



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

Tópicos em Observação da Terra

**FELIPE CARVALHO DE SOUZA
LUIZ FERNANDO FERREIRA GOMES DE ASSIS**

**ANÁLISE DE AMOSTRAS VOLUNTÁRIAS DE USO E
COBERTURA DO SOLO DA PLATAFORMA *OPEN
STREET MAP* UTILIZANDO MAPAS
AUTO-ORGANIZÁVEIS**

São José dos Campos

2019

Lista de ilustrações

Figura 1 – Arquitetura da rede SOM com a topologia hexagonal - Adaptado de Kohonen (2013)	9
Figura 2 – Metodologia adotada neste trabalho em diagrama - Adaptado de Santos et al. (2019)	10
Figura 3 – Área de estudo	11
Figura 4 – Visualização em coordenadas paralelas da série temporal de cobertura do solo utilizada neste trabalho	12
Figura 5 – Erro de quantização das topologias avaliadas	13
Figura 6 – Visualização em coordenadas paralelas em matriz produzida pela ferramenta <i>ggsom</i> para a rede SOM 20x20	14
Figura 7 – Matriz de confusão	16
Figura 8 – Porcentagem de confusão entre clusters	16

Lista de abreviaturas e siglas

VGI *Volunteered geographic information*

SOM *Self organizing-maps*

BMU *best match unit*

Sumário

1	INTRODUÇÃO	4
2	FUNDAMENTAÇÃO TEÓRICA	6
2.1	Informação Geográfica Voluntária	6
2.2	<i>Open Street Map</i>	8
2.3	Mapas Auto-organizáveis	8
3	MATERIAIS E MÉTODOS	10
3.1	Área de estudo	11
3.2	Dados	12
4	RESULTADOS	13
4.1	Avaliação visual com a ferramenta <i>ggsom</i>	13
4.2	Avaliação estatística com o pacote SITS	15
5	CONCLUSÃO	17
	REFERÊNCIAS	18

1 Introdução

Em consequência do crescimento populacional e de mudanças climáticas, o monitoramento de mudança de cobertura e uso do solo é fundamental para assegurar a qualidade do meio ambiente e melhorar a gestão de recursos naturais (GIULIANI et al., 2017). Aplicações de monitoramento do solo requerem observações frequentes, como fenologia da vegetação, resposta a desastres naturais e qualidade da água (YANG et al., 2017).

Com as recentes políticas de dados abertos de agências governamentais e espaciais, ampliou-se a facilidade de aquisição de imagens de observação da terra em grande volume e frequência, provindos de diferentes satélites, como Landsat-8, Sentinel-2 e CBERS. Assim, torna-se cada vez mais necessária a extração de informações dessa ampla quantidade de imagens. No entanto, imagens oriundas de satélites possuem características que dificultam sua manipulação, por exemplo, bandas com diferentes resoluções espaciais, séries temporais irregulares e projeções cartográficas desiguais em cada parte do mundo.

Uma forma de solucionar as complexidades apresentadas anteriormente é o uso de cubo de dados de observação da terra, o qual possui uma interface simples e intuitiva, com o objetivo de proporcionar uma estrutura de dados organizada para o usuário final (APPEL; PEBESMA, 2019). Lu, Appel e Pebesma (2018) definem cubo de dados como uma matriz de quatro dimensões: X (longitude), Y (latitude), Z (tempo), e as bandas espectrais. Diversos trabalhos usam cubos de dados para realizar a seleção de amostras de uso e cobertura do solo (LEWIS et al., 2017), (SANTOS et al., 2019). Essas arquiteturas, no entanto, sofrem de transferibilidade e generalização limitados quando aplicados em diferentes áreas e tempos (ZOU; ZHONG, 2018). Até agora, esse tem sido um dos principais obstáculos para tirar o máximo proveito do enorme volume e diversidade de dados de sensoriamento remoto disponíveis para o monitoramento de uso e cobertura da Terra (ASSIS et al., 2018a).

Nesse sentido, as iniciativas de informações geográficas de crowdsourcing (CGI) (NEWMAN et al., 2012; BONNEY et al., 2016) podem ser um fator-chave para alavancar o monitoramento do uso e cobertura da Terra por meio da análise de cubos de dados. Exploradas com sucesso em uma infinidade de aplicações, variando de monitoramento da biodiversidade (SULLIVAN et al., 2014; RADDICK et al., 2013) a gerenciamento de desastres (ASSIS et al., 2018b), essas iniciativas podem fornecer aos cientistas um conhecimento espacial valioso e local que seria de outra forma caro para obter. Além disso, eles podem beneficiar diretamente a sociedade nos casos

em que os cidadãos também são receptores das informações obtidas por cientistas profissionais (HAKLAY, 2010). Poucas iniciativas, no entanto, envolvem voluntários na produção de dados de uso e cobertura da Terra (FRITZ et al., 2009; ARSANJANI et al., 2013; SCHULTZ et al., 2017), servindo principalmente como conjuntos de dados de treinamento e validação para geração de mapas automatizados de classificação de grandes áreas (FONTE et al., 2017a; LEINENKUGEL et al., 2019; HERFORT et al., 2019). Além disso, eles podem evitar certos vieses em relação aos interesses de campanhas específicas, refletidos pela existência de diferentes sistemas de classificação (FRIEDL et al., 2010; TEAM, 2017; GLOBCOVER, 2017).

Posto que avanços científicos têm se realizado com uso do cubo de dados, o objetivo deste trabalho é avaliar as amostras de cobertura e uso do solo fornecidas por voluntários com o uso de imagens de baixa resolução, a ferramenta de visualização de dados *ggsom*¹ e o pacote SITS².

O presente trabalho é composto pelas seguintes seções: A Seção 2 apresenta a fundamentação teórica, na qual compreende a explicação sobre os conceitos necessário para entender o presente trabalho. A Seção 3 apresenta os materiais e métodos, cuja seção compreende a área de estudo e os dados utilizados. A Seção 4 apresenta os resultados e discussões. Por fim, a Seção 5 apresenta as conclusões e sugestões para trabalhos futuros.

¹ <https://cran.r-project.org/web/packages/ggsom/index.html>

² <https://github.com/e-sensing/sits>

2 Fundamentação Teórica

Nesta seção é apresentado os conceitos fundamentais deste trabalho, tais como Informação Geográfica Voluntária, *Open Street Map* e Mapas Auto-organizáveis.

2.1 Informação Geográfica Voluntária

Ainda hoje, dados precisos de Observação da Terra (OBT) são coletados e disponibilizados na maioria das vezes por especialistas que trabalham em instituições de pesquisa e agências estaduais. No entanto, o número de publicações científicas e projetos de pesquisa que demonstram as vantagens e o potencial da integração de dados voluntários em programas de OBT está aumentando rapidamente (SEE et al., 2017). Do ponto de vista da sociedade, os principais fatores que impulsionam essa mudança de paradigma na OBT e em outras ciências relacionadas são: 1) tecnologias móveis e da Internet que permitem que as pessoas se expressem, interajam e produzam dados relevantes; 2) a quantidade crescente de pessoas com acesso a essas tecnologias e a capacidade de operá-las, assim como a presença dessas tecnologias em nosso cotidiano; e 3) as pessoas mais conscientes que concordam em fornecer dados que produzem de maneira passiva e ativa ao usar essas tecnologias. Esses dados servem a propósitos diferentes, por exemplo, pesquisa científica, medidas de segurança e otimização de serviços personalizados.

Do ponto de vista da pesquisa científica, a opinião da multidão e os dados que ela pode voluntariamente produzir e disponibilizar são cada vez mais valiosos devido aos seguintes fatores principais: 1) respondendo a muitas questões relevantes de pesquisa, e responder às necessidades atuais da sociedade requer a combinação de dados de diferentes fontes, 2) descobrindo as complexas relações entre questões socioeconômicas, culturais e ambientais; estão surgindo conexões mais fortes com fluxos dinâmicos de influência entre cientistas e a sociedade; e 3) o conhecimento teórico para lidar com esses dados grandes e com várias fontes, bem como as infra-estruturas de processamento de dados necessárias para isso, estão atualmente bem desenvolvidos e estabelecidos.

Até o momento, os voluntários conseguiram contribuir para o monitoramento do uso e cobertura do solo geralmente por: 1) fornecendo dados de uso e cobertura coletados in situ ou com base em conjuntos de dados auxiliares, 2) resolvendo desacordos entre os mapas oficiais de uso e cobertura em locais específicos ou 3) disponibilizando abertamente as amostras de sua própria campanha de validação de mapa de uso e cobertura da Terra. O primeiro caso, os dados de uso e cobertura da Terra são nor-

malmente coletados por meio de um aplicativo de smartphone que guia os usuários a pontos específicos, instrua-os a fotografar a paisagem circundante e solicita que eles respondam algumas perguntas. Um exemplo representativo deste aplicativo é o FotoQuest Go ¹ (MOORTHY et al., 2018). No segundo caso, os voluntários podem visualizar em imagens de alta resolução do Google Earth ou Bing, além de diferentes mapas de uso e cobertura da Terra autoritativos para áreas em que os mapas de múltiplas discordam em relação à verdadeira classe de uso e cobertura da Terra. É o caso de GeoWiki.org ² (FRITZ et al., 2009), pioneira em plataformas para o mapeamento de uso e cobertura da Terra assistido por voluntários. No terceiro caso, os voluntários podem manter online um mapa de uso e cobertura da Terra autoritativo ou produzido por si próprio para definir a quantidade e a distribuição espacial das amostras de validação por meio de um dos diferentes modelos de amostragem disponíveis (ou seja, aleatórios, estratificados etc.). Em seguida, visualizando as imagens do Google Earth e os dados do Open Street Map (OSM), os usuários podem definir a verdadeira classe de uso e cobertura da Terra para esses pontos de validação, validando assim consistentemente o mapa carregado. Os usuários podem compartilhar suas amostras de validação com outros usuários e a comunidade científica, se assim o desejarem.

LACO-WIKI (SEE et al., 2017) é a plataforma que possui esses recursos e permite que os usuários realizem essas operações por meio de uma interface gráfica amigável. O Collect Earth³ (BEY et al., 2016) também é um software livre e de código aberto desenvolvido recentemente para monitoramento de uso e cobertura da Terra. Possui uma interface de usuário para visualizar simultaneamente imagens multitemporais de diferentes sensores e imagens de alta resolução do Google Earth e Bing Maps. Ao visualizar essas imagens de várias fontes e multitemporais, junto com perfis inter e intra-anuais de índices de vegetação (i.e., NDVI, NDWI e EVI), os usuários podem identificar a classe uso e cobertura do solo de pontos de amostra definidos por meio de um projeto de amostragem estatisticamente consistente. As ferramentas para o monitoramento voluntário de uso e cobertura da Terra mencionado acima têm, portanto, os seguintes recursos principais: interfaces gráficas amigáveis do usuário, geração automática de conjuntos estatisticamente consistentes de amostras de validação, cálculo de métricas de análise de precisão das amostras validadas voluntárias e a possibilidade de validar as suas próprias, bem como mapas oficiais de uso e cobertura da Terra, por exemplo CORINE ⁴, UrbanAtlas ⁵, GlobeLand 30 ⁶, etc.

¹ <http://fotoquest-go.org/en/>

² <https://www.geo-wiki.org/>

³ <http://www.openforis.org/tools/collect-earth.html>

⁴ <https://www.umweltbundesamt.de/en/topics/soil-agriculture/land-use-reduction/the-corine-land-cover-clc-program>

⁵ <https://land.copernicus.eu/local/urban-atlas>

⁶ <https://www.gim-international.com/content/article/globeland30-a-milestone-in-earth-observation-and-open-access-espacial-informação>

2.2 Open Street Map

O OSM (*Open Street Map*) é o exemplo mais proeminente de uma plataforma para a produção participativa de informações geográficas. Ele permite que especialistas locais e voluntários envolvidos contribuam com informações geográficas que frequentemente não estão disponíveis em outros lugares e que seriam difíceis de obter. Esses voluntários entendem que a tarefa se refere não apenas a alimentar o aplicativo, mas também a melhorar a eficácia do aplicativo e sua curva de aprendizado sobre o tópico com experiência.

Iniciativas como OSM Land-Use está tentando envolver voluntários na produção de valiosos dados de uso e cobertura da Terra. No entanto, apesar de propor estratégias interessantes para o envolvimento dos cidadãos, as iniciativas não exploram completamente o potencial dos voluntários para produzir informações sobre o uso e cobertura da terra nem beneficiam diretamente os colaboradores.

As principais desvantagens desse projeto são as restrições aos métodos de garantia de qualidade (COMBER et al., 2013; FOODY et al., 2013; ARSANJANI et al., 2013) e seu engajamento limitado dos cidadãos (FRITZ et al., 2012) em algumas áreas. Esta última edição está relacionada à contradição *querer vs. ter que*. Quando os cientistas cidadãos são capazes de definir livremente o tópico e a área geográfica de suas contribuições, a produção de dados geralmente diminui de áreas de interesse para áreas menos relevantes e vibrantes. Por outro lado, a qualidade dos dados pode ser menos confiável se os voluntários não tiverem experiência e conhecimento local sobre tópicos e áreas definidos pelo sistema a serem pesquisados. A consistência temporal e da estrutura de dados (FONTE et al., 2017b) também são fatores que precisam ser resolvidos com firmeza pelos aplicativos de ciência cidadã.

2.3 Mapas Auto-organizáveis

O Mapa Auto-organizável (SOM, do inglês *Self Organizing-maps*) (KOHONEN, 1982) é uma rede neural não supervisionada que, usa o procedimento de aprendizado competitivo para mapear os vetores de entrada multidimensionais em uma grade bidimensional retangular ou hexagonal de baixa dimensão. Um mapa auto-organizado é, portanto, caracterizado pela formação de um mapa topográfico dos padrões de entrada, no qual as localizações espaciais dos neurônios na grade são indicativas das características estatísticas intrínsecas contidas nos padrões de entrada, originando o nome "mapa auto-organizado" (HAYKIN, 2010).

Em resumo, na rede SOM, os nós de saída competem entre si pelos vetores de entrada, e ao final de cada iteração é determinado o nó vencedor (BMU, do inglês

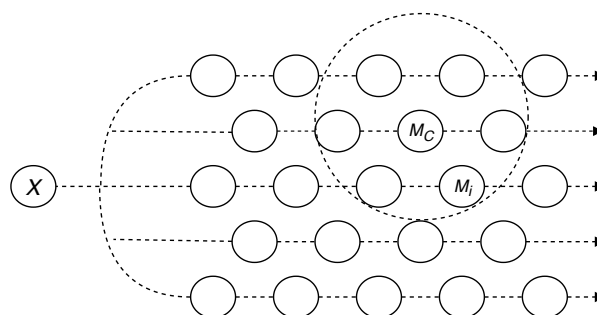


Figura 1 – Arquitetura da rede SOM com a topologia hexagonal - Adaptado de Kohonen (2013)

best match unit), aquele que possui a menor distância, comumente euclidiana, com o vetor de entrada. Após a escolha do BMU, todos os nós vizinhos de um determinado raio, atualizam seus valores, de forma a se aproximar do padrão escolhido (KOHONEN, 2013). A Figura 1 apresenta um exemplo de arquitetura do SOM com topologia hexagonal, na qual é possível observar o vetor de entrada X , neurônio vencedor M_c , o raio de vizinhança (círculo) e o neurônio atingido pela taxa de atualização M_i .

3 Materiais e Métodos

A estrutura metodológica deste trabalho é apresentada na Figura 2 em forma de fluxograma. Para realizar a avaliação de amostras, utilizou-se duas abordagens: análise visual com o uso da ferramenta *ggsom* e análise estatística com o pacote SITS (SIMOES et al., 2019). A ferramenta *ggsom* baseia-se em dois pacotes da linguagem de programação R: *Kohonen* (WEHRENS; BUYDENS et al., 2007) e *ggplot2* (WICKHAM, 2011). O pacote *Kohonen* é usado para treinar o SOM e o *ggplot2* para a criação do gráfico de coordenadas paralelas. Desta forma, a ferramenta *ggsom* opera como um utilitário entre os dois pacotes supracitados, de forma a modelar o dado gerado pelo *Kohonen* e visualizá-lo no *ggplot2*.

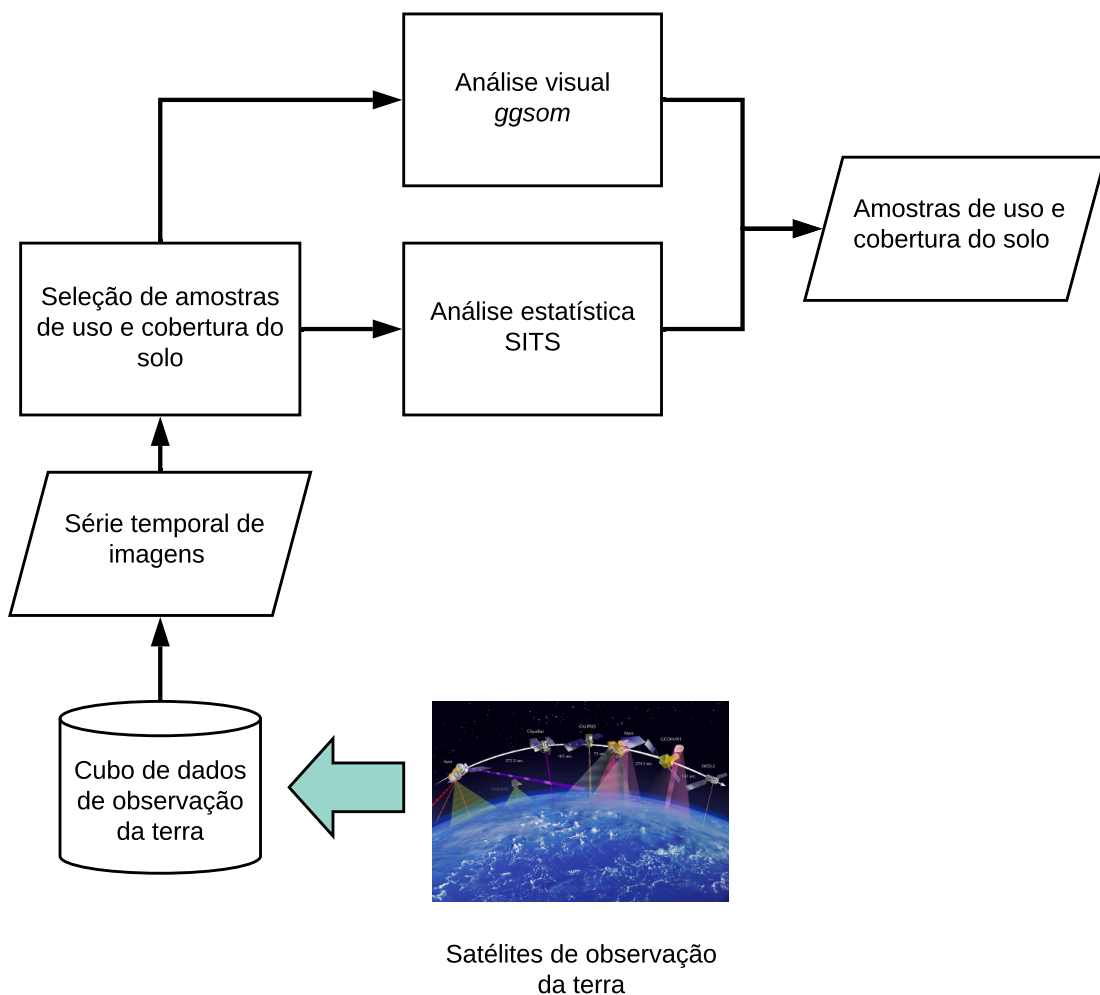


Figura 2 – Metodologia adotada neste trabalho em diagrama - Adaptado de Santos et al. (2019)

O pacote SITS implementa a metodologia proposta por Santos et al. (2019), na qual um *cluster* é formado por um ou mais neurônios que pertencem a mesma classe. Os neurônios são rotulados mediante a técnica de voto majoritário, onde é definido à priori um limiar de probabilidade para alterar o rótulo de cada neurônio. Desta forma, caso o neurônio não possua amostras mapeadas, atribui-se o rótulo de "NoClass".

3.1 Área de estudo

A área de estudo apresentada na Figura 3 compreende os estados do norte, nordeste e centro-oeste do Brasil, localizados na região da Amazônia Legal. A região de estudo foi escolhida com base no conjunto de 6335 amostras coletadas na plataforma *Open Street Map*, com as seguintes classes: Floresta, Pastagem e Cultura permanente. A paleta de cores foi definida manualmente de forma que as classes mais parecidas espectralmente compreendam cores mais próximas, assim, segundo o Manual Técnico de Uso da Terra (IBGE, 2013) formam-se dois grupos: Áreas de Vegetação Natural (Floresta) e Área Antrópicas Agrícolas (Pastagem e Cultura Permanente).

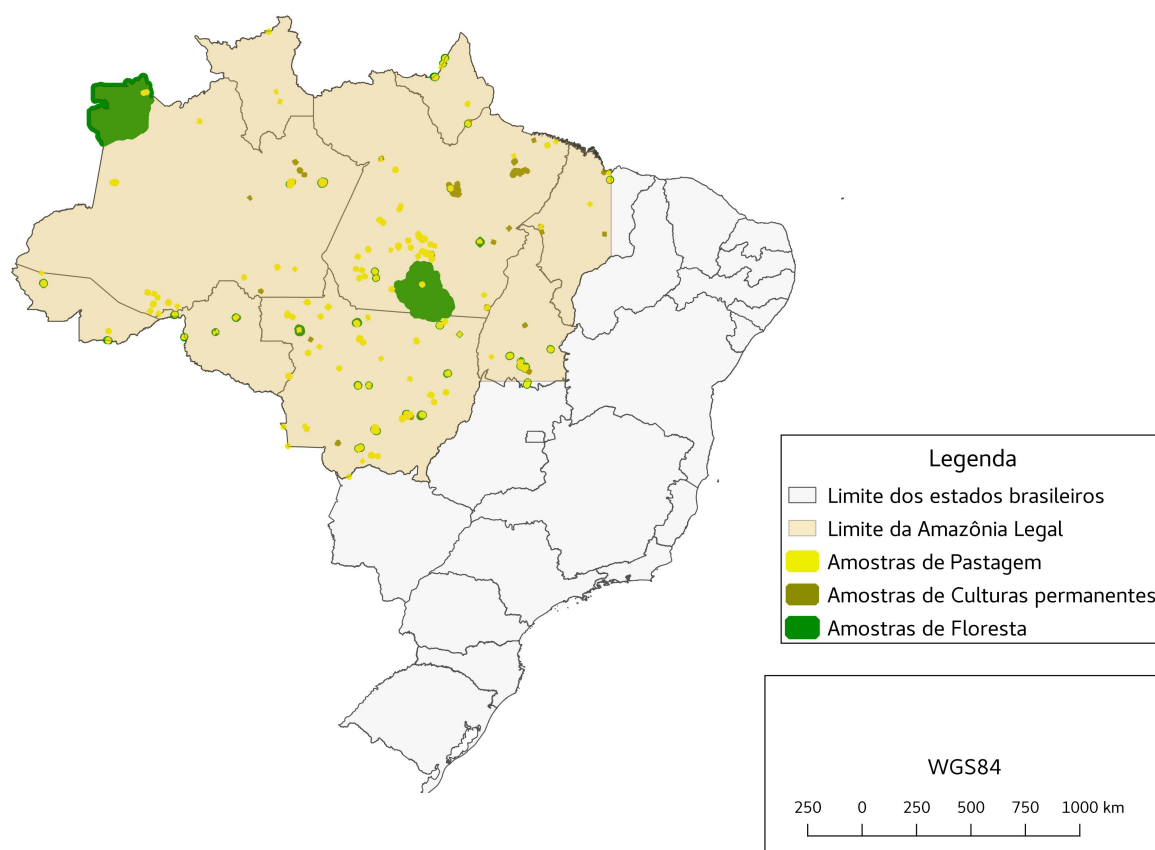


Figura 3 – Área de estudo

3.2 Dados

Os dados usados neste estudo foram extraídos do sensor MODIS (*The Terra Moderate Resolution Imaging Spectroradiometer*) a bordo do satélite Terra desenvolvido pela NASA¹. O produto MOD13Q1 possui a resolução espacial de 250m, gera duas bandas de vegetação primária com período de revisita de 16 dias. Para nosso estudo, as séries temporais correspondem a valores medidos em cada revisita, obtidos em pixels diferentes, usando o Índice de Vegetação por Diferença Normalizada (do inglês, *normalized difference vegetation index* - NDVI).

A Figura 4 apresenta em forma de coordenadas paralelas às séries temporais extraídas do conjunto de imagens correspondentes ao período de março de 2019 a novembro de 2019. É possível observar a sobreposição de linhas entre as classes e alguns padrões temporais, por exemplo as quedas no período de março a maio.

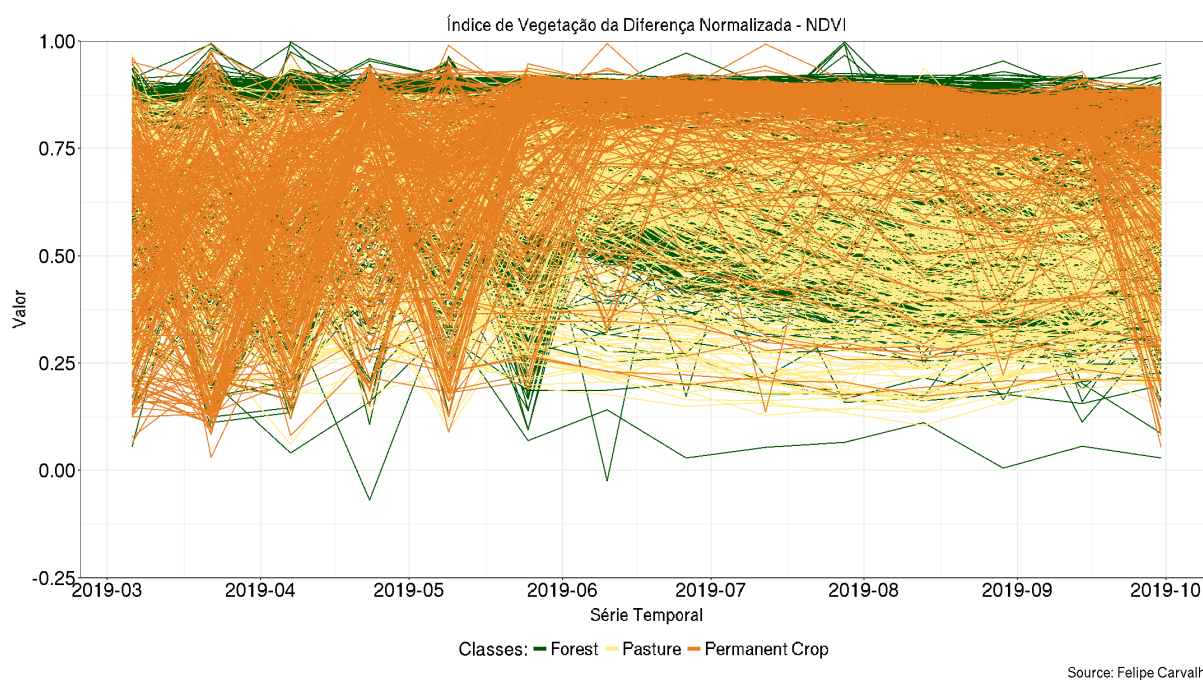


Figura 4 – Visualização em coordenadas paralelas da série temporal de cobertura do solo utilizada neste trabalho

¹ <https://terra.nasa.gov/about/terra-instruments/modis>

4 Resultados

Com o objetivo de obter grupos homogêneos no agrupamento, várias configurações do SOM foram geradas, sendo elas: topologia retangular e variações de 10x10, 15x15 e 20x20 de neurônios. Os resultados são apresentados abaixo, a Figura 5 mostra o erro de quantização das topologias supracitadas. O erro de quantização mostra a distância mínima entre os vetores de entrada e seus respectivos BMUs por época. Em cada topologia avaliada, utilizou-se dois métodos de treinamento: *online* e em *batch*. O aprendizado em *batch* refere-se ao uso de um conjunto de padrões na fase de aprendizado, diferentemente, o aprendizado *online* contempla apenas uma observação por iteração. Assim, a topologia com o menor erro mínimo foi de 20x20, treinada em modo *batch*.

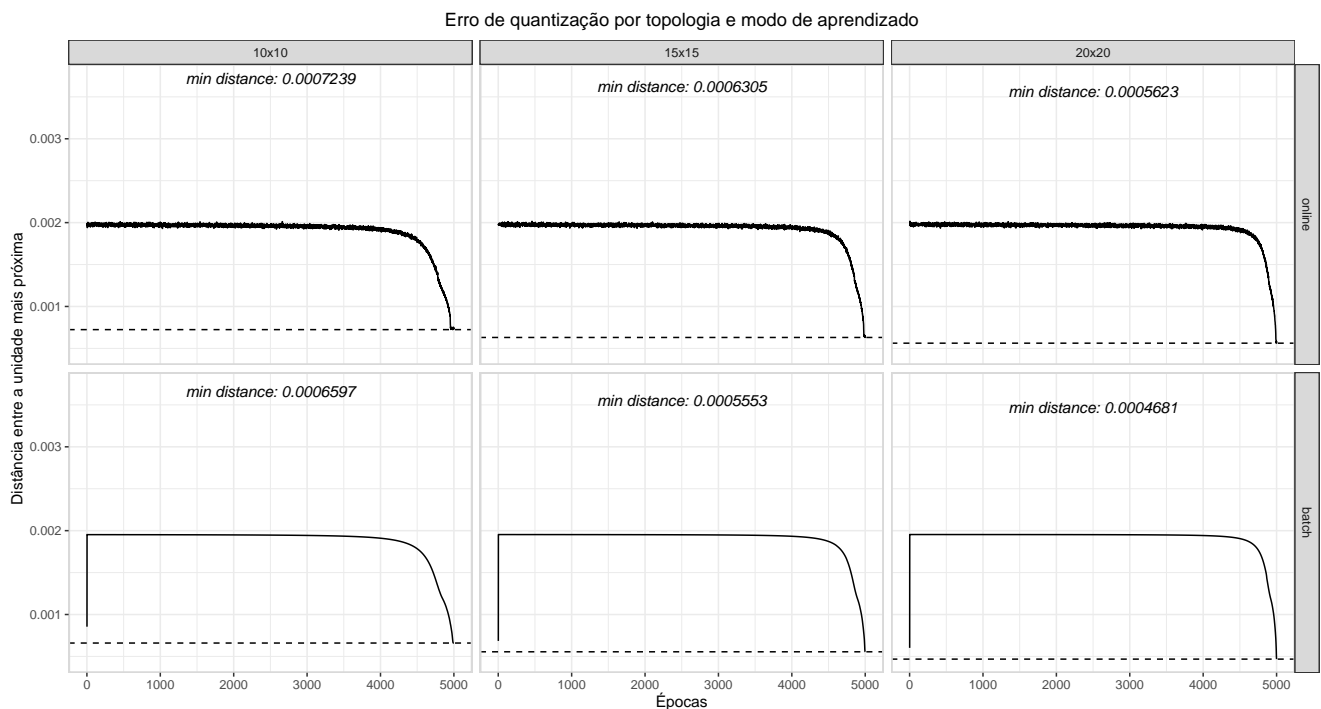


Figura 5 – Erro de quantização das topologias avaliadas

4.1 Avaliação visual com a ferramenta *ggsom*

A Figura 6 apresenta a visualização da série temporal de cobertura e uso do solo utilizada neste trabalho, por meio da ferramenta *ggsom*. No canto superior esquerdo é apresentado a quantidade de observações associadas a cada neurônio (N) e sua respectiva pureza medida pela entropia (P). Para facilitar a visualização, é referido como (X, Y) cada neurônio da grade.

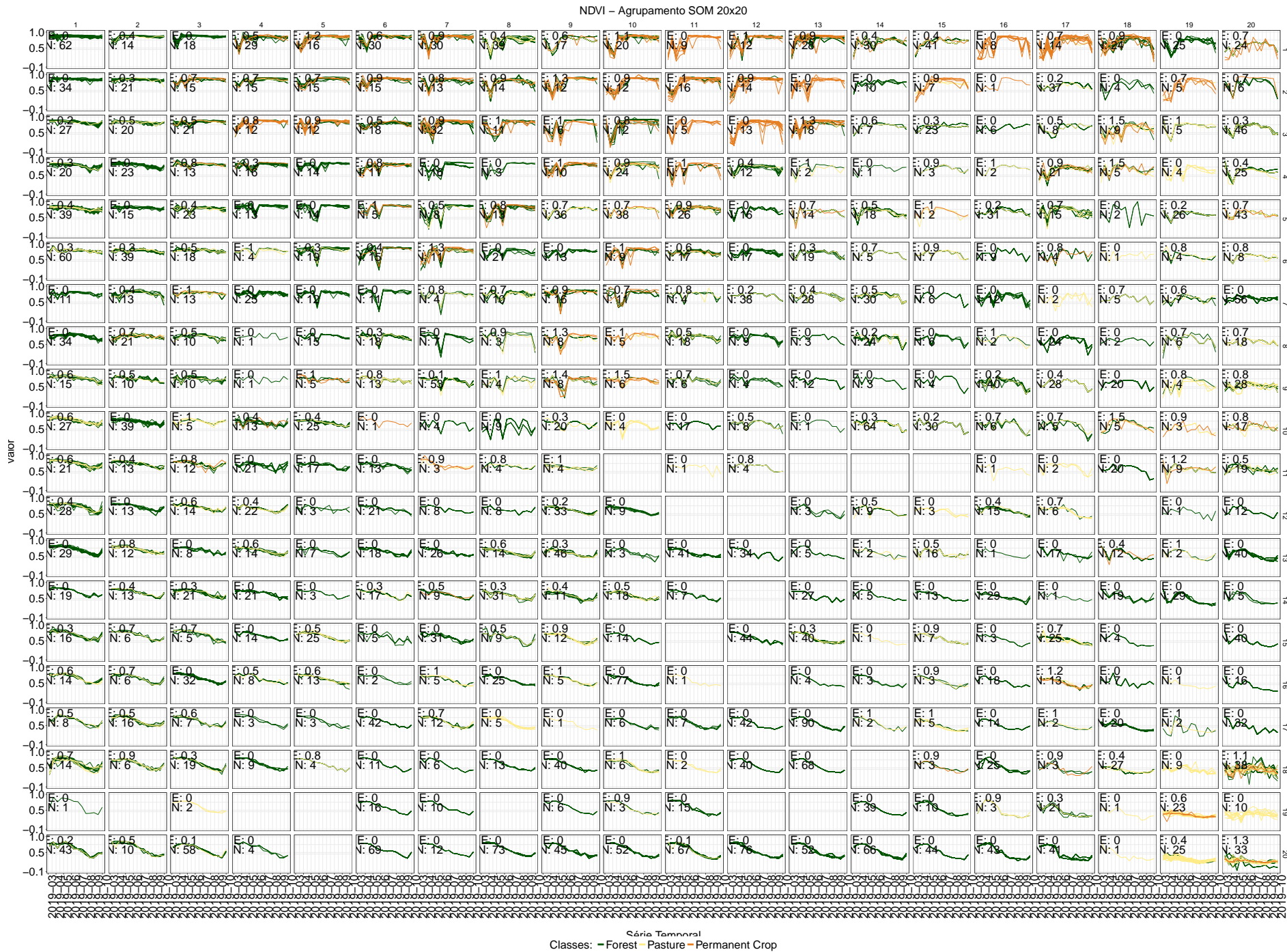


Figura 6 – Visualização em coordenadas paralelas em matriz produzida pela ferramenta *ggsom* para a rede SOM 20x20

É possível detectar diferentes padrões e características através da análise visual por meio da ferramenta *gsom*. Tais como: Separabilidade de classes, padrões de vizinhança e a confusão espectro-temporal em cada neurônio. De acordo com a Figura 6, apenas alguns neurônios obtiveram confusão espectro-temporal entre classes, por exemplo: (20x20), (20x18), (18x4) e (17x4). Diversos neurônios apresentaram o comportamento espectral esperado, como Floresta (1x1), (1x2) e (2x4), os quais apresentam o NDVI constante próximo a 1. Em resumo, os piores resultados encontram-se em neurônios com mais de duas classes mapeadas, mas, como resultado positivo indicamos que os neurônios homogêneos são os que contém mais observações.

Em relação a separabilidade, as amostras da classe Floresta possuem maior pureza em relação às demais, seguidas de Cultura Permanente e, por fim, Pastagem. As amostras de Cultura Permanente e Pastagem se encontram, por muitas vezes, no mesmo neurônio que as amostras de Floresta. A partir da análise feita, é possível concluir que diversas quedas na série temporal não pertencem aos períodos de plantação ou colheita, pois diversos neurônios confundiram classes espectralmente distintas, por exemplo Pastagem e Floresta (7x5). Outro fato são as quedas e os picos nas amostras de Floresta (6x7), é impossível uma recuperação florestal em um ciclo de tempo tão curto, logo, uma possível explicação para tais ruídos é a presença de nuvens nas regiões onde as amostras se localizam.

4.2 Avaliação estatística com o pacote SITS

Para garantir a reprodutibilidade deste trabalho, utilizou-se os mesmos parâmetros para treinar a rede SOM pelo pacote SITS, as análises estão publicadas em um jupyter notebook¹.

A acurácia geral da metodologia proposta por Santos et al. (2019) foi de 91.81%, indicando uma boa separabilidade entre as classes. A Figura 7 apresenta a matriz de confusão, em sua diagonal secundária é apresentado os verdadeiros positivos, onde é possível visualizar que entre 6021 amostras de Floresta aproximadamente 7% foram preditas como Pastagem e Cultura Permanente. É interessante observar a confusão entre as classes Floresta e Pastagem, a qual 290 amostras foram mapeadas como Pastagem, no entanto, pertencem a classe Floresta. Assim como a confusão anterior, a classe de Cultura Permanente mapeou 38 amostras como Floresta. Por fim, a classe com a maior taxa de verdadeiros positivos (*Recall*) foi Floresta, devido a quantidade de amostras, e a amostras de Cultura Permanente atingiram a menor taxa.

¹ github.com/oldlipe/cap378

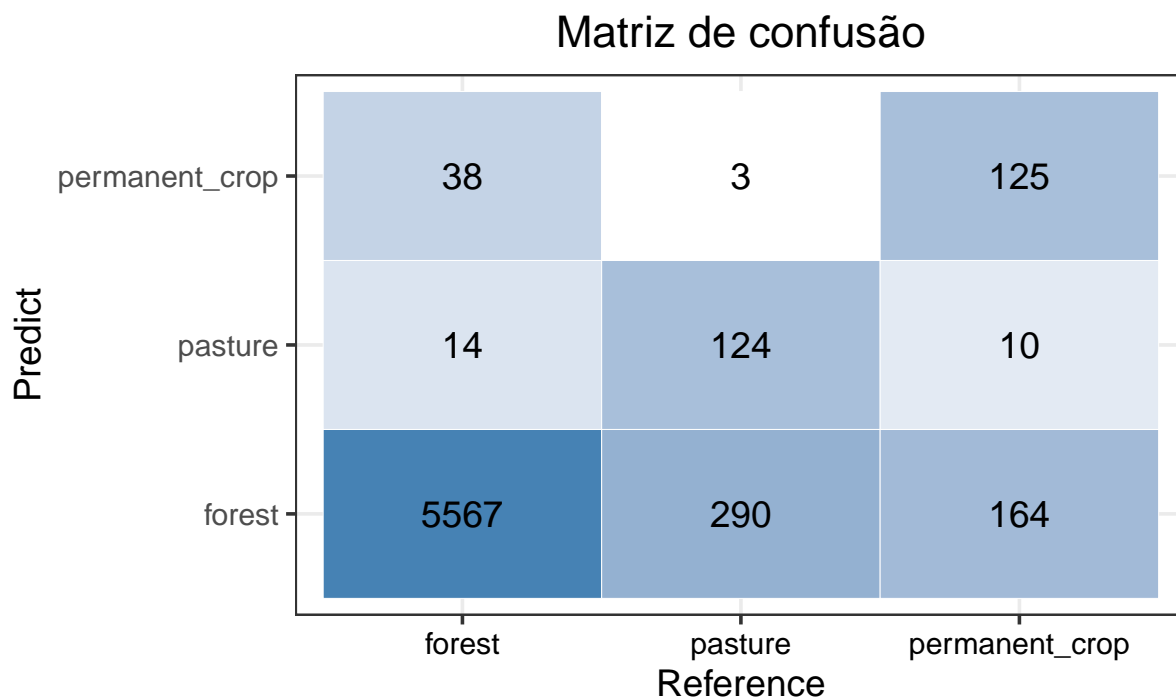


Figura 7 – Matriz de confusão

A Figura 8 apresenta as classes avaliadas e as confusões cometidas em cada uma, destaca-se a confusão entre as classes Cultura Permanente e Floresta, no entanto, ainda assim, as amostras se mostram fidedignas e passíveis para utilização de algoritmos de classificação.

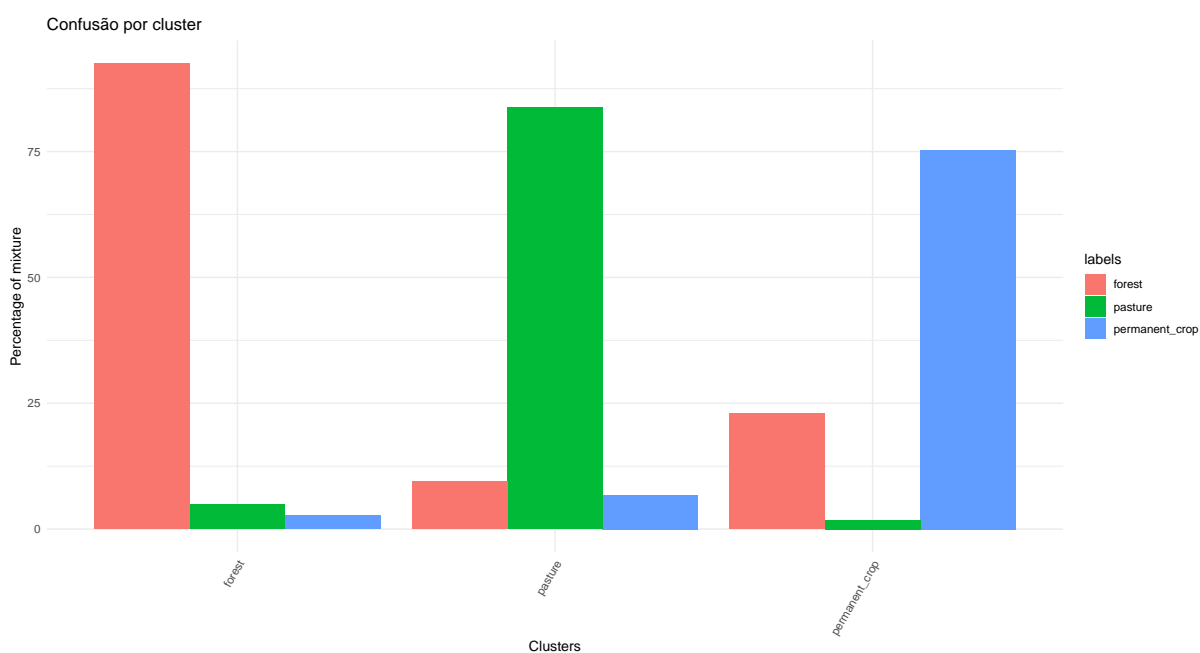


Figura 8 – Porcentagem de confusão entre clusters

5 Conclusão

Neste trabalho foi apresentado duas análises aplicadas para a seleção de amostras de uso e cobertura do solo, ambas utilizaram a rede neural SOM. A ferramenta *ggsom* usa como técnica de visualização as coordenadas paralelas. Por outro lado, o pacote SITS aplica a técnica de voto majoritário, onde é definido à priori um limiar de probabilidade para alterar o rótulo de cada neurônio.

Através da análise visual foi possível identificar padrões nas séries temporais, assim, avaliando o comportamento espectral de cada classe e concluindo que os picos e quedas apresentados em alguns neurônios representam nuvens. Outra informação obtida foi a homogeneidade de algumas classes, por exemplo Floresta, informação útil para a aplicação algoritmos de classificação. A análise estatística confirmou as informações levantadas através ferramenta *ggsom*, comprovando que as amostras com maior confusão entre *clusters* foram de Cultura Permanente.

Como a análise de cubo de dados exige mais amostras no treinamento e na validação de modelos de classificação de imagem, a entrada de dados dos voluntários é vantajosa na medida em que eles podem gerar automaticamente mapas de classificação de uso e cobertura da Terra como resultado de suas ações. Eles podem contribuir com amostras em áreas com alto desacordo, onde um projeto de amostragem automatizado garante uma significância estatística com base em suas forças individuais. O monitoramento do uso e cobertura da Terra, conforme necessário e realizado em todas as escalas, ou seja, do local ao global, do pequeno ao grande conjunto de dados, deve responder de acordo com as diferentes questões e demandas, fazendo com que a coleta de dados produzidos pelos cidadãos detenha o conhecimento local e as habilidades técnicas necessárias e motivações para contribuir com relevância. Não apenas em relação aos métodos tradicionais de amostragem e validação, discutimos neste trabalho como os voluntários por meio de técnicas de redes neurais podem ser organizados para tarefas de monitoramento relacionadas ao uso e cobertura da Terra. A análise subjetiva intrínseca à qualitativa e o fato de fazer inferências baseadas na literatura existente que não utilizam uma abordagem sistemática podem ser qualificadas como uma limitação comum deste estudo. É difícil, então, generalizar nossa abordagem, embora a teoria fundamentada e experiências e perspectivas anteriores tenham sido usadas para conter o problema. Tentamos minimizar isso através da descrição de nossos processos de análise. Estudos futuros serão orientados para o desenvolvimento e implantação das ideias discutidas neste documento.

Referências

- APPEL, M.; PEBESMA, E. On-demand processing of data cubes from satellite image collections with the gdalcubes library. *Data*, Multidisciplinary Digital Publishing Institute, v. 4, n. 3, p. 92, 2019. Citado na página 4.
- ARSANJANI, J. J. et al. Toward mapping land-use patterns from volunteered geographic information. *International Journal of Geographical Information Science*, Taylor & Francis, v. 27, n. 12, p. 2264–2278, 2013. Citado nas páginas 5 e 8.
- ASSIS, L. et al. A discussion of crowdsourced geographic information initiatives and big earth observation data architectures for land-use and land-cover change monitoring. In: *VGI-ALIVE - AnaLysis, Integration, Vision, Engagement Workshop on 21st Association of Geographic Information Laboratories in Europe (AGILE) Conference*. Lund, Sweden: [s.n.], 2018. Citado na página 4.
- ASSIS, L. F. F. d. et al. A service-oriented middleware for integrated management of crowdsourced and sensor data streams in disaster management. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 18, n. 6, p. 1689, 2018. Citado na página 4.
- BEY, A. et al. Collect earth: Land use and land cover assessment through augmented visual interpretation. *Remote Sensing*, v. 8, n. 10, 2016. ISSN 2072-4292. Disponível em: <<http://www.mdpi.com/2072-4292/8/10/807>>. Citado na página 7.
- BONNEY, R. et al. Can citizen science enhance public understanding of science? *Public Understanding of Science*, SAGE Publications Sage UK: London, England, v. 25, n. 1, p. 2–16, 2016. Citado na página 4.
- COMBER, A. et al. Using control data to determine the reliability of volunteered geographic information about land cover. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 23, p. 37–48, 2013. Citado na página 8.
- FONTE, C. et al. Generating up-to-date and detailed land use and land cover maps using openstreetmap and globeland30. *ISPRS International Journal of Geo-Information*, Multidisciplinary Digital Publishing Institute, v. 6, n. 4, p. 125, 2017. Citado na página 5.
- FONTE, C. C. et al. Assessing vgi data quality. *Citizen Sensor*, p. 137, 2017. Citado na página 8.
- FOODY, G. M. et al. Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. *Transactions in GIS*, Wiley Online Library, v. 17, n. 6, p. 847–860, 2013. Citado na página 8.
- FRIEDL, M. A. et al. Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote sensing of Environment*, Elsevier, v. 114, n. 1, p. 168–182, 2010. Citado na página 5.

FRITZ, S. et al. Geo-wiki. org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, Molecular Diversity Preservation International, v. 1, n. 3, p. 345–354, 2009. Citado nas páginas 5 e 7.

FRITZ, S. et al. Geo-wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, Elsevier, v. 31, p. 110–123, 2012. Citado na página 8.

GIULIANI, G. et al. Building an earth observations data cube: lessons learned from the swiss data cube (sdc) on generating analysis ready data (ard). *Big Earth Data*, Taylor & Francis, v. 1, n. 1-2, p. 100–117, 2017. Citado na página 4.

GLOBCOVER. *ESA Data User Element*. 2017. Available at: <http://due.esrin.esa.int/page_globcover.php>. Citado na página 5.

HAKLAY, M. How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environment and planning B: Planning and design*, SAGE Publications Sage UK: London, England, v. 37, n. 4, p. 682–703, 2010. Citado na página 5.

HAYKIN, S. *Neural Networks and Learning Machines, 3/E*. [S.l.]: Pearson Education India, 2010. Citado na página 8.

HERFORT, B. et al. Mapping human settlements with higher accuracy and less volunteer efforts by combining crowdsourcing and deep learning. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 11, n. 15, p. 1799, 2019. Citado na página 5.

IBGE. *Manual técnico de uso da terra*. [S.l.]: IBGE, 2013. v. 3. Citado na página 11.

KOHONEN, T. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, Springer, v. 43, n. 1, p. 59–69, 1982. Citado na página 8.

KOHONEN, T. Essentials of the self-organizing map. *Neural Networks*, v. 37, p. 52 – 65, 2013. Citado nas páginas 1 e 9.

LEINENKUGEL, P. et al. The potential of open geodata for automated large-scale land use and land cover classification. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 11, n. 19, p. 2249, 2019. Citado na página 5.

LEWIS, A. et al. The australian geoscience data cube—foundations and lessons learned. *Remote Sensing of Environment*, Elsevier, v. 202, p. 276–292, 2017. Citado na página 4.

LU, M.; APPEL, M.; PEBESMA, E. Multidimensional arrays for analysing geoscientific data. *ISPRS International Journal of Geo-Information*, Multidisciplinary Digital Publishing Institute, v. 7, n. 8, p. 313, 2018. Citado na página 4.

MOORTHY, I. et al. Fotoquest go: A citizen science tool for in-situ land use and land cover monitoring. In: *EGU General Assembly Conference Abstracts*. [S.l.: s.n.], 2018. v. 20, p. 8124. Citado na página 7.

- NEWMAN, G. et al. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, Wiley Online Library, v. 10, n. 6, p. 298–304, 2012. Citado na página 4.
- RADDICK, M. J. et al. Galaxy zoo: Motivations of citizen scientists. *arXiv preprint arXiv:1303.6886*, 2013. Citado na página 4.
- SANTOS, L. et al. Self-organizing maps in earth observation data cubes analysis. In: SPRINGER. *International Workshop on Self-Organizing Maps*. [S.l.], 2019. p. 70–79. Citado nas páginas 1, 4, 10, 11 e 15.
- SCHULTZ, M. et al. Open land cover from openstreetmap and remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 63, p. 206–213, 2017. Citado na página 5.
- SEE, L. et al. Laco-wiki: A new online land cover validation tool demonstrated using globeland30 for kenya. *Remote Sensing*, v. 9, n. 7, 2017. ISSN 2072-4292. Disponível em: <<http://www.mdpi.com/2072-4292/9/7/754>>. Citado nas páginas 6 e 7.
- SIMÕES, R. et al. *SITS: Data Analysis and Machine Learning using Satellite Image Time Series*. Sao Jose dos Campos, Brazil, 2019. Disponível em: <<https://github.com/e-sensing/sits>>. Citado na página 10.
- SULLIVAN, B. L. et al. The ebird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation*, Elsevier, v. 169, p. 31–40, 2014. Citado na página 4.
- TEAM, G. *GeoNetwork opensource portal to spatial data and information*. 2017. Available at: <<http://www.fao.org/geonetwork/srv/en/main.home?uuid=ba4526fd-cdbf-4028-a1bd-5a559c4bff38>>. Citado na página 5.
- WEHRENS, R.; BUYDENS, L. M. et al. Self-and super-organizing maps in r: the kohonen package. *Journal of Statistical Software*, v. 21, n. 5, p. 1–19, 2007. Citado na página 10.
- WICKHAM, H. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, Wiley Online Library, v. 3, n. 2, p. 180–185, 2011. Citado na página 10.
- YANG, H. et al. The standardization and harmonization of land cover classification systems towards harmonized datasets: a review. *ISPRS International Journal of Geo-Information*, Multidisciplinary Digital Publishing Institute, v. 6, n. 5, p. 154, 2017. Citado na página 4.
- ZOU, M.; ZHONG, Y. Transfer learning for classification of optical satellite image. *Sensing and Imaging*, Springer, v. 19, n. 1, p. 6, 2018. Citado na página 4.