

RUMO A ESTIMATIVA OBJETIVAS DE SAFRAS AGRÍCOLAS

Marcos Adami¹, Patrícia Mauricio Campos², Lucas Barbosa Fernandes², Rafaela dos Santos Souza², Fernando Arthur Santos Lima², Priscilla Azevedo dos Santos¹, Cleverton Tiago Carneiro de Santana^{1,2}, Candice Mello Romero Santos², Ieda Del'Arco Sanches¹

¹ Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos - SP, {marcos.adami; priscilla.santos; ieda.sanches}@inpe.br; ² Companhia Nacional de Abastecimento - Conab, Brasília-DF, {patricia.campos; lucas.barbosa; rafaela.souza; fernando.a.lima; cleverton.santana; candice.santos}@conab.gov.br

RESUMO

As estimativas de safras devem atender no mínimo a três critérios: i) oportunidade, ii) confiabilidade, e iii) transparência. Neste sentido, este trabalho tem por objetivo apresentar uma metodologia para realizar a estimativa de área cultivada com soja para os estados do Maranhão, Tocantins, Piauí e Bahia. As primeiras estimativas oriundas deste procedimento ocorrem ainda ao longo da safra, logo que o trabalho de campo se encerra, atendendo assim ao critério de oportunidade. Além disto, as estimativas tem associadas uma estimativa de incerteza, o que atende ao critério de confiabilidade e todas são passíveis de serem auditadas, atendendo ao critério de transparência.

Palavras-chave — Mapeamento, Soja, MATOPIBA, Amostragem.

ABSTRACT

Crop estimates should comply with at least three criteria: i) timely, ii) reliability and iii) accountability. In this way, the present work aims at presenting a methodology to estimate soybean crop area in the states of Maranhão, Tocantins, Piauí, and Bahia. First estimates are obtained during the crop season, therefore achieving the criteria of timely. Besides that, all estimates have an uncertainty value attached to them, that meets the reliability criteria. And all estimates are auditable, therefore, meeting the accountability criteria.

Key words — Mapping, Soybeans, MATOPIBA, Sampling.

1. INTRODUÇÃO

A produção agrícola é estimada utilizando duas variáveis, a área e a produtividade (rendimento médio por unidade de área). Estes dados de estimativa de safras são acompanhados com muito interesse durante todo o andamento da safra e são utilizados para uma série de decisões públicas e privadas que vão desde a definição de preços, estoques, dentre outros, até a segurança alimentar. Dada a sua importância, no mínimo três critérios devem ser atendidos: i) oportunidade, ii) confiabilidade, e iii) transparência. O critério oportunidade

está relacionado ao momento da divulgação destas estimativas, em tempo hábil que permita a tomada de decisões. O critério confiabilidade está na associação destas estimativas à uma medida de incerteza [1]. O critério de transparência está associado com a capacidade de verificar e checar as informações e os resultados, possibilitando a governança do processo.

O critério de oportunidade é mais facilmente obtido utilizando o método subjetivo. O método subjetivo não utiliza o rigor científico e obtém as informações necessárias para o levantamento de safras tendo por base a opinião de agentes técnicos ligados ao agronegócio. Por conta disto é muito mais fácil de ser utilizado e barato, o que o torna ágil no momento da aquisição da informação, atendendo ao critério de oportunidade. Esta modalidade de levantamento também possibilita a transparência, pois todos os dados podem ser verificados junto aos informantes. Entretanto, devido à falta de embasamento científico, este método não possibilita a estimativa da incerteza da informação que podem estar embasadas em opiniões enviesadas, de acordo com a opinião ou interesse do agente, não atendendo ao critério de confiabilidade. Por outro lado, para atender ao critério de confiabilidade, o levantamento deve ocorrer de maneira objetiva, utilizando rigor científico. Entretanto, estes levantamentos são mais demorados e tem um custo maior, o que dificulta o entendimento do critério de oportunidade [2]. Como tem por base método científico, este método atende a necessidade de transparência da informação. Desta maneira, o balanço e a integração entre os métodos, subjetivo e objetivo, deve ser levado em conta para a estimativa de safras dando mais segurança e transparência ao processo de geração de estatísticas agrícolas.

Atualmente, a maioria das estimativas de safras brasileiras utilizam os métodos subjetivos. Apesar de que seus resultados são apoiados por alguns métodos objetivos, poucos resultados apresentam medidas de incertezas associadas aos levantamentos. Assim, procurando um balanço para atender aos critérios citados, deve ser levado em conta os avanços que ocorreram em todas as áreas, mas principalmente os avanços na área de geotecnologias, como as políticas de livre acesso às imagens de resolução espacial mediana [3], computação em nuvem [4], dentre outros. Pois o uso das geotecnologias facilita o atendimento de todos os critérios listados.

Neste sentido, associando os avanços na área de geotecnologias com as necessidades de atender aos critérios de oportunidade, confiabilidade e transparência, necessários para estimativa de safras agrícolas, este trabalho tem por objetivo apresentar uma metodologia para realizar a estimativa de área cultivada com soja para os estados do Maranhão, Tocantins, Piauí e Bahia, não restritivos apenas à região do MATOPIBA, mas para toda a área dos estados listados.

2. MATERIAL E MÉTODOS

Para o desenvolvimento deste trabalho foram utilizadas imagens obtidas pelo TERRA/MODIS, LANDSAT/OLI e SENTINEL/MSI. Esses dados foram processados utilizando os seguintes aplicativos: *Google Earth Engine* (GEE) [4], R [5] com o pacote *SamplingStrata* [6]. Inicialmente foram obtidos os mapeamentos das áreas cultivada dos três anos anteriores à pesquisa, posteriormente, foi sobreposta uma grade celular de 20km por 20 km de lado (blocos) e foram obtidos os percentuais de área cultivada em cada um dos blocos, bem como o percentual de área que foi agricultura em pelo menos dois mapeamentos. Estes dados foram utilizados para estratificação. Foi considerado como estrato nulo os blocos com menos de 0,5% da área utilizada para agricultura. Os detalhes metodológicos estão apresentados na Figura 1.

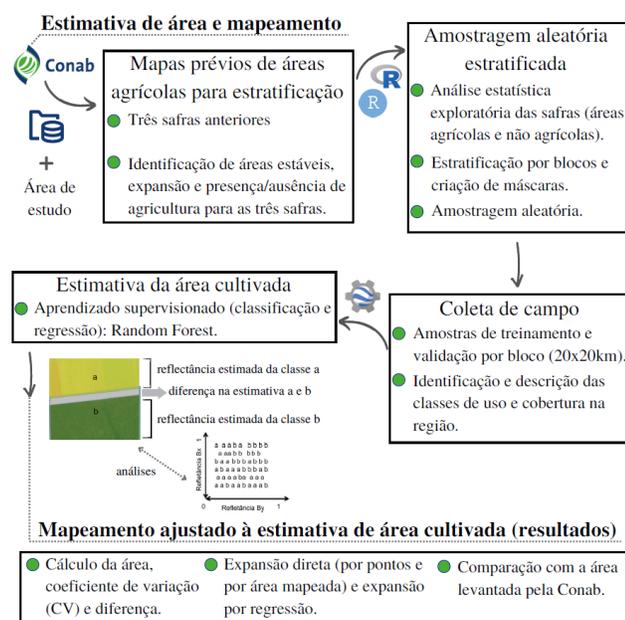


Figura 1. Fluxograma para estimativa de área objetiva.

Uma vez obtido os estratos, foi realizado o sorteio aleatório em duplo estágio, sendo que no primeiro estágio foram sorteados os blocos e posteriormente, como segundo estágio foram sorteados 20 pontos aleatoriamente em cada um dos blocos. Estes pontos tem dois objetivos. O primeiro deles é realizar a estimativa da área cultivada por expansão direta [1], [2], [7] e o segundo é o de realizar a validação do

mapeamento [8], [9]. Para cada bloco sorteado é preparado um material de campo envolvendo uma planilha eletrônica em que são obtidas diversas informações a respeito de cada um dos pontos sorteados, com destaque o uso e a cobertura do solo no ponto. Também são selecionadas imagens de satélite, de datas próximas ao trabalho de campo e livres de nuvens, para permitir a navegação no campo. Assim, depois de finalizado o plantio, equipes são selecionadas para ir ao campo coletar informações a respeito de cada um dos pontos. Durante o trajeto no interior destes blocos são coletados dados de treinamento para o mapeamento.

O mapeamento é realizado por meio do GEE [4] de cada um dos blocos utiliza todas as imagens Landsat/OLI e Sentinel/MSI disponíveis durante o período da safra e como classificador utiliza o Random Forest [10]. Este processo de mapeamento dos blocos ocorre de maneira iterativa, sendo que o fotointerprete utiliza os pontos de treinamento coletados no campo e imagens livres de nuvens para avaliar visualmente o mapeamento de cada bloco. Além disto, índices de precisão de mapeamento são informados para o fotointerprete, possibilitando a decisão de adicionar ou retirar treinamentos para minimizar os erros de omissão e inclusão. Posteriormente, após todos os blocos serem mapeados é realizada a estimativa da área para a região de estudo utilizando uma subamostra dos treinamentos para cada um dos blocos e é realizada uma segunda estimativa de área por expansão direta para cada um dos blocos e posteriormente a estimativa por regressão [1], [2], [7].

3. RESULTADOS E DISCUSSÃO

O processo para criação do design amostral envolveu inicialmente a criação das máscaras agrícolas e a obtenção do percentual cultivado em cada bloco para a estratificação. Ao todo a região é composta por 3.627 blocos, dos quais, aproximadamente 80% não tinham mais de 0,5% da área ocupada com agricultura e foram considerados como estrato não amostral. Os demais blocos, foram estratificados em 4 estratos e nos quais foram sorteados 30 blocos (Figura 2). A distribuição de blocos por estrato se deu da seguinte maneira, 10 blocos para o estrato 1, 9 para o estrato 2, 6 para o estrato 3 e 5 para o estrato 4. Em cada um destes blocos foram sorteados 20 pontos aleatoriamente, totalizando 600 pontos, dos quais todos os pontos com características agrícolas foram visitados. Para a realização do trabalho de campo nos 30 blocos sorteados a Companhia Nacional de Abastecimento (Conab) investiu aproximadamente R\$ 70.000,00. Este investimento envolveu aluguel de veículos, combustível, deslocamento e diárias dos técnicos para realizar o levantamento em campo.

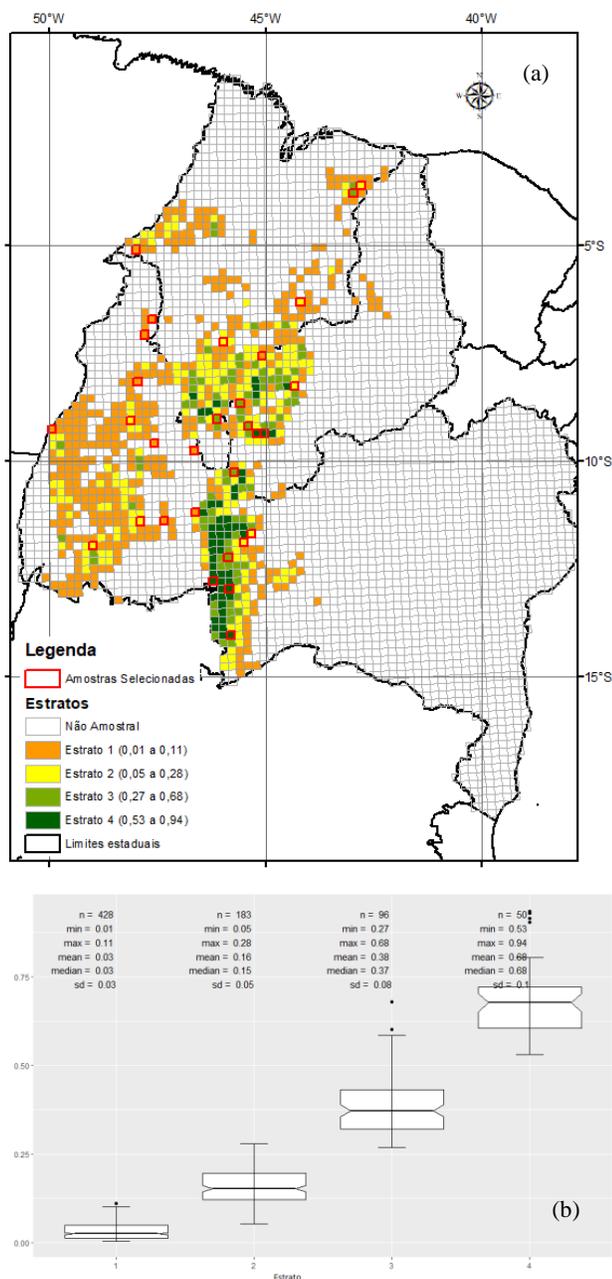


Figura 2. a) Estratificação e amostra selecionada e b) Percentual de área agrícola em cada estrato.

Logo após o término do trabalho de campo foi realizada a primeira estimativa de safras, utilizando a expansão direta (Tabela 1). Para toda essa região a Conab estimou uma área ocupada com soja de 4.660,50 mil ha. A primeira estimativa, por expansão direta a partir dos pontos, estimou uma área de 5.719,31, em torno de 22% menor do que a área estimada pela Conab. O coeficiente de variação (CV) desta estimativa foi de 10,6%, um coeficiente alto. Com o final do mapeamento foi realizada a segunda estimativa, utilizando a área de soja mapeada em cada um dos blocos. A segunda estimativa se aproximou mais das estimativas da Conab, com uma

diferença de aproximadamente 6% e um CV de 6%. E por fim, a estimativa por regressão ficou mais próxima ainda da estimativa realizada pela Conab, 2,2%, com um CV um pouco maior, de 6,8%. Estes resultados corroboram os resultados obtidos por [1], [2], [7][9], obtendo resultados similares aos obtidos por esses autores.

Modalidade	Área (mil ha)	CV (%)	Diferença Conab (%)
Pontos	5.719,31	10,59	22,72
Mapeamento	4.930,09	5,95	5,78
Regressão	4.761,54	6,78	2,17

Tabela 1. Área estimada (mil ha), coeficiente de variação (CV%) e diferença com o valor estimado pela Conab (%)

Posteriormente, utilizando todas as amostras de treinamento foi mapeada a área de soja nestes estados. Como pode ser observado na Figura 3, grande maioria da área concentra-se na região do MATOPIBA, apesar de ter algumas áreas expandindo para regiões mais ao Norte, no estado do Maranhão e ao Leste no estado do Piauí. Os resultados deste mapeamento contribuem para os levantamentos subjetivos de safra e fornecem insumos para os técnicos da Conab validar os levantamentos de safra.

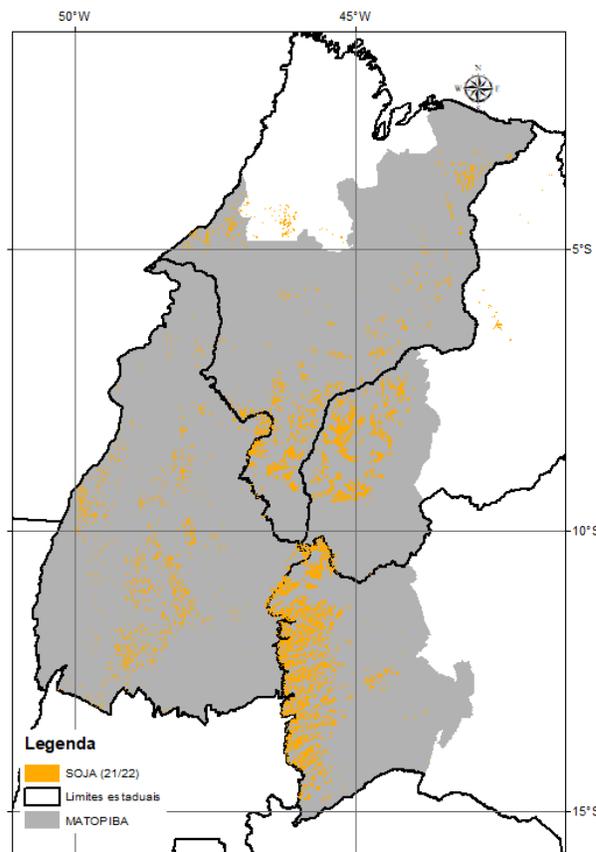


Figura 3. Área de soja na região de estudo.

Além disto, os resultados da validação do mapeamento apontaram para uma precisão geral de 98,7% ($\pm 1,2$), precisão

do usuário e do consumidor da ordem de 98% e 93%, respectivamente. Com isso possibilita a estimativa de área por estado, sendo que o estado que teve maior área de soja foi a Bahia, com 1,8 milhões de ha (Tabela 2). Entretanto, a maior diferença ocorreu no estado do Tocantins, com 9%. O estado em que ocorreu a menor diferença foi o estado do Maranhão, com 2% de diferença. No total, a diferença entre o estimado pela Conab e o mapeado foi da ordem de 1%.

UF	Área de soja (mil ha)		
	Conab	Mapeamento	Diferença (%)
MA	1.038,26	1.057,33	1,8%
PI	840,40	902,96	7,4%
BA	1.916,50	1.819,36	-5,1%
TO	1.057,46	1.150,43	8,8%
Total	4.876,80	4.930,09	1,1%

Tabela 1. Área estimada pela Conab e pelo mapeamento (mil ha) e a diferença entre ambas as estimativas (%).

Cabe aqui ressaltar que estas estimativas de área atendem aos critérios estabelecidos de oportunidade, confiabilidade, e transparência [1], [2], [7]. Pois, as estimativas foram obtidas em épocas oportunas, a primeira delas, a estimativa por pontos, foi realizada ainda com a safra em andamento. Todos os dados levantados e os mapeamentos apresentam as estimativas de erro associadas e todos os dados são passíveis de serem auditados. Além disto, a Conab está trabalhando juntamente com o INPE e a Embrapa no desenvolvimento de um método de estimativa de produtividade que integra modelos agrometeorológicos e dados espectrais para estimativa de produtividade. Conciliando estes dados com levantamento objetivos de produtividade, o que representa um avanço na estimativa de safra brasileira.

4. CONCLUSÕES

Este trabalho apresenta os dados preliminares de um sistema que integra informações de campo e associa os avanços na área de geotecnologias com as necessidades de atender aos critérios de oportunidade, confiabilidade e transparência, necessários para a realização de estimativa de safras agrícolas.

5. REFERÊNCIAS

[1] M. Adami, M. A. Moreira, B. F. T. Rudorff, C. da C. Freitas, and R. T. de Faria, "EXPANSÃO DIRETA NA ESTIMATIVA DE CULTURAS AGRÍCOLAS POR MEIO DE SEGMENTOS REGULARES," *Revista Brasileira de Cartografia*, vol. 57, no. 1, Nov. 2009, [Online]. Available: <https://seer.ufu.br/index.php/revistabrasileiracartografia/article/view/44959>

[2] M. Adami, M. A. Moreira, B. F. T. Rudorff, C. D. C. Freitas, R. T. de Faria, and F. Deppe, "Sampling

frame for crop area estimation," *Pesqui Agropecu Bras*, vol. 42, no. 1, 2007, doi: 10.1590/S0100-204X2007000100011.

[3] L. Zeng, B. D. Wardlow, D. Xiang, S. Hu, and D. Li, "A review of vegetation phenological metrics extraction using time-series, multispectral satellite data," *Remote Sens Environ*, vol. 237, p. 111511, Feb. 2020, doi: 10.1016/j.rse.2019.111511.

[4] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens Environ*, vol. 202, 2017, doi: 10.1016/j.rse.2017.06.031.

[5] R Core Team, "R: A Language and Environment for Statistical Computing." Vienna, Austria, 2021. [Online]. Available: <https://www.R-project.org/>

[6] G. Barcaroli, "SamplingStrata: An R Package for the Optimization of Stratified Sampling," *J Stat Softw*, vol. 61, no. 4, 2014, doi: 10.18637/jss.v061.i04.

[7] M. Adami, M. A. Moreira, B. F. T. Rudorff, and C. da C. FREITAS, "Análise da eficiência dos estimadores de expansão direta e de regressão par a áreas cultivadas com café, milho e soja no município de Cornélio Procópio, estado do Paraná," *Agricultura em São Paulo*, vol. 51, no. 2, pp. 5–13, 2004.

[8] P. Olofsson, G. M. Foody, M. Herold, S. v. Stehman, C. E. Woodcock, and M. A. Wulder, "Good practices for estimating area and assessing accuracy of land change," *Remote Sens Environ*, vol. 148, pp. 42–57, May 2014, doi: 10.1016/j.rse.2014.02.015.

[9] X.-P. Song *et al.*, "Massive soybean expansion in South America since 2000 and implications for conservation," *Nat Sustain*, vol. 4, no. 9, pp. 784–792, Sep. 2021, doi: 10.1038/s41893-021-00729-z.

[10] L. Breiman, "Random Forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.