatística	Wald	2126,9833
Valor-p do teste Wald estatística	Wald-Probabilidade	0,0000
Estatística estudentizada Breusch-Pagan de Koenker	K(BP)	16,2765
Valor-p da estatística estudentizada K(BP)	K(BP)-Probabilidade	0,0010
Estatística Jarque-Bera	JB	81,9962
Valor-p da estatística Jarque-Bera	JB-Probabilidade	0,0000
Sigma ao quadrado dos erros (variância)	σ^{2}_{Erro}	0,2101
Índice global de Moran	IGM	0,1068
Índice esperado de Moran	IGEM	-0,0091
Variância	σ^{2}_{IGM}	0,0052
Z-score	Z-score	1,5976
Valor-p	Valor-p	0,1101

O modelo proposto inicialmente, apesar de apresentar um excelente ajuste pelo coeficiente de determinação, correlação, VIF e nível aceitável pelo teste *Joint Wald*, não correspondeu aos critérios que condicionam a validação de um modelo de regressão linear, tendo o mesmo apresentado heterocedasticidade, não estacionáriedade, não significância da variável explicativa longitude ao nível de probabilidade especificada, ausência de distribuição normal dos resíduos padronizados e autocorrelação espacial dos resíduos, portanto trata-se de um modelo tendencioso e mal especificado, consequentemente não apto para uso.

A identificação de variáveis adequadas, em número suficiente, corretamente transformadas e combinadas, que satisfaçam todas as condições de construção e validação de um modelo de regressão linear múltipla, trata-se de um trabalho árduo e muitas vezes inconcluso, ao ponto de em alguns casos permitirem-se a permanência de uma variável considerada de elevada importância teórica mesmo não sento estatisticamente significativa, ou ainda a adoção de outras técnicas de regressão, como a regressão geograficamente ponderada.

O total de combinações possíveis testadas com as 23 variáveis submetidas a regressão exploratória foi de 1.362.156.864, resultando nas possibilidades de construção dos modelos com as combinações apresentadas na Tabela 5. Nota-se

que todos os possíveis modelos comtemplam a variável altitude submetida à transformação logarítmica para normalização de uma distribuição de caráter exponencial como é a desta variável, outros modelos ainda fazem uso desta mesma variável sob efeito de outros tipos de transformações. Os possíveis modelos de regressão linear múltipla ficam sujeita, portanto apenas as variáveis: altitude (Alt), excedente hídrico (Exc), deficiência hídrica (Def), comprimento do caminho de fluxo (FPL) e índice de energia do fluxo (SPI).

Tabela 5 – Índices e testes diagnósticos das possíveis combinações obtidas durante a análise exploratória dos modelos de regressão linear.

N°	Categoria	$R^{2}_{Ajustado}$	AICc⁺	JB [∓]	KBP [*]	VIF*	SA [*]	X1	X2	Х3	X4
1	Tamp	0.8772	220,29	0,068	0,056	2,179	0,139	Exc	Alt ³	FPL ³	Log Alt
2	Temp	0.8772	220,29	0,068	0,056	2,179	0,139	Exc BC	Alt ³	FPL³	Log Alt
3	Tomp BC	0.8772	220,29	0,068	0,056	2,179	0,139	Exc	Alt ³	FPL ³	Log Alt
4	Temp BC	0.8772	220,29	0,068	0,056	2,179	0,139	Exc BC	Alt ³	FPL ³	Log Alt
5	Evp(Tomp)	0.8769	-1294,19	0,069	0,056	2,179	0,139	Exc	Alt ³	FPL ³	Log Alt
6	Exb(Temp)	0.8769	-1294,19	0,069	0,056	2,179	0,139	Exc BC	Alt ³	FPL ³	Log Alt
7		0.8642	1069,97	0,130	0,073	2,179	0,136	Exc	Alt ³	FPL ³	Log Alt
8		0.8642	1069,97	0,130	0,073	2,179	0,136	Exc BC	Alt ³	FPL ³	Log Alt
9	Temp ²	0.8624	1071,45	0,071	0,064	2,171	0,170	Exc	SPI ²	Alt ³	Log Alt
10		0.8624	1071,45	0,071	0,064	2,171	0,170	Exc BC	SPI ²	Alt ³	Log Alt
11		0.8573	1075,47	0,080	0,061	2,200	0,102	Alt ³	FPL ³	Exc ^{1/3}	Log Alt
12		0.8510	1856,85	0,191	0,096	2,179	0,132	Exc	Alt ³	FPL ³	Log Alt
13		0.8510	1856,85	0,191	0,096	2,179	0,132	Exc BC	Alt ³	FPL ³	Log Alt
14		0.8493	1858,09	0,115	0,080	2,171	0,158	Exc	SPI ²	Alt ³	Log Alt
15	Temp ³	0.8493	1858,09	0,115	0,080	2,171	0,158	Exc BC	SPI ²	Alt ³	Log Alt
16	-	0.8403	1864,50	0,104	0,058	2,638	0,139	Exc ²	Alt ³	Exc ^{1/3}	Log Alt
17		0.8400	1864,70	0,057	0,135	1,871	0,113	Def ²	Exc ²	Alt ³	Log Alt
18		0.8361	1867,37	0,085	0,057	2,305	0,114	Alt ³	Exc ³	Exc ^{1/3}	Log Alt

[†]Corrected Akaike Information Criterion (AICc)

*Koenker's studentized Bruesch-Pagan statistic/Godfrey

*Fator de Inflação da Variância

[¥]Autocorrelação espacial

Como todos os modelos apresentaram um r² ajustado superior a 0,80, o critério de seleção foi baseado nos maiores valores de JB, KBP e SA respectivamente, no entanto fica a critério do pesquisador determinar os modelos de interesse da forma que melhor lhe convier já que todos os 18 modelos são considerados satisfatórios. Apesar dos dois modelos eleitos (em destaque) serem aparentemente iguais, o modelo que apresenta a variável explicativa "excedente hídrico" sob efeito da transformação boxcox ocasionou um melhor ajuste ao intercepto do modelo, e mesmo que tal ajuste tenha sido pouco expressivo, serviu de critério de desempate. Três valores considerados como outliers, foram

[‡]Jarque-Bera

identificados para a variável explicativa FPL³ e um valor para a variável altitude³, acarretados durante o processo de transformação. Após a remoção de tais outliers, a variável FPL³ apresentou-se não significativa pelo teste t robusto, e o modelo construído sem a inclusão desta variável (Equação 3) resultou no aumento da eficiência de previsão e na melhora significativa do valor-p para o teste de autocorrelação espacial residual com padrão aleatório de distribuição dos dados, portanto indicativo de correta especificação das variáveis explicativas do modelo.

Temperatura³ =
$$17468,50 - 3,53$$
.ExcBC - $0,000008$.Altitude³ - (3)
1247,31.Log(Altitude)

A validação dos coeficientes do modelo final e o resumo dos índices estatísticos encontram-se apresentados nas Tabelas 6 e 7. Nota-se que o valor-p do teste t robusto apresentou um valor praticamente igual a 0, o que reafirma a excelente especificação das variáveis explicativas, e apesar da repetição de uma destas variáveis, os métodos de transformação proporcionaram ausência de multicolinearidade.

Tabela 6 – Coeficientes e respectivas estatísticas para o modelo de regressão linear múltipla identificado por regressão esploratória.

Variával	Cooficiente	Erro	Tooto t	Valor n	EP	Teste t	Valor-p	Coeficiente	VIF
variavei	Coenciente	Padrão	Teste i	valor-p	Robusto	Robusto	Robusto	Padronizado	
Intercepto	17468,50314	426,0010	41,0058	0,0000*	464,5972	37,5992	0,0000*	0,0000	-
Exc BC	-3,534387	0,7669	-4,6086	0,0000*	0,6856	-5,1550	0,0000*	-0,2507	2,14
Alt ³	0,00008	0,0000	-9,9687	0,0000*	0,0000	-8,6880	0,0000*	-0,5250	2,01
Log(Alt)	-1247,31710	223,7459	-5,5747	0,0000*	234,1611	-5,3267	0,0000*	-0,2804	1,83
*Significati	vo ao nível de	5% do nr	obabilida	do					

Significativo ao nivel de 5% de probabilidade

Tabela 7 – Resumo estatístico dos parâmetros diagnósticos adotados para validação do modelo de regressão linear múltipla identificado por regressão esploratória.

Diagnóstico	Símbolo/ Abreviação	Valor
Critério de informação Akaike	AIC	1779,76563
Critério de informação Akaike corrigido	AICc	1780,36563
Coeficiente de determinação	r ²	0,85936
Coeficiente de determinação ajustado	۲ ² Ajustado	0,85522
F-estatística conjunta	F-Estatística	207,75017
Valor-p do teste F-estatística conjunta	F-Probabilidade	0,00000
Wald estatística	Wald	443,49609
Valor-p do teste Wald estatística	Wald-Probabilidade	0,0000
Estatística estudentizada Breusch-Pagan de Koenker	K(BP)	4,10740
Valor-p da estatística estudentizada K(BP)	K(BP)-Probabilidade	0,25010
Estatística Jarque-Bera	JB	4,06952
Valor-p da estatística Jarque-Bera	JB-Probabilidade	0,13071
Sigma ao quadrado dos erros (variância)	σ^{2}_{Erro}	1084350,0671
Índice global de Moran	IGM	0,0480
Índice esperado de Moran	IGEM	-0,0095
Variância	σ^{2}_{GM}	0,0051
Z-score	Z-score	0,8060
Valor-p	Valor-p	0,4202



Figura 6 – Temperatura média anual para o estado do Espírito Santo.

4.3. ANÁLISE DE COMPONENTES PRINCIPAIS

A melhoria das técnicas de modelagem do solo e da variação espacial de suas propriedades, a disponibilidade de ferramentas para análise da paisagem, a acessibilidade de atributos do terreno em alta resolução espacial e a crescente necessidade de abordagens simplificadas capazes de capturar padrões espaciais significativos de propriedades hidráulicas do solo, tanto em grandes escalas como em escalas refinadas, são motivos que têm direcionado muitas pesquisas ao estudo da influência da topografia nas propriedades e processos do solo (ROMANO; CHIRICO, 2004).

Uma das principais intensões na aplicação da metodológica de geração das classes de pedopaisagem consistiu na identificação e seleção de um número mínimo de produtos que se enquadrassem na descrição e cumprisse com os propósitos citados por Romano e Chirico (2004), uma vez que é visível o apelo para que estas propostas sejam continuadas e novas iniciativas nesta direção sejam tomadas.

Além do MDE, dificilmente encontra-se tão prontamente disponível um produto que reúna tantos requisitos necessários ao cumprimento do objetivo proposto, dos quais citam-se: a facilidade de acesso, qualidade em resolução espacial e precisão, disponibilidade de ferramentas de geoprocessamento e adequabilidade desse produto a tais ferramentas, diversidade em funcionalidade e usabilidade, e sobre tudo derivação em níveis categóricos de subprodutos com propriedades distintas e discriminantes do terreno. Entende-se, porém, que se por um âmbito o grande número de atributos que podem ser derivados do MDE constitui vantagem, por outro, a redundância quanto suas origens pode certamente implicar em desvantagem pelo seu uso conjunto na busca de variabilidade máxima de cada produto.

Para contornar o problema associado aos subprodutos derivados do MDE, a utilização de AA apresentou-se muito eficiente para separação de grupos com elevada redundância, seguida da seleção da variável mais explicativa, dentro de cada grupo, por meio da ACP.

A amplitude máxima na distância de ligação entre grupos, obtida pela matriz de dissimilaridade plotada no dendrograma (Figura 7), foi de 1,48, fazendo distinção entre os pares mais distantes, compreendidos pelas variáveis índice de energia do fluxo e área de captação modificada correspondentes aos segmentos do dendrograma 10 e 13 e fator Is e declividade correspondentes aos segmentos 6 e 7, sendo estas ultimas as variáveis de maior semelhança entre si, com uma variação de apenas 0,17 de distância de ligação, ou 11,55% de dissimilaridade.



Figura 7 – Dendrograma da análise de agrupamento dos atributos topográficos primários, secundários e climáticos pedogenéticos.

Segundo a ordem de remoção das variáveis (Tabela 7) conservaram-se apenas a 11, 8, 14, 16, 3, 1, 17 que correspondendo respectivamente a: índice topográfico de umidade, aspecto, mde, curvatura plana, precipitação, distância euclidiana e curvatura do perfil.

Pela análise dos resultados da classificação dessas sete variáveis submetidas à *Iso Cluster Unsupervised Classification* presenciou-se uma representação irrealista de classes pedológicas do solo por consequência do uso da variável Aspecto, acarretando em um elevado número de feições de dimensão diminuta quando comparado ao mapa temático de solos por análise da classes pedológicas quanto a sua alocação, distribuição, abrangência e padrões de ocorrência, portanto essa variável foi removida do grupo inicialmente selecionado para classificação.

Após nova simulação com exclusão da variável Aspecto, identificaram-se no novo produto, representações irrealistas provenientes de fortes tendências à geração de classes com padrões da variável precipitação em grandes áreas de ocorrência, justificando-se também sua remoção. Por fim a simulação seguinte revelou-se estar em conformidade com os padrões naturais das classes edáficas, notando-se a presença de regiões com elevada propensão de representação de certas classes de solo definidas pelo método convencional de classificação dos solos (Figura 8).

Grupos da AA	Variáveis *	Descrição	Ordem de Remoção pela ACP
1	1	Distância Euclidiana	NR ^{*1}
	2	Relevo Relativo	8
	6	Fator LS	9
2	7	Declividade	7
	12	Declive de Captação Modificada	3
	14	Modelo Digital de Elevação	12 [*]
3	8	Aspecto	6 [*]
4	16	Curvatura Plana	11 [*]
	3	Precipitação	15 [*]
5	5	Radiação Solar	13
	4	Temperatura	14
6	9	Comprimento do Caminho de Fluxo	5
0	15	Curvatura Tangencial	10
	17	Curvatura do Perfil	NR ^{*2}
	10	Índice de Energia do Fluxo	1
7	13	Área de Captação Modificada	2
	11	Índice Topográfico de Umidade	4 ^{**}

Tabela 7 – Ordem de remoção dos atributos topográficos primários, secundários e climáticos pedogenéticos.

*Para melhor disposição, a ordenação das variáveis segue a sequência do dendrogama. *Última variável removida dentro do grupo e, portanto, selecionada para classificação do mapa de pedopaisagens.

^{1/2} Variáveis mais explicativas pela ACP e "Não Removidas" por empate no método.

Ressalta-se que a intensão na elaboração das classes de pedopaisagens do solo não foi a definição de classes pedológicas como usualmente tem sido feito, mas sim o delineamento de áreas inominadas com características físico-hídricas e termodinâmicas semelhantes, o que foi certamente alcançado pela intrínseca correlação da seleção final das variáveis com tais características, com base nos seguintes processos físicos: gradiente de fluxo lateral e aceleração em direção aos corpos d'água, gradiente climático local, convergência e divergência de fluxo d'água, previsão de zonas de saturação, e aceleração dos fluxos laterais pelo gradiente natural do relevo.



Figura 8 – Mapa de classes de solo do Sul do ES (A) e de pedopaisagem (B).

5. CONCLUSÃO

- As camadas matriciais de caráter edáfico mostraram-se representativas da variabilidade espacial das propriedades físico-hídricas e termodinâmicas para tematização de classes categóricas do mapa de pedopaisagem.
- O uso de técnicas de análise de agrupamento associada a componentes principais mostrou-se eficiente para eliminação de variáveis redundantes e, portanto pouco explicativas do fenômeno em estudo.
- Para o presente estudo a altitude mostrou-se como variável de maior importância devido ao grande número de vaiáveis que podem ser derivadas dessa e por ser um produto de fácil aquisição e sem ônus.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, M.; PENA, S. B. Permeabilidade e Áreas de Máxima Infiltração. In: Estrutura Ecológica Nacional, Uma Proposta de Delimitação e Regulamentação. Lisboa: Isapress. 121-136 p. 2013.

AB'SÁBER, A. N. O domínio dos mares de morro no Brasil. São Paulo, USP/Instituto de geografia, (**Série Geomorfologia**, 2), 9 p. 1966.

BALL, G. H.; HALL, D. J. **ISODATA, A novel method of data analysis and pattern classification**. Tecnical Report. 1965.

BERTONI, J.; LOMBARDI NETO, F. Conservação do solo. 8.ed. São Paulo: Ícone, 2012. 355p.

BOEHNER, J.; KOETHE, R.; CONRAD, O.; GROSS, J.; RINGELER, A.; SELIGE, T. Soil regionalisation by means of terrain analysis and process parameterisation. In: MICHELI, E., NACHTERGAELE, F., MONTANARELLA, L. **Soil Classification** 2001. Luxembourg, European Soil Bureau, Research Report, n. 7, p. 213-222, 2002.

BÖHNER, J.; SELIGE, T. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. **Göttinger Geographische Abhandlungen**, [S.L.], v. 15, p. 13-27, 2006.

BURROUGH P. A; MCDONNELL R. A. **Principles of Geographical Information Systems.** Oxford University Press; 1998.

CARVALHO JUNIOR, W. de; CHAGAS, C. da S.; FERNANDES FILHO, E. I.; VIEIRA, C. A. O.; SCHAEFER, C. E. G.; BHERING, S. B.; FRANCELINO, M. R. Digital soilscape mapping of tropical hillslope areas by neural networks. **Scientia Agrícola**, Piracicaba, v. 68, n. 6, 691-696 p., nov./dec. 2011.

CASTRO, F. da S. **Zoneamento Agroclimático para a cultura do Pinus do Estado do Espírito Santo.** 2008. Dissertação (Mestre em Produção Vegetal). Universidade Federal do Espírito Santo. Alegre, 2008.

CHILÈS, J. P.; DELFINER, P. **Geostatistics: Modeling Spatial Uncertainty**. New York: John Wiley & Sons. 695 p. 1999.

DEMATTÊ, J. A. M.; GENÚ, A. M.; FIORIO, P. R.; ORTIZ, J. L.; MAZZA, J. A.; LEONARDO, H. C. L. Comparação de mapas de solos obtidos por sensoriamento remoto espectral e pelo método convencional. **Revista Agropecuária Brasileira,** Brasília, v. 39, n. 12, 1219-1229 p., dez. 2004.

EMPRESA BRASILEIRA DE PESQUISA AGOPECUÁRIA – EMBRAPA, Centro Nacional de Pesquisa de Solos. Sistema brasileiro de classificação de solos. 6 ed. Rio de Janeiro, 412 p. 1999.

ESRI. Classification trees of the interpolation methods offered in Geostatistical Analyst Disponível em: <http://resources.arcgis.com/en/help/main/10.2/#/Classification_trees_of_the_in terpolation_methods_offered_in_Geostatistical_Analyst/003100000000000/ >. Acesso em: 07 fev. 2013.

FLORINSKY, I. V. Accuracy of local topographic variables derived from digital elevation models. **International Journal of Geographical Information Science**, v. 12, n.1, 47-61 p., 1998.

FLORINSKY, I. V.; EILERS, R. G.; MANNING, G. R.; FULLER, L. G. Prediction of soil properties by digital terrain modeling. **Environmental Modelling & Software**, v. 17, 295-311 p., 2002.

GREENLEE, D. D. Raster and vector processing for scanned linework. **Photogrammetric Engineering and Remote Sensing**, v. 53, n. 10, p.1383–1387, 1987.

HABTEC. **Diagnóstico e plano diretor das bacias dos rios Santa Maria da Vitória e Jucu**. Plano Diretor. 2 v. Espírito Santo, nov./1997a.

HARTMANN, D.L. **Global Physical Climatology**. London: Academic Press, 1994. 408 p.

INSTITUTO JONES DOS SANTOS NEVES – IJSN: Banco de dados. Disponível em: <http://www.ijsn.es.gov.br/Mapas/Consulta/?module=Search&action=Results&l ookfor=&type=AllFields&filter%5B%5D=format%3A%22Arquivo+Shapefile%22 &view=list>. Acesso em: 01 jul. 2014.

JENSON S. K., DOMINGUE, J. O. Extracting topographic structure from digital elevation data for geographic information system analysis. **Photogrammetric Engineering and Remote Sensing**, v. 54, n. 11, p. 1593-1600, 1988.

JOLLIFFE, I. T. Discarding Variables in a Principal Component Analysis. I: Artificial Data. **Jounarl of the Royal Statistical Society**, v. 21, n. 2, 160-173 p, 1972.

JOLLIFFE, I. T. Discarding Variables in a Principal Component Analysis. II: Artificial Data. **Jounarl of the Royal Statistical Society**, v. 22, n. 1, 21-31 p, 1973.

LAND PROCESSES DISTRIBUTED ACTIVE ARCHIVE CENTER – LP DAAC, 2011. ASTER Global Digital Elevation Model Version 2 – Summary of Validation Results. 26 p. Available in: <https://igskmncnwb001.cr.usgs.gov/aster/GDEM/Summary_GDEM2_validation _report_final.pdf>. 5 Apr./2012. Acessado em: 03 Out. 2013.

MCCUNE, B.; DYLAN K. Equations for potential annual direct incident radiation and heat load index. **Journal of Vegetation Science**, v. 13, p. 603-606, 2002.

MOORE, I. D.; GRAUSON, R. B.; LADSON, A. R. Digital terrain modelling: a review of hydrological, geomorphological and biological applications. 1991. In: BEVEN, K, J.; MOORE, I. D. Terrain analysis and distributed modelling in hydrology. J. Wiley & Sons, 1991.

NAGHETTINI, M.; PINTO, É. J. de A. **Hidrologia Estatística**. Belo Horizonte: CPRM, 2007. 552 p. ISBN: 978 -85-7499-023-1.

O'CALLAGHAN, J. F.; MARK, D. M. "The extraction of drainage networks from digital elevation data," **Computer Vision, Graphics and Image Processing**, v. 28, p. 328-344, 1984.

OLIVEIRA, M. M.; LOBO FERREIRA, J. P. C. **Proposta de uma metodologia para definição de áreas de infiltração máxima**. In: CONGRESSO DA ÁGUA, 6., Porto: Associação Portuguesa dos Recursos Hídricos. 2002.

OTTONI, M. V. **Classificação físico-hídrica de solos e determinação da capacidade de campo** *in situ* a partir de testes de infiltração. 141 p. 2005. Dissertação (Mestrado em Ciências em Engenharia Civil). Universidade Federal do Rio de Janeiro. Rio de Janeiro. 2005.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world mape of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, 1633-1644 p., 2007.

PEREIRA, A. R.; ANGELOCCI, L. R.; SENTELHAS, P. C. **Agrometeorologia:** fundamentos e aplicações práticas. Guaíba: Agropecuária, 2002. 478 p.

PLANCHON, O. DARBOUX, F. "A fast, simple and versatile algorithm to fill the depressions of digital elevation models," **Catena**, v. 46, p. 159-176, 2001.

PILZ, J.; SPÖCK, G. Why Do We Need and How Should We Implement Bayesian Kriging Methods. Stochastic Environmental Research and Risk Assessment. [S.L.], 2007.

PINEDA, L. A. Estudo observacional e de modelagem hidrológica de uma micro-bacia em floresta não perturbada na Amazônia Central. São José dos Campos: INPE, 17 de mar. 2008. 236 f. Tese (Doutorado do Curso de Pós-Graduação em Meteorologia). Instituto Nacional de Pesquisas Espaciais – INPE.

QUINN, P.; BEVEN, P.; CHEVALLIER, P.; PLANCHON, O. The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models. **Hydrol. Proc.**, v. 5, p. 59-80, 1991.

ROMANO, N.; CHIRICO, G. B. The role of Terrain Analysis in using and Developing Pedotransfer Functions. **Developments in Soil Science**, v. 30, 2004.

RICHARDS, J. A. **Remote Sensing Digital Image Analysis an introduction**. 4 ed. Edition. New York: Springer. 2013. ISBN-978-3-642-30061-5.

TRICART, J. As relações entre a morfogênese e a pedogênese. **Notícia Geomorfológica**, Campinas-SP, 8 (15) : p.5-18, jun. 1968.

VALERIANO, M. de M.; PICINI, A. G. Geoprocessamento de informações agroclimatológicas. INPE: São José dos Campos. 2003.

VICINI, L.; SOUZA, A. M. **Análise multivariada da teoria à prática.** Monografia (Especialização), Universidade Federal de Santa Maria. 2005.

ZEVERBERGEN, L. W. THORNE, C. R. Quantitative analysis of land surfasse topography Earth surfasse processes and landoforms. v. 12, p. 47-56, 1987.

WILSON, J. P.; GALLANT, J. C. **Terrain Analysis: Principles and Applications**. Nova Jersey: Jonh Wiley & Sons, Inc. jan./2000. ISBN 0-471-32188-5.



APENDICE A – Tangential Curvatura.



APENDICE B – Stream Power Index



APENDICE C – Solar Radiation.











APENDICE F – Profile Curvature.



APENDICE G – Topographic Wetness Index.



APENDICE H – Plan Curvature.







APENDICE J – Catchment Area.







APENDICE L – Euclidean Distance.



APENDICE M – Catchment Slope.



APENDICE N – Aspect.