ENTRELUGAR

O USO DE POLÍGONOS DE AMOSTRAGEM EM CLASSIFICAÇÕES SUPERVISIONADAS DE IMAGENS DE SATÉLITE

USE OF SAMPLING POLYGONS IN SUPERVISIONED CLASSIFICATIONS OF SATELLITE IMAGES

EL USO DE POLÍGONOS DE MUESTREO EN CLASIFICACIONES SUPERVISADAS DE IMÁGENES DE SATÉLITE

Paulo Roberto Fitz
Professor Doutor Visitante – FCH/PPGG – UFGD
paulofitz@ufgd.edu.br

Jeferson Cordeiro Vieira
Acadêmico do Curso de Geografia da UFGD
vieirajcufgd@gmail.com

Mirlla Casimiro Soares
Acadêmica do Curso de Geografia da UFGD
mih_cass@hotmail.com

Resumo: A dinâmica existente no território sempre ocupou destaque em pesquisas geográficas e, mais precisamente, ambientais. Diferentes métodos e técnicas podem ser utilizados para a realização destes estudos e, entre eles, pode-se destacar a aplicação das técnicas do sensoriamento remoto com o uso de imagens de satélites orbitais. Logo, este estudo buscou, a partir da classificação supervisionada de imagens do satélite *Landsat* com o uso de áreas de treinamento, ou clusters de polígonos, analisar os dados obtidos através da relação entre as dimensões das áreas classificadas com o incremento de polígonos de amostragem. Foram realizadas experimentações com três classes para os anos de 1985 e de 2018, a saber, corpos d'água, áreas antropizadas e cobertura vegetal "original". As simulações realizadas confirmaram a hipótese apresentada, ou seja, a de que haveria uma tendência de estabilização nos dados de acordo com o incremento de

áreas de treinamento. Como recomendação para futuras classificações, sugere-se a adoção de, pelo menos, cinquenta polígonos amostrais por classe, que configurem, da melhor forma possível, a região a ser classificada uma vez que tais áreas de treinamento deverão abarcar todas as características relativas às classes a serem adotadas.
Palavras-chave: sensoriamento remoto, classificação de imagens de satélite, áreas de treinamento, cluster de polígonos, número de amostras.

Abstract: The territory dynamics has always been prominent in geographic and, more precisely, environmental research. Different methods and techniques can be used to perform these studies, among them, it is possible to highlight the application of remote sensing techniques with the use of orbital satellite images. This study aimed to analyze the data obtained through the relationship between the dimensions of the classified areas and the increasing number of sampling polygons, based on the supervised classification of satellite images using training areas or polygon clusters. Experiments were carried out with three classes for the years 1985 and 2018, namely water bodies, anthropic areas and "original" vegetation cover. The simulations carried out confirmed the hypothesis presented, i.e., that there would be a trend of stabilization in the data according to the increase of training areas. As a recommendation for future classifications, it is suggested to adopt at least fifty sample polygons per class, which best defines the region to be classified, since such areas of training should cover all the characteristics related to the classes to be adopted.
Keywords: remote sensing, classification of satellite images, training areas, cluster of polygons, number of samples.

Resumen: La dinámica existente en el territorio siempre ocupó destaque en investigaciones geográficas y, más precisamente, ambientales. Diferentes métodos y técnicas pueden ser utilizados para la realización de estos estudios y, entre ellos, se puede destacar la aplicación de las técnicas del sensoriamiento remoto con el uso de imágenes de satélites orbitales. Este estudio buscó, a partir de la clasificación supervisada de imágenes de satélite con el uso de áreas de entrenamiento, o clusters de polígonos, analizar los datos obtenidos a través de la relación entre las dimensiones de las áreas clasificadas con el incremento de polígonos de muestreo. Se realizaron

experimentos con tres clases para los años 1985 y 2018, a saber, cuerpos de agua, áreas antropizadas y cobertura vegetal "original". Las simulaciones realizadas confirmaron la hipótesis presentada, o sea, la de que habría una tendencia de estabilización en los datos de acuerdo con el incremento de áreas de entrenamiento. Como recomendación para futuras clasificaciones, se sugiere la adopción de al menos cincuenta polígonos muestrales por clase, que configuren de la mejor manera posible la región a ser clasificada, ya que tales áreas de entrenamiento deberán abarcar todas las características relativas a las clases a ser adoptadas.

Palabras clave: teledetección, clasificación de imágenes por satélite, áreas de entrenamiento, cluster de polígonos, número de muestras.

INTRODUÇÃO

A dinâmica existente no território sempre ocupou destaque em pesquisas geográficas e, mais precisamente, ambientais. As modificações espaciais verificadas em dadas regiões, especialmente em função de atividades antrópicas, estão relacionadas aos cenários físicos, políticos, econômicos e ambientais desses territórios em determinado momento. Tais abordagens estão presentes nas obras de autores como Claval (1987), Corrêa (2010), George (1986), Lacoste (1978, 2002), Santos (1998, 2006), entre outros.

Para se estudar estes processos, diferentes métodos, técnicas e ferramentas podem ser empregados. Em função da escala de trabalho, muitas vezes são utilizadas imagens de satélite para a realização de estudos e análises com acurácia relativa às características das mesmas. A cobertura vegetal presente na superfície terrestre, por exemplo, denota diferentes comportamentos quando do uso das técnicas de sensoriamento remoto, especialmente no tocante à utilização de imagens de sensores multiespectrais. Conforme suas propriedades físicas e químicas, a vegetação responderá à radiação solar incidente de acordo com sua própria natureza. A análise de imagens multiespectrais refletirá, consequentemente, o estado físico-químico das plantas de acordo com seus atributos inerentes em determinado espaço e momento.

Nessa perspectiva, são efetuadas composições *coloridas* e "*falsa-cor*", a partir das diferentes faixas espectrais disponibilizadas pelas imagens, para facilitar a visualização e a compreensão dos fenômenos que ocorrem na superfície terrestre. Crósta (1993, p. 63) alerta para o fato de que, com relação às bandas espectrais, "uma

cuidadosa seleção das mesmas faz-se necessária, para que a combinação de 3 bandas contenha a informação espectral realmente desejada."

Num segundo momento, em geral, são realizadas classificações das mesmas para a geração de mapas temáticos relativos ao cenário regional exibido. Segundo é apontado por Fitz (2008, p. 129) "a classificação de uma imagem nada mais é do que a identificação de determinados elementos nelas presentes pela associação de cada um de seus pixels a uma determinada classe preestabelecida."

Vários algoritmos presentes em diferentes softwares podem ser utilizados para classificar imagens de satélite. As classificações podem ser *supervisionadas*, quando o usuário interfere diretamente na mesma através da escolha das categorias de amostragem; e *não supervisionadas*, quando a classificação é automática, ou seja, o usuário somente define o número de classes a serem reconhecidas. (FITZ, 2008).

Em classificações supervisionadas, são utilizadas unidades de amostragem que podem ser: "(1) um único pixel, (2) um cluster de pixels (geralmente um quadrado de 3 x 3 pixels, (3) um polígono (ou objeto) e (4) um cluster de polígonos" (CONGALTON; GREEN, 2009, p. 70, tradução própria). Neste trabalho foram utilizadas áreas de treinamento dadas por polígonos desenhados pelo usuário, os quais abrangem amostras das feições presentes na superfície e reconhecidas como pertencentes a um dado conjunto predeterminado, ou seja, um "cluster de polígonos".

Os polígonos de amostragem escolhidos dizem respeito, portanto, à assinatura espectral presente nas áreas previamente escolhidas como modelos para a obtenção das classes de cobertura da terra. A partir deste momento o *software* se encarregará de agrupar pixels semelhantes de acordo com as áreas de treinamento adotadas.

Desta maneira, Richards; Jia (2006, p. 199) destacam que:

> pixels contidos em áreas de treinamento devem ser suficientes para permitir estimativas razoáveis obtidas pelos elementos da média condicional da classe e da matriz de covariância vetorial.

Analisando esta questão de uma maneira um pouco mais detalhada, Congalton; Green (2009, p. 75) indicam que:

> Devido ao grande número de amostras potenciais (isto é, pixels, aglomerados de pixels, polígonos e aglomerados de polígonos)

em uma imagem de sensoriamento remoto, não é incomum a tradicional proposta de se aplicar uma amostragem de 2%, ou mesmo 5%.

Ainda nesta perspectiva, cabe apresentar que, conforme Eastman (1998, p. 133) "o número de *pixels* de cada conjunto de treinamento (i.e., todas as áreas de treinamento para uma única classe de cobertura do solo) não deveria ser menor do que dez vezes o número de bandas". Esta informação é interessante para que se tenha uma noção da quantidade de polígonos necessários para uma classificação suficientemente confiável. Entretanto, ela carece de maiores detalhamentos, especialmente quando se trabalha com imagens *Landsat* e se deve decidir sobre a pertinência ou não de determinada feição à dada classe. De acordo com Crósta (1993, p. 115): "Várias áreas de treinamento podem ser definidas para uma mesma classe, para assegurar que os pixels a ela pertencentes são realmente representativos dessa classe."

Para elucidar possíveis dúvidas quanto ao assunto, autores como Genderen et al. (1978), Aronoff (1985) e Congalton (1991) propuseram soluções mais ou menos elaboradas. Em geral, os estudos desenvolvidos nessa área ao longo dos tempos, vinculam-se às estatísticas relativas à aceitação ou não das amostras em relação à verdade de campo. A proposição de Genderen et al. (1978, p. 7), adotada neste trabalho, apresenta a função de probabilidade de *f* erros em *x* amostras da seguinte forma (1):

$$P\ (f \text{ erros em } x \text{ amostras}) = (f^x)p^f q^{x-f} \qquad (1)$$

Onde: $p = P\ [de\ erro]$

$q = P\ [de\ não\ cometer\ erro]$

No caso de não ser cometido erro na amostragem, a equação fica:

$$P\ (0 \text{ erros em } x \text{ amostras}) = q^x. \qquad (2)$$

As modelagens realizadas pelo autor diziam respeito a pontos amostrais e a possibilidade de cometer ou não erros, dada pela probabilidade de rejeição da amostragem. Trabalhando questões semelhantes e citando os trabalhos de Rosenfield et al. e Genderen et al., Richards e Jia (2006, p. 306) sustentam que, "na prática, pode-se escolher entre 30 e 60 amostras por categoria".

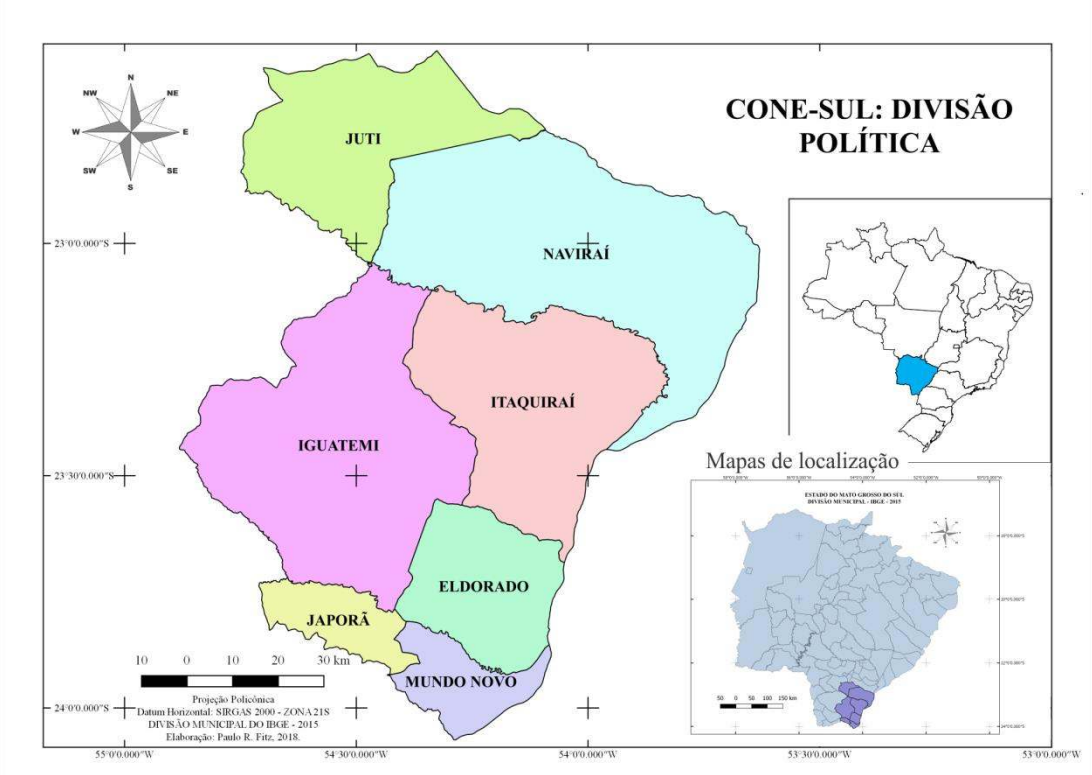
Mesmo que a proposta do presente estudo seja a de utilizar polígonos de amostragem, ou áreas de treinamento, será utilizada a modelação proposta por Genderen et al. (1978, p. 7). Neste sentido, Aronoff (1985, p. 101) alerta para o fato de que, "dependendo da localização geográfica e do tamanho do estudo, a forma e o tamanho dos polígonos individuais do mapa podem mostrar tendências significativas."

Isto posto, ressalta-se que o número e o tamanho dos polígonos de amostragem surgem como principais indicadores do refinamento da divisão em classes de uma imagem. O presente trabalho procurou, portanto, analisar a área de estudo a partir da fixação de uma quantidade razoável de polígonos de amostragem necessários para uma boa compreensão da área trabalhada e sua representação cartográfica.

Nesta investigação foram utilizadas imagens da região do Cone-Sul, dentro do Estado do Mato Grosso do Sul. Esta nomenclatura diz respeito ao "Estudo da Dimensão Territorial do Estado de Mato Grosso do Sul: Regiões de Planejamento", de janeiro de 2015. (MATO GROSSO DO SUL, 2015). A Figura 1 apresenta o mapa de localização da região utilizada para o estudo[1].

[1] Na realidade, foi utilizado um retângulo máscara que abrangia a região do Cone-Sul e seu entorno, uma vez que as feições existentes na natureza não respeitam a divisão política estabelecida.

Figura 1 – Mapa de localização



Fonte: Adaptado da Divisão Municipal IBGE, 2018.

A hipótese presumia que deveria haver uma tendência de estabilização nos dados de acordo com o incremento de áreas de treinamento dadas por polígonos de amostragem. Assim sendo, no limite, um número infinito de polígonos representaria com exatidão as diferentes classes. Este estudo buscou, por conseguinte, analisar os dados obtidos através da relação entre as dimensões das áreas classificadas com a crescente quantidade de polígonos de amostragem.

METODOLOGIA

A execução do trabalho se deu a partir do descarregamento gratuito dos arquivos com as imagens de uma região do sul do Estado do Mato Grosso do Sul desde os *sites* da *USGS (United States Geological Survey* – Serviço Geológico dos Estados Unidos) e do INPE (Instituto Nacional de Pesquisas Espaciais) do Brasil. Optou-se pelo uso das imagens do instrumento TM (*Thematic Mapper*) do *Landsat* 5, de 22/08/1985, e do sensor OLI (*Operational Land Imager*) do *Landsat* 8, de 29/08/2018, uma vez que ambos possuem várias faixas ou bandas espectrais com comprimentos de onda

semelhantes e idênticas resoluções espaciais. O quadro 1 apresenta uma síntese dos sensores utilizados e algumas de suas características.

Quadro 1 – Sensores e suas características

SATÉLITE	SENSOR	BANDAS ESPECTRAIS	RESOLUÇÃO ESPECTRAL	RESOLUÇÃO ESPACIAL
LANDSAT 5	TM	AZUL (B1)	0,45 – 0,52 µm	30m
		VERDE (B2)	0,52 – 0,60 µm	
		VERMELHO (B3)	0,63 – 0,69 µm	
		IVP (B4)	0,76 – 0,90 µm	
LANDSAT 8	OLI	AZUL (B2)	0,45 – 0,51 µm	30m
		VERDE (B3)	0,53 – 0,59 µm	
		VERMELHO (B4)	0,64 – 0,67 µm	
		IVP (B5)	0,85 – 0,88 µm	

Fonte: https://landsat.gsfc.nasa.gov; http://www.dgi.inpe.br. Acessos em 22/01/2018 e 23/10/2018.

As imagens foram reprojetadas para o sistema de referência atualmente em uso no Brasil, o SIRGAS 2000. Como uma só imagem não contemplava toda a área de estudo, foi agregada a imagem vizinha para a constituição de um mosaico que englobasse sua totalidade. Assim, foram utilizados os pontos 76 e 77 da órbita 224.

De posse das imagens, e com o uso do software QGIS versão 2.18[2], foram realizadas composições coloridas *falsa-cor* para posterior classificação das mesmas. Basicamente foram utilizadas combinações 4R-3G-2B para o ano de 1985 (*Landsat 5*) e 5R-4G-3B para o ano de 2018 (*Landsat 8*). Escolheu-se dois anos diferentes contemplando a mesma região de estudo para confrontação dos resultados e observação de possíveis discrepâncias entre os dados.

Os principais aspectos do território devem ser trabalhados de acordo com a escala pretendida pelo usuário, caso se faça necessária uma representação temática impressa desta superfície. É importante o leitor ter em consideração que, como se está trabalhando com imagens contemplando 30 x 30 pixels, a escala de trabalho deverá se ajustar a este padrão. Neste sentido, para o que diz respeito à possível impressão de um produto (mapa) dever-se-á trabalhar com o conceito da GSD (*Ground Sample Distance*)

[2] Disponível para download em: https://www.qgis.org

ou Distância Correspondente do Terreno, a qual, segundo Fitz (2008, p. 97) "refere-se ao tamanho real (no terreno) de um determinado *pixel* com relação à resolução de uma imagem e à sua escala". Conforme o autor, GSD = N/R, onde N = denominador da escala e R = Resolução da imagem, em dpi. Trabalhando-se com 300 dpi, valor indicado pelo mesmo autor como mínimo a ser utilizado, e com GSD = 30m, o próprio valor do pixel, o qual deve ser transformado para polegadas, isto é, aproximadamente, 1.181", pode-se estabelecer que N = GSD . R, ou seja, N = 1.181" . 300 dpi = 354.300. Consequentemente, se fosse o caso, o mapa resultante a ser impresso não poderia possuir escala maior do que 1:354.000. Torna-se oportuno ressaltar que o intuito do presente trabalho não é o de produzir material gráfico impresso e, neste sentido, esta demanda torna-se dispensável.

Quanto às classes de cobertura da terra, pode-se citar Lang (2009, p. 41) que afirma que:

> A segmentação de imagens em multiescalas objetivando um sistema hierarquizado de objetos relacionados entre si, apresenta nítidos paralelos ao conceito de holismo. Lá como cá a característica essencial está numa discretização da informação altamente complexa em objetos elementares, os quais, num entrelaçamento de relações estruturadas de forma relativamente simples, formam diferentes planos, uma estrutura.

Neste ponto, pode-se mencionar o Manual Técnico de Uso da Terra de 2013 do IBGE – Instituto Brasileiro de Geografia e Estatística – que exibe o "Sistema básico de classificação da cobertura e do uso da terra – SCUT". Nele, a cobertura da terra é entendida como "os *elementos da natureza como a vegetação (natural e plantada), água, gelo, rocha nua, areia e superfícies similares, além das construções artificiais criadas pelo homem, que recobrem a superfície da terra*." (IBGE, 2013, p. 44. Grifo dos autores). Adiante, o manual destaca que dados de sensoriamento remoto podem ser correlacionados com a cobertura da terra e mapear o tema a partir da interpretação dos dados a ela relativos.

Utilizando imagens de média resolução, optou-se pelo uso de apenas três classes de uso da Terra e cobertura do solo. Estas foram escolhidas em função de suas características bastante visíveis nas imagens *Landsat* com resolução espacial de 30m.

Salienta-se, entretanto, que determinadas feições podem sofrer mascaramentos em função de suas características:

- classe *corpos d'água*, que inclui todos os conjuntos de feições que caracterizem acumulação de água, como rios, lagos, açudes;
- *áreas antropizadas*, que dizem respeito às feições reconhecidas como aquelas alteradas pela ação do homem tais como áreas urbanas, de pastagem, com agricultura; e
- *cobertura vegetal "original"*, que engloba, além da vegetação nativa composta principalmente por mata atlântica e de várzea, porções com plantações de florestas de espécies exóticas, uma vez que, nesta escala, torna-se impraticável distinguir a introdução de tais matas.

Salienta-se, outrossim, que certas feições como corpos d'água de menor expressão, dotados de mata ciliar, sequer são percebidos como tal, uma vez que, em geral, são acobertados pela vegetação mais densa. Áreas úmidas também poderão ser confundidas em função de respostas espectrais semelhantes, ora da vegetação, ora de cursos d'água, por exemplo.

As classificações foram realizadas a partir do algoritmo *dzetsaka* disponibilizado pelo QGIS. Trata-se de um método de classificação supervisionada desenvolvido por Nicolas Karasiak[3], o qual utiliza um classificador GMM (*Gaussian Mixture Model*), ou Modelo de Mistura Gaussiana. Conforme Portela (20015, p. 31), esses modelos podem ser entendidos como:

> famílias de distribuições formadas pela composição de mais de uma distribuição básica. Esses modelos são formados pela distribuição de probabilidade $D1$ com probabilidade $w1$, pela distribuição de probabilidade $D2$ com probabilidade $w2$, etc.. No caso de uma mistura de Gaussianas, grupos são descritos estatisticamente pelos parâmetros média, μ, e matriz de covariância, Σ.

Como base de análise, foram utilizadas as composições falsa-cor 4R-3G-2B para o ano de 1985 (*Landsat* 5) e 5R-4G-3B para o ano de 2018 (*Landsat* 8). As áreas

[3] https://plugins.qgis.org/plugins/dzetsaka/

das três classes, para cada simulação, foram calculadas através do *plugin r.report* do QGIS, uma vez que o propósito deste trabalho era o de relacionar as áreas das classes dos mapas temáticos gerados com a quantidade de polígonos utilizados como amostragem. Para isto, os resultados obtidos foram exportados e retrabalhados em uma planilha *Excel*. A partir da planilha foram confeccionados gráficos relativos a cada ensaio conforme as três classes definidas em cada ano examinado.

Seguindo os preceitos da hipótese inicial de trabalho, foram realizadas diversas experimentações com o intuito de verificar sua validade em termos de quantidade de áreas de treinamento necessárias para a obtenção de bons resultados em uma classificação. Três usuários, com diferentes percepções quanto à fisionomia da região de estudo, efetuaram distintos ensaios com imagens de uma mesma região dos anos de 1985 e de 2018. As áreas de treinamento foram delimitadas por polígonos de amostragem que pudessem representar da melhor forma possível a classe escolhida. Cada conjunto de polígonos representativos deveria englobar mais do que trinta pixels.

Num primeiro momento, os usuários realizaram uma classificação com um conjunto contemplando um polígono relativo a cada classe escolhida, a saber, *corpos d'água*; *áreas antropizadas*; e *áreas com cobertura vegetal "original"*. Em seguida, foram agregados mais três polígonos, um para cada uma destas classes, totalizando seis áreas de treinamento (três clusters com dois polígonos). Dessa maneira, os polígonos representativos foram sendo adicionados um a um conforme as classes predeterminadas e cada novo grupo apresentavam três novas áreas de treinamento que se somavam às anteriores. Fugindo ao recomendado, este procedimento foi realizado, num primeiro momento, até completarem-se vinte simulações, dadas por vinte clusters com três classes cada, num total de sessenta áreas de treinamento. Somente após a constatação da ineficácia de uso dessa quantidade de áreas de treinamento é que foram sendo agregados novos polígonos até atingir um total de sessenta clusters, cada um com três classes, ou seja, cento e oitenta áreas de treinamento.

RESULTADOS, ANÁLISES E DISCUSSÕES

Os resultados obtidos variaram de acordo com a escolha das áreas de treinamento consideradas por cada usuário. As maiores variações se deram quando as amostras eram completamente aleatórias, ou seja, quando não havia correspondência

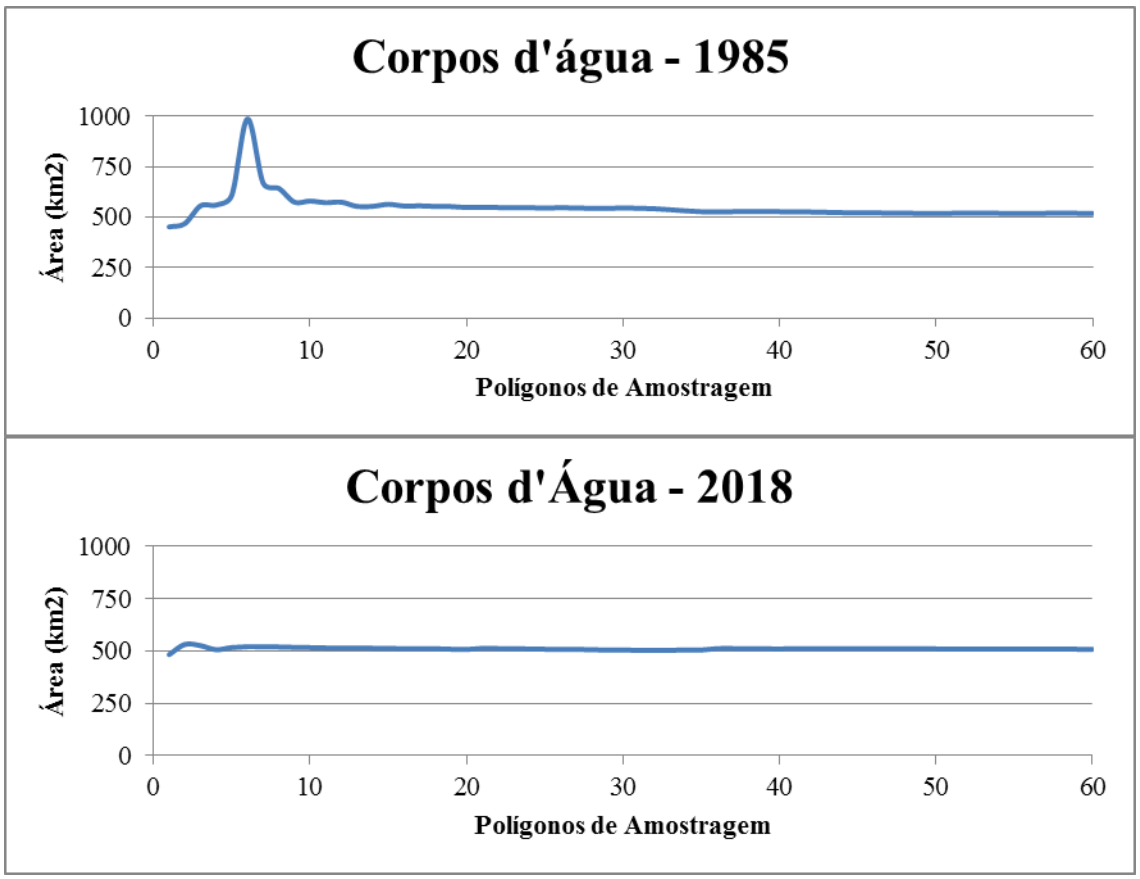
entre os polígonos utilizados inicialmente com os seguintes e assim por diante.

As primeiras simulações levaram a novas conjecturas, uma vez que foram percebidas intensas perturbações nos dados iniciais dos gráficos gerados e de pequenas tendências a partir de então. Como exposto anteriormente, a proposta inicial de vinte polígonos de amostragem, que fugia às recomendações, foi aumentada para trinta, depois para quarenta e, finalmente, para um total de sessenta áreas de treinamento por classe, ou cluster de polígonos, seguindo os apontamentos já citados. Foram confeccionados um total de cento e oitenta polígonos de amostragem para cada ano. Os gráficos decorrentes de um dos experimentos podem ser observados pelas figuras a seguir que apresentam os dois anos analisados – 1985 e 2018 – para que o leitor tenha uma ideia da dinâmica verificada.

A figura 2 permite constatar que os comportamentos das curvas representativas para ambos os anos são bastante similares, especialmente a partir do décimo polígono de amostragem, quando há uma tendência de estabilização da curva no valor de cerca de 500 km^2 de área[4]. As perturbações iniciais mais acentuadas no gráfico relativo ao ano de 1985 devem-se a ajustes quanto à agregação de corpos d'água providos ou não de sedimentos. A estabilização da curva, para ambas as datas, se deu somente a partir, aproximadamente, do trigésimo quinto polígono de amostragem, mais notoriamente para o ano de 1985.

[4] Valores arredondados, próximos aos valores reais encontrados.

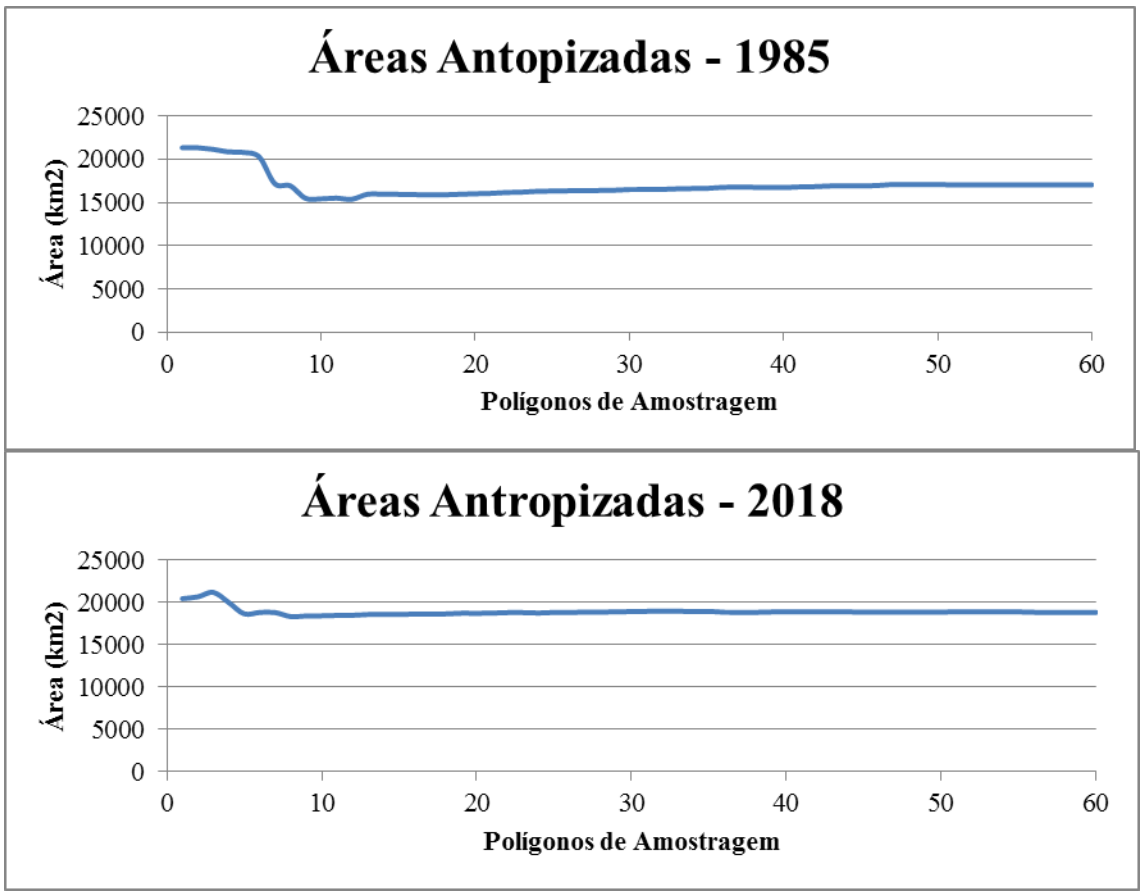
Figura 2 – Áreas Mapeadas como Corpos d'água – 1985 e 2018



Fonte: Autoria própria, 2019.

Assim como na classe corpos d'água, as áreas tidas como "antropizadas" (figura 3) apresentaram uma maior variação inicial, até quando do uso de cerca de dez polígonos/áreas de treinamento. A partir daí ocorreu uma certa estabilização com uma leve tendência de incremento de áreas observada de maneira mais consistente no ano de 1985, quando os valores se aproximaram de 17.000 km^2 de área. A real estabilização da curva para este ano, entretanto, só ocorreu a partir do trigésimo quinto polígono. O ano de 2018 também apresentou um ligeiro crescimento a partir do oitavo polígono de amostragem para experimentar uma estabilização a partir da vigésima segunda simulação, quando atingiu um valor em torno de 19.000 km^2 de área.

Figura 3 – Áreas Mapeadas como Antropizadas – 1985 e 2018

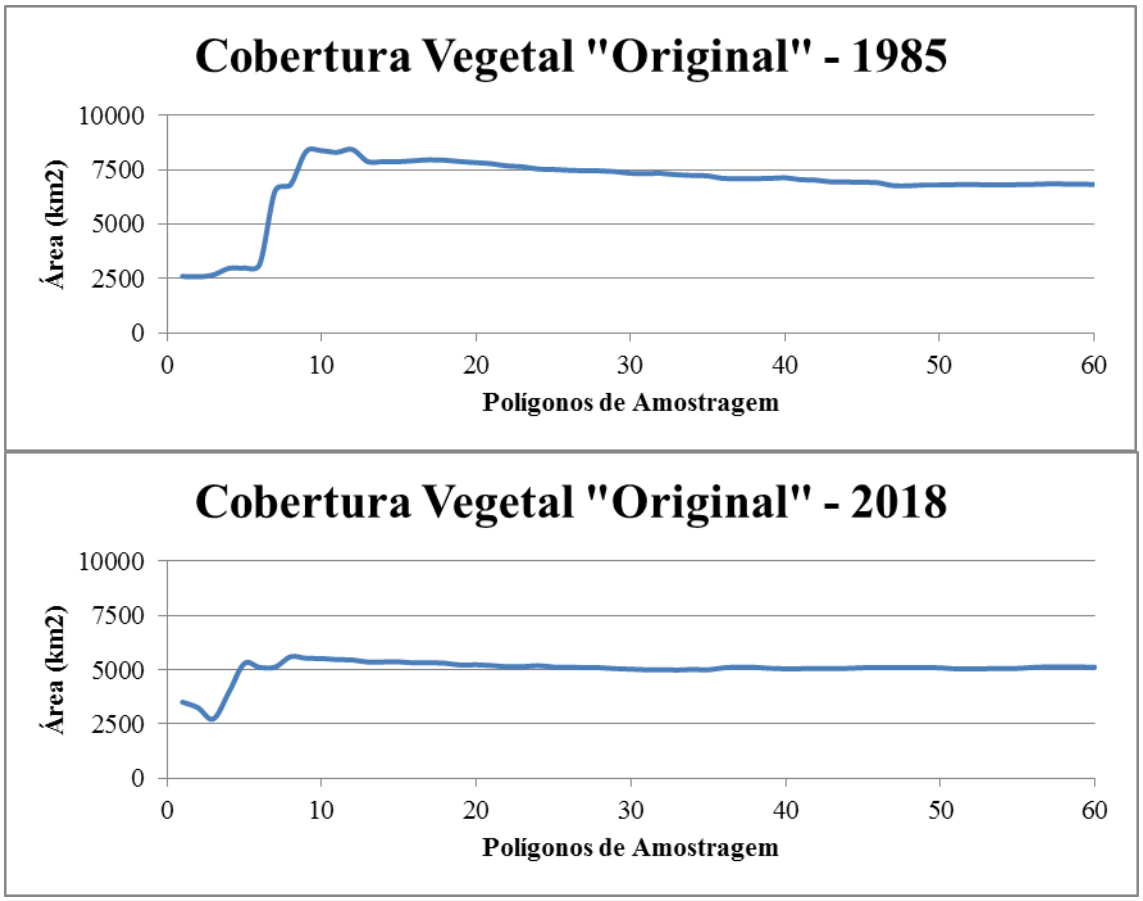


Fonte: Autoria própria, 2019.

A Cobertura Vegetal "Original" (figura 4) foi, sem dúvida, a classe que apresentou as maiores alterações em ambos os anos trabalhados. Um certo equilíbrio só começou a ser observado a partir do décimo segundo polígono de amostragem e, mesmo assim, com algumas variações. A estabilização, para o ano de 1985, se deu somente a partir do quinquagésimo polígono, onde os valores de área consolidaram-se em torno de 6.800 km^2. Novamente percebeu-se uma harmonização mais evidente em 2018, praticamente a partir do oitavo polígono de amostragem, com seus valores estabilizando-se, de fato, em torno de 5.100 km^2, a partir da trigésima oitava simulação realizada. Tais variações podem estar relacionadas às características das imagens quando de suas aquisições. Uma quantidade maior ou menor de umidade na superfície, por exemplo, implica em respostas espectrais podem apresentar padrões assemelhados que podem confundir as classificações, as quais tendem a se estabilizar somente a partir

de um grande número de simulações.

Figura 4 – Áreas Mapeadas como Cobertura Vegetal "Original" – 1985 e 2018



Fonte: Autoria própria, 2019.

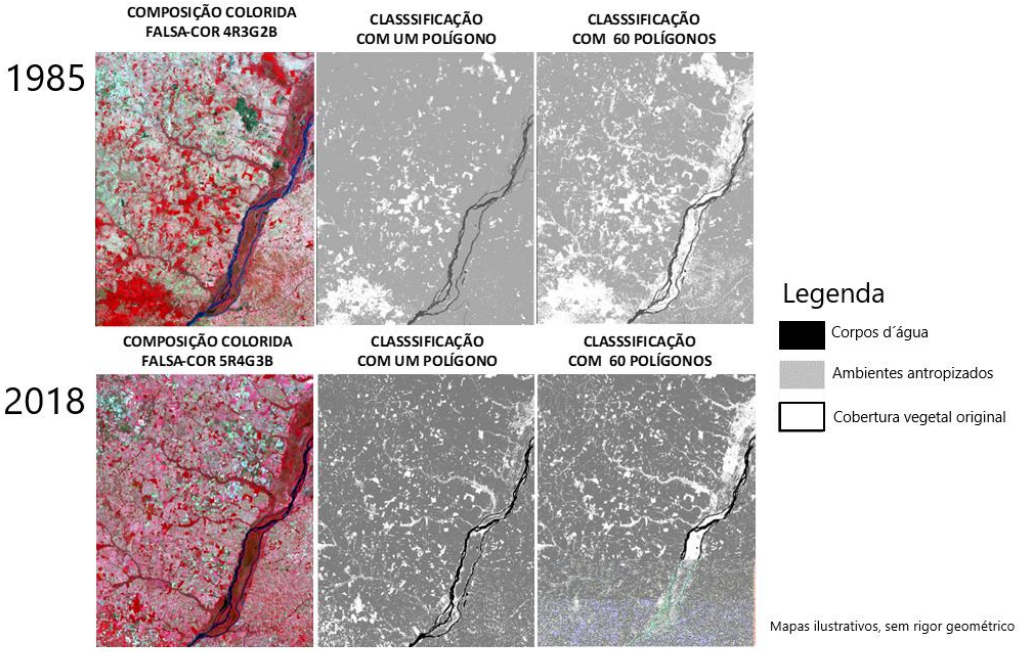
De um modo geral, como pode ser observado pelas figuras, até a simulação com o uso de cerca de dez polígonos de amostragem por classe, ocorreram variações significativas em termos de áreas das classes para ambas as datas pré-selecionadas. A partir daí, houve uma tendência de harmonização nos valores, porém com ligeiras variações. Conforme são inseridas novas áreas de treinamento, determinadas feições cujas respostas espectrais são assemelhadas, podem trocar de classe em função da agregação ou desagregação de elementos nas mesmas.

Assim sendo, o uso de poucas áreas de treinamento pode refletir imprecisões no sentido de mascarar determinadas feições. As variações iniciais se devem, portanto, a ajustes cognitivos que vão sendo assimilados pelo usuário ao longo

das simulações. Com o andar da modelagem, feições ligeiramente diferenciadas vão sendo agregadas e contribuem na acurácia do produto.

As classificações realizadas apresentaram resultados ligeiramente diferenciados de acordo com a percepção de cada usuário, ou seja, em conformidade com suas escolhas frente a diferentes polígonos de amostragem. Entretanto, conforme a quantidade de amostras crescia, havia uma tendência de aproximação dos resultados. O principal problema registrado deu-se em função do conflito existente entre áreas de várzea e ambientes antropizados. A figura 5 a seguir retrata a interpretação realizada a partir das composições coloridas com 1(um) e 60 (sessenta) polígonos, respectivamente.

Figura 5 – Mapas de Referência – 1985 e 2018



Fonte: Autoria própria, 2019.

Os questionamentos a serem realizados dizem respeito, portanto, às diferentes percepções dos usuários com relação às feições classificadas. Nestas condições, os valores relativos às áreas quando do uso de poucos polígonos de amostragem podem variar substancialmente.

A agregação de diferentes feições quando da escolha das áreas de treinamento também sugere desvios na interpretação geral das classes. Assim, um

polígono que engloba um corpo d'água que apresenta uma quantidade maior de sedimentos, em dada localização, terá uma reflectância diferente de outro que se apresenta mais límpido. Conforme são agregados novos polígonos de amostragem, feições como corpos d'água, com respostas espectrais diferenciadas, tendem a ajustar-se frente a outros elementos que apresentam semelhante resposta espectral mediante o algoritmo utilizado (modelagem gaussiana).

O incremento de polígonos em diferentes áreas de treinamento tende a harmonizar tais ocorrências e, no limite, encaminha-se para o estabelecimento de um equilíbrio. Observou-se que, conforme eram acrescentados polígonos de amostragem com pequenas distinções com relação aos anteriores, as classificações adaptavam-se à nova realidade.

Algumas das variações (máximos e mínimos) observadas nos gráficos se devem, portanto, a tais cenários. Como exemplo, pode-se citar o caso de áreas baixas e sujeitas à inundação, em preparo para cultivo, que apresentam reflectâncias idênticas às várzeas de cursos d'água. Assim sendo, algumas feições entendidas em dada categoria, no decorrer das diversas inferências, vão perdendo relevância em detrimento de outras e acabam sendo representadas em classes distintas.

Destarte, não se pode relevar questões como as ligadas ao regime pluviométrico da região trabalhada. A anterior presença ou ausência de precipitações antes da coleta das imagens pode gerar um mascaramento do contexto.

Percebe-se que, no limite, o número de polígonos exigidos para uma classificação ideal dirige-se ao infinito. Isto significa que somente teríamos uma "realidade" se a imagem fosse integralmente vetorizada. Entretanto, pôde-se notar que, a partir do décimo polígono, ocorre uma certa equalização da curva representativa, mas ainda com algumas tendências como fora comentado. Esta condição deve-se à inclusão ou exclusão de elementos na área de treinamento em análise conforme estes vão sendo associados ou não a dada classe.

Para que se tenha uma ideia da confiabilidade dos dados apresentados, foram realizados cálculos relativos à acurácia das interpretações com relação ao tamanho das amostras. Para tal, foi utilizada a formulação proposta por Genderen (1978), comentada anteriormente, e foi estabelecido que as amostras estavam corretas, ou seja, $P\ (0\ erros\ em\ x\ amostras) = q^x$.

A tabela 1 apresenta os resultados de probabilidades de não haver erros de interpretação, ou a probabilidade de rejeição, de 85%, 90%, 95% e 99% de precisão, para amostras de 10,20, 30, 40, 50 e 60 polígonos.

Tabela 1 – Probabilidade de não haver erros (rejeição) em amostras de tamanhos variados.

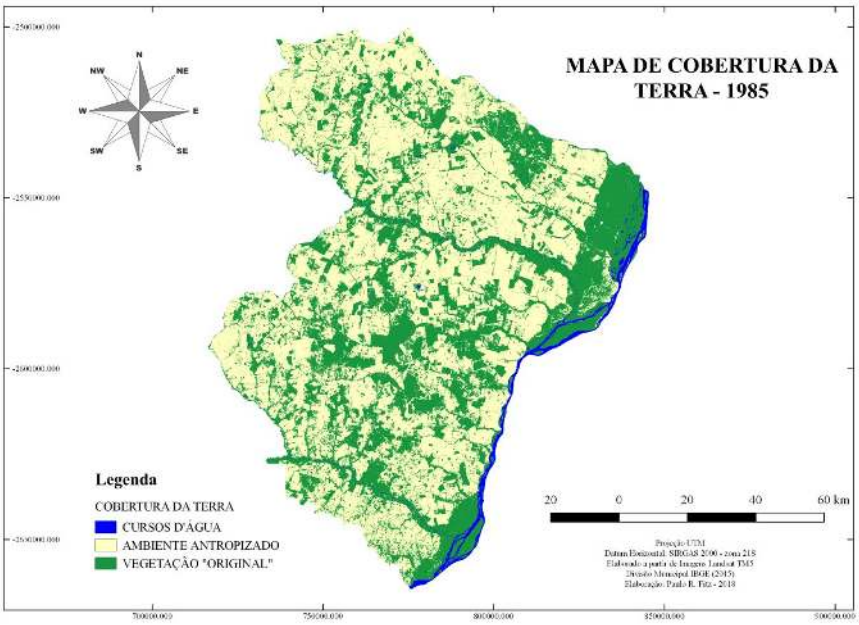
PRECISÃO DE INTERPRETAÇÃO	AMOSTRAS					
	10	**20**	**30**	**40**	**50**	**60**
0,99	0,904382	0,817907	0,739700	0,668972	0,605006	0,547157
0,95	0,598737	0,358486	0,214639	0,128512	0,076945	0,046070
0,90	0,348678	0,121577	0,042391	0,014781	0,005154	0,001797
0,85	0,196874	0,038760	0,007631	0,001502	0,000296	0,000058

Fonte: Adaptado de Genderen (1978).

Como pode ser observado pela tabela, conforme vão sendo acrescentadas áreas de treinamento, a probabilidade de rejeição dos resultados decresce. Por conseguinte, para uma precisão de 95%, teríamos uma rejeição de 59,87% no caso de 10 amostras e de 4,6% para 60 amostras. Assim sendo, em sessenta amostras teríamos menos de três que não pertenceriam à respectiva classe.

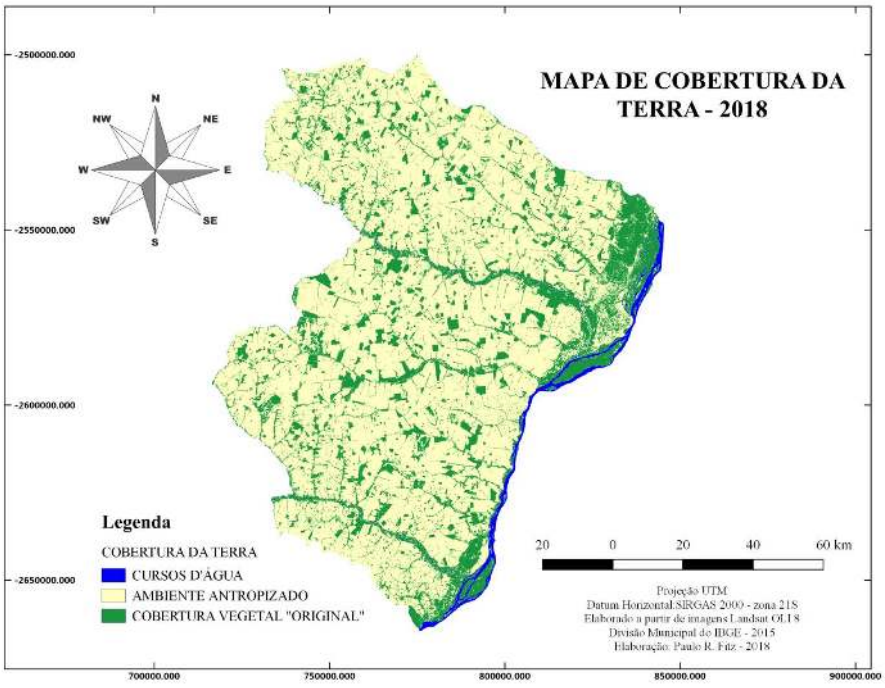
Destaca-se, entrementes, que o intuito deste trabalho não era o de analisar questões como as relativas à expansão das ações antrópicas na região. No entanto, esta relação fica clara a partir das considerações visualizadas pelos gráficos anteriores, bem como pelas figuras 6 e 7 a seguir que contemplam especificamente a região do Cone-Sul:

Figura 6 – Mapa de Cobertura da Terra – 1985.



Fonte: Autoria própria, 2019.

Figura 7 – Mapa de Cobertura da Terra – 2018.



Fonte: Autoria própria, 2019.

CONSIDERAÇÕES FINAIS

As simulações realizadas confirmaram a hipótese apresentada, a saber, a de que haveria uma tendência de estabilização nos dados de acordo com o incremento de áreas de treinamento. Entretanto, conforme ocorria o acréscimo de polígonos representativos, observaram-se algumas pequenas tendências nas classes escolhidas.

Deve-se considerar que, ao se eleger as classes, determinadas áreas de treinamento podem sofrer pequenos ajustes ao longo do processo. Esta condição diz respeito à percepção de cada usuário e às diferentes apresentações das feições trabalhadas. Assim, na medida em que se extraem áreas de treinamento com respostas espectrais ligeiramente diferenciadas, os resultados vão se alterando e se ajustando ao propósito final: um mapa temático.

O principal desafio, quando da classificação de uma imagem diz respeito, portanto, ao cuidado com a escolha das áreas de treinamento, que está diretamente relacionado à cognição de cada usuário. Dessa maneira, em função do entendimento individual, alguns polígonos englobarão pixels cujas respostas espectrais poderão pertencer a distintas classes conforme sua própria percepção. O conhecimento prévio da área de estudo torna-se fundamental para a realização de uma boa classificação. Saídas a campo para checagem das informações tornam-se, portanto, imprescindíveis para um refinamento da modelagem.

Uma maior quantidade de áreas de treinamento, dadas por polígonos de amostragem, tende a minimizar erros e aumentar a acurácia da modelagem. Conforme mencionado, um número tendendo ao infinito geraria, hipoteticamente, uma classificação precisa.

Como sugestão para futuras classificações, sugere-se a adoção de, pelo menos, cinquenta polígonos amostrais por classe, que configurem, da melhor forma possível, a região a ser classificada. Estas áreas de treinamento deverão abarcar todas as características relativas às classes a serem adotadas. Outra proposta diz respeito ao estabelecimento de um número maior de classes. Neste sentido, tem-se que mais classes implicarão em maiores detalhamentos, o que pode minimizar determinadas tendências observadas, como, por exemplo, a distinção entre áreas alagadiças e em preparo para cultivos.

Finalmente é importante lembrar que um mapa temático é uma

representação da configuração da superfície terrestre. Neste sentido, ele não deve ser visto como uma reprodução real da mesma, mas como uma interpretação o mais fidedigna possível.

REFERÊNCIAS

ALMEIDA, Flávio Gomes de; SOARES, Luiz Antônio Alves. **Ordenamento Territorial.** Rio de Janeiro: Bertrand Brasil, 2009.

ARONOFF, Stan. **The Minimum Accuracy Value as an Index of Classification Accuracy.** In: PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING, Vol. **51,** No. **1,** January 1985, p. 99-111.

CASTRO, Iná Elias de.; GOMES, Paulo Cesar da Costa; CORRÊA, Roberto Lobato. **Geografia: conceitos e temas.** Rio de Janeiro: Bertrand Brasil, 1995.

CLAVAL, Paul. **A Nova Geografia**. Coimbra, Portugal: Livraria Almedina, 1987

CONGALTON, Russel G. **A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data**. ***Remote sensing of environment 37.*** Elsevier North-Holland Inc., 1991. p. 35-46.

CONGALTON, Russel G.; GREEN, Kass. **Assessing the accuracy of remotely sensed data: principles and practices.** 2 ed. CRC Press: Boca Raton, 2009.

CORRÊA, Roberto Lobato. **Trajetórias Geográficas**. Rio de Janeiro: Bertrand Brasil, 2010.

CRÓSTA, Alvaro P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Campinas: Unicamp, 1993.

EASTMAN, J. Ronald. **IDRISI for Windows:** Introdução e exercícios tutoriais. Porto Alegre, RS: Centro de Recursos Idrisi Brasil – UFRGS, 1998.

FITZ, Paulo R. **Geoprocessamento sem complicação.** São Paulo: Oficina de Textos, 2008.

GENDEREN, J.L. van; LOCK, B. F.; VASS, P. A. **Remote Sensing: Statistical Testing of Thematic Map Accuracy.** ***Remote sensing of environment 7.*** Elsevier North-Holland Inc., 1978. p. 3-14.

GEORGE, Pierre. **Os métodos da geografia**. São Paulo: Difel, 1986.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Manual Técnico de Uso da Terra**. 3 ed. IBGE: Rio de Janeiro, 2013.

INPE. Instituto Nacional de Pesquisas Espaciais. **Imagens do satélite *Landsat***. Disponíveis em www.inpe.br. Acessos entre julho de 2018 e janeiro de 2019.

LACOSTE, Yves. **Geografia do Subdesenvolvimento.** 5 ed. Difel: Rio de Janeiro, 1978.

_____. **A Geografia, isto serve, em primeiro lugar, para fazer a guerra.** 6 ed. Papirus: Campinas, 2002.

LANG, Stefan. A Utilização do Conceito de Totalidade (Holismo) na Geração de Planos de Segmentação Regionalizados em Conjuntos de Imagens de Altíssima Resolução. In: BLASCHKE, Thomas; KUX, Hermann (org.). **Sensoriamento Remoto e SIG Avançados.** 2 ed. São Paulo: Oficina de Textos, 2009.

MATO GROSSO DO SUL. Semade. **Estudo da Dimensão Territorial do Estado de Mato Grosso do Sul: Regiões de Planejamento**. 2015. Disponível em: http://www.semade.ms.gov.br/wp-content/uploads/sites/20/2015/03/estudo_dimensao_territorial_2015.pdf. Acesso em 06/01/2018,

PORTELA, Nara Miranda. **Modelo de mistura de gaussianas fuzzy contextual**. Tese (Doutorado) – Universidade Federal de Pernambuco. UFPE: Recife, 2015.

RICHARDS, John A.; JIA, Xiuping. **Remote Sensing Digital Image Analysis:** An Introduction. 4 ed. Berlin: Springer-Verlag, 2006.

SANTOS, Milton. **Técnica, Espaço, Tempo: Globalização e meio técnico-científico informacional.** 4.ed. São Paulo: Hucitec, 1998.

_____. **A Natureza do Espaço.** 4.ed. São Paulo: EDUSP, 2006.

USGS. U.S. Geological Survey **Imagens do Satélite *Landsat***. Disponíveis em https://www.usgs.gov/. Acesso entre julho de 2018 e janeiro de 2019.

Recebido para publicação em março de 2019

Aceito para publicação em julho de 2019