

MAPEAMENTO DA EVOLUÇÃO ESPACIAL E TEMPORAL DE ÁREAS AGRÍCOLAS NO BIOMA PAMPA UTILIZANDO FENOLOGIA DA VEGETAÇÃO E GOOGLE EARTH ENGINE

Rafael Henrique Bloedow Kayser¹, Anderson Ruhoff²

¹ Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, rafael.hbkayser@gmail.com; ² Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, andersonruhoff@gmail.com;

RESUMO

Este trabalho apresenta uma metodologia simples, porém eficiente de identificação de áreas agrícolas a partir da análise da variação temporal de índices de vegetação utilizando imagens Landsat. Através do cálculo do desvio padrão em cada pixel em um conjunto de imagens que abrangem um ciclo de cultivo de verão, foi possível identificar valores limiares nos quais acima destes a região pode ser caracterizada como cultivos de arroz ou soja. Essa metodologia foi aplicada na região do bioma pampa gaúcho. Nesse trabalho foi utilizada uma tecnologia de computação em nuvem chamada Google Earth Engine, facilitando o processamento de um grande conjunto de dados. A metodologia foi validada a partir de dados de censo agropecuário, sendo avaliada anualmente no período de 1996 a 2017. O erro relativo total neste período foi de -1,43% e o R^2 igual a 0,82.

Palavras-chave — classificação, áreas agrícolas, índices de vegetação, fenologia da vegetação, Google Earth Engine

ABSTRACT

This work presents a simple but efficient approach for crop mapping using vegetation phenology indices extracted from Landsat images. By calculating the standard deviation in each pixel in a set of images that cover a summer crop cycle, it was possible to identify threshold values in which above the region could be characterized as rice or soybean crops, being applied in the Pampa biome, in the South of Brazil. In this work a cloud computing technology called Google Earth Engine was used, facilitating the processing of a large dataset. The methodology was validated from agricultural census data, being evaluated annually in the period from 1996 to 2017. The total relative error in this period was -1.43% and the R^2 was equal to 0.82.

Key words — classification, cropland, vegetation indices, phenology vegetation, Google Earth Engine

1. INTRODUÇÃO

O aumento populacional das últimas décadas tem acarretado em uma maior necessidade de produção de alimentos, e

consequentemente, numa maior pressão sobre os sistemas ambientais, aumentando a ocupação agrícola e o uso de água para a irrigação. Nesse sentido, estudos e levantamentos das áreas agrícolas são imprescindíveis, a fim de garantir o crescimento sustentável, preservando os ecossistemas e os recursos hídricos. O sensoriamento remoto possibilita oferecer importantes informações relacionadas à agricultura. Através dessas ferramentas, é possível fornecer informações de identificação das áreas irrigadas, tipo de cultivo, desenvolvimento de biomassa, demanda hídrica, evapotranspiração, entre outros produtos [1].

Existem diferentes métodos para o mapeamento de áreas agrícolas. Para uma escala regional, há abordagens que utilizam métodos de classificação supervisionada e não-supervisionada. Estes métodos tendem a apresentar diferentes resultados pois dependem de uma boa acurácia na seleção de amostras de treinamento, além de demandar um alto consumo de tempo. Recentemente, surgiram abordagens baseadas na análise das séries temporais da reflectância espectral ou de índices de vegetação, valendo-se de propriedades específicas apresentadas por determinada feição em um período de tempo [2].

Nos últimos anos, uma nova plataforma de processamento de produtos de sensoriamento remoto foi disponibilizada através da plataforma Google Earth Engine (GEE) [3]. O GEE consiste em um banco de dados na escala de petabytes, onde estão catalogados grande parte dos produtos de sensoriamento remoto disponíveis gratuitamente, como toda a coleção Landsat. Além disso, o GEE também proporciona o acesso a este banco de dados de maneira remota, através da tecnologia de computação em nuvem, sem a necessidade de download, aumentando exponencialmente a capacidade de processamento de séries temporais de imagens.

O objetivo deste trabalho é desenvolver uma metodologia para mapeamento de áreas agrícolas, especialmente irrigadas, utilizando a variação temporal de índices de vegetação (fenologia da vegetação), valendo-se da facilidade de processamento das imagens Landsat a partir da ferramenta GEE. Propõe-se que esta metodologia seja facilmente aplicada para todas as imagens disponíveis, a fim de que se seja possível a obtenção de séries históricas anuais de áreas ocupadas por agricultura ao longo do tempo.

2. MATERIAIS E MÉTODOS

2.1. Área de estudo

A área de estudo corresponde à metade sul do Rio Grande do Sul, com uma área de aproximadamente 160.000 km². Representa uma região com predominância do bioma Pampa, compreendido por formações campestres em sua maior extensão, além de florestas semi-decíduais ao longo das planícies e várzeas dos principais rios. Devido às condições naturais, essa área caracteriza-se como uma das regiões com maior desenvolvimento agrícola do país. De acordo com o censo nacional de agricultura de 2016 [4], aproximadamente 1,08 milhões de hectares de arroz e 1,86 hectares de soja foram cultivados, o equivalente a 54,2 e 6,6% do total cultivado no Brasil. Juntas, as duas culturas representam mais de 85% da área cultivada na região, com o restante sendo ocupado por culturas de sequeiro, como o milho, e também culturas de inverno, como o trigo e a aveia. A região de estudo foi dividida em seis zonas, conforme a divisão adotada pelo Instituto Rio-Grandense do Arroz - IRGA (Figura 1).

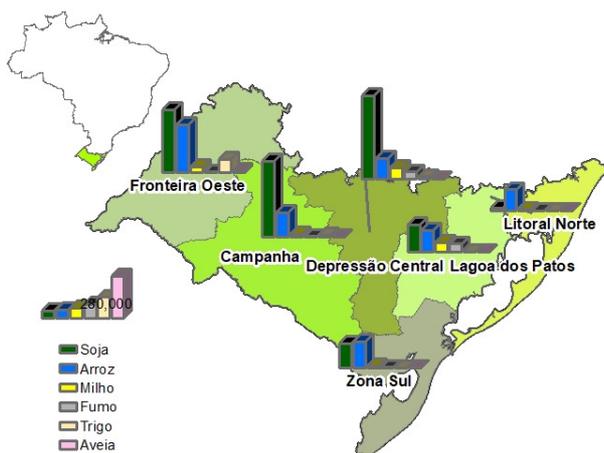


Figura 1. Localização da área de estudo no sul do Brasil, indicando os principais cultivos agrícolas em cada região, de acordo com o censo agrícola de 2016.

A região é historicamente conhecida pelo cultivo irrigado do arroz. No entanto, nos últimos anos a região passou por uma expansão significativa da soja, cultivada em sequeiro. A conversão acelerada das áreas de campo nativo para o cultivo da soja e outras monoculturas anuais e o uso demasiado de agrotóxicos, configuram uma realidade preocupante no Pampa na atualidade [5].

2.2. Seleção das imagens Landsat e análise de disponibilidade de dados

A principal fonte de dados deste estudo está relacionada às coleções Landsat 5, 7 e 8, incluídas no catálogo do GEE. Foram utilizadas as coleções Landsat de reflectância de

superfície (com correção atmosférica) ortorretificadas, obtidas dos sensores ETM (Landsat 5), ETM+ (Landsat 7) e OLI (Landsat 8). A missão Landsat 5 coletou dados de 1984 a 2012, e as coleções Landsat 7 e 8, respectivamente a partir de 1999 e 2013, seguindo atualmente em atividade. A área de estudo abrange um total de 18 cenas das missões Landsat, incluídas total ou parcialmente.

As séries anuais de áreas agrícolas a nível municipal estão disponíveis a partir do ano de 1987, sendo definido como ponto de partida para a análise das imagens. Inicialmente analisou-se a disponibilidade de pixels na área de estudo em cada mês, uma vez que é necessária a exclusão de áreas cobertas por nuvens. Posteriormente, imagens com baixa qualidade radiométrica foram excluídas das análises. Como a resolução temporal do Landsat é de 16 dias, numa situação ideal sem a presença de nuvens, teria-se a disponibilidade de uma a duas imagens por mês. Entretanto, a ocorrência de nuvens reduz drasticamente a quantidades de imagens disponíveis para análise. A Figura 2 apresenta os percentuais da área de estudo com pelo menos um pixel no mês de janeiro, período importante na caracterização agrícola, no qual ocorrem os picos de desenvolvimento da vegetação e floração. Na ausência de informações nesse período, a identificação das áreas agrícolas resulta em baixa acurácia. Observa-se que até o ano de 1998 existem muitas áreas desprovidas de informações, aumentando a partir deste ponto devido à entrada de dados do Landsat 7.

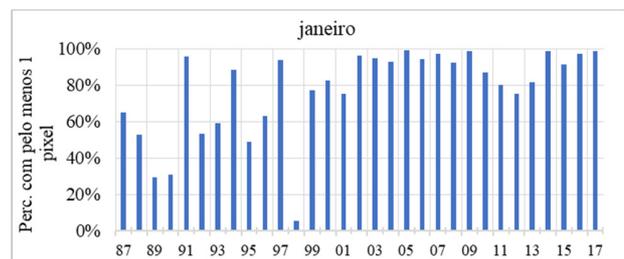


Figura 2. Percentual da área de estudo com pelo menos um pixel disponível no mês de janeiro.

2.4. Definição de critério para distinção entre áreas agrícolas e não agrícolas

A partir de metodologias similares, baseadas na fenologia da vegetação, procurou-se desenvolver um critério de separação entre as áreas agrícolas e não-agrícolas/naturais a partir da análise da variação temporal de índices de vegetação. Inicialmente foram extraídas amostras de áreas identificadas como cultivos agrícolas de arroz e soja, além de áreas de campo, florestas, áreas úmidas e banhados, a partir do mapeamento da cobertura vegetal do bioma Pampa em relação ao ano de 2009 [6]. Posteriormente, analisou-se o comportamento fenológico de cada tipo de cobertura da terra a partir de índices de vegetação. A partir das bandas de reflectância superficial, foram calculados os índices NDVI (Normalized Difference Vegetation Index) e LSWI (Land Surface Water Index). O LSWI foi utilizado para identificar

áreas com vegetação e água. Verificou-se uma tendência consistente de variação dos índices de vegetação nas áreas agrícolas de arroz e soja, enquanto que os demais tipos de cobertura mantiveram-se constantes ou com uma menor variação ao longo de um ciclo de cultivo. A Figura 3 apresenta a média do NDVI nas amostras extraídas em áreas de arroz, soja e áreas de campo aberto, cujas imagens foram selecionadas entre os anos de 2008 e 2009.

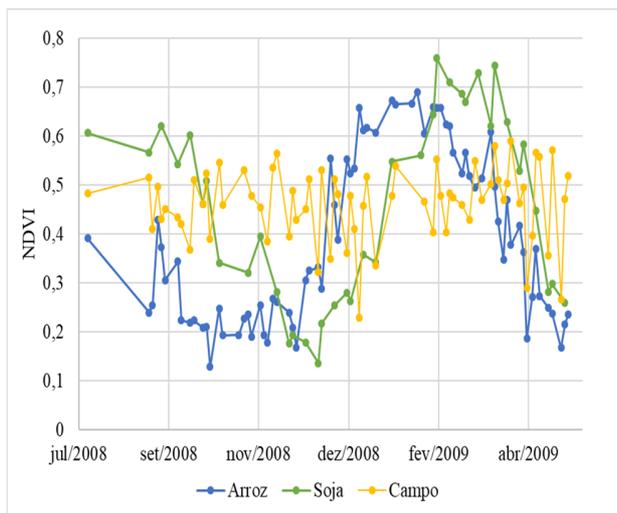


Figura 3. Variação temporal do NDVI ao longo de um ciclo agrícola em amostras do arroz, soja e áreas de campo.

Verificou-se que as áreas agrícolas apresentam um desvio-padrão bastante elevado em relação às demais classes, considerando-se um conjunto de imagens ao longo de um ciclo de safra de verão, iniciando em setembro até maio do ano seguinte. No entanto, algumas áreas cobertas sazonalmente por água, áreas úmidas/banhados também podem apresentar um desvio-padrão elevado. Assim, foram utilizadas bases de dados globais (na mesma resolução espacial de imagens Landsat) para descartar áreas não-agrícolas. Utilizou-se um produto global de cobertura florestal [7] e um produto de cobertura de água [8], ambos desenvolvidos a partir da classificação multi-temporal de imagens Landsat. Adicionalmente, também foi utilizada uma máscara de áreas úmidas (banhados), a partir de um levantamento realizado em escala local [9]. A área total ocupada por banhados na área de estudo não é significativa, no entanto está concentrada em algumas regiões, especialmente no litoral sul.

A plataforma do GEE possui algumas ferramentas que permitem delimitar de forma mais eficaz os polígonos ocupados por áreas agrícolas, como o filtro de suavização de contornos, além do filtro de pixels conectados, o qual elimina um conjunto de pixels cujo agrupamento é menor do que uma determinada quantidade, útil para a filtragem de áreas esparsas menores do que a área mínima de desenvolvimento de um cultivar. A Figura 4 apresenta um esquema da metodologia a ser adotada no trabalho.

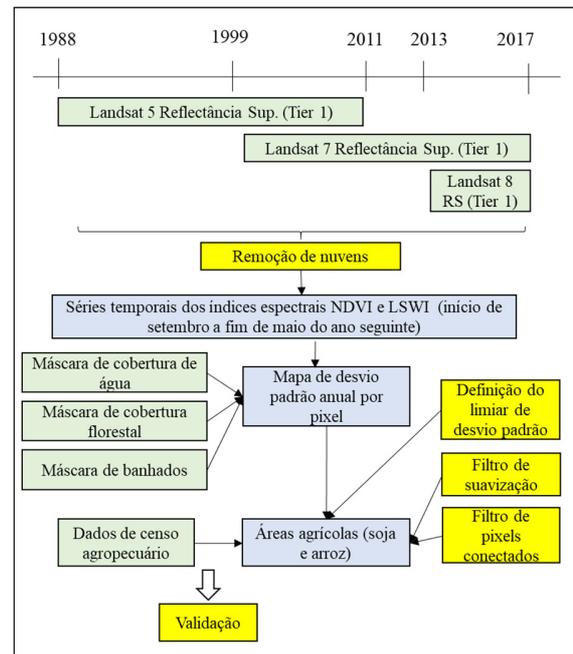


Figura 4. Fluxograma com a metodologia adotada para a identificação de áreas agrícolas com base na fenologia e índices de vegetação.

3. RESULTADOS E DISCUSSÃO

3.1. Discussão do limiar de desvio padrão

Inicialmente foram selecionados cinco anos/safra com boa disponibilidade de dados para avaliar uma faixa de limiares de desvio-padrão, com valores pré-estabelecidos entre 0,13 e 0,17. Esses limiares foram posteriormente aplicados a metodologia proposta e os resultados foram comparados às áreas agrícolas de arroz e soja fornecidos pelo Censo Agropecuário Brasileiro. Foram testados três cenários: o primeiro com o desvio-padrão anual do NDVI, o segundo com o desvio-padrão do LSWI, e o terceiro com desvio-padrão máximo entre os dois índices. Verificou-se que os limiares que minimizam os erros relativos correspondem aos valores de 0,15 e 0,16 (Figura 5).

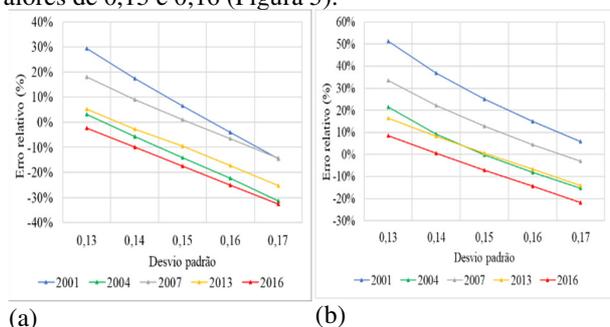


Figura 5 – Variação do limiar de desvio padrão entre os cinco anos testados e cinco faixas de desvio padrão para os cenários: a) NDVI, b) máximo entre NDVI e LSWI.

3.2. Comparação anual com dados de censo agropecuário

Os limiares de desvios padrão de 0,15 e 0,16 foram testados anualmente, de 1996 a 2017, para os três cenários de índices espectrais estabelecidos. Foram verificados os erros relativos em relação às áreas de cada uma das seis regiões, e também em relação ao total da área estudada. Verificou-se menores erros para o limiar de 0,15 utilizando o NDVI e o LSWI, e de 0,16 para cenário de máximo entre os dois índices (Tabela 1)

Tabela 1. Erro relativo por região e em toda a área de estudo para os melhores limiares de desvio padrão por cenário

Região	Erro relativo total (%)		
	NDVI (dp. = 0,15)	LSWI (dp. = 0,15)	Máximo (dp. = 0,16)
Front. Oeste	3,9%	2,6%	0,4%
Campanha	3,3%	7,4%	10,8%
Dep. Central	-19,1%	-18,9%	-18,3%
Lagoa dos Patos	-19,5%	-18,4%	-12,9%
Lit. Norte	-19,5%	-18,4%	-12,9%
Zona Sul	-5,2%	-1,4%	9,5%
Total	-5,3%	-4,2%	-1,4%

Considerando-se a acurácia dos limiares, adotou-se o cenário de máximo desvio-padrão entre NDVI e LSWI para o limiar de 0,16 como o mais adequado, embora as diferenças entre os limiares e cenários não seja muito significativa. A Figura 6 apresenta o resultado da série temporal de área agrícola comparada com as áreas informadas por censo agropecuário de soja, arroz, além do total de todas as áreas cultivadas, indicando que de fato, a área identificada a partir da classificação de imagens é mais próxima do somatório entre as áreas cultivadas de soja e arroz. Analisando os resultados anualmente, é possível verificar que para 60% dos anos/safra analisados houve uma diferença de no máximo $\pm 10\%$ em relação ao censo agropecuário. Os valores percentuais indicados na figura representam o erro relativo de cada ano/safra em relação aos dados de censo agropecuário.

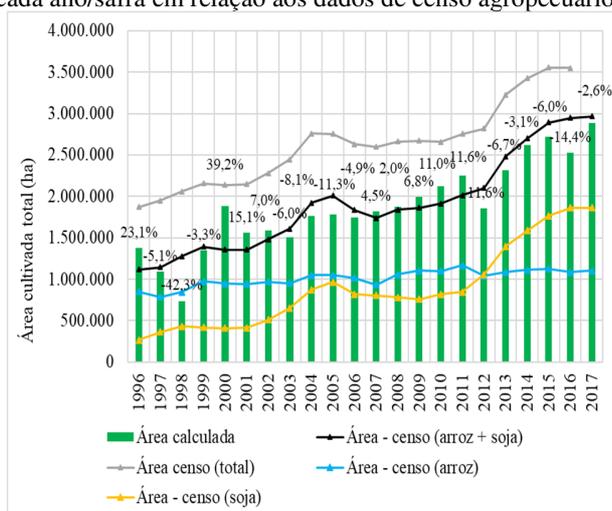


Figura 6 – Série temporal de áreas agrícolas estimadas a partir de índices de vegetação e comparada com as áreas anuais de censo agropecuário para toda a área de estudo.

4. CONCLUSÕES

A área de estudo mostrou-se bastante adequada à aplicação da metodologia testada. Entre os anos de 1997 e 2017 houve um aumento de mais de 130% da área cultivada na área do bioma Pampa gaúcho, conforme verificado tanto pela classificação de imagens quanto pelos dados de censo agropecuário. É possível que em outras regiões esta metodologia não seja adequada, especialmente em áreas com diferentes cultivos agrícolas ou com grande variação dos índices de vegetação ao longo do tempo. Não foi possível até o momento desenvolver uma metodologia de separação entre os cultivos de soja e arroz, dadas as semelhanças fenológicas apresentadas entre as duas culturas. Entretanto, é possível que áreas ocupadas pela rizicultura apresentem um maior índice de evapotranspiração, dada à condição de que esta é cultivada sob regime de inundação, necessitando de uma quantidade muito maior de água. Esta hipótese será testada em trabalhos futuros.

É importante destacar também a grande capacidade de processamento de imagens apresentada pela ferramenta do Google Earth Engine. Foi analisado um período total de 22 anos, e uma média de 350 imagens ao ano. Considerando que cada imagem adquirida manualmente possua em torno de 1 GB, teríamos a necessidade de realizar um download de mais de 7 mil GB de dados a fim de contemplar tudo o que foi analisado na nuvem via GEE. A análise de um conjunto de imagens ao longo um ano/safra tem um tempo de processamento em torno de 20 minutos, sendo extraídas das imagens apenas os resultados finais da classificação.

6. REFERÊNCIAS

- [1] Molden, B. and David J. "Remote sensing for irrigated agriculture: examples from research and possible applications" *Agric. Water Manag.* v. 46, pp. 137-155, 2000.
- [2] Dong, J. et al. "Tracking the dynamics of paddy rice planting area in 1986 – 2010 through time series Landsat images and phenology-based algorithms." *Remote Sensing of Environment.* v. 160, p. 99–113, 2015.
- [3] Gorelick, N. et al. "Google Earth Engine: Planetary-scale geospatial analysis for everyone". *Remote Sensing of Environment.* v. 202, p. 18–27, 2017.
- [4] IBGE. "Produção agrícola municipal: culturas temporárias e permanentes". 2017.
- [5] Capoane, V. "Expansão da agricultura no bioma Pampa." *8º Região de Estudos Ambientais*, 2018.
- [6] Hasenack, H. et al. "Mapeamento da cobertura vegetal do Bioma Pampa: Ano-base 2009" UFRGS - Centro de Ecologia. 2016.
- [7] Hansen, M. C. et al. "High-Resolution Global Maps of 21st-Century Forest Cover Change". *Science*, v. 342, n. 6160, p. 850-853, 2013.
- [8] Pekel, J et al. "High-resolution mapping of global surface water and its long-term changes". *Nature*, v. 540, n. 7633, p. 418–422, 2016.
- [9] Ramos, R. et al. "Mapeamento e diagnóstico de áreas úmidas no Rio Grande do Sul com o uso ferramentas de geoprocessamento". FZB/RS. Porto Alegre. 2002.