

## ESTIMATIVA E MAPEAMENTO DE ÁREAS CULTIVADAS COM SOJA

*Patrícia Mauricio Campos<sup>1</sup>, Rafaela dos Santos Souza<sup>1</sup>, Fernando Arthur Santos Lima<sup>1</sup>, Lucas Barbosa Fernandes<sup>1</sup>, Marcos Adami<sup>2</sup>, Eunice Costa Gontijo<sup>1</sup>, Társis Rodrigo de Oliveira Lima<sup>1</sup>, Cleverton Tiago Carneiro de Santana<sup>1,2</sup>, Candice Mello Romero Santos<sup>1</sup>, Ieda Del'Arco Sanches<sup>2</sup>*

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

### RESUMO

Entre as ações para o aprimoramento das estatísticas agrícolas nacionais, a Conab e o INPE estão realizando o mapeamento de áreas cultivadas com soja por meio do sensoriamento remoto, geoestatística e informações de campo para estimar a área plantada e geolocalizar os cultivos. Até o momento, o trabalho compreendeu os estados de Mato Grosso, Mato Grosso do Sul, Goiás, Rondônia e a região do Matopiba. Foram gerados mapas de percentuais de área agrícola considerando uma grade de blocos de 20 km x 20 km de área e as máscaras de cultivo anuais de três anos/safra anteriores. Em blocos selecionados, foram coletadas amostras em campo para a classificação de áreas cultivadas com soja e outros cultivos. Esses dados serviram de base para a estimativa de área e o mapeamento da soja, que apresentaram boa exatidão. Concluiu-se que as técnicas aplicadas apresentaram resultados relevantes para o acompanhamento da safra agrícola.

**Palavras-chave** — estatística agrícola, aprendizado de máquina, classificação supervisionada, sensoriamento remoto, random forest.

### ABSTRACT

*To improve national agricultural statistics, Conab and INPE are mapping soybeans areas based on remote sensing, geostatistics and field data to estimate the planted area and spatialize the crop areas. So far, the study covered the Mato Grosso, Mato Grosso do Sul, Goiás and Rondônia states and the Matopiba region. Agricultural area percentages maps were generated crossing a grid of 20km x 20 km blocks and the annual crop masks from three previous crop seasons. A randomly group of blocks was used to sample and collect in loco data field classifying areas as soybeans areas and non-soybeans areas. These data contributed to the area estimation and the soybean mapping, showing a good precision and accuracy. Up to now, the applied techniques have been relevant to crop monitoring and forecasting.*

**Key words** — Agricultural statistics, machine learning, supervised classification, remote sensing, random forest.

### 1. INTRODUÇÃO

A importância da agricultura brasileira para a economia deve-se principalmente ao seu notável desempenho no cultivo de *commodities* para exportação como a soja. Na última safra 2021/2022, o Brasil teve destaque como maior produtor mundial dessa leguminosa [1], seguido dos Estados Unidos [2]. A produção brasileira foi de 124,0 milhões de toneladas, em uma área de 41,4 milhões de hectares, e nos EUA foi de 120,7 milhões de toneladas, em 33,4 milhões de hectares.

Em atendimento à demanda da comunidade nacional e internacional por melhorias nas estatísticas agrícolas, principalmente em função da necessidade de informações confiáveis referentes à produção de alimentos no mundo, a Companhia Nacional de Abastecimento (Conab), como instituição responsável pelas informações agrícolas no Brasil, tem trabalhado para que novas metodologias possam contribuir com dados de safra e de monitoramento agrícola com maior agilidade, exatidão e transparência.

Nesse contexto, a Conab e o Instituto Nacional de Pesquisas Espaciais (INPE) têm estudado e aplicado sensoriamento remoto, técnicas de aprendizado de máquina, geoestatística e dados de campo para a estimativa de área cultivada e a avaliação do comportamento da cultura ao longo do seu ciclo de desenvolvimento [3, 4].

O presente trabalho tem como objetivo apresentar a metodologia em estudo e seus resultados parciais na estimativa de área plantada e na geolocalização de cultivos de soja.

### 2. MATERIAL E MÉTODOS

A área de estudo compreendeu os estados de Mato Grosso, Mato Grosso do Sul, Goiás e Rondônia no ano/safra 2020/2021, e a região do Matopiba, composta por Maranhão, Tocantins, Piauí e Bahia no ano/safra 2021/2022. A metodologia foi aplicada em duas etapas, adaptada de [5, 6]. A primeira etapa consistiu na criação de um painel amostral que utilizou amostragem aleatória estratificada em duplo estágio para estimar a área cultivada e validar o mapeamento. Na segunda etapa, foi realizada a classificação automática, que considera a tecnologia de aprendizado de máquina, para mapear automaticamente as áreas cultivadas com soja.

## 2.1. Amostragem e mapeamento de cultivos anuais

Foram elaboradas máscaras de cultivos anuais dos 3 anos/safra anteriores ao ano a ser mapeado, analisado a partir das imagens produzidas pelo satélite/sensor Terra/MODIS (*Moderate Resolution Imaging Spectroradiometer*), além da observação de imagens Landsat 8/OLI (*Operational Land Imager*) e Sentinel-2/MSI (*Multispectral Instrument*) para contribuir na identificação dos alvos e na amostragem das áreas como agrícola e não agrícola.

Nos estados de Mato Grosso, Mato Grosso do Sul, Goiás e Rondônia, foram consideradas as imagens obtidas entre agosto/2017 e abril/2020. Para a região do Matopiba, foram consideradas as imagens obtidas entre agosto de 2018 e abril de 2021.

Com base nessas máscaras e na grade de blocos de 20 km x 20 km, foram calculadas as áreas ocupadas por agricultura em cada bloco nos estados de Mato Grosso, Mato Grosso do Sul, Goiás, Rondônia e na região do Matopiba. Após esse cálculo, foram gerados os estratos compostos pelos blocos e seus respectivos valores percentuais de área agrícola. O universo amostral contemplou os blocos com área agrícola superior a 400 ha e o painel amostral foi composto por blocos distribuídos aleatoriamente com base no percentual da área de cada bloco ocupada por agricultura. Dentro de cada bloco selecionado, também foram sorteados 20 pontos aleatórios para a verificação *in loco* das características do uso agrícola ou não agrícola de cada ponto amostral.

Os pontos previamente sorteados também foram utilizados para a validação do mapeamento. Além desses pontos, foram coletados pontos de treinamento para auxiliar na classificação dos blocos selecionados.

## 2.2. Mapeamento de área cultivada com soja

O classificador automático Random Forest [7] foi aplicado em cada bloco, gerando dois produtos: a área classificada e a probabilidade de cada um dos pixels de pertencer à classe de soja. Para a classificação, foram utilizadas todas as imagens disponíveis dos sensores OLI/Landsat 8 e MSI/Sentinel-2 no período entre outubro/2020 e abril/2021 nos estados de Mato Grosso, Mato Grosso do Sul, Goiás e Rondônia. Na região do Matopiba, foram utilizadas todas as imagens disponíveis dos sensores OLI (Landsat 8) e MSI (Sentinel-2) no período entre outubro/2021 e abril/2022.

Durante o deslocamento até os pontos de validação previamente sorteados, foram coletados pontos para treinamento do classificador.

De posse da probabilidade de ocorrência de soja para toda a região, foi calculada a probabilidade de corte da classificação para que a área mapeada fosse igual a área estimada pela regressão. Assim, foi gerado o mapeamento de áreas cultivadas com soja nos estados analisados, com área similar àquela resultante da estimativa por regressão.

Foram realizados ajustes manuais para inserir áreas não identificadas pelo classificador, mas identificadas pela equipe

de campo da Conab. Também foi realizada a análise e exclusão de áreas que foram consideradas como ruídos no mapeamento por apresentarem padrão distante do que poderia ser considerado como de uso agrícola.

## 2.3. Estimativa de área

Foram realizadas duas estimativas de área cultivada com soja [4, 5, 6]. A primeira foi calculada considerando os pontos amostrais por meio de expansão direta. A segunda estimativa de área foi calculada por regressão linear considerando o mapeamento dos blocos sorteados.

## 3. RESULTADOS

Foram gerados os mapeamentos de áreas cultivadas com soja nos estados de Mato Grosso, Mato Grosso do Sul, Goiás, Rondônia e na região do Matopiba, conforme apresentado nas Figuras 1 e 2, respectivamente.

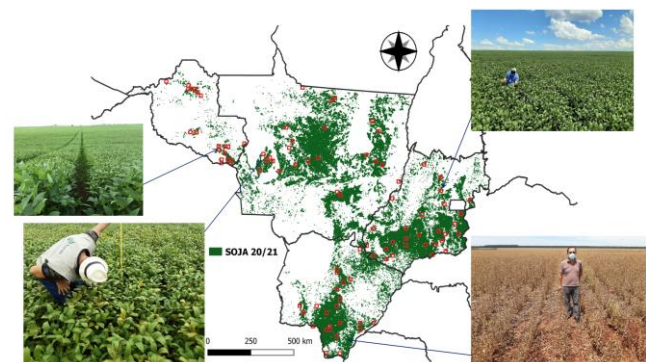


Figura 1. Localização dos blocos, fotos e áreas cultivadas com soja nos estados de Mato Grosso, Mato Grosso do Sul, Goiás, Rondônia.

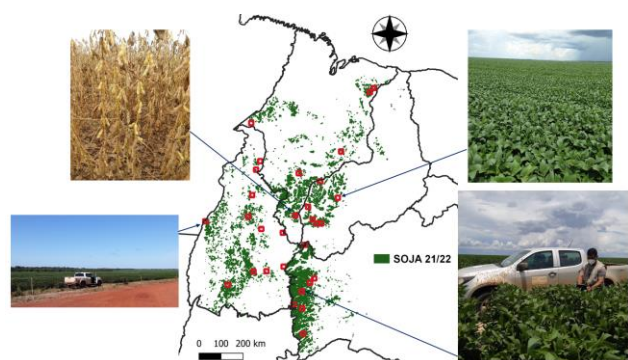


Figura 2. Localização dos blocos, fotos e áreas cultivadas com soja na região do Matopiba.

A validação dos mapeamentos foi realizada a partir de medidas obtidas por meio da Matriz de Confusão. Na Tabela 1, é apresentada a exatidão global, a exatidão do usuário e a exatidão do produtor para a cultura da soja em cada estado/região.

UF/Região	Exatidão Total (%)		Exatidão do Usuário (%)	Exatidão do Produtor (%)
Goiás	97,00	±1,05	97,64	98,79
Mato Grosso	97,82	±0,99	97,65	99,69
Mato Grosso do Sul	96,12	±1,3	96,23	98,93
Rondônia	98,58	±0,81	98,54	99,95
Matopiba	98,71	±1,2	98,18	93,68

Tabela 1. Métricas de validação do mapeamento.

A exatidão do usuário é associada à confiabilidade do mapeamento e é definida como a probabilidade de um ponto de soja em campo ser mapeado corretamente. Enquanto a exatidão do produtor é associada à sensibilidade do mapeamento. Ela é definida como a probabilidade de um ponto mapeado como soja realmente ser soja em campo.

UF/Região	Área (mil ha)	CV (%)	Diferença (%)
Goiás	3.683,93	9,2	-0,27
Mato Grosso	11.146,68	5,9	8,28
Mato Grosso do Sul	3.265,65	7,8	3,70
Rondônia	334,80	13,3	-15,56
Matopiba	5.728,28	10,56	17,02

Tabela 2. Estimativa de área cultivada com soja por expansão direta, coeficiente de variação e diferença entre a área estimada e o dado levantado pela Conab.

UF/Região	Área (mil ha)	CV (%)	Diferença (%)
Goiás	4.317,63	4,7	16,88
Mato Grosso	10.093,14	2,8	-1,95
Mato Grosso do Sul	3.543,29	4,6	12,52
Rondônia	381,72	7,1	0,16
Matopiba	4.761,54	6,78	- 3,73

Tabela 3. Estimativa de área cultivada com soja por regressão linear, coeficiente de variação e diferença entre a área estimada e o dado levantado pela Conab.

Os produtos desse trabalho foram disponibilizados na projeção SIRGAS 2000 e o cálculo de área por polígono utilizou a Projeção de *Albers Equal Area*. A escala do mapeamento é de 1:100.000 e foi empregada a base municipal do IBGE de 2019 [8] para definir os limites estaduais e municipais.

O resultado das estimativas de área cultivada com soja e

respectivos coeficientes de variação (CV), bem como a diferença entre o dado gerado com o uso de sensoriamento remoto e a área levantada em campo pela Conab, é apresentado nas Tabelas 2 e 3.

## 4. DISCUSSÃO

O uso de sensoriamento remoto aliado a técnicas de aprendizado de máquina e geoestatística apresentou resultados relevantes e tempestivos para a estimativa de área cultivada com soja. Sendo assim, foi possível constatar seu potencial no aprimoramento das estimativas agrícolas, conferindo maior transparência na formação de dados de safra.

Durante o processo de mapeamento foram identificados alguns pontos de atenção, que são discutidos nos próximos tópicos.

### 4.1. Mapeamento de áreas cultivadas com soja

A seleção de amostras de área agrícola e não agrícola para a elaboração das máscaras foi realizada com base na interpretação da composição de imagens, na análise do perfil espectral e no conhecimento de campo. Visando uma amostragem de boa qualidade, foi observada a importância da inserção de amostras representativas em diferentes perfis de uso e ocupação do solo. Outro fator importante foi em relação ao quantitativo de amostras, que variou conforme a existência desses diferentes perfis citados anteriormente, ficando a critério do observado pelo analista para oferecer ao classificador a correta interpretação do alvo a ser mapeado.

As máscaras geradas foram consideradas representativas para as regiões analisadas. No entanto, devido à diversidade de cultivos e o amplo calendário de plantio e colheita principalmente na região do Matopiba, foi observado que as amostras inseridas não foram suficientes para representar parte da região. Portanto, o mapeamento do Matopiba exigiu maior esforço para inserção de novas áreas e exclusão de áreas interpretadas como ruídos da classificação.

Apesar dessa necessidade de ajustes, a classificação automática se mostrou uma ótima ferramenta para a identificação de áreas agrícolas e de áreas cultivadas com soja, conforme exatidões apresentadas na Tabela 1. Considerou-se que houve bom desempenho, resultado também observado em outros trabalhos [3, 4, 5, 6].

### 4.2. Estimativa de área

As estimativas de área também apresentaram bom desempenho. Na tabela 2, apesar do coeficiente de variação ser superior à segunda estimativa (Tabela1), o esforço e custo para se obter esse resultado foi inferior. No entanto, em função do tamanho da área mapeada, é importante buscar reduzir essa variação. Assim, por meio da segunda estimativa, obteve-se um coeficiente de variação aceitável, permitindo ao analista ter uma margem para a tomada de

decisão quanto à área a ser adotada.

A amplitude do coeficiente de variação pode ser explicada pela diversidade de cultivos na área de estudo. Foi possível observar menor amplitude em Mato Grosso devido à maior homogeneidade de padrões de cultivo no estado, de modo que na primeira safra o cultivo de soja é o predominante nas áreas. No Matopiba, a maior amplitude pode ser explicada pela considerável coexistência de lavouras de soja e milho na primeira safra. Além disso, o calendário de plantio também é mais amplo nessa região. Essas características contribuíram para a maior amplitude do coeficiente de variação (Tabelas 2 e 3).

A estimativa de área nos estados de Mato Grosso, Mato Grosso do Sul, Goiás e Rondônia foram suficientes para o acompanhamento da safra e a tomada de decisão de forma tempestiva. Considerando que a Conab consolida as estimativas por Unidade da Federação, é importante manter esse recorte nos mapeamentos e nas estimativas de área resultantes desses mapeamentos. Dessa forma, espera-se facilitar o acompanhamento e a percepção de possíveis variações entre a área mapeada e a área levantada.

No Matopiba, foi possível ter a visão de área por estado apenas após a conclusão do trabalho, que foi obtida pelo recorte do mapeamento considerando a base cartográfica estadual IBGE [7], além do cálculo realizado por regressão. Portanto, foi necessário mais tempo para se obter esse dado. A estimativa por expansão e por regressão foi útil no sentido de se acompanhar a tendência de área na região, mas constatou-se a dificuldade de aplicação direta para a Conab.

## 5. CONCLUSÕES

As previsões de safra são fundamentais para a formulação de políticas públicas e a tomada de decisão pelo Governo Federal, agentes do setor e pela sociedade. Desse modo, é necessário que as estimativas expressem a realidade e estejam disponíveis em época oportuna para a maior assertividade da política agrícola brasileira, o desenvolvimento do setor agropecuário e a geração de renda do País.

Esse trabalho contribuiu para as estimativas de área e de produtividade, oferecendo informações precisas sobre a distribuição geográfica em cada estado. Na estimativa de área, o resultado do mapeamento auxiliou na análise da informação declarada, como um dado passível de verificação em campo, e na maior transparência na geração do dado. O conhecimento da localização das áreas de cultivo também possibilitou o monitoramento das áreas produtivas e o cruzamento de parâmetros agrometeorológicos, oferecendo indicativos sobre a previsão de rendimento das lavouras e aprimorando a metodologia de estimativa de produtividade da Companhia.

Conclui-se que o uso de sensoriamento remoto e a adoção de técnicas de geoestatística apresentaram bons resultados para o mapeamento e a estimativa de áreas cultivadas com soja. Portanto, o presente trabalho é escopo de pesquisa científica em andamento pelo INPE e a Conab

para ampliar a aplicação da metodologia em outras áreas de estudo.

## 6. REFERÊNCIAS

- [1] Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira de grãos. Volume 9, número 12, 87p, Setembro, 2022.
- [2] United States Department of Agriculture. World Agricultural Production. Circular Series WAP 9-22. 47 p. September 2022.
- [3] Song, X.P., Hansen, M.C., Potapov, P.V., Adusei, B., Pickering, J., Adami, M., Lima, A., Zalles, V., Stehman, S.V., Di Bella, C.M., Conde, M.C., Copati, E.J., Fernandes, L.B., Hernandez-Serna, A., Jantz, S.M., Pickens, A.H., Turubanova, S., Tyukavina, A. Massive soybean expansion in South America since 2000 and implications for conservation. *Nat Sustain* 4, 784–792 (2021). <https://doi.org/10.1038/s41893-021-00729-z>.
- [4] Fernandes, L.B., Lima, F.A.S., Santos, C.M.R., Piffer, T.R.O., Nascimento, J.L.S. Two-Stage Sampling Area Estimate Summer-Planted Maize in Federal District - Brazil - Season 2017/18. Poster apresentado no Living Planet Symposium 2019 - ESA, Milão, Itália, Maio de 2019.
- [5] Song, X.P., Potapov, P.V., Krylov, A., King, L.A., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, A.V., Hansen, M.C. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. *Remote Sensing of Environment*, 190:383 – 395, 2017.
- [6] Potgieter, A. B. , Lawson, K. , Huetec, A. R. Determining crop acreage estimates for specific winter crops using shape attributes from sequential MODIS imagery. *International Journal of Applied Earth Observation and Geoinformation*. Volume 23, 254-263, August 2013.
- [7] Breiman, L. Random Forests, *Machine Learning*, vol. 45, pp. 5–32. 2001. <https://doi.org/10.1023/A:1010933404324>
- [8] Instituto Brasileiro de Geografia e Estatística. Malha Municipal Digital da Divisão Político- Administrativa Brasileira. 2019.