MICHEL EUSTÁQUIO DANTAS CHAVES

USO DAS SÉRIES TEMPORAIS DE ÍNDICES DE VEGETAÇÃO PARA O MONITORAMENTO AGRÍCOLA NO ESTADO DE MATO GROSSO

LAVRAS-MG
2019

MICHEL EUSTÁQUIO DANTAS CHAVES

USO DAS SÉRIES TEMPORAIS DE ÍNDICES DE VEGETAÇÃO PARA O MONITORAMENTO AGRÍCOLA NO ESTADO DE MATO GROSSO

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, área de concentração em Engenharia Agrícola, para a obtenção do título de Doutor.

Prof. Dr. Marcelo de Carvalho Alves
Orientador

LAVRAS-MG
2019

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Chaves, Michel Eustáquio Dantas.
Uso das séries temporais de índices de vegetação para o monitoramento agrícola no Estado de Mato Grosso / Michel Eustáquio Dantas Chaves. - 2018.
148 p.

Orientador: Marcelo de Carvalho Alves.
Tese (doutorado) - Universidade Federal de Lavras, 2018.
Bibliografia.

1. Agricultura. 2. Sensoriamento Remoto. 3.Geoprocessamento. I. Alves, Marcelo de Carvalho. II. Título.

MICHEL EUSTÁQUIO DANTAS CHAVES

USO DAS SÉRIES TEMPORAIS DE ÍNDICES DE VEGETAÇÃO PARA O MONITORAMENTO AGRÍCOLA NO ESTADO DE MATO GROSSO

USE OF VEGETATION INDICES TIME SERIES FOR AGRICULTURAL MONITORING IN THE STATE OF MATO GROSSO

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, área de concentração em Engenharia Agrícola, para a obtenção do título de Doutor.

APROVADA em 05 de dezembro de 2018.

Dr. Marcelo Silva de Oliveira	UFLA
Drª. Thelma Sáfadi	UFLA
Drª. Elizabeth Ferreira	UFLA
Drª. Margarete Marin Lordelo Volpato	EPAMIG

Prof. Dr. Marcelo de Carvalho Alves
Orientador

LAVRAS-MG
2019

A Deus,

À minha Mãe, Ansheridan Dantas de Maria, ao meu Pai, José Eustáquio Chaves (in memorian), ao meu Irmão, Andrey Eustáquio Dantas Chaves e à minha namorada, Katyanne Viana da Conceição pelo amor, confiança, companheirismo, respeito e exemplo de vida, sempre fornecendo incentivo e suporte para o meu desenvolvimento pessoal e profissional,

Dedico.

AGRADECIMENTOS

Em primeiro lugar, a Deus, pelo dom da vida e por me dar companhia, sabedoria e discernimento em todos os momentos.

De forma especial, à minha Mãe Ansheridan e ao meu Irmão Andrey, pela confiança e pelos exemplos de honestidade e integridade, e à minha namorada Katyanne, pela lealdade e presença constante, sempre me encorajando a buscar os meus objetivos.

Ao amigo Padre Arnoldo e sua Família pelos conselhos e ensinamentos.

Em âmbito acadêmico, à Universidade Federal de Lavras – UFLA e ao Departamento de Engenharia – DEG pela oportunidade, bem como à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela concessão de bolsa de estudos.

Especificamente, ao meu orientador, Dr. Marcelo de Carvalho Alves pela orientação, confiança, conhecimento e amizade, bem como à Drª. Thelma Sáfadi e ao Dr. Marcelo Silva de Oliveira pelo conhecimento e respeito. Às Doutoras Elizabeth Ferreira e Margarete Marin Lordello Volpato pela participação na banca examinadora, contribuições ao trabalho e ensinamentos.

Aos Professores de inglês do Núcleo de idiomas da UFLA, Semírames Ávila, Evandro Furtado e Ernani Augusto, pelos ensinamentos de inglês que me ajudaram na escrita e nas pesquisas acadêmicas.

Aos amigos João Vitor Guerrero, Guilherme Mataveli, Rodrigo Vasconcelos, Alex Bocoli, João Paulo, José Augusto, Amarildo Ferreira, Jonathan Rocha, Marcus André, Matheus Cortez, Pedro Arthur, Igor de Andrade, Diego Marin, Rômulo Gandia, Filipe Trindade, Júlia Rodrigues, Miryan Pires, Jade Alacoque, Jefferson Soares, Gleydson Campos, Lucas Santana, amigos do futebol, bem como outros que, com incentivo e conhecimento, contribuíram com o desenvolvimento desta tese.

“Todas as vitórias ocultam uma abdicação”. (Simone de Beauvoir).

RESUMO GERAL

A agricultura é um dos principais setores da balança comercial brasileira, especialmente a partir da virada do século. Neste cenário, o Estado de Mato Grosso se destaca. Chamado de “Celeiro do mundo”, teve sua paisagem alterada com a expansão da fronteira agrícola sobre os biomas Cerrado e Amazônia, transformando-se em potência mundial do setor, com grandes glebas de produção em seu território. Considerando as dimensões estaduais, a distribuição e a dinâmica das atividades agropecuárias, é importante destacar a utilidade de satélites e sensores remotos para o monitoramento de grandes extensões territoriais agrícolas, por sua capacidade de coletar informações detalhadas e confiáveis com alta frequência de revisita, podendo ser associados aos sistemas de obtenção de dados agrícolas convencionais. As séries temporais provenientes de produtos compostos voltados à vegetação, tais como o MOD13Q1 e o MYD13Q1 do sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS), possibilitam avaliar o comportamento da vegetação por meio da identificação de sazonalidades e tendências inerentes aos ciclos fenológicos de culturas agrícolas. Na parte espacial, técnicas de geoestatística aliadas às informações *in situ* e de censo permitem quebrar as barreiras de escala e criar o efeito de *downscaling*, fator que apresenta potencial para identificar áreas de culturas, bem como estimar a produtividade agrícola com menor incerteza. Diante das inovações tecnológicas, este estudo apresenta a utilização de técnicas de geoestatística e de análise de séries temporais de índices de vegetação dos produtos supracitados para a derivação de parâmetros do ciclo fenológico, com o intuito de identificar, interpretar e monitorar a evolução de áreas agrícolas em Mato Grosso entre os anos de 2000 e 2012, que representam o período de crescimento exponencial da agricultura após a virada do século. Em escala estadual, foi possível identificar as áreas de soja e sua intensificação ao longo do período com Acurácia Global de 92,1% e índice Kappa de 0,84, bem como a produtividade em cinco aglomerados de fazendas de diferentes mesorregiões com 95,09% de acurácia, considerando o desvio padrão e o erro provável. Para disseminar os mapas de área e produtividade obtidos, foi desenvolvida uma plataforma web intitulada SojaSAT. Em nível de aglomerado de fazendas e de talhão, foi possível identificar culturas na safra e na safrinha com Acurácia Global de 89,5% e índice Kappa de 0,80. Os resultados obtidos mostram a utilidade das séries temporais MODIS aliadas às técnicas de geoestatística e de análise de séries temporais para monitorar o ciclo fenológico das culturas, tornando possível identificar as áreas destinadas ao plantio de diferentes cultivos de acordo com as respostas espectro-temporais detectadas.

Palavras-chave: Agricultura. Sensoriamento Remoto. Geoprocessamento. Análise de séries temporais. Geoestatística. Reconhecimento de padrões. Índices de vegetação.

GENERAL ABSTRACT

The agriculture is one of main sectors of the Brazilian trade balance, especially from the century begin. In this scenario, the State of Mato Grosso stands out. Called the "Barn of the World", its landscape changed with the agricultural frontier expansion over the Cerrado and Amazon biomes, becoming a world power of the sector, with large areas of production in its territory. Considering the state dimensions, distribution, and dynamics of agricultural practices, it is important to highlight the usefulness of remote satellites and sensors for the monitoring of large agricultural territorial extensions, for their ability to collect detailed and reliable information with a high revisit frequency, associated with conventional agricultural data collection systems. The time series from vegetation-oriented composite products such as MOD13Q1 and MYD13Q1 from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor make it possible to evaluate vegetation dynamic by identifying seasonalities and trends inherent to phenological cycles of agricultural crops. In the spatial part, geostatistical techniques combined with in situ and census data allow to break down the scale barriers and create the downscaling effect, a factor that presents potential to identify agricultural crops and to estimate yield with less uncertainty. In face of technological innovations, this study presents the use of geostatistics techniques and vegetation indices time series analysis for the derivation of phenological cycle parameters, in order to identify, interpret and monitor the evolution of agricultural areas in Mato Grosso between 2000 and 2012, period of exponential agricultural expansion after the turn of the century. At the state level, it was possible to identify soybean areas and their intensification throughout the period with a Global Accuracy of 92.1% and a Kappa index of 0.84, as well as productivity in five agglomerations of farms of different mesoregions with 95.09% of accuracy, considering the standard deviation and probable error. In order to disseminate the area and yield maps obtained, a web platform was developed entitled SojaSAT. At the level of an agglomerate of farms and field, it was possible to identify crops in the harvest and in the safrinha with Global Accuracy of 89.5% and Kappa index of 0.80. The results obtained show the usefulness of the MODIS time series combined with geostatistical techniques and time series analysis to monitor the phenological cycle of the crops, making it possible to identify areas for planting different crops according to the detected spectro-temporal responses.

Keywords: Agriculture. Remote Sensing. Geoprocessing. Time series analysis. Geoestatistics. Patterns recognition. Vegetation indices.

SUMÁRIO

1 INTRODUÇÃO

A agricultura brasileira é um dos principais campos de análise em que são empregadas técnicas de sensoriamento remoto. Bastante diversificada, abrange culturas consideradas vitais para a alimentação humana, bem como se destaca em outras atividades, tais como a produção de fibras, de grãos para alimentação animal e de energia, em grande pluralidade de ambientes, regimes climáticos e sistemas de manejo. Diante dessa diversidade, o setor agrícola necessita de informações precisas e contínuas sobre a condição das culturas, a localização dos campos de cultivo, a medição de suas áreas, as estimativas de produção, a identificação da sucessão entre as culturas e o monitoramento da dinâmica agrícola para diversos fins, sobretudo para o planejamento da cadeia produtiva, em distintos níveis.

A dinâmica desta diversificada agricultura nacional teve aumento exponencial a partir da virada do século XX para o XXI, e as transformações abruptas neste setor a partir desse momento atingiram diversas esferas da cadeia agrícola nacional, causando impactos que vão desde os tratos culturais iniciais à precificação de produtos no mercado internacional, acarretando consequências existentes até os dias atuais.

Rico em biodiversidade, sendo composto pelos biomas Amazônico, Cerrado e Pantanal e, ao mesmo tempo, chamado de “Celeiro do mundo”, Mato Grosso pode ser considerado o Estado brasileiro que teve a sua paisagem mais alterada ao longo dos anos por conta do avanço da agricultura, o que resultou em um cenário extremamente heterogêneo, esculpido pelo ritmo intensivo de conversão de uso e cobertura da terra. Com a expansão da fronteira agrícola sobre o Cerrado e a Amazônia, Mato Grosso transformou-se em potência mundial do setor, com grandes glebas de produção intensiva espalhadas por seu território.

Pela importância da agricultura, discute-se que a utilização de sistemas inovadores para identificar culturas agrícolas e estimar safras é essencial para orientar políticas que contemplem a determinação de preços, a manutenção de estoques reguladores e o direcionamento de investimentos, e que a eliminação de subjetividades e imprecisões pode otimizar a produção nas lavouras, do plantio à comercialização, indicando informações sobre impactos negativos ou ocasiões favoráveis que podem refletir na economia.

Ao mesmo tempo, discute-se que estimativas de safra conduzidas por agentes nacionais se apresentam como um mecanismo de proteção da produção nacional, pois geram um referencial capaz de impedir ou diminuir a ação especulativa causada por estimativas geradas por agentes externos ao país, muitas vezes pertencentes a países concorrentes no mercado internacional. Todo este cenário faz com que seja necessário eliminar a subjetividade e a imprecisão das estimativas de área e produtividade de culturas agrícolas, que, aliadas à falta de informações em tempo hábil, causam problemas na tomada de decisão em todos os níveis da cadeia produtiva nacional, afetando o armazenamento, o escoamento, e a definição de preços, fazendo com que o Brasil perca dinheiro em todos os níveis do setor agrícola.

No Brasil, as estatísticas agrícolas oficiais são feitas e divulgadas pela Companhia Nacional de Abastecimento (CONAB) e pelo Instituto Brasileiro de Geografia e Estatística (IBGE). No entanto, os métodos utilizados por estes órgãos, embora asseverados, se valem da aplicação de questionários, entrevistas com representantes do setor agropecuário e amostragens como dado de entrada, o que aumenta o risco da subjetividade e pode gerar um produto final com maior grau de imprecisão. Além disso, a aquisição de informações não apresenta agilidade e os custos para o emprego dos métodos são altos. Assim, programas de melhoria e aperfeiçoamento metodológico têm sido incentivados, visando aprimorar a obtenção de informações em tempo ágil e sob custos menores.

As inspeções *in situ*, com o acompanhamento do dia-a-dia em cada talhão, apresentam utilidade para o setor. Porém, tendo em conta as dimensões territoriais de Mato Grosso e a dinâmica com que ocorrem as atividades agropecuárias no Estado, é oneroso e praticamente inviável o acompanhamento *in situ* de toda a produção. Por isso, é importante que os responsáveis pela cadeia produtiva de Mato Grosso disponham de tecnologias de estudo e monitoramento de suas culturas, visando a eliminação de incertezas e especulações. Este panorama é, igualmente, importante em nível de aglomerado de fazendas e de talhão.

Diante disso, novas metodologias têm sido propostas para aumentar a eficiência do setor de estimativas de área e produtividade de safras no Brasil. Para tanto, são vastas as aplicações que podem ser úteis à detecção da dinâmica agrícola e à indicação de mudanças gradativas no desenvolvimento fenológico. Especificamente, o uso de sensoriamento remoto aliado a técnicas de geoestatística, análise de séries temporais, linguagem de programação e inteligência artificial, pode favorecer a identificação de processos na superfície terrestre.

O sensoriamento remoto pode fornecer dados precisos sobre o setor agrícola, sendo as geotecnologias provavelmente os melhores meios, em termos de custo-benefício, para a coleta de informações detalhadas e confiáveis sobre grandes áreas, com alta frequência de revisita, o que favorece o monitoramento e a interpretação da dinâmica do comportamento fenológico das culturas. Isso ocorre, pois, os sistemas sensores registram a interação entre a radiação eletromagnética e as substâncias que compõem a superfície terrestre, e detectam processos e fenômenos sem contato direto com os alvos, gerando produtos e informações que o olho humano não seria capaz de observar, em uma forma adequada para a interpretação do usuário.

Em virtude de sua elevada capacidade de fornecer dados com agilidade, repetitividade, abrangência geográfica, baixo custo operacional e objetividade, os sistemas satelitários de sensoriamento remoto para a observação terrestre, possuem potencial para serem associados aos sistemas de obtenção de dados agrícolas convencionais que já vêm sendo utilizados há décadas no caso brasileiro por instituições como a CONAB e o IBGE.

Entre os sensores disponíveis, um de particular importância para a avaliação da agricultura é o *Moderate Resolution Imaging Spectroradiometer* (MODIS), que está a bordo dos satélites TERRA e AQUA. O sensor MODIS foi desenvolvido para subsidiar investigações acerca das mudanças no uso e ocupação da superfície terrestre, e possui frequência temporal quase diária e resolução espacial de 250 metros, características compatíveis com o tamanho das lavouras mato-grossenses e sua dinâmica agrícola, além de fornecer produtos compostos com qualidade geométrica e com correção atmosférica.

O MODIS está diretamente relacionado à agricultura devido ao emprego de séries temporais de índices de vegetação por ele fornecidos, o *Enhanced Vegetation Index* (EVI) e o *Normalized Difference Vegetation Index* (NDVI), que são úteis como parâmetros para avaliar o comportamento espectral da vegetação agrícola durante o transcorrer dos ciclos fenológicos. A análise de séries temporais de índices de vegetação MODIS visando detectar mudanças no uso da terra e monitorar o avanço de culturas, bem como a elaboração de algoritmos para a eliminação de ruídos que causam alterações na interpretação dos resultados e para a extração de padrões de trajetórias espectro-temporais, vem apresentando bons resultados para a estimativa de áreas e de produtividade de culturas com precisão e acurácia, diminuindo a necessidade e os custos de ir a campo.

Justamente por possuírem características como a frequência diária de disponibilização de dados, as séries temporais de índices de vegetação MODIS estão diretamente relacionadas com os ciclos fenológicos. As grandes vantagens do MODIS para o monitoramento das modificações da superfície terrestre estão na qualidade geométrica das imagens, na melhoria na qualidade dos detectores e do sistema de imageamento, no maior número de bandas e nos algoritmos específicos para a geração dos produtos compostos. Outra vantagem é a relação direta de índices de vegetação com os ciclos fenológicos das culturas. No início do ciclo fenológico, quando a quantidade de fitomassa é escassa e a resposta espectral é influenciada pelo solo, os valores de EVI e NDVI são baixos. À medida em que a cultura se desenvolve e a produção de fitomassa aumenta, os valores se elevam até o pico vegetativo. Com a senescência e a colheita, eles diminuem, até atingirem os níveis encontrados no início.

A possibilidade de ordenamento cronológico de imagens, permitindo a análise em série, é outro fator que oportunizou o uso de dados fornecidos pelo sensor MODIS para avaliar campos de cultivo. A análise de uma série temporal permite verificar a existência de tendências e periodicidades nos dados, o que não é permitido com a análise de imagens e parâmetros não-sequenciais. Especificamente em relação à agricultura, as séries temporais provenientes de índices de vegetação do sensor MODIS podem favorecer a observação de padrões típicos de distintos usos agrícolas.

Contribuindo com a análise de séries temporais, estão a utilização de métricas fenológicas ou parâmetros sazonais e de reconhecimento de padrões, que aumentam a captação de diferenças sutis e o grau de separabilidade entre classes. A aplicação destas técnicas visa melhorar a identificação da área das culturas, e, por consequência, aprimorar os modelos de estimativa de safra que possuem esta informação como dado de entrada.

As alternâncias de práticas na lavoura a cada safra e a rotação e sucessão de culturas são fatores que dificultam a adequada contextualização da agricultura e a formulação de estimativas de safras mais precisas, sendo necessário abordálas na classificação de áreas. Além das mudanças anuais de ciclos fenológicos causados pelo clima ou por variações nas práticas agrícolas, o mapeamento da agricultura baseado em análise de séries temporais é igualmente dificultado pela falta de amostras usadas para treinar o algoritmo supervisionado e pelo obscurecimento causado por nuvens. Porém, com o avanço tecnológico e a implementação de novas técnicas, o reconhecimento de padrões em séries temporais provou ser uma solução eficiente para minimizar estes problemas e lidar com distorções temporais que ocorrem de um ciclo fenológico para o outro.

Neste contexto, esta tese aborda a aplicação de técnicas de sensoriamento remoto, geoprocessamento, geoestatística e análise de séries temporais para avaliar a evolução espacial e temporal de culturas agrícolas e a geração de estimativas de produtividade na primeira década do século XXI no Estado de Mato Grosso, com destaque para a cultura da soja, por ser a mais plantada. As análises possuem o suporte de dados de campo de aglomerados de fazendas mato-grossenses. Estes dados contém as informações, em nível de talhão, de datas de plantio, germinação e colheita, da textura do solo de cada campo de cultivo, da área total plantada e do total de sacas de 60 kg colhidas em cada talhão ao final de cada período de colheita.

Os resultados obtidos representam contribuições a respeito da utilização do sensoriamento remoto, especialmente do sensor MODIS e de seus produtos compostos de índices de vegetação, dados obtidos *in situ* e técnicas de geoestatística, análise de séries temporais e classificação de imagens de satélite para otimizar para a identificação de áreas e a estimativa de produtividade de culturas agrícolas em diferentes níveis, desde o talhão de uma fazenda ao total de área destinada à agricultura no Estado de Mato Grosso. Outra contribuição está na abordagem utilizada para elaboração e o funcionamento de uma plataforma virtual valendo-se de técnicas e aplicações gratuitas para a difusão de conhecimento.

No campo computacional, esta tese oferece ao campo científico mais uma confirmação de que o futuro das análises acerca da agricultura, em diferentes escalas, está diretamente ligado ao trabalho com grande volume de dados (*big data*) e à utilização de técnicas de linguagem de programação computacional e algoritmos específicos para a geração de informações a respeito das lavouras sem grandes custos e em curto período de tempo, fator que confere competitividade aos métodos desenvolvidos e/ou aplicados. Do mesmo modo, com a plataforma virtual desenvolvida, a disseminação de conteúdo economiza tempo de processamento e esforços computacionais aos usuários, facilitando a averiguação das condições da agricultura em Mato Grosso.

Do mesmo modo, os resultados obtidos contribuem para o entendimento do avanço e da dinâmica da agricultura em Mato Grosso, sendo eficazes para auxiliar processos de tomada de decisão que abrangem o monitoramento agrícola no contexto conceitual de uma agricultura com alta produção comercial e heterogênea.

1.1 Pergunta científica

A pergunta científica que motivou a escrita desta tese é:

Poderia a utilização e análise das séries temporais de índices de vegetação do sensor MODIS, em conjunto com a aplicação de técnicas de geoestatística, permitir a identificação de áreas agrícolas e não agrícolas, a estimativa de produtividade da cultura da soja e, consequentemente o monitoramento agrícola no Estado de Mato Grosso, nas escalas estadual, mesorregional, microrregional, municipal, de aglomerado de fazendas, de fazendas e de talhão?

1.2 Objetivo geral

O objetivo geral desta tese é identificar as áreas cultivadas com soja, algodão e milho no Estado de Mato Grosso, compreender a dinâmica das lavouras, realizando o monitoramento de culturas em diferentes safras, e gerar estimativas de produtividade com base em técnicas de sensoriamento remoto, análise das séries temporais e geoestatística.

1.3 Objetivos específicos

Os objetivos específicos desta tese foram:

a) Propor um índice de vegetação específico para a identificação das áreas de cultivo de soja em Mato Grosso, o *Soybean Enhanced Index* (SEI);
b) Identificar as áreas da expansão do setor produtivo agrícola ao longo das safras avaliadas;
c) Estimar a produtividade de talhões de soja nas safras avaliadas, com base nos valores referentes ao vigor da biomassa obtido pelos índices de vegetação, por meio de técnicas de geoestatística;
d) Mitigar a interferência dos problemas *Modifiable Area Unit Problem* (MAUP) e *Change Of Support Problem* (COSP), que estão relacionados à escolha adequada da escala de estudo;
e) Obter o efeito de *Downscaling* nos resultados de identificação de área e produtividade, quebrando a barreira de dependência do tamanho dos *pixels*;
f) Caracterizar a evolução temporal do ciclo fenológico das culturas de soja, milho e algodão em Mato Grosso e estabelecer comparações e relações entre os perfis temporais gerados, os calendários agrícolas

das culturas e os dados oficiais, com o intuito de avaliar se a informação derivada de satélite corresponde aos dados *in situ*;

g) Detectar variações interanuais nas lavouras dos aglomerados de fazendas utilizados para a validação do método.
h) Disseminar o conteúdo gerado na tese, disponibilizando os resultados obtidos em plataforma virtual desenvolvida, a plataforma virtual SojaSAT.

1.4 Organização

Seguindo as normas da Universidade Federal de Lavras – UFLA para a elaboração de trabalhos acadêmicos no modelo de estrutura de artigo, esta tese está organizada em duas partes, como segue:

A primeira é esta, a Introdução geral. Nela, estão contidos o Resumo geral e a Introdução geral da tese, contendo a contextualização do problema observado, a hipótese, o objetivo geral e os objetivos específicos, bem como a organização da tese.

A segunda parte da tese consiste na apresentação de quatro artigos científicos, cada qual redigido com as normas específicas dos periódicos escolhidos para envio e publicação. O primeiro artigo, intitulado ***"Recent applications of the MODIS sensor for soybean crop monitoring and deforestation detection in Mato Grosso, Brazil"***, apresenta uma revisão de literatura com o intuito de exibir a pesquisa de *background* sobre a aplicação prática do sensor *MODerate resolution Imaging Spectroradiometer* (MODIS) como ferramenta para avaliações na agricultura mato-grossense, atestando seu potencial para uso. Este artigo foi publicado no periódico *CAB Reviews - Perspectives in agriculture, veterinary science, nutrition and natural resources,* ISSN: 1749-8848.

No segundo artigo, intitulado **“A geostatistical approach for modeling soybean crop area and yield based on census and remote sensing data”**, os problemas Modifiable Area Unit Problem (MAUP) e Change Of Support Problem (COSP), relacionados à escolha adequada da escala de estudo, foram abordados com a proposição de um índice de vegetação específico para a identificação das áreas de cultivo de soja em Mato Grosso, o Soybean Enhanced Index (SEI), e a aplicação de técnicas de geoestatística para a estimativa de produtividade nas áreas identificadas. Estes problemas afetam as estimativas, e estas técnicas se complementaram no monitoramento da área e da produção em nível estadual. Ademais, o uso de dados in situ na etapa de validação permitiu a obtenção de valores sobre a produtividade de um conjunto de unidades espaciais a partir de dados de outro conjunto de unidades espaciais, proporcionando o efeito de downscaling nos resultados, quebrando a barreira de dependência do tamanho do pixel.

Esta aplicação conjunta permitiu o monitoramento da distribuição espacial destes fenômenos, em nível de talhão, ao longo do período de maior intensificação da sojicultura em Mato Grosso, entre 2000 e 2011. Este artigo foi publicado no periódico *Remote Sensing,* ISSN: 2072-4292 (http://www.mdpi.com/2072-4292/10/5/680).

O terceiro artigo, intitulado ***“Agricultural dynamic detected by pattern recognition in MODIS satellite image time series of Mato Grosso”***, apresenta uma abordagem para o uso de séries temporais de índices de vegetação, que evidenciam a evolução do vigor vegetativo ao longo dos ciclos fenológicos, como dado de entrada para resolver o problema da identificação de culturas agrícolas, tanto na safra quanto na safrinha, de forma interanual, em um aglomerado de fazendas com intensa dinâmica agrícola, o Santa Luzia, em Sapezal-MT. A identificação de cultivos na safra e, também, na safrinha, em regiões com intensa dinâmica agrícola é um problema para o sensoriamento remoto por conta da falta de uniformidade nos plantios e nos ciclos fenológicos das variedades e da semelhança espectral entre culturas.

Esta análise foi feita com o uso do algoritmo *Time-Weighted Dynamic Time Warping* (TWDTW) para a análise de trajetórias de séries temporais e demonstra como o uso de métricas representativas de momentos importantes dos ciclos fenológicos e técnicas de reconhecimento de padrões em séries temporais pode auxiliar na detecção de culturas na safra e na safrinha. Este método demonstrou, em artigos encontrados na literatura e em testes prévios, sensibilidade às mudanças sazonais da vegetação e considerar a variabilidade climática e sazonal interanual, detectando padrões temporais, aplicando critérios para a separabilidade de classes e lidando com distorções temporais, fatores que lhe conferem potencial para avaliar as culturas plantadas em Mato Grosso. Este artigo está em tramitação no periódico *Land Use Policy,* ISSN: 0264-8377.

O quarto artigo, intitulado ***"SojaSAT: web platform to monitoring soybean crops in Mato Grosso"***, aborda a construção de uma plataforma virtual, a SojaSAT, para a disseminação de parte dos resultados obtidos durante a tese, que correspondem à identificação de áreas de cultivo de soja e à estimativa de produtividade em Mato Grosso.

A disseminação deste conteúdo é importante, pois as informações obtidas podem servir como fonte de informação auxiliar para fins que vão desde o acompanhamento histórico da produção de soja a partir da virada do século até a compreensão da expansão ocorrida e o seu escoamento, podendo ser comparados com dados atuais e evitar uma classificação desta época para análise temporal. Ressalta-se que a plataforma SojaSAT foi desenvolvida com tecnologias gratuitas e livres, e que o artigo está em fase final de preparação para submissão.

SEGUNDA PARTE - ARTIGOS

Nesta parte, são apresentados os artigos de resultados obtidos da tese, redigidos com as normas específicas dos periódicos escolhidos para envio e publicação.

O primeiro artigo apresentado, intitulado *"Recent applications of the MODIS sensor for soybean crop monitoring and deforestation detection in Mato Grosso, Brazil"*, foi publicado no volume 14, número 007, do periódico *CAB Reviews - Perspectives in agriculture, veterinary science, nutrition and natural resources*, ISSN: 1749-8848.

O segundo artigo, intitulado "*A geostatistical approach for modeling soybean crop area and yield based on census and remote sensing data*", foi publicado no volume 10, número 5, periódico *Remote Sensing*, ISSN: 2072-4292 (http://www.mdpi.com/2072-4292/10/5/680).

O terceiro, intitulado *"Agricultural dynamic detected by pattern recognition in MODIS satellite image time series of Mato Grosso"*, foi submetido e está em tramitação no periódico *Land Use Policy*, ISSN: 0264-8377.

Por fim, o artigo intitulado *"SojaSAT: web platform to monitoring soybean crops in Mato Grosso"*, foi submetido e está em tramitação no periódico *Applied Geomatics*, ISSN: 1866-9298.

ARTIGO 1 - RECENT APPLICATIONS OF THE MODIS SENSOR FOR SOYBEAN CROP MONITORING AND DEFORESTATION DETECTION IN MATO GROSSO, BRAZIL

Normas do periódico *CAB Reviews - Perspectives in agriculture, veterinary science, nutrition and natural resources*, ISSN: 1749-8848.
Versão publicada: (https://www.cabdirect.org/cabdirect/abstract/20193059813)

M E D Chaves, M de C Alves.

Authors ORCID.org: 0000-0002-1498-6830; 0000-0003-3957-328X.

Address: Engineering Department, Federal University of Lavras. Campus Universitário, PO Box 3037, ZIP code 37200-000, Lavras, Brazil.

Correspondence: Marcelo de C Alves. Email: marcelo.alves@deg.ufla.br

Abstract:

The intensive dynamic of land use and land cover changes (LULC) in Mato Grosso, Brazil, and the environmental costs of the adopted agricultural practices along the years attract the attention of researchers and institutions. In order to evaluate these aspects, moderate resolution remote sensing data and techniques have been applied. By the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, compatible with the farms size and seasonal dynamic of Mato Grosso agriculture, it is possible acquire repeatedly data with large coverage, freely. In the context of Mato Grosso, which implies in a large geographical area with dynamic processes operating at multiple scales, this sensor allows vegetation monitoring and seasonal detection of changes in canopy. This review paper presents the use of MODIS sensor data to improve soybean crop detection, yield prediction and deforestation monitoring in Mato Grosso over the time. Different data of this sensor were applied along the years, especially the vegetation indices, and a large number of methods were developed through the MODIS derived information. The presented results show how the MODIS sensor was successfully applied to local and regional soybean crop and deforestation monitoring, offering new opportunities for land cover mapping in Mato Grosso.
Keywords: Remote sensing; Change detection; Yield estimates; Agriculture; Environment.

Review methodology: This study was carried out to evaluate the use of moderate resolution remote sensing and statistical modeling for investigations about soybean crops in Mato Grosso. We searched the databases: ADS - Astrophysics Data System; AGORA (FAO); CAB Abstracts (CABI); Current Contents – Physical; Chemical & Earth Sciences (Clarivate Analytics); DBLP Computer Science Bibliography (Universität Trier); DOAJ - Directory of Open Access Journals; Ei Compendex/Engineering Village (Elsevier); Genamics JournalSeek; HINARI (WHO); Inspec (IET); Journal Citation Reports/Science Edition (Clarivate Analytics, formerly Thomson Reuters‘ IP & Science branch); Julkaisufoorumi Publication Forum (Federation of Finnish Learned Societies); Norwegian Register for Scientific Journals, Series and Publishers (NSD); Science Citation Index Expanded - Web of Science (Clarivate Analytics); Scopus (Elsevier) and Web of Science (Clarivate Analytics). The articles were organized in the Mendeley platform in order of topics: “Soybean crop area identification”, “Soybean crop yield model”, “Deforestation detection”, “Economic importance of soybean crops”. In addition, we used the references from the articles obtained by this method to check for additional relevant material.

Introduction

Decades of scientific research have shown considerable progress towards assessing soybean crop monitoring with moderate resolution remote sensing data, especially by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, particularly due to its ability to provide regular spatially and temporally-explicit data across large areas when compared to field-based assessments. Remote sensors operate on a variety of basic physical principles, recording the electromagnetic properties of a land surface (either the energy reflected, emitted or scattered) and, hence, provide a variety of information on land properties [1].

The use of remote sensing techniques for detection of soybean areas and yield was expanded in recent years. The combined use of different methods of analysis, such as vegetation indices, geostatistics and machine learning bring a new perspective for these investigations. Across the time, the State of Mato Grosso (MT), Brazil, became one of the most study areas evaluated, because have a great relevance in the Brazilian and global agribusiness market, being a great soybean producer and exporter, since the early 2000’s. The soybean crop area in Mato Grosso increased by 275% between 2000 and 2015 [2]. The increase in soybean crop area during this period is accompanied by the increase in double cropping of soybean with another commercial crop. During this period, soybean production was found to shift from predominantly single-crop systems to majority double-crop systems, increasing from 33,288 to 78,434 ha [3]. Likewise, the area under intensive soy-corn double-cropping rotations expanded from 460,000 ha to over 4,300,000 ha between 2001 and 2016 [4].

Another extensive use of MODIS sensor in recent decades is related to deforestation process. The economic and environmental implications of increasing land area for soybean cultivation from deforestation and massive conversion of land use have become a major issue at the international policy agenda level [5]. Soybeans are much more damaging than other crops because they justify massive transportation infrastructure projects that unleash a chain of events leading to destruction of natural habitats over wide areas in addition to what is directly cultivated for soybeans. Because this, the improve of conversion rate of land use transition could be affected the environment in Mato Grosso [6].

Mato Grosso is located in the central–west region of Brazil and covers an area of approximately 905,000 km^2. The landscape is heterogeneous, which is a result of an active pioneer frontier shaped by different populations (public and private colonists, logging industries, indigenous societies), land uses, practices, and varying natural conditions (climate, soil, vegetation) [7]. The southern region is a tropical wetland known as Pantanal (61,000 km^2). In the north are the moist forests of Amazonia (484,000 km^2). The central region is dominated by vast tropical savannas known as Cerrado (360,000 km^2), where agribusiness is concentrated [3]. Mato Grosso's climate (Köppen Aw) is hot - semi-humid to humid - with pronounced seasonality marked by a dry season from May to October [8]. The rainy season occurs from October to May [9]. The climatic gradient is largely coincident with a gradient in land-use change, indicating the interconnectedness of biophysical and socio-economic processes [10]. Many of Mato Grosso's soils are old, deep, and nutrient-poor [8]. Nevertheless, these soils have the highest agricultural potential of Brazil, due to the use of modern agricultural techniques, the development of adapted soybean varieties, and favorable world markets [11].

By 2000, the widespread use of monoculture rendered the crops highly vulnerable to "Asian rust" (*Phakopsora pachyrhizi Sydow & P. Sydow*), a soybean disease that appeared during the 2002/2003 harvest. Furthermore, the unfavorable exchange rate between the Brazilian currency (Real) and the United States (US) dollar from 2005 to 2007 highlighted the economic vulnerability of the region's producers to soybean monoculture [12]. To reduce their vulnerability, the soybean producers changed their agricultural management practices. Since Asian rust makes soybean crops unfeasible between June and September, a soybean host-free period (*Vazio Sanitário*) was established in 2007. Additionally, a second crop (*Safrinha*) following the soybean harvest was introduced to prevent soil erosion, improve soil quality, break pest cycles, maintain soil moisture, and set the conditions for high-quality no-tillage operations [13]. This strategy enables growers to take advantage of the long tropical growing season and produce two crops per year [7].

Mato Grosso, as an agricultural State, has benefitted from a number of geographic and institutional conditions that have increased the capacity of the state's agriculture sector to serve as a growth engine [14], constituting in a separate case, with three peculiarities: First, is a relatively recent agricultural frontier. Even today, properties are still being cleared or converted to agriculture as investment capital is made available. Cities located in the midst of agriculture, are also growing rapidly, largely fueled by in-migration from elsewhere in Brazil. Second, Mato Grosso's agricultural sector is tightly coupled with local, urban-based supply chains [15]. Due the strength of the local supply chain, and the presence of downstream processing facilities in Mato Grosso, the local economy captures and circulates a larger proportion of the potential value of each harvest. Third, and finally, in Mato Grosso many farm owners live and spend locally [16].

This context converted Mato Grosso into a globally important center of agricultural production [17, 18]. The total land-use shift into soybean from 2001 to 2011 was almost 8.7 million ha, of which almost 3.5 million ha belonged to the Brazilian Amazon biome [9]. Most of these areas are currently being gradually abandoned [17, 18].

In 2000, [6] explained that the soybean crops are driven by global market forces, making them different from many of the land-use changes that have dominated the scene in Brazil so far, particularly in Amazonia. The global market for soybeans, which propels the advance of this crop, is really composed of three markets: whole soybeans, soy oil and soy meal. Currently, [19] also explain that the soybean market has an influence on land use, investigating the association of economic environment indicators as drivers for soybean crop dynamics of land use and expansion. The environmental costs of this growth have been carefully documented, being observed also in the deforestation process of Amazon biome. Soybean production in Mato Grosso has been widely tied to deforestation, directly, via the conversion of forest areas to cropland, and indirectly, through the sector's impact on regional land markets and investment decisions [20, 21, 22]. The land use changes related to large scale agricultural activities have enabled social development but also raised new concerns about human health [23].

Different answers emerge in this context. These circumstances are promising or threatening the region in future? Is this type of land use change a good sign to the people living there? Should it be more encouraged? Or has to be stopped by any means? Observing the studies about municipalities along the time, may be the economic growth represents a mandatory stage before environmental preservation, and the environmental preservation can be a source of socio-economic development in Mato Grosso, because we note that the environmental degradation was more extensive in the pre-industrial phase, achieve the peak in industrial phase and decreases in post industries economies.

Given these perceptions, the application of methodologies using geospatial intelligence and analysis have become necessary for the implementation of a synergy between sustainable agricultural practices and decreasing environmental impacts.

This review is structured as follows. First, we present recent applications of the MODIS sensor for soybean crop monitoring and yield estimates. Second, the economic importance of soybeans crops. Finally, the MODIS sensor application for deforestation detection, the great problem caused by the soybean crops expansion in Mato Grosso.

Moderate resolution remote sensing applied to crop area and yield estimates

The use of remote sensing for soybean crop monitoring in Mato Grosso was extensive in the last decade. The major number of these investigations have been carried out on the agricultural dynamics related to soybean crop expansion toward the Amazon biome as a major driver of land-use and land-cover changes. The wide timeframe and broad spatial coverage provided by remote sensing data have potential to address these issues because the enhanced temporal resolution allows for the production of near-real-time estimates of agricultural statistics [19].

One factor for this use is the official data imprecision. Different authors [19, 24, 25] have reported that official data released by two Brazilian agencies, namely, CONAB

(National Company of Food Supply) and IBGE, suffer from three main issues: (1) municipality statistics are not released shortly after a harvest, but nearly 18 months after the soybean season end; (2) official statistics lack associated digital/logical georeferenced information that can be used to provide spatial analysis of land-use and land-cover dynamics [9]; and (3) a historical compilation of crop management and crop rotation is impossible to provide, as there is no spatial information.

The IBGE yield survey is not completely consistent. The “yield” is defined in the IBGE dataset as the average ratio of production (kg) per harvested area (ha) for each crop, rather than sowed area [26-27]. Because they are municipal administrative limits data, IBGE's information did not capture the occurrences from which the crops occur in one municipality and its commercialization in another [28]. They have explained that many farms were distributed in more than one municipality, and the final yield was taken to the municipality where the farm headquarters, changing the municipal statistics. Many multinationals are located in these municipalities, and some agricultural negotiations are carried out by producers who sow them in neighboring communes.

Large-scale commercial cropping of soybeans expanded in the tropical Amazon and Cerrado biomes of Brazil after 1990. More recently, cropping intensified from single-cropping of soybeans to double-cropping of soybeans with corn or cotton [29]. The intensification of cropping practices followed quickly on the heels of this cropland expansion. By 2000, producers began shifting from single-cropping of soybeans to double-cropping in which a second late-season crop (*Safrinha*), most commonly corn and cotton, follows soybean harvest within a single 6-month rainy season. In 2001, only 5,000 km^2 of Mato Grosso’s 33,000km^2 of commercial cropland were double-cropped, but by 2011 that area had expanded more than five-fold to 28,000km^2 of the state’s 58,000km^2 of cropland. The soy–corn rotation accounts for nearly 92% of double-cropping in Mato Grosso [30].

In face to this, the MODIS sensor aboard the Terra and Aqua satellites was applied to identify soybean crops in Mato Grosso. The MODIS provides an adequate imaging configuration for crop monitoring based on (1) an almost-daily revisit time; (2) a spatial resolution of 250-m, considered adequate for mapping large-scale agricultural fields; and (3) a geometric quality that is high enough for time series analysis [31]. The

MODIS images have considerable advantages in the extensive agricultural crops characterization, mainly due to their higher temporal resolution. Many studies focused on Mato Grosso have highlighted the efficiency of MODIS vegetation indices time series for mapping cropland and crop expansion [32, 9, 33] and cropping system management [8, 13, 34], making possible the area identification and yield prediction.

Different techniques are used to soybean crop detection and monitoring. The decision tree classifier based on MODIS Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) time-series data to classify agricultural land use data [8], wavelet analysis based on MODIS EVI time-series data to map cropland distribution [34] and linear regression model based on MODIS EVI time-series and Landsat data to produce fractional cropland maps [32]. The MODIS EVI images from the soybean-sowing period allow to identify soybean crops [24]. The MODIS EVI images from the maximum crop development period presented satisfactory application for soybean crop classification [35], can be used with official data to identify soybean croplands presented good agreement with large cultivated areas [37].

In other hand, MODIS data was used to examine patterns of cropland expansion, cropland abandonment, changing cropping frequency from 2001 to 2011, and detect that in 2001, 3.3 million hectares of mechanized agriculture were cultivated in Mato Grosso, of which 500.000 hectares had two commercial crops per growing season (double-cropping) [18]. By 2011, Mato Grosso had 5.8 million hectares of mechanized agriculture, of which 2.9 million hectares were double-cropped. The patterns of vegetation dynamics identified from MODIS data also was used to map double-cropping, single cropping, forest and pasture through the Time-Weighted Dynamic Time Warping (TWDTW) method [38], classifying crops with various vegetation dynamics. The TWDTW analysis performed well in identifying single cropping, double-cropping, forest and pasture from EVI derived from MODIS data. Other new approach was developed to identify crop types and cropping patterns (Soy-maize, Soy-cotton, Soy-pasture, Soy-fallow, Fallow-cotton and Single crop), using MODIS NDVI time series data and field survey data [39].

These studies used 16-day composites derived only from MOD13Q1 images. Furthermore, the aforementioned studies did not consider large-scale interannual variations resulting from the agricultural dynamics in Mato Grosso. In contrast, [26] used eight-day composites derived through combining MOD13Q1 and MYD13Q1 images and address this issue with the proposed SEI*Pixel*, which covers the crop phenological cycle between 2000/2001 and 2010/2011 crop seasons. The analysis of 8-day rather than 16-day composites, as well as the consideration of large-scale interannual variations, allows the detection of subtle spectral-temporal differences between crop types, improving the method's crop area identification.

In general, the approaches adopted in these studies include: (1) reconstruction of time series data by removing the cloud-contaminated pixels, (2) development of indices and phenological parameters to distinguish croplands from other land cover types, and (3) extraction of cropping patterns using vegetation indices time-series data and *in situ* data.

Regards to the yield, recently models based on remote sensing, census and climate data have been developed to generate estimates in Mato Grosso. The reliability of the MODIS EVI data physiological meaning was tested to develop a remote sensing-based procedure to estimate soybean production prior to crop harvest was evaluated in different levels [19]. At the state, municipality and local levels, the results obtained confirmed the reliability of remote sensing-based models for soybean production forecast. Other method, the Coupled Model (CM), combines the MODIS Crop Detection Algorithm (MCDA) map with the resulting MODIS Productivity Detection Model (MPDM) to estimate soybean yield and production [40].

The sensitivity of agricultural output to climate change has often been estimated by modelling crop yields under climate change scenarios or with statistical analysis of the impacts of year-to-year climatic variability on crop yields. However, the cropland area and the number of crops harvested per growing season (cropping frequency) both also affect agricultural output and both also show sensitivity to climate variability and change.

The remote sensing, official statistics and climate data to model the change in agricultural output associated with the response of crop yield, crop frequency and crop area to year-to-year climate variability. Roughly 70% of the change in agricultural output caused by climate was determined by changes in frequency and/or changes in area [17].

Environmental impacts of soybean crop expansion

The Brazilian Amazon has experienced one of the world's highest deforestation rates in last decades. Cattle ranching and soy expansion constitute the major drivers of deforestation, both through direct conversion and indirectly by land use displacement, especially in Mato Grosso. However, deforestation rates decreased significantly after the implementation of the Action Plan to Prevent and Control Deforestation in Legal Amazon (PPCDAm) in 2004 [41].

The soybean market exerts the principal influence on land use. The profit exerts a direct and non-negligible influence on the evolution of consolidated land use for soybean in Mato Grosso [19]. The tropical agriculture has been widely identified as a driver of environmental change, and has been linked broadly to deforestation, because during the peak years of soybean profitability (2002-early 2004) deforestation reached new heights [14]. From 2001-2006 nearly 60,000km^2 of tropical Amazon forest were cleared in the state; approximately 10% of these clearings were converted directly to agriculture. The deforestation rates are indeed responsive to agricultural output prices; and the changes to conservation policies implemented beginning in 2004 and 2008 significantly contributed to the curbing of deforestation rates, avoiding substantial forest clearings [42].

The generalization of double-cropping systems in Mato Grosso is a major driver of the decoupling of deforestation and crop production [43, 13, 21]. This is because it became more profitable for producers to invest in new technologies for use in already opened areas than to finance new forest clearings. This situation emerged for three reasons: (1) efficient command-and-control and market-oriented policies discouraged producers from investing their capital in the acquisition of new areas to be cleared, (2) the economic situation marked by very high oil prices at the 2000s end did not stimulate crop expansion [44] and (3) the scarcity of high quality remaining agricultural land available for agricultural expansion, especially at the southern Amazon’s agricultural frontier [18].

When [6] treated the soybean cultivation as a threat to the environment in Brazil, explain that multiple adverse impacts of soybean expansion on biodiversity will be mitigated, and other development considerations substantially addressed, if many actions are taken by policy makers: (1) create protected areas in advance of soybean frontiers, (2) encourage elimination of many subsidies that speed soybean expansion beyond what would occur otherwise from market forces, (3) rapidly carry out studies to assess the costs of social and environmental impacts of soybean expansion, including opportunity costs of money and land, (4) strengthen the environmental-impact regulatory system, including assessment of the indirect impacts of infrastructure in stimulating other economic activities that are often destructive, (5) create mechanisms such that commitments can be made not to implant specific infrastructure projects that are judged to have excessive impacts and (6) encourage decision-making based on the full roster of costs and benefits, in contrast to the present system exemplified by the ‘Forward Brazil’ Program.

Particularly during the soybean ‘boom’ in the early 2000s, factors that land speculation, appreciation in land value, and expansion of cropland on pasture were linked to the displacement of cattle production, which led to the increased deforestation in the Amazon biome [41]. The demand for commodities by China increases deforestation in the Brazilian Legal Amazon [45]. Even though deforestation rates were lower in 2010 than in 2000, the estimates confirmed the effect of soybean-planted area in increasing both exports and deforestation over this period. In 1997, Mato Grosso exported less than 2 million Mt of soybeans. By 2012, that total rose to more than 10 million Mt, with nearly 7 million Mt destined for China [46].

In this context, the launch of the Action Plan to Prevent and Control Deforestation in Legal Amazon (PPCDAm) in 2004 can be considered as an important milestone that reflected the federal government ambition to hamper the progress of the Amazon agricultural frontier after the 2004 deforestation peak [42, 41]. This program enabled increased participation of municipalities, institutions and civil society in the establishment of a new governance model focused on three main areas: (1) territorial management and land use, (2) legislation compliance and command-and-control tools, such as monitoring or licensing systems, and (3) promotion of sustainable practices.

Brazil has demonstrated in the Amazon region that intensification combined with laws and policies to reduce deforestation is feasible at a very large scale. Its success in reducing deforestation was achieved while cattle and soy production continued to increase, largely because of rapid increases in the productivity of cattle operations. Mato Grosso farmer today in the Amazon biome is subjected to at least eight dialogues involving forests and deforestation (Reducing Emissions from Deforestation and Forest Degradation - REDD+, Consumer Goods Forum, ABC credit, Soy moratorium, Beef moratorium, Municipality Black List, Roundtable for Responsible Soy - RTRS and Forest code), each with its own different approaches [47].

Although the relation between soybean cultivation and deforestation has decreased in recent years due to new government policies [18, 13], there has been continuous land-use transition to soybean covering previously uncultivated areas [17], as pastures, whose drivers are still not well understood. In other hand, [48] identified 54 municipalities that were not in compliance with Soy Moratorium; concluding that the soy continues to be a significant driver of forest conversion, especially in municipalities where soy is the main economic activity.

Moderate resolution remote sensing applied to deforestation monitoring

Remote sensing also has long been used for detecting deforestation because it can acquire data repeatedly with large coverage, with images available at no cost. Therefore, much research explored MODIS sensor approaches to accurately detect the changes in the Brazilian Amazon and Cerrado [3, 49-52]. In the context of Amazonia, which implies in a large geographical area with dynamic processes operating at multiple scales, remote sensing is a key tool that allows the quantification of carbon stocks in forest biomass, the evaluation of seasonal changes in forest canopy, which ultimately will determine the direction and magnitude of carbon, water and energy flows between the canopy and the atmosphere [52], and the monitoring of stressing factors that controls the changes in forest biomass through time [53].

The importance of remote sensing on these issues is reflected in the increased number of publications since the 1980s [54]. The principal problem detected is related with the indirect land use changes (ILUC) in the agricultural sector. ILUC refers to land use changes that arise from displacement effects when one land use (e.g. soybeans) encroaches on another land use (e.g. pasture), causing the impacted land use to encroach on wildlands [55]. Many studies have focused on the ILUC impacts on the Amazon forest. Statistical evidence of ILUC associated with soy expansion in Legal Amazon during the first half of the 2000s [22]. This study was later updated by [20], who confirmed that soy expansion contributed to frontier deforestation during 2002–2011 in Brazil displacing cattle ranching to the forest frontier.

Other studies also noted that the soybean expansion was not directly correlated with area expansion. The agricultural intensification has evolved rapidly in Mato Grosso, as the proportion of the net cropped area cultivated with double-cropping systems harvesting two successive commercial crops [13]. The soybean expansion areas were mainly conducted within formerly deforested areas [26]. The ILUC is associated with soy alongside the BR-163 highway during 2001–2004, but found no connection between soy expansion in Mato Grosso and deforestation alongside BR-163 during 2005-2012 [41].

The impact of soybean expansion on deforestation in Mato Grosso was indirect, due the deforested areas in this region were first converted to pasture before being used for soybean production [56, 13, 28]. This expansion in formerly deforested areas in Mato Grosso as process encouraged by the Soy Moratorium established by producers, Non-Governmental Organizations (NGOs), State administrations and private enterprises in 2006 [57]. The Soy Moratorium was the first voluntary zero-deforestation supply-chain commitment implemented in the tropics [58]. It defines the agreement of Brazil's major soybean trading companies not to purchase soybeans produced on areas deforested after June 2006, which was changed to 2008 during the Moratorium renewal in 2014 [58].

Observing the land trajectories patterns, it is also possible to identify the ILUC influence in Mato Grosso. Before the establishment of the double-crop, the area was first converted from forest to pasture [26]. Also, it was observed the single crop in the first sowing after pasture. The farmers convert natural areas to pastures before the agricultural land use to dribble the Soy Moratorium [58]. From 2001 to 2005, 74% of soybean expansion was into previously cleared pasture areas and that from 2006 to 2010 the total reached 91% [21]. The total of 23.6 thousand km^2 (17.6% of deforestation) was attributable to the soy industry during 2002-2012 [59].

Other studies have been conducted in Mato Grosso using MODIS sensor to examine deforestation and agricultural expansion [18, 60-64]. Few areas remain for the legal expansion of croplands, indicating that the major share of expansion has to be onto cattle pastures [65]. Relative to cattle ranching, crop farming might promote regional economic development; cultivating crops on low-productivity pastures might also help to restore soil fertility, impel neighboring pastures into more intensive cultivation, and spare land from deforestation [17]. The greenhouse gas (GHG) emissions from soybean produced [11] affect the local environment. While economic growth has come at a significant cost to the environment, value added by the agriculture sector, directly and indirectly, has surpassed the CO_2-e value emitted through land clearings [14].

In general, most change detection studies cover two broad categories: selection of suitable remote sensing parameters and use of proper algorithms. The parameters can be pixel-based, object-based and subpixel-based [64]. The change detection algorithm can also be separated into two broad categories: detecting the binary change and non-change and detecting the detailed change trajectories. However, the change and non-change information cannot provide sufficient information required for particular purposes (e.g., driving forces of deforestation, land-cover and land-use change modelling); detailed change trajectories are often needed, for which the post-classification comparison approach is commonly used [64].

The post-classification comparison method involves classification of individual images into land-cover maps at multitemporal scale and comparison of classified land-cover maps pixel by pixel. This method provides detailed “from-to” change trajectories and identifies where the change occurred and how much change occurred. The accuracy change detection is dependent on the accuracy of each individual classification [66].

The observed patterns in the literature provide evidence that it is possible to achieve the dual objectives of forest conservation and agricultural production in contexts where there is a sufficient supply of previously cleared land and incentives that encourage productive use of that land instead of expansion into forests. The efforts to model Brazil’s low-carbon development alternatives indicate that the implementation of existing technologies to restore degraded lands and increase pasture productivity could free enough additional land to accommodate projected growth, although achieving this would be challenging and require substantial private and public investments [21].

There is a large deforested area that is underused or abandoned in the Amazon [67]. About 20% (78 million hectares) of the Amazon forest area has already been cleared and it is estimated that about 17.3 million ha (3% of the Amazonian biome) are underused, abandoned, or under conditions other than forest regeneration [68]. Some of these areas, if recovered, can serve the Brazilian agriculture expansion. Even in rural settlements, where food production depends on eventual new deforestation, it is estimated that the already cleared area is of 12.7 million ha [69].

Conclusions

Through this review we have demonstrated the utility and advances of moderate resolution remote sensing data for generating land cover products to support increasingly complex information needs for soybean crop and deforestation monitoring. The MODIS sensor data offers new opportunities and challenges for land cover mapping in Mato Grosso. Although studying the land use cover changes (LUCC) derived by intensive agricultural dynamic in Mato Grosso remains a challenge, partly since the dynamics and trajectories of change are complex and fast-evolving and partly since robust methods for analyses are still in development for many LUCC processes, every moment improved classification algorithms continue to emerge.

Furthermore, the literature exposed that the use of field datasets is very important to complement the spectral information. The incorporation of the temporal element into classification approaches is already benefiting from sophisticated spatial and spectral algorithms, pointing to the potential for further improvements in land cover classification outcomes. The approaches that incorporate spatial, spectral, and temporal data, and the knowledge of the land use and land cover changes allow more accurate assessments.

References

1. Joshi N, Baumann M, Ehammer A, Fensholt R, Grogan K, Hostert P, et al. A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. Remote Sens. 2016;8:70.

2. Brazilian Institute of Geography and Statistics (IBGE). 2017. Produção agrícola municipal - Automatic Data Recovery System – SIDRA. Available from: (www.sidra.ibge.gov.br/) (accessed on 02 Jan 2019).

3. Kastens JH, Brown JC, Coutinho AC, Bishop CR, Esquerdo JCDM. Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. PLoS ONE. 2017;12(4): e0176168.

4. Spera SA, Galford GL, Coe MT, Macedo MN, Mustard JF. Land use change affects water recycling in Brazil’s last agricultural frontier. Global Change Biol. 2016;22(10):3405-3413.

5. Food and Agriculture Organization (FAO). 2011. World Food and Agriculture in Review (http://www.fao.org/docrep/013/i2050e/i2050e07.pdf).

6. Fearnside PM. Soybean cultivation as a threat to the environment in Brazil. Environ Conserv. 2001;28(01):23–38.

7. Arvor D, Jonathan M, Meirelles MSP, Dubreuil V, Durieux L. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. Int J Remote Sens. 2011;32(22):7847–71.

8. Brown JC, Kastens JH, Coutinho AC, Victoria D de C, Bishop CR. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. Remote Sens Environ. 2013;130:39–50.

9. Gusso A, Arvor D, Ricardo Ducati J, Veronez MR, Da Silveira LG. Assessing the Modis crop detection algorithm for soybean crop area mapping and expansion in the Mato Grosso state, Brazil. Sci World J. 2014;2014(February 2016): 863141.

10. Davidson EA, Araújo AC de, Artaxo P, Balch JK, Brown IF, Bustamante MMC, et al. The Amazon basin in transition. Nature. 2012;481(7381):321–8.

11. Raucci GS, Moreira CS, Alves PA, Mello FFC, Frazão LA, Cerri CEP, et al. Greenhouse gas assessment of Brazilian soybean production: A case study of Mato Grosso State. J Clean Prod. 2015;96:419–25.

12. Arvor D, Dubreuil V, Mendez Del Villar P, Ferreira CM, Meirelles MSP. Développement, crises et adaptation des territoires du soja au Mato Grosso: l'exemple de Sorriso. Confins. 2009;6.

13. Arvor D, Meirelles MSP, Dubreuil V, Bégué A, Shimabukuro YE. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. Appl Geogr. 2012;32(2):702–13.

14. Richards P, Pellegrina H, VanWey L, Spera S. Soybean development the impact of a decade of agricultural change on urban and economic growth in Mato Grosso, Brazil. PLoS One. 2015;10(4): e0122510.

15. Garrett RD, Lambin EF, Naylor RL. Land institutions and supply chain configurations as determinants of soybean planted area and yields in Brazil. Land use policy. 2013;31:385–96.

16. Richards P, VanWey L. Where Deforestation Leads to Urbanization: How Resource Extraction Is Leading to Urban Growth in the Brazilian Amazon. Ann Assoc Am Geogr. 2015;105(4):806–23.

17. Cohn AS, VanWey LK, Spera SA, Mustard JF. Cropping frequency and area response to climate variability can exceed yield response. Nat Clim Chang. 2016;6(6):601–4.

18. Spera SA, Cohn AS, VanWey LK, Mustard JF, Rudorff BF, Risso J, et al. Recent cropping frequency, expansion, and abandonment in Mato Grosso, Brazil had selective land characteristics. Environ Res Lett. 2014;9(6):064010.

19. Gusso A, Ducati Jr, Bortolotto VC. Analysis of soybean cropland expansion in the southern Brazilian Amazon and its relation to economic drivers. Acta Amaz. 2017;47(4):281–92.

20. Richards PD, Walker RT, Arima EY. Spatially complex land change: The indirect effect of Brazil's agricultural sector on land use in Amazonia. Glob Environ Chang. 2014;29:1–9.

21. Macedo MN, DeFries RS, Morton DC, Stickler CM, Galford GL, Shimabukuro YE. Decoupling of deforestation and soy production in the southern Amazon during the late 2000s. Proc Natl Acad Sci. 2012;109(4):1341–6.

22. Arima EY, Richards PD, Walker R, Caldas MM. Statistical confirmation of indirect land use change in the Brazilian Amazon. Environ Res Lett. 2011;6(2):024010.

23. Arvor D, Tritsch T, Barcellos C, Jegou N, Dubreuil V. Land use sustainability on the South-Eastern Amazon agricultural frontier: Recent progress and the challenges ahead. Appl Geogr. 2017;80:86–97.

24. Gusso A, Formaggio AR, Rizzi R, Adami M, Rudorff BFT. Soybean crop area estimation by Modis/Evi data. Pesqui Agropecu Bras. 2012;47(3):425–35.

25. Johann JA, Rocha JV, Duft DG, Lamparelli RAC. Estimativa de áreas com culturas de verão no Paraná, por meio de imagens multitemporais EVI/Modis. Pesqui Agropecu Bras. 2012;47(9):1295–306.

26. Chaves MED, Alves M de C, de Oliveira MS, Sáfadi T. A geostatistical approach for modeling soybean crop area and yield based on census and remote sensing data. Remote Sens. 2018;10(5):680.

27. Anderson MC, Zolin CA, Sentelhas PC, Hain CR, Semmens K, Tugrul Yilmaz M, et al. The Evaporative Stress Index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. Remote Sens Environ. 2016;174:82–99.

28. Dubreuil V, Laques AÉ, Nédélec V, Arvor D, Gurgel H. Paysages et fronts pionniers amazoniens sous le regard des satellites: L'exemple du Mato Grosso. Espac Geogr. 2008;37(1):57–74.

29. Neill C, Jankowski K, Brando PM, Coe MT, Deegan LA, Macedo MN, et al. Surprisingly Modest Water Quality Impacts From Expansion and Intensification of Large-Sscale Commercial Agriculture in the Brazilian Amazon-Cerrado Region. Trop Conserv Sci. 2017;10:194008291772066.

30. Spera S. Agricultural Intensification Can Preserve the Brazilian Cerrado: Applying Lessons From Mato Grosso and Goiás to Brazil's Last Agricultural Frontier. Trop Conserv Sci. 2017;10:194008291772066.

31. Justice CO, Townshend JRG, Vermote EF, Masuoka E, Wolfe RE, Saleous N, et al. An overview of MODIS Land data processing and product status. Remote Sens Environ. 2002;83(1–2):3–15.

32. Zhu C, Lu D, Victoria D, Dutra LV. Mapping fractional cropland distribution in Mato Grosso, Brazil using time series MODIS enhanced vegetation index and Landsat Thematic Mapper data. Remote Sens. 2016;8(1):22-36.

33. Morton DC, DeFries RS, Shimabukuro YE, Anderson LO, Arai E, del Bon Espirito-Santo F, et al. Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. Proc Natl Acad Sci. 2006;103(39):14637–41.

34. Galford GL, Mustard JF, Melillo J, Gendrin A, Cerri CC, Cerri CEP. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. Remote Sens Environ. 2008;112(2):576–87.

35. Risso J, Rizzi R, Rudorff BFT, Adami M., Shimabukuro YE, Formaggio AR et al. Índices de vegetação Modis aplicados na discriminação de áreas de soja. Pesqui Agropecu Bras. 2012;47(9):1317–26.

36. Souza CHW, Mercante E, Johann JA, Lamparelli RAC, Uribe-Opazo MA. Mapping and discrimination of soya bean and corn crops using spectro-temporal profiles of vegetation indices. Int J Remote Sens. 2015;36(7):1809–1824.

37. Victoria D de C, Paz AR da, Coutinho AC, Kastens J., Brown JC. Cropland area estimates using Modis NDVI time series in the state of Mato Grosso, Brazil. Pesq. Agropec. Bras. 2012;47(9):1270–8.

38. Maus V, Câmara G, Cartaxo R, Sanchez A, Ramos FM, Queiroz GR de. A Time-Weighted Dynamic Time Warping method for land use and land cover mapping. Ieee J Sel Top Appl Earth Obs Remote Sensing. 2015;XX(X):1–10.

39. Chen Y, Lu D, Moran E, Batistella M, Dutra LV, Sanches ID, et al. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. Int J Appl Earth Obs Geoinf. 2018;69(March):133–47.

40. Gusso A, Ducati JR, Veronez MR, Arvor D, Jr LGDS. Spectral Model for Soybean Yield Estimate Using MODIS EVI data. Int J Geosci. 2013;4:1233–41.

41. Gollnow F, Lakes T. Policy change, land use, and agriculture: The case of soy production and cattle ranching in Brazil, 2001-2012. Appl Geogr. 2014;55:203–11.

42. Assunção J, Gandour C, Rocha R. Deforestation slowdown in the Brazilian Amazon: Prices or policies? Environ Dev Econ. 2015;20(6):697–722.

43. Arvor D, Dubreuil V, Simões M, Bégué A. Mapping and spatial analysis of the soybean agricultural frontier in Mato Grosso, Brazil, using remote sensing data. GeoJournal. 2013;78(5):833–50.

44. Richards PD, Myers RJ, Swinton SM, Walker RT. Exchange rates, soybean supply response, and deforestation in South America. Glob Environ Chang. 2012;22(2):454–62.

45. Fearnside PM, Figueiredo AMR, Bonjour SCM. Amazonian forest loss and the long reach of China’s influence. Environ Dev Sustain. 2013;15(2):325–38.

46. Neill C, Macedo MN. The rise of Brazil's globally-connected Amazon soybean agriculture. In: M. Gutmann M, Lesser J, Editors. Global Latin America: Into the Twenty-first Century. Berkeley, USA: University of California Press; 2016. p. 167-186.

47. Nepstad D, Irawan S, Bezerra T, Boyd W, Stickler C, Shimada J, et al. More food, more forests, fewer emissions, better livelihoods: Linking REDD+, sustainable supply chains and domestic policy in Brazil, Indonesia and Colombia. Carbon Manag. 2013;4(6):639–58.

48. Silva Júnior CA, Lima M. Soy Moratorium in Mato Grosso: Deforestation undermines the agreement. Land use policy. 2017;(September):0–1.

49. Beuchle R, Grecchi RC, Shimabukuro YE, Seliger R, Eva HD, Sano E, et al. Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. Appl Geogr. 2015;58(March):116–27.

50. Grecchi RC, Gwyn QHJ, Bénié GB, Formaggio AR, Fahl FC. Land use and land cover changes in the Brazilian Cerrado: A multidisciplinary approach to assess the impacts of agricultural expansion. Appl Geogr. 2014;55:300–12.

51. Graesser J, Aide TM, Grau HR, Ramankutty N. Cropland/pastureland dynamics and the slowdown of deforestation in Latin America. Environ Res Lett [Internet]. 2015;10(3):034017.

52. Aragão LEOC, Anderson LO, Fonseca MG, Rosan TM, Vedovato LB, Wagner FH, et al. 21st Century drought-related fires counteract the decline of Amazon deforestation carbon emissions. Nat Commun [Internet]. 2018;9(536):1-12.

53. Hmimina G, Dufrene E, Pontailler JY, Delpierre N, Aubinet M, Caquet B. Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements. Remote Sens. Environ. 2013;132:145–158;

54. Santos JR dos, Galvão LS, Aragão LEO e C de. Remote sensing of Amazonian forests: Monitoring structure, phenology and responses to environmental changes. Rev Bras Cartogr. 2014;66(7):1413–36.

55. Walker R. Sparing Land for Nature in the Brazilian Amazon: Implications from location rent theory. Geogr Anal. 2014;46(1):18–36.

56. Sentelhas PC, Battisti R, Câmara GMS, Farias JRB, Hampf, AC. The soybean yield gap in Brazil – magnitude, causes and possible solutions for sustainable production. J Agric Sci. 2015;65(7):1–18.

57. Rudorff BFT, Adami M, Aguiar DA, Moreira MA, Mello MP, Fabiani L, et al. The soy moratorium in the Amazon biome monitored by remote sensing images. Remote Sens. 2011;3:185–202.

58. Gibbs HK, Rausch L, Munger J, Schelly I, Morton DC, Noojipady P, et al. Brazil's Soy Moratorium. Science (80-). 2015;347(6220):377–8.

59. Jusys T. A confirmation of the indirect impact of sugarcane on deforestation in the Amazon. J Land Use Sci. 2017;12(2–3):125–37.

60. Grecchi RC, Beuchle R, Shimabukuro YE, Aragão LEOC, Arai E, Simonetti D, et al. An integrated remote sensing and GIS approach for monitoring areas affected by selective logging: A case study in northern Mato Grosso, Brazilian Amazon. Int J Appl Earth Obs Geoinf. 2017;61(May):70–80.

61. Müller H, Griffiths P, Hostert P. Long-term deforestation dynamics in the Brazilian Amazon—Uncovering historic frontier development along the Cuiabá–Santarém highway. Int J Appl Earth Obs Geoinf. 2016;44:61–9.

62. Chen G, Powers RP, de Carvalho LMT, Mora B. Spatiotemporal patterns of tropical deforestation and forest degradation in response to the operation of the Tucuruí hydroelectricdam in the Amazon basin. Appl Geogr. 2015;63:1–8.

63. Imbach P, Manrow M, Barona E, Barretto A, Hyman G, Ciais P. Spatial and temporal contrasts in the distribution of crops and pastures across Amazonia: A new agricultural land use data set from census data since 1950. Global Biogeochem Cycles. 2015;29(6):898–916.

64. Lu D, Li G, Moran E. Current situation and needs of change detection techniques. Int J Image Data Fusion. 2014;5(1):13–38.

65. Morton DC, Noojipady P, Macedo MM, Gibbs H, Victoria D de C, Bolfe EL. Reevaluating suitability estimates based on dynamics of cropland expansion in the Brazilian Amazon. Glob Environ Chang. 2016;37:92–101.

66. Li G, Lu D, Moran E, Calvi MF, Dutra V, Batistella M, et al. Examining deforestation and agropasture dynamics along the Brazilian TransAmazon Highway using multitemporal Landsat imagery. GIScience Remote Sens. 2018;00(00):1–23.

67. Moutinho P, Guerra R, Azevedo-Ramos C. Achieving zero deforestation in the Brazilian Amazon: What is missing? Elem Sci Anthr. 2016;4:000125.

68. Terra Class 2010. INPE. Available from: http://www.inpe.br/cra/projetos_pesquisas/terraclass2010.php.

69. Alencar A, Pereira C, Castro I, Cardoso A, Souza L, Costa R, et al. Desmatamento nos Assentamentos da Amazônia: histórico, tendências e oportunidades. IPAM - Inst Pesqui Ambient da Amaz. 2016;93.

ARTIGO 2 - A GEOSTATISTICAL APPROACH FOR MODELING SOYBEAN CROP AREA AND YIELD BASED ON CENSUS AND REMOTE SENSING DATA

Normas do periódico *Remote Sensing*, ISSN: 2072-4292.
Versão publicada: (http://www.mdpi.com/2072-4292/10/5/680).

Michel Eustáquio Dantas Chaves[1,*], Marcelo de Carvalho Alves[1], Marcelo Silva de Oliveira[2] and Thelma Sáfadi[2]

1 Department of Engineering, Federal University of Lavras, University Campus, P.O. Box 3037, Lavras 37200-000, Brazil; marcelo.alves@deg.ufla.br
2 Department of Statistics, Federal University of Lavras, University Campus, P.O. Box 3037,
Lavras 37200-000, Brazil; marcelo.oliveira@des.ufla.br (M.S.d.O.); safadi@des.ufla.br (T.S.)
***** Correspondence: medchaves@posgrad.ufla.br; Tel.: +55-35-99968-2078
Received: 16 March 2018; Accepted: 21 April 2018; Published: 27 April 2018

Abstract: Advances in satellite imagery and remote sensing have enabled the acquisition of spatial data at several different resolutions. Geographic information systems (GIS) and geostatistics can be used to link geographic data from different sources. This article discusses the need to improve soybean crop detection and yield prediction by linking census data, GIS, remote sensing, and geostatistics. The proposed approach combines Brazilian Institute of Geography and Statistics (IBGE) census data with an eight-day enhanced vegetation index (EVI) time series derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data to monitor soybean areas and yields in Mato Grosso State, Brazil. *In situ* data from farms were used to validate the obtained results. Binomial areal kriging was used to generate maps of soybean occurrence over the years, and Gaussian areal kriging was used to predict soybean crop yield census data inside detected soybean areas, which had a downscaling effect on the results. The global accuracy and the Kappa index for the soybean crop detection were 92.1% and 0.84, respectively. The yield prediction presented 95.09% accuracy considering the standard deviation and probable error. Soybean crop detection and yield monitoring can be improved by this approach.

Keywords: MODIS EVI time series; downscaling; soybean crop monitoring

---

1. Introduction

Advances in satellite imagery and remote sensing now enable scientists to acquire spatial data at several different resolutions [1]. An increase in spatial resolution is known as *downscaling*. In the context of remote sensing, downscaling refers to a decrease in the pixel size of remotely sensed images [2]. Data also can be downscaled for area-to-point prediction or for areal interpolation, a kriging-based disaggregation technique [3]. Area-to-point prediction uses a downscaling process to predict the same continuous variable at a finer spatial resolution than the input. Once a prediction surface has been created, predictions can be aggregated back to a new set of polygons [2,3].

The issue of scaling is important in various fields, and is one of the most challenging and fascinating aspects of spatial statistics. The choice of an appropriate measurement scale is essential for understanding spatial processes, because mechanisms that are crucial to the spatial dynamics of a process at one scale may be unimportant at another. Relationships between variables at one scale may be obscured or distorted when viewed at another scale [1]. The measurement process involves the important aspect of uncertainty in both sampling and measurement. The sampling process is determined by the support—i.e., the space over which each observation is defined—and the spatial extent; the components of this process are the sampling scheme, the number of observations, and the density of the sample [4]. The support is the parameter of interest in downscaling.

Change-of-support issues are a concern within the field of geostatistics. The change-of-support problem (COSP) refers to the challenge of making a valid inference about a spatial variable at one support based on data at a different support [5]. Most COSPs and their solutions have been concerned with upscaling the prediction of a variable whose support is larger than that of the observed data. Block kriging prediction is commonly used to upscale spatial processes, and is one solution to the point-to-area COSP [5].

In geostatistics, the regularization of the semivariogram changes the support, or measurement scale, of the semivariogram function that is used to describe the structure of spatial variation, without requiring new data on the new support. Changing the support of a variable creates a new variable with different statistical properties. The issue of how the spatial variation in one variable that is associated with a given support relates to that of the new variable with a different support is a change-of-support problem [5].

The modifiable areal unit problem (MAUP) can change the support of soybean crop variables [6], since it is impossible to grow both soybeans and other crops in the same space simultaneously. The MAUP comprises two problems: scale effect and zoning effect. The scale or aggregation effect concerns the different inferences obtained when the same set of data is grouped into increasingly larger areal units. The grouping or zoning effect refers to the variability in results due to alternative configurations of the areal units, leading to differences in unit shape at the same or similar scales.

Geostatistical methods offer an approach to generating the best linear unbiased prediction (BLUP), creating a general method for combining data collected from different geographical units. BLUP can be used to provide a smooth surface of point support based on aggregated data, and make predictions for one set of geographic units based on another set of either nested or overlapping geographic units [5]. This kriging approach accounts for the shape and size of geographical units, accommodates different spatial supports for the data and the prediction without restriction to a single type of areal data at one time, and provides a measure of prediction error variance. Its recent generalization as area-and-point kriging allows the mapping of attribute values within each sampled geographical unit under the constraint that the average of point estimates returns the areal data with coherency [7].

This paper presents an uncertainty approach to the geostatistical mapping of harvested soybean areas and yields in the State of Mato Grosso (MT), Brazil, combining geographic sets of area-based data with remotely sensed data provided by an eight-day Moderate Resolution Imaging Spectroradiometer (MODIS) enhanced vegetation index (EVI) time series. This approach provides a measure of uncertainty in the results, and incorporates relevant covariate information that may be used to improve predictions.

2. Materials and Methods

2.1. Study Area

The study area is the State of Mato Grosso (MT), which is located in the central–west region of Brazil (Figure 1) and covers an area of approximately 905,000 km^2. Mato Grosso was chosen for this study because its landscape is heterogeneous, which is a result of an active pioneer frontier shaped by different populations (public and private colonists, logging industries, indigenous societies), land uses, practices, and varying natural conditions (climate, soil, vegetation) [8].

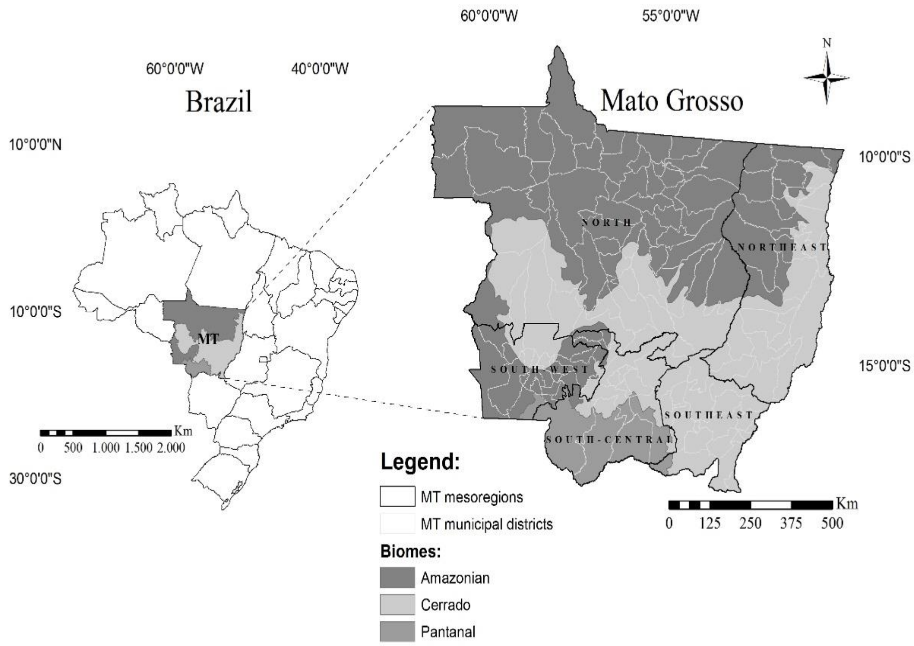


Figure 1. Geographic location of the State of Mato Grosso, Brazil, with its mesoregions and biomes. Source: Brazilian Institute of Geography and Statistics (IBGE) (2017).

The southern region of the State is a tropical wetland known as the Pantanal (61,000 km^2). In the north are the moist forests of Amazonia (484,000 km^2). The central region is dominated by vast tropical savannas known as Cerrado (360,000 km^2), where agribusiness is concentrated [9]. According to Brown et al. [10], Mato Grosso's climate (Köppen Aw) is hot—semi-humid to humid—with pronounced seasonality marked by a dry season from May to October. The rainy season occurs from October to May [11]. The climatic gradient is largely coincident with a gradient in land-use change, indicating the interconnectedness of biophysical and socio-economic processes [12]. Many of the State's soils are old, deep, and nutrient-poor [10]. Nevertheless, these soils have the highest agricultural potential in the country, due to the use of modern agricultural techniques, the development of adapted soybean varieties, and favorable world markets [13]. According to Brazilian Institute of Geography and Statistics (IBGE) [14], the total planted area in Mato Grosso increased by 297% (from 4.74 million to 14.10 million ha) between 2000 and 2015.

By 2000, the widespread use of monoculture rendered the crops highly vulnerable to "Asian rust" (*Phakopsora pachyrhizi* Sydow & P. Sydow), which is a new soybean disease that appeared during the 2002/2003 harvest. Furthermore, the unfavorable exchange rate between the Brazilian currency (Real) and the United States (US) dollar from 2005 to 2007 highlighted the economic vulnerability of the region's producers to soybean monoculture [15]. To reduce their vulnerability, the soybean producers changed their agricultural management practices. Since Asian rust makes soybean crops unfeasible between June and September, a soybean host-free period (*Vazio Sanitário*) was established in 2007. Additionally, a second crop (*Safrinha*) following the soybean harvest was introduced to prevent soil erosion, improve soil quality, break pest cycles, maintain soil moisture, and set the conditions for high-quality no-tillage operations [16]. This strategy enables growers to take advantage of the long tropical growing season and produce two crops per year [8].

2.2. Data Acquisition and Processing

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor provides a reliable reflectance time series with acceptable resolution and frequency, showing high potential for large-scale mapping. Two of the most suitable products are MOD13Q1 and MYD13Q1, which are provided every 16 days at a 250-meter spatial resolution, presenting the enhanced vegetation index (EVI) and the normalized difference vegetation index (NDVI). These products start on different days (one starts in the middle of the compositing period of the other). This situation effectively creates an eight-day product, which improves temporal change detection [17]. Previous research has successfully used MODIS time series data as an input to classify land cover types [10,16,18,19] and monitor the soybean crops in Mato Grosso [11,20,21] over the last decade. Due to the singular characteristics of the soybean-producing farms in the State, the MODIS data, with a near daily temporal resolution and 250-m spatial resolution, can be considered as adequate for crop mapping.

Between the EVI and NDVI vegetation indexes, the EVI was chosen due to its potential to reduce atmospheric and soil background effects [22]. Different studies, such as Zhu et al. [20] and Zhong et al. [23], consider EVI to be the most suitable index to represent crop growth with sufficient sensitivity in high biomass periods, especially for soybean discrimination during maximum crop development. According to Risso et al. [18], the EVI performs the best for soybean discrimination during the crop seasons in Mato Grosso, while the NDVI is unfit for this objective. The EVI time series profiles gradually increase as the vegetative growth stage progresses over time; it reaches the peak value on the heading date, and then begins to decline in the reproductive growth stage until the mature crop is harvested [24]. The EVI is determined by:

$$\mathrm{EVI} = \frac{G(\rho_{NIR} - \rho_{red})}{(\rho_{NIR} + C_1)(\rho_{red} - C_2)(\rho_{blue} + L)} \tag{1}$$

where the ρ values are partially atmospherically corrected (Rayleigh and ozone absorption) surface reflectance; L is the canopy background adjustment (L=1), C_1 and C_2 are coefficients of the aerosol resistance term that uses the blue band (458-479 nm) of

MODIS to correct for aerosol influences in the red band ($C_1 = 6$ and $C_2 = 7.5$), and G is a gain factor ($G = 2.5$) [25].

However, some limitations may occur when attempting to apply this approach to the large-scale classification and monitoring of crops [16,20,24]. The EVI time series generated from the MODIS surface reflectance data include various noise components, such as aerosols and bidirectional reflectance distribution factors; thus, it is necessary to reprocess these data to remove the noise components before using them to discriminate crop growth profiles [24]. The MOD13Q1 and MYD13Q1 have quality assessment (QA) fields that indicate cloud state and snow/ice presence; these quality control fields are divided into VI Quality and Pixel Reliability, respectively. This information allows MODIS database users to verify the reliability of a pixel in relation to the atmospheric conditions (clouds, ozone, dust, and other aerosols), and check the bidirectional effects related to data acquisition geometry (source–target–sensor) and possible failure of instruments or data processing methods [26]. In this paper, the EVI data were obtained from the United States Geological Survey's Land Processes Distributed Active Archive Center (LP DAAC) [27]. We acquired one MODIS tile (h12v10), which covered all of the main soybean cultivation areas. The images were re-projected into GeoTIFF format, using the MODIS Reprojection Tool [27]. The QA fields were used to eliminate noisy pixels.

2.3. Proposition of the Soybean Enhanced Index SEI_{Pixel}

The EVI images obtained for each crop season were chronologically organized. We then proposed a calculation of the total change magnitude per pixel, called the Soybean Enhanced Index (SEI_{Pixel}). The Euclidean distance between endpoints (dates 1 and 2) was calculated based on the change in dimensional space characterized by the MODIS EVI tile h12v10 for each crop season, according to Equation (2), which was modified from Jensen (2005):

$$\mathrm{SEI}_{Pixel} = \sum_{g=1}^{n} [\mathrm{EVI}_{ijk(date\ 2)} - \mathrm{EVI}_{ijk(date\ 1)}]^2 \tag{2}$$

where i and j are the latitude and longitude of the two-dimensional space of each MODIS tile, and $\mathrm{EVI}_{ijk(date\ 1)}$ and $\mathrm{EVI}_{ijk(date\ 2)}$ are the values of dates 1 and 2 in EVI band k, for the n seasonal growth profile.

For this procedure, the *in situ* data were stratified according to the sowing date described for each crop field. The EVI dates 1 and 2 refer respectively to the minimum and maximum vegetative vigor of soybean in a seasonal growth profile.

The SEI_{Pixel} for each crop season was classified into binary (0 and 1) values, where 0 and 1 were "Non-soybean" and "Soybean" cropland, respectively. The threshold values for the classification were defined based on the time series responses and on the Brazilian Institute of Geography and Statistics (IBGE) survey of soybean crop production and harvested area [28] for all of the crop seasons in all of the Mato Grosso municipalities. The harvested soybean crop areas estimated by the SEI_{Pixel} were recorded and compared with the IBGE survey, defining the best threshold values based on the highest similarity between the "Soybean" and "Non-soybean" cropland classification results of both surveying methods.

The Mato Grosso municipalities vector boundaries—2007 reference year, 1:2,500,000 scale, SIRGAS 2000 geographic latitude and longitude—were obtained from IBGE [14]. The vector boundaries were converted to WGS84 geographic latitude and longitude. The vector boundary zones for each of the 141 municipalities corresponded to the harvested area and yield data obtained for the 2000/2001 to 2011/2012 crop seasons. Zonal statistics were used to calculate the number of soybean crop pixels, according to the SEI_{Pixel}, inside each vector boundary zone of the 141 municipalities.

The intensification of soybean cropland in Mato Grosso, per pixel (iSEI_{Pixel}), was computed by summing the (SEI_{Pixel}) binary images (0 or 1) for the n crop seasons, generating information about the number of years harvested (Equation 3):

$$\mathrm{iSEI}_{Pixel} = \sum^{n} (\mathrm{SEI}_{Pixel\ ijk}) \tag{3}$$

The R-squares goodness of fit criteria, which were derived from the polynomial regression linear fit, were used to compare the classified results of the (SEI_{Pixel}) and IBGE [28] data.

2.4. Geostatistical Analysis

2.4.1. Binomial Areal Kriging Model

The Mato Grosso region (*D)* was composed of N areas determined by a vector that was used to sample the SEI_{Pixel} imagery for each evaluated crop season, and the iSEI_{Pixel} imagery for all of the crop seasons collectively. The sampled pixels of the SEI_{Pixel} imagery were classified as either "Non-soybean" (0) or "Soybean" (1) binary values. In the iSEI_{Pixel} imagery, each pixel was assigned a value between 0–12, where 12 represented the agroecosystems cultivated with soybean crops for 12 consecutive years. The value of the soybean crop areas and the total quantity of soybean crop areas were denoted by $l(u_i)$ and $n(u_i)$, respectively. The vector of spatial coordinates $u_i = (x_i, y_i)$, i =

1, ..., N, was georeferenced to the centroid of the *i-th* area. The random variable $Z(u_i)$ was the percentage of soybean crop area occurrence, and one of the realizations of $Z(u_i)$ was $z(u_i)$, which was defined by $l(u_i)/n(u_i)$.

Soybean crop areas within region *A* were assumed to be influenced by an underlying event—denoted by $R(u)$—that was caused by a stochastic process $\{R(u), u \in D \subset \Re^2\}$, where $R(u)$ was a random variable with continuous variation and correlation over space and whose values were not directly observed. Therefore, there are two distinct geographic supports. The first relates to the geographic areas inside region *D*, with $z(u_i)$ observed values of the random variable $Z(u_i)$. The second relates to the nature of the evaluated process, which was characterized by the continuous surface $R(u)$. Considering $R(u_i)$, $i = 1, ..., N$, the average event associated with the centroid of the *i-th* geographical support of the area v_i is defined by [29]:

$$R(u_i) = \frac{1}{v_i} \int_{v_i} R(u)\,du \tag{4}$$

The $R(u_i)$ values are associated with the percentage of occurrence of the event $P(u_i)$ in the *i-th* geographical support area v_i, in the time interval δt [30]:

$$R(u_i) = \frac{P(u_i)}{\delta t} \tag{5}$$

where for $\delta t = 1$, $R(u_i) = P(u_i)$.

The random variable $L(u_i)$ relates to the number of occurrences of the event in $n(u_i)$ repetitions of the *i-th* geographical support and presents binomial distribution as:

$$L(u_i) \sim Bin\left[r(u_i), n(u_i)\right] \tag{6}$$

where *Bin* denotes a binomial distribution, the parameter $r(u_i) = \mathrm{p}(u_i)$, and $n(u_i)$ is the total quantity of soybean crop areas in the *i-th* geographical support.

The different cases of the risk of occurrence of soybean crop areas are independent when $R(u)$ is fixed, because $R(u)$ is the only source of correlation between the cases. Thus, the observed rates are variables with distribution [29]:

$$Z(u_i) \mid R \sim \frac{1}{n(u_i)} Bi\left[r(u_i), n(u_i)\right] \tag{7}$$

The following conditional moments can be established based on Equation (7) [3]:

$$E\left[Z(u_i) \mid R\right] = R(u_i) \tag{8}$$

$$E\left[Z^2(u_i) \mid R\right] = \frac{n(u_i)-1}{n(u_i)} R^2(u_i) + \frac{R(u_i)}{n(u_i)} \quad (9)$$

$$E\left[Z(u_i)Z(u_i) \mid R\right] = R(u_i)R(u_i) \quad (10)$$

The semivariogram of the binomial kriging was established based on the conditional moments [29]:

$$\gamma^R_{(u_i,u_j)} = \gamma^Z_{(u_i,u_j)} - \frac{1}{2}\left\{\frac{1}{n(u_i)} + \frac{1}{n(u_j)}\right\}\mu(1-\mu) + \frac{1}{2}\left\{\frac{\sigma^2_{R(u_i)}}{n(u_i)} + \frac{\sigma^2_{R(u_j)}}{n(u_j)}\right\} E\left[Z(u_i)Z(u_i) \mid R\right] = R(u_i)R(u_i) \quad (11)$$

where $\gamma^R_{(u_i,u_j)} = \frac{1}{2}E\left[R(u_i) - R(u_j)\right]^2$ and is defined as the semivariogram function of the risk of soybean crop area occurrence; $\gamma^Z_{(u_i,u_j)} = \frac{1}{2}E\left[Z(u_i) - Z(u_j)\right]^2$ is the semivariogram function of the rates; and n(u_i) and n(u_j) are the total number of soybean crop areas in the *i-th* and *j-th* geographic supports, with centroids in u_i and u_j, respectively. The average risk μ is E[R(u)]. The risk variances in the *i-th* and *j-th* geographic supports are $\sigma^2_{R(u_i)}$ and $\sigma^2_{R(u_j)}$, respectively. If the variogram is bounded by a finite value $\gamma(\infty)$, a covariance function can be found, such as [31]:

$$C(u) = \gamma(\infty) - \gamma(u) \quad (12)$$

When an experimental variogram exhibits a sill, it can be fit to a theoretical variogram that is a covariance function C(u), and based on the formula for bounded variograms [31]:

$$\gamma(u) = C(0) - C(u) \quad (13)$$

where $C(u)$ is the stationary covariance and $C(0)$ = Var{Z(u)} is the value at the origin of the covariance function.

The risk $R(u)$ was estimated based on the information of $Z(u_i)$. Binomial kriging was done based on the direct covariance between $Z(u_i)$ and $Z(u_j)$. Binomial kriging in an unknown location u_0 was estimated by a linear combination of neighbor rates $z(u_i)$ [32]:

$$\hat{R}(u_0) = \sum_{i=1}^{N} \lambda(u_i) z(u_i) \tag{14}$$

where $z(u_i)$ is the rate observed in the *i-th* geographical support with a centroid in u_i, N is the total number of centroids used in the calculus of $\hat{R}(u_0)$, and $\lambda(u_i)$ is the weight attributed to the *i-th* observation of $z(u_i)$.

The $\lambda(u_i)$ was determined according to the spatial correlation of the risk and the properties of non-bias and minimum variance. Non-bias implies that the difference between the estimated value and the true value at the same point must be null:

$$E\left[\hat{R}(u_0) - R(u_0)\right] = 0 \tag{15}$$

$$E\left[\sum_{i=1}^{N} \lambda(u_i) z(u_i) - R(u_0)\right] = 0 \tag{16}$$

$$E\sum_{i=1}^{N} \lambda(u_i) E\left[z(u_i)\right] - E\left[R(u_0)\right] = 0 \tag{17}$$

$$E\sum_{i=1}^{N} \lambda(u_i)\mu - \mu = 0 \tag{18}$$

$$\mu\left[\sum_{i=1}^{N} \lambda(u_i) - 1\right] = 0 \rightarrow \sum_{i=1}^{N} \lambda(u_i) = 1 \tag{19}$$

The non-bias property requires that the sum of the weights equals 1.

The minimum variance property means that the estimator has minimum variance when compared to other linear estimators:

$$\sigma^2_{R(u_0)} = E\left\{\left[\hat{R}(u_0) - R(u_0)\right]^2\right\} \text{ is minimal} \tag{20}$$

Equation (18) must be minimized under the restriction that $\sum_{i=1}^{N} \lambda(u_i) = 1$. Minimization is achieved using Lagrange techniques. Minimizing $\sigma^2_{R(u_0)}$, the weights were obtained by applying [29]:

$$\begin{cases} \sum_{i=1}^{N} \lambda(u_i) C^{Z}_{(u_i, u_j)} + \varphi = C^{ZR}_{(u_j, u_0)} \text{ for all } j = 1, \ldots, N \\ \sum_{i=1}^{N} \lambda(u_i) = 1 \end{cases} \tag{21}$$

where $C^{Z}_{(u_i,u_j)}$ is the covariance of the frequencies, and $C^{ZR}_{(u_j,u_0)}$ is the covariance between frequency and risk. The covariance depends not only on the spatial separations, x_i - x_j, between the centroids of the fishnet, but also on the number of soybean crop areas therein. φ is the Lagrange multiplier.

The dispersion around the resulting estimate of risk, or the kriging variance, was obtained by [32]:

$$\begin{aligned}\hat{\sigma}^2_R\left(u_0\right) = C^{Z}_{(0)} - \sum_{i=1}^{N} \lambda\left(u_i\right) C^{ZR}_{(u_i,u_0)} + \varphi = \\ C^{R}_{(0)} - \sum_{i=1}^{N} \lambda\left(u_i\right) C^{R}_{(u_i,u_0)} + \varphi \text{ if } i \neq 0\end{aligned} \tag{22}$$

The binomial kriging prediction values depend on the structure of the semivariogram or the covariance function.

2.4.2. Gaussian Areal Kriging Model

The soybean crop yield model does not observe a realization of the point-level process, $Z(s)$, but instead uses census data (B) pertaining to the 141 vector boundaries of the Mato Grosso municipalities, $Z(B) = (Z(B_1),..., Z(B_n))'$, and the prediction of $Z(A)$ is of interest. The A can be general, leading to several different types of change-of-support problems. For example, when B_i values are points instead of areas and A_i values are areas, the point-area change-of-support problem arises. If A_i values are points and B_i values are areas, then the problem becomes one of the spatial disaggregation, an area-to-point COSP, requiring the prediction of a spatial intensity function from aggregated data. A_i and B_i may be different geographic regions at the same spatial scale or different spatial scales. In such cases, A may be wholly contained in one of the B_i, or it may overlap with two or more of the B units. Thus, the best linear unbiased predictor had the form [5]:

$$\hat{Z}\left(A\right) = w\left(A\right)' Z\left(B\right), \text{ with } w\left(A\right)' = \left(w_1\left(A\right),...,w_n\left(A\right)\right) \tag{23}$$

where each weight $w_i(A)$ measures the influence of datum $Z(B_i)$ on the prediction of $Z(A)$.

Gotway and Young [5] provides the optimal weights, as follows:

$$w_u = \left(\sum u\right)^{-1} \sigma_u \tag{24}$$

where $w_u = (w_1(A),...,w_n(A), m')' \equiv (w(A)'m')'$, and m is a $(p \times 1)$ vector of Lagrange multipliers.

The $(n + p) \times (n + p)$ matrix Σ_u is based on the data and covariates of the B units [5]:

$$\sum u = \begin{bmatrix} \sum BB & x(B) \\ x(B)' & \phi \end{bmatrix} \tag{25}$$

The elements of the ($n \times n$) matrix Σ_{YY} are the (B_i, B_j) covariance:

$$C(B_i, B_j) = \text{cov}(Z(B_i), Z(B_j)) = \frac{1}{|B_i|}\frac{1}{|B_j|}\int_{B_i}\int_{B_j} C(u-v)\,du\,dv \tag{26}$$

The matrix σ_u comprises information on the A units, as well as their spatial relationships to the B-units, and is equal to:

$$\sum u = \begin{bmatrix} \sigma_{AB} \\ x(A) \end{bmatrix} \tag{27}$$

where the elements of the ($n \times 1$) vector σ_{AB} are the (A, B_i) covariance:

$$C(A, B_i) = \text{cov}(Z(A), Z(B_i)) = \frac{1}{|A|}\frac{1}{|B_i|}\int_{A}\int_{B_i} C(u-v)\,du\,dv \tag{28}$$

The ($p \times 1$) vector $x(A)$ comprises the covariate information associated with A, $x(A) = (x_1(A), x_2(A),..., x_P(A))'$ [5].

Gotway and Young [5] sets forth the corresponding mean-squared prediction errors:

$$\sigma_{AA} - w_u' \, \sigma_u \tag{29}$$

with $\sigma_{AA} = \text{cov}(Z(A), Z(A))$.

The measurement error in the data $Z(B_i)$ can be accommodated by assuming that:

$$Z(B) = \int_B Z(s)\,ds + \varepsilon(B) \tag{30}$$

in addition to using a filtered version of this predictor [5,33].

When it is necessary to predict $Z(s)$ associated with a point s using data $Z(B)$, the predictor becomes:

$$\hat{Z}(s) = \sum_{i=1}^{N} w_i(s) Z(B_i) \tag{31}$$

The weights $w_i(s)$ can be obtained using Equation (24) with σ_u replaced by $[\sigma_{sB}'\ x(s)'\ ]'$, where σ_{sB} is the ($n \times 1$) vector with ith element cov(s, B_i).

The relationship between the semivariogram associated with $Z(B)$ and the one associated with the underlying point-support process, $Z(s)$, is given by Cressie [33] and Gotway and Young [5]:

$$2\gamma\left(B_i, B_j\right) \equiv \mathrm{var}\left(Z\left(B_i\right) - Z\left(B_j\right)\right) = -\frac{1}{\left|B_i\right|\left|B_i\right|}\int_{B_i}\int_{B_i}\gamma\left(s-u\right) \\ -\frac{1}{\left|B_j\right|\left|B_j\right|}\int_{B_j}\int_{B_j}\gamma\left(s-u\right)dsdu + \frac{2}{\left|B_i\right|\left|B_j\right|}\int_{B_i}\int_{B_j}\gamma\left(s-u\right)dsdu \tag{32}$$

where $\gamma(s - u) = (1/2)\mathrm{var}(Z(s) - Z(u))$ is the semivariogram of the point-support process $\{Z(s)\}$.

The binomial and Gaussian kriging implementations were synthesized in Figure 2.

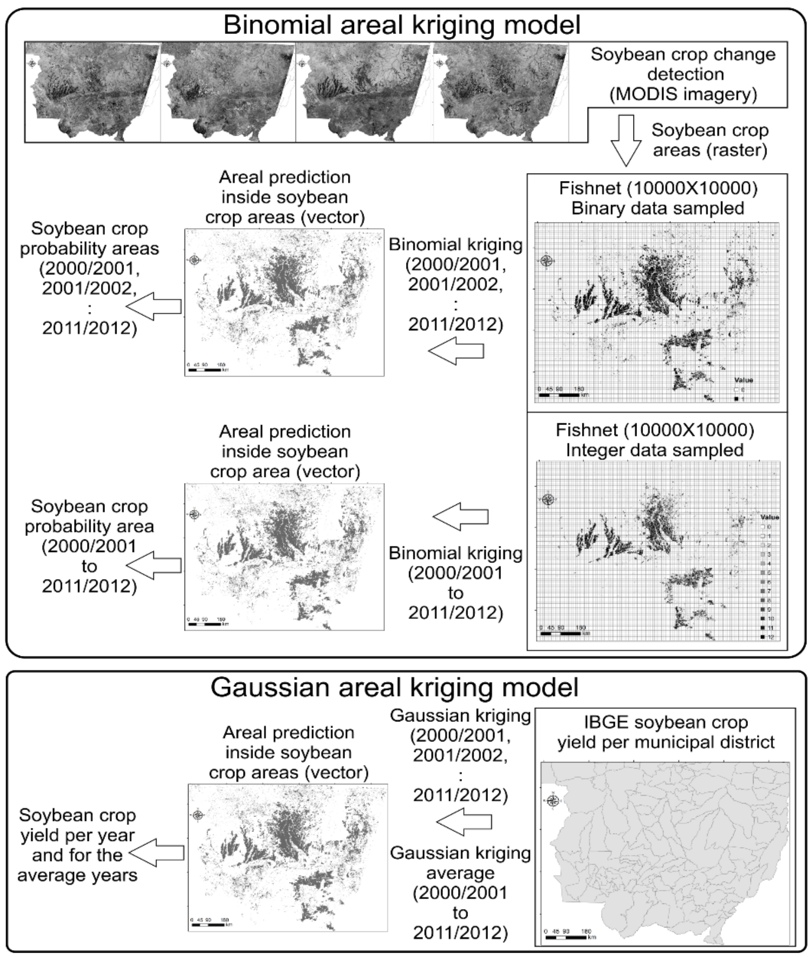


Figure 2. Scheme of binomial and Gaussian areal kriging implementation.

2.4.3. Block Kriging Model

The prediction results of the binomial and Gaussian areal kriging at locations with point support were predicted in the soybean crop areas detected by remote sensing. The percentage of soybean crop occurrence and yield were predicted by block kriging, which was used to predict the average value of the process at a larger scale, accounting for not only the size but also the shape and orientation of volume *B*. The prediction of the spatial average from point-support measurements $Z(s_1)$, $Z(s_2)$,..., $Z(s_n)$ was:

$$Z(B) = \frac{1}{|B|}\int_B Z(s)ds \tag{33}$$

The set *B* is a soybean cropland region that forms the spatial support of $Z(B)$. The resulting predictor is given by:

$$\hat{Z}(B) = \sum_{i=1}^{N} w_i Z(s_i) \tag{34}$$

The optimal weights (w_i) were obtained by applying [5,33,34]:

$$\begin{aligned} &\sum_{i=1}^{N} w_k C(s_i, s_k) + \sum_{j=1}^{P} m_j x_j(s_i) = C(B, s_i) \qquad i = 1,\dots,N \\ &\sum_{j=1}^{P} w_i x_j(s_i) = x_j(B) \qquad j = 1,\dots,P \end{aligned} \tag{35}$$

where $x_j(B) = \frac{1}{|B|}\int_B x_j(u)du, \quad j = 1,2,\dots,p$; m_j represents the Lagrange multipliers from the constrained minimization, and $C(B, s_i)$ is $C(B,s_i) = \mathrm{cov}(Z(B), Z(s_i)) = \int_B C(u,v)dudv / |B|$.

The advantage of predicting $Z(B)$—rather than predicting $Z(u_j)$ for many points u_j in *B* and averaging the resulting predictions—is the ease of obtaining an uncertainty measure of the resulting predictor. The mean-squared prediction error associated with $\hat{Z}(B)$ was [5,33,34]:

$$C(B,B) - \sum_{i=1}^{N} C(B, s_i) - \sum_{j=1}^{P} m_j x_j(B) \tag{36}$$

where $C(B,B) = \frac{1}{|B|^2}\int_B\int_B C(u-v)dudv$.

The point-to-point covariance function, $C(u - v)$, is assumedly known for theoretical derivations, but is estimated and modeled with a valid positive definite function based on the point-support data.

2.4.4. Modeling Semivariogram and Covariance Function

The estimation of the covariance terms in the kriging system requires knowledge of the point-support covariance or point-support semivariogram model. Given the availability of areal or polygon data only, it was necessary to compute and model the variogram of the areal data and deconvolute the block-support model to derive the point-support semivariogram. The point-support variogram was inferred using iterative deconvolution [35]. An experimental semivariogram and a covariogram were computed and then modeled by a weighted least-squares fitting procedure. A candidate point-support model, or deconvoluted model, was then chosen and regularized. The regularized model was then compared to the model fitted to the experimental variogram of the areal or polygon data. Based on the differences between the regularized model and the areal model, the optimal point-support model was rescaled, providing a new candidate model for the next iteration. The deconvoluted model was closest to the model for the areal data, and was used for the area-to-point kriging. This procedure considered the irregular shape and size of areal units [36].

Each polygon was overlaid with a square lattice, and a point was assigned to each intersection in the lattice to produce the semivariogram or covariance function. The lattice spacing parameter specifies the horizontal and vertical distance between each point. The horizontal axis of the semivariogram or covariance function presents the average distance between the polygons, which was calculated using the cells of the overlapping grid. The crosses represent the empirical semivariances, or covariance values, which were calculated using the available averaged data for the polygons.

The line is the estimated point semivariogram or covariance function models. The bars are the confidence intervals, which are calculated assuming re-estimated empirical semivariances or normally distributed and uncorrelated covariance. For the properly specified semivariogram or covariance model, 90% of the empirical covariance is expected to fall within the confidence intervals, indicating that the model fits the data and that the results can be trusted. The line of the semivariogram model might not pass through the confidence intervals, but this is not a problem, and the criteria for a good model does not change. If a large percentage of the empirical semivariances fall within the confidence intervals, the areal kriging can produce more accurate predictions and prediction standard errors than point kriging, with values assigned to the centroids of polygons [3].

The stable covariance function and semivariogram models were used to characterize the soybean harvested areas and yields, respectively. The stable model has the same behavior near the origin and reaches a sill [34]:

$$C(u) = b\exp\left(-\left(\frac{u}{a}\right)^{\alpha}\right) \quad (a > 0,\ \ 0 < \alpha \leq 2,\ \ b > 0) \tag{37}$$

where *b* is the sill, *a* is the range, and *α* is the parameter.

Wackernagel [31] determines the stable semivariogram model:

$$\gamma(u) = b\left(1 - \exp\left(-\frac{|u|^{\alpha}}{a}\right)\right) \quad (a > 0,\ \ 0 < \alpha < 2,\ \ b > 0) \tag{38}$$

2.5. Validation Phase

The *in situ* data were kindly provided by the farmers of five visited agglomerates. This data referenced the cultivation practices applied at the crop field level in the 2010/2011 crop season, and included: variety, sowing, germination, and harvesting dates, soil texture, total planted area, and the number of 60-kg bags of harvested crop per hectare (ha) of each crop field. The 2010/2011 crop season was chosen for analysis because it sustained minimal impact from climatic variations caused by La Niña (rainfall reduced) or El Niño (excessive rainfalls) [37], which are phenomena that directly affect Mato Grosso, causing losses in yield [38].

The soybean crop expansion occurred in four main agricultural regions: (1) along the BR-163 highway that links Cuiabá to Santarém (State of Pará); (2) on the Parecis plateau, in the western region, including Sapezal and Diamantino; (3) the eastern region, near Santo Antônio do Leverger and Querência; and (4) the southeastern region, near Rondonópolis and Primavera do Leste. In this context, five agglomerates with high soybean production were chosen for the validation: Vale do Rio Verde (northern region, in Tapurah/Ipiranga do Norte); Santa Luzia and Colorado (western region, in Sapezal and Diamantino, respectively); Malu (eastern region, in Bom Jesus do Araguaia/Ribeirão Cascalheira); and Colibri (southeastern region, Santo Antônio do Leverger) (Figure 3).

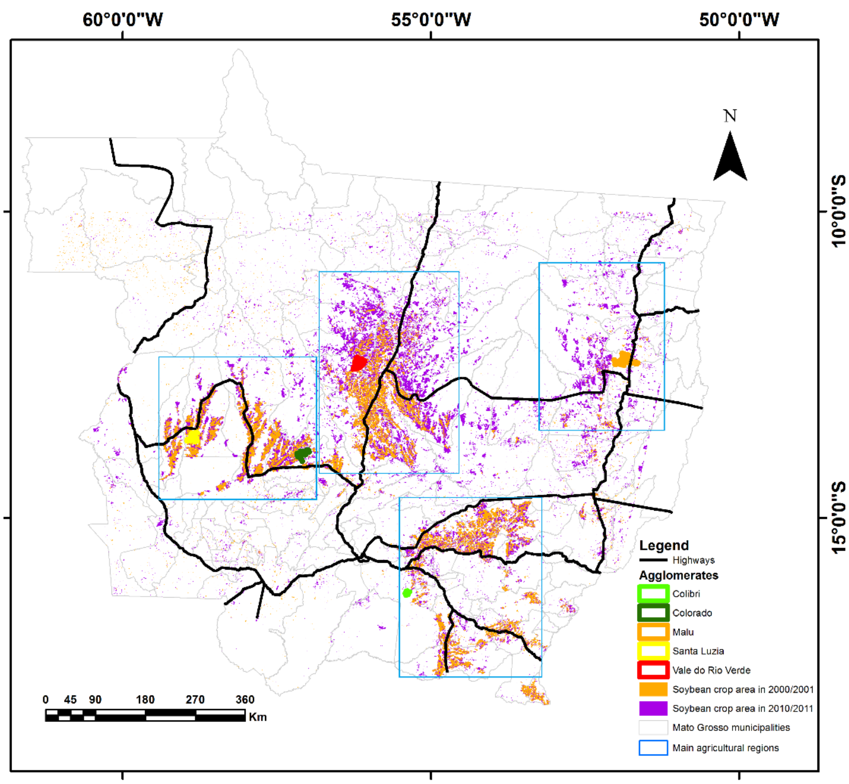


Figure 3. Geographic location of the Colibri, Colorado, Malu, Santa Luzia and Vale do Rio Verde agglomerates used for the validation phase. The four main agricultural regions of Mato Grosso are indicated in blue.

2.5.1. Soybean Crop Area Validation Phase

Punctual crop field analyses were applied to verify the accuracy of the SEI_{Pixel} mapping. To validate the “Soybean” class, one point was chosen from each crop field, totaling 386 samples. To validate the "Non-soybean" areas, the TerraClass land use/land cover map of 2012 [39] was used. TerraClass is a partnership between the National Institute for Space Research (INPE) and the Brazilian Company for Agricultural Research (EMBRAPA) that aimed to classify land use and occupation in the deforested areas of the Legal Amazon. The TerraClass map had the same number of sampling points, but these points were distributed in “Non-soybean” areas. The classification was assessed using an error matrix, overall accuracy, and Kappa statistics [40].

2.5.2. Soybean Crop Yield Validation Phase

To effectuate downscaling at the crop field level, the validation of soybean crop yields involved the verification of the soybean crop yield for each crop field in the agglomerates relative to the yield predicted for 2010/2011, computing the number of 60-kg bags obtained per ha^{-1}. For this, each soybean crop field detected on the farms was sampled once, and the yields obtained from the *in situ* data were compared with the map of yield classes generated by the Gaussian areal kriging model. For each crop field, the total number of 60-kg bags of soybean harvested per ha^{-1} at the end of the harvest period was recorded and compared with the prediction. The results were presented in five classes of 60-kg bags per ha^{-1}: less than or equal to 0, from 0.1 to 25, from 25.01 to 45, 45.01 to 55, and equal to or greater than 55.1.

The approach adopted in this study is synthesized in Figure 4.

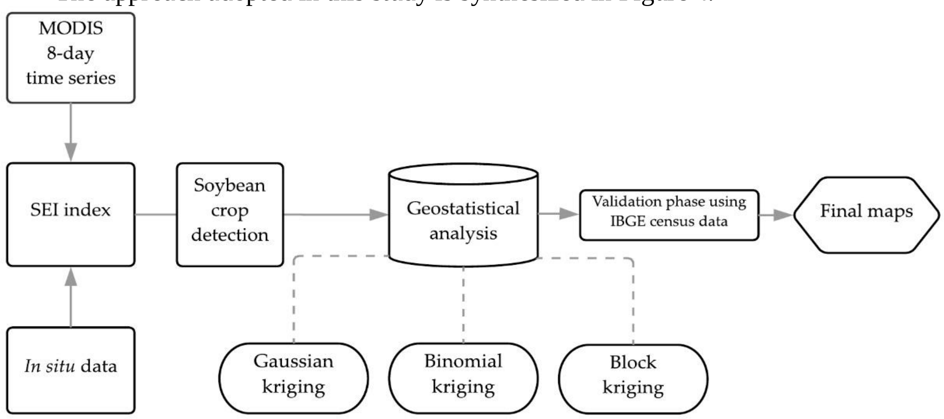
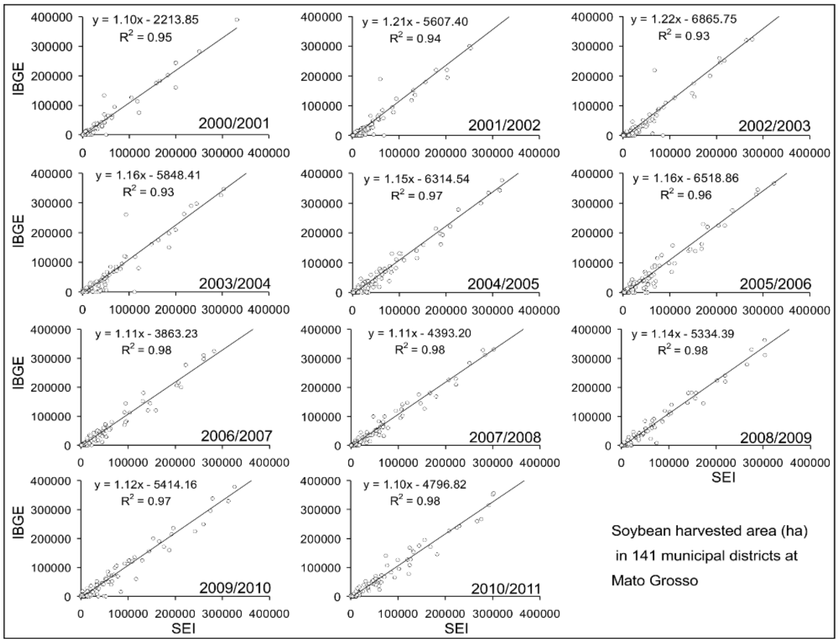


Figure 4. Technical flowchart.

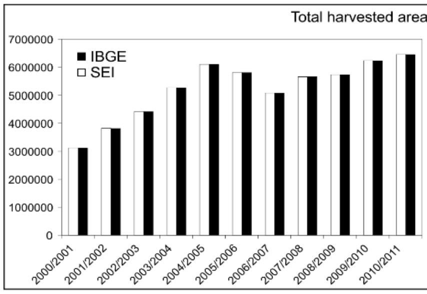
3. Results

3.1. Soybean Crop Area Results

The SEI_{Pixel} enabled the detection of soybean harvested areas when compared with IBGE data for the municipalities of Mato Grosso. The regression models between the SEI_{Pixel} and IBGE presented strong correlations, which were represented by R^2 values above 93% (Figure 5).



Total harvested area (ha) in Mato Grosso



Harvest period	IBGE (ha)	SEI (ha)	Threshold
2010/2011	6454331	6460429	0.5284
2009/2010	6226452	6238955	0.404
2008/2009	5730502	5735283	0.557
2007/2008	5659149	5661801	0.613
2006/2007	5075079	5077324	0.4735
2005/2006	5811907	5814920	0.497
2004/2005	6106654	6110462	0.483
2003/2004	5263428	5270469	0.489
2002/2003	4413271	4415837	0.518
2001/2002	3818231	3821444	0.284
2000/2001	3121353	3121944	0.306

Figure 5. Evaluation of the model for soybean harvested area estimation based on Moderate Resolution Imaging Spectroradiometer (MODIS) data, when compared to IBGE statistics per municipality and the total harvested area in Mato Grosso.

3.1.1. Binomial Areal Kriging Model for Soybean Crop Area Identification

The covariance function described how soybean harvested areas vary, on average, in Mato Grosso. Since the spatial variability does not change with direction, omnidirectional covariograms were computed, and stable models were fitted using the least-square algorithm (Figure 6).

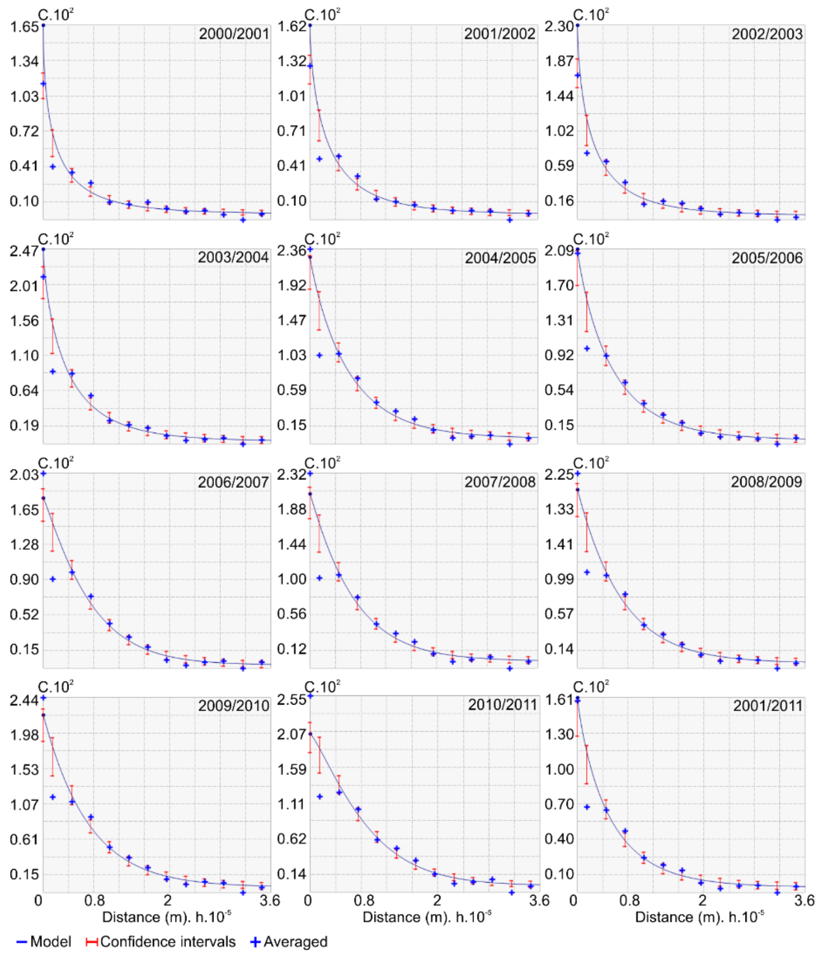


Figure 6. Empirical covariance values (crosses), estimated point of the stable covariance model (line), and bars with the 90% confidence intervals of the binomial areal kriging of the probability of occurrence of soybean crop areas per harvest period and the averaged years (2001/2011).

The values for the binomial kriging equations were obtained based on the mathematical expression of the covariance function. The kriging equations, the estimated percentage of soybean crop areas, kriging variance, and standard errors were solved. Then, the binomial kriging results were predicted for the soybean harvested vector areas detected by remote sensing using block kriging. The maps were classified into percentage of occurrence classes (Figure 7).

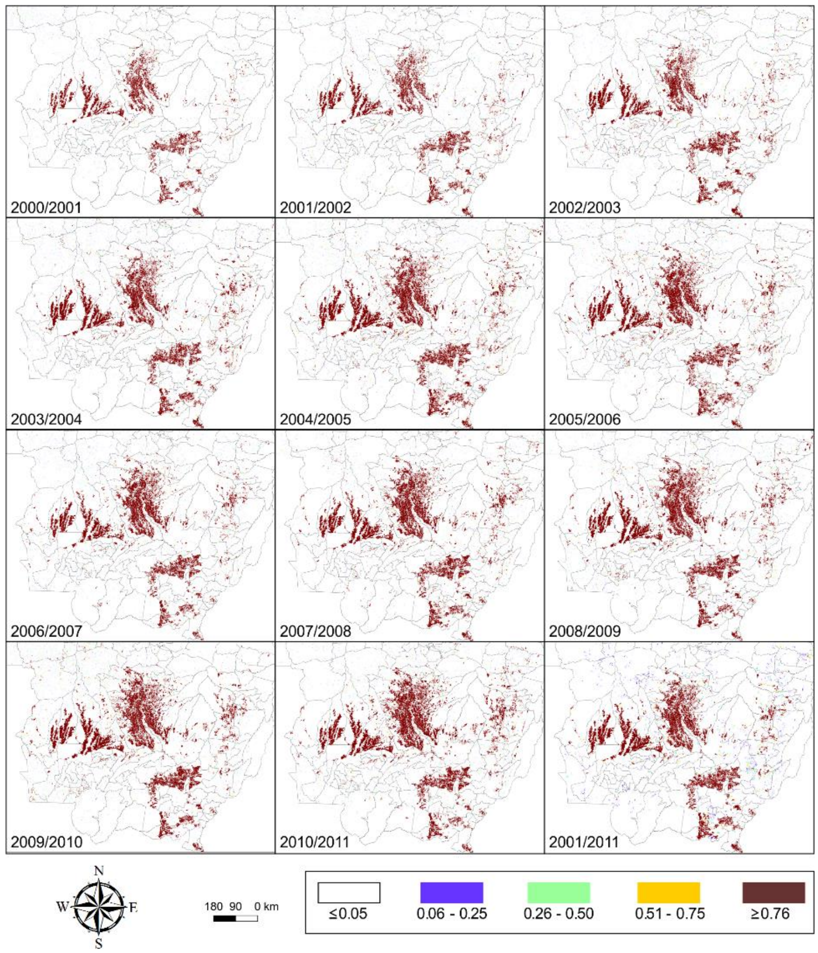


Figure 7. Binomial areal kriging prediction of the percentage of occurrence of soybean crop areas inside soybean harvested areas detected by remote sensing, per harvest period (2000/2001 to 2010/2011) and the averaged years (2001/2011).

The kriging errors were mapped in the same way as the percentage of soybean crop areas. In general, the errors associated with the estimates were small. The largest errors, which were those above 3%, occurred mainly in sparse rural areas; however, highly aggregated areas in regions of Sorriso, Lucas do Rio Verde, Nova Mutum, Nova Maringá, and Sapezal also presented errors above 3%. The periods 2006/2007 and 2010/2011 presented low errors in the sparse rural areas. The period 2010/2011 also presented low errors in highly aggregated soybean crop areas (Figure 8).

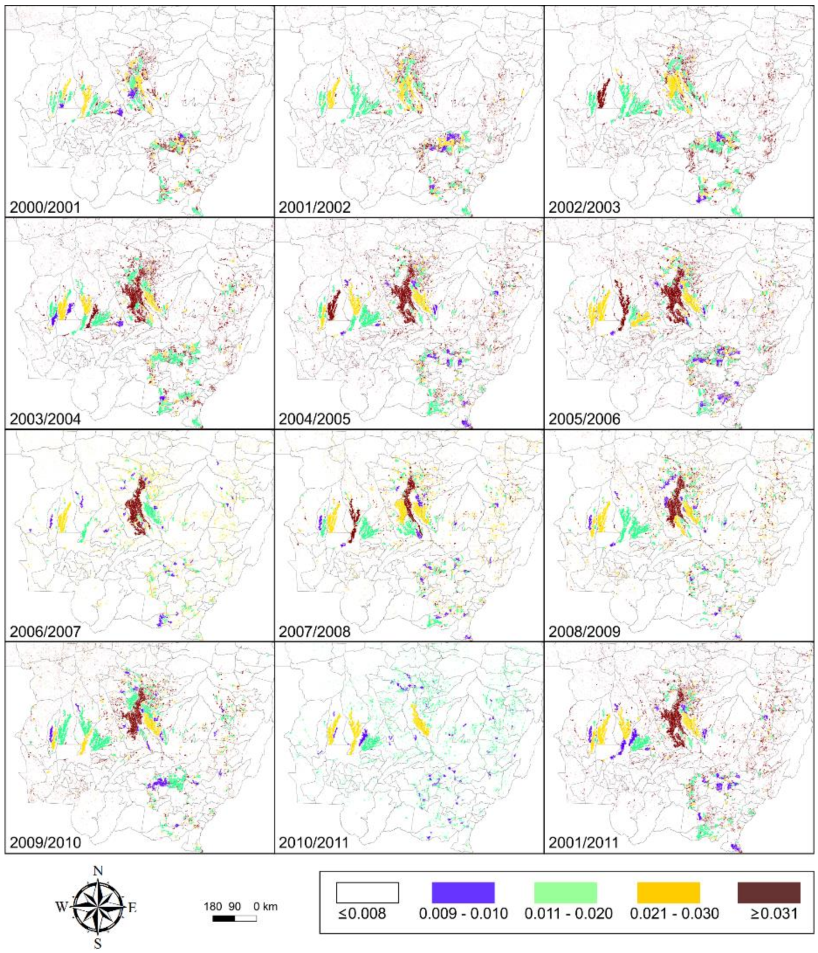


Figure 8. Binomial areal kriging prediction standard error of the percentage of occurrence of soybean crop areas inside soybean harvested areas detected by remote sensing, per harvest period (2000/2001 to 2010/2011) and the averaged years (2001/2011).

3.2. Accuracy Assessment of Soybean Crop Area Results

The SEI_{Pixel} classification accuracy was evaluated by the analysis of the soybean mapping and the 2010/2011 *in situ* checkpoint data (Table 1).

Table 1. Classification errors detected by the punctual crop field area analysis.

Agglomerate	**Soybean classified as Soybean**	**Soybean classified as Non-soybean**	**Non-soybean classified as Soybean**	**Crop fields in 2010/2011**	**Accuracy Ratio**
Santa Luzia	57	0	10	67	85.1%
Colorado	61	1	6	68	89.8%
Vale do Rio Verde	77	0	12	89	86.6%
Colibri	21	1	0	22	95.5%
Malu	151	17	14	182	83.0%
Total	**367**	**19**	**42**	**428**	**85.8%**

The absolute identification of other land uses as being "Soybean" occurred on a major scale, which was possibly due to the similarity in temporal behavior between soybean and other summer crops such as cotton and corn, which were well represented in the region. Subsequently, an accuracy assessment of soybean classification was conducted using the reference data (Table 2).

Table 2. Error matrix with SEI_{Pixel} and *in situ* checkpoint data, and mapping accuracy for the 2010/2011 crop season.

Classified data	**Reference data**				
	Soybean	Non-soybean	References	Commission errors	User's accuracy
Soybean	**367**	42	409	0.10	0.90
Non-soybean	19	**344**	363	0.05	0.95
References	386	386	772		
Omission errors	0.05	0.11			
Producer's accuracy	0.95	0.89			
Overall accuracy	***0.92***				
Kappa Index	***0.84***				

The SEI_{Pixel} classification was compared with *in situ* data in the analysis of the spatial distribution of the probability of soybean occurrence in the agglomerates Vale do Rio Verde, Colorado, Santa Luzia, Colibri, and Malu in the 2010/2011 crop season (Figure 9).

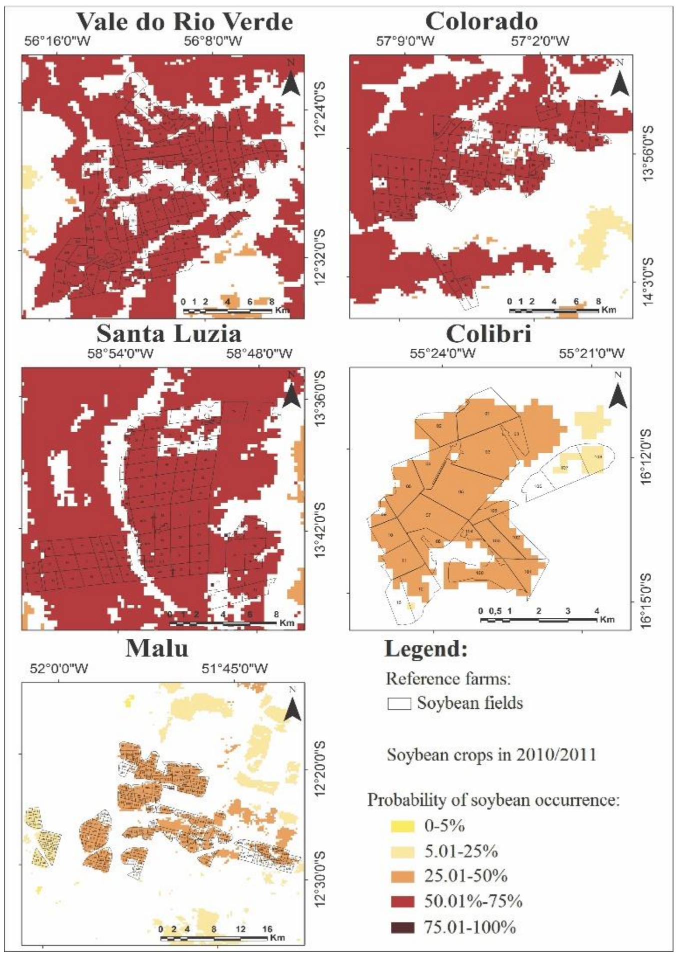


Figure 9. Spatial distribution of the probability of soybean occurrence in the agglomerates Vale do Rio Verde, Colorado, Santa Luzia, Colibri, and Malu in the 2010/2011 crop season.

3.3. Soybean Crop Yield Results

3.3.1. Gaussian Areal Kriging Model for Yield Prediction

While the covariance function has described the variation among the soybean-harvested areas in Mato Grosso, the semivariogram can also be used to describe soybean crop yield throughout the space. Omnidirectional semivariograms have been computed using stable models fitted by the least-squares algorithm. The deconvoluted point semivariogram (line) differed from the estimated empirical semivariogram values for the polygons. In this case, areal kriging, in which values are assigned to polygons, can produce more accurate predictions and fewer prediction standard errors than point kriging. Almost all of the re-estimated empirical semivariogram values for polygons (crosses) were inside the 90% confidence intervals (vertical lines) (Figure 10).

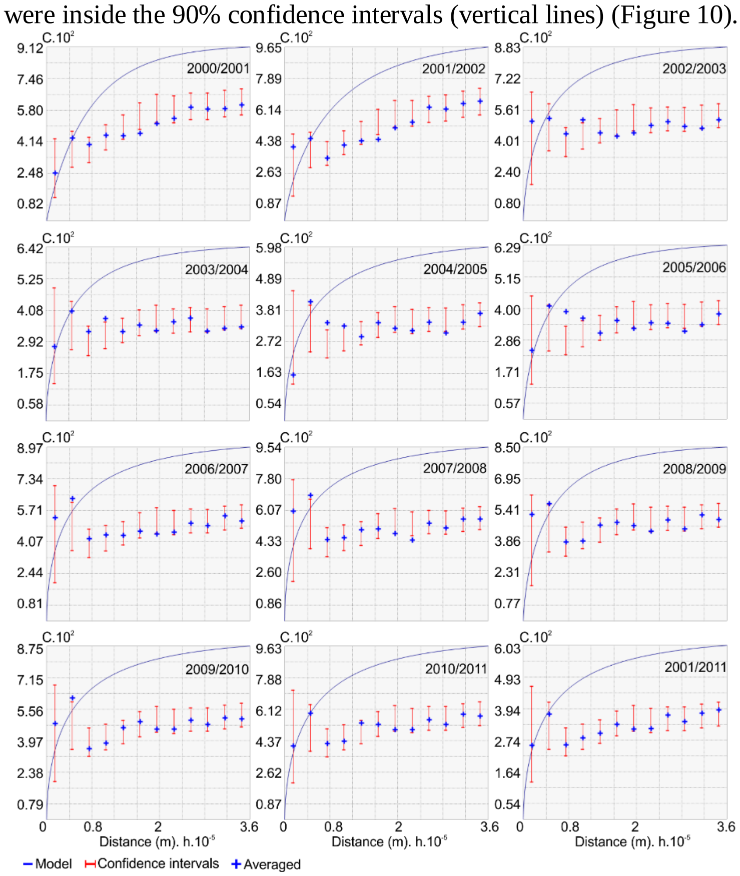


Figure 10. Empirical semivariogram values (crosses), estimated point of the stable semivariogram model (line), and bars with the 90% confidence intervals of the Gaussian areal kriging of the soybean crop yield (total number of 60-kg bags per ha^{-1}) per harvest period (2000/2001 to 2010/2011) and the averaged years (2001/2011).

The Gaussian areal kriging results were predicted for the soybean harvested vector areas that were detected by remote sensing using block kriging. The highest yield occurred not only in the aggregated soybean crop areas but also in sparse rural areas. However, major producer areas represented a higher proportion of the total yield, and those with smaller areas represented a lower proportion of the total yield. The major producer-aggregated areas presented high yields, particularly in 2010/2011 (Figure 11).

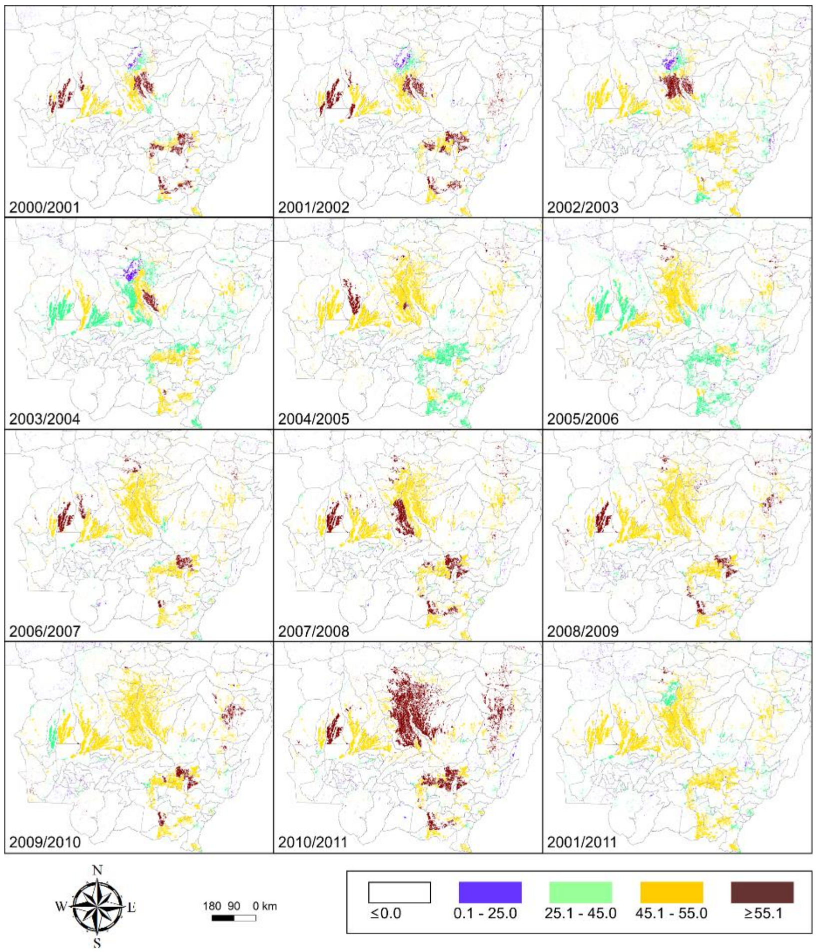


Figure 11. Gaussian areal kriging prediction of the soybean crop yield (total number of 60-kg bags per ha^{-1}) inside the soybean-harvested areas that were detected by remote sensing, per harvest period (2000/2001 to 2010/2011) and the averaged years (2001/2011).

Yield predictions with the most uncertainty did not necessarily correspond to units with low-yield predictions, but the smallest uncertainty values tended to be associated with the predicted high-yield regions (Figure 12).

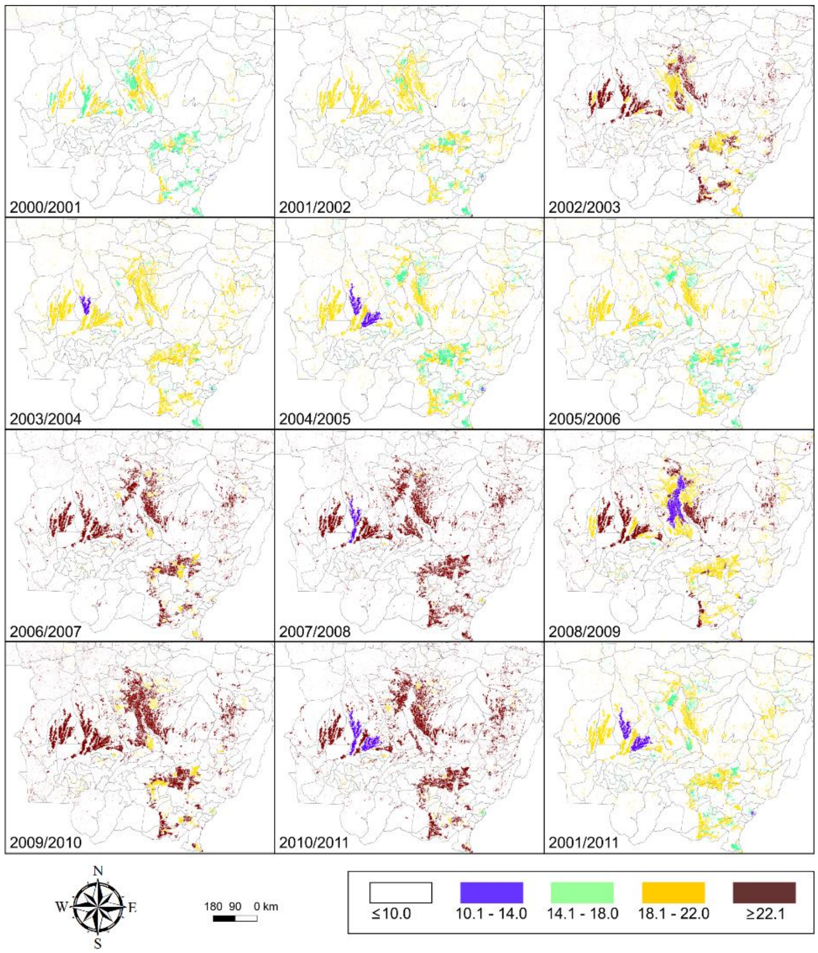


Figure 12. Gaussian areal kriging prediction standard error of the soybean crop yield (total number of 60-kg bags per ha^{-1}) inside the soybean-harvested areas that were detected by remote sensing, per harvest period (2000/2001 to 2010/2011) and the averaged years (2001/2011).

3.4. Accuracy Assessment of Soybean Crop Yield Results

A comparison of yield estimates (60-kg bags per ha^{-1}) for each agglomerate and the *in situ* data is presented in Table 3.

Table 3. Yield classification errors detected by the punctual crop field analysis.

Agglomerate	Soybean crop fields	Correct class of soybean crop yield	Standard deviation	Accuracy ratio
Santa Luzia	57	13	23.33	22.80%
Colorado	61	16	23.69	26.22%
Vale do Rio Verde	77	38	23.33	49.35%
Colibri	21	10	23.69	47.61%
Malu	151	129	23.33	85.43%
Total	**367**	**206**	**23.54**	**56.13%**

Given the limitations, the standard deviation component derived from the interpolation (Equation 40) is evaluated [41]:

$$E_{50} = \pm 0.6745.\sigma \tag{39}$$

The probable error was considered in the yield evaluation, because the defined yield classes do not include the extensive crop fields for which the high yield was approximated. The soybean crop yield prediction analysis was improved by accounting for the probable error (Table 4).

Table 4. Yield classification considering the probable error (E_{50}) in the interpretation of the standard deviation.

Agglomerate	Crop fields	Standard deviation	50% error ($E_{50} = \pm 0.6745.\sigma$)	Correct class of soybean crop yield	Accuracy Ratio
Santa Luzia	67	23.33	±15.74	57	85.00%
Colorado	68	23.69	±15.98	67	98.53%
Vale do Rio Verde	89	23.33	±15.74	85	95.51%
Colibri	22	23.69	±15.98	21	95.45%
Malu	182	23.33	±15.74	177	97.25%
Total	**428**	**23.54**	**±15.84**	**407**	**95.09%**

The spatial distribution of the soybean crop yield in the agglomerates Vale do Rio Verde, Colorado, Santa Luzia, Colibri, and Malu in the 2010/2011 crop season (Figure

13) highlights the predominance of the soybean crop in the agricultural regions of Mato Grosso.

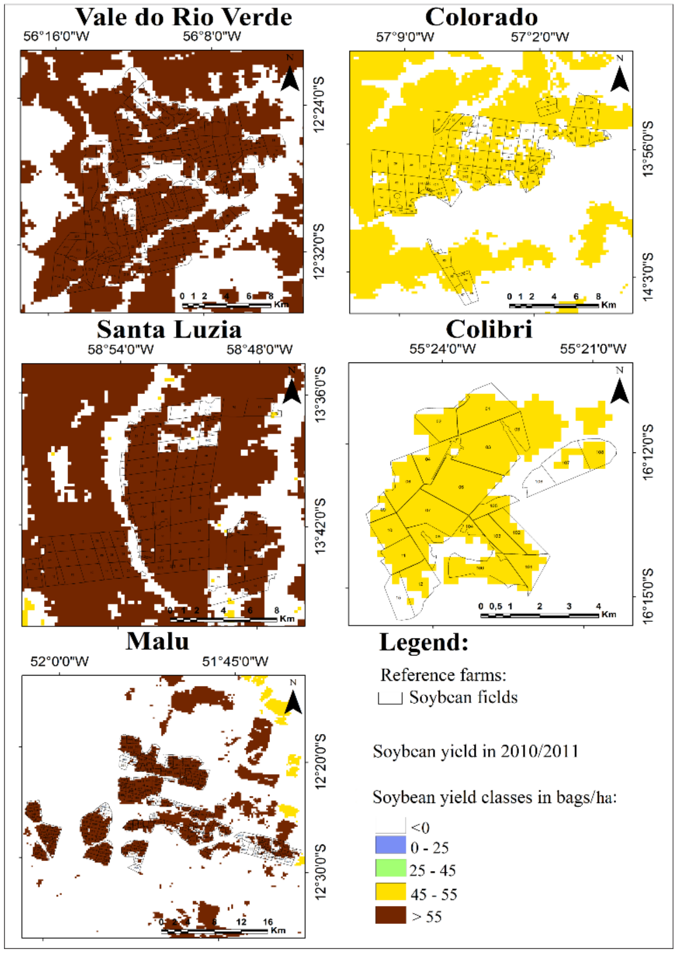


Figure 13. Spatial distribution of the soybean crop yield (60-kg bags per ha^{-1}) in the agglomerates Vale do Rio Verde, Colorado, Santa Luzia, Colibri, and Malu in the 2010/2011 crop season.

4. Discussion

4.1. Soybean Crop Area Results

Gusso et al. [42] used MODIS EVI images from the soybean-sowing period in Mato Grosso to observe their approximated thresholds (Figure 5). Arvor et al. [8], using

MODIS EVI for crop mapping in Mato Grosso, observed accuracies higher than 95% in 2006/2007. Risso et al. [18], evaluating MODIS EVI and NDVI for soybean crop area discrimination, also observed a satisfactory application of EVI images from the maximum crop development period for soybean classification in Mato Grosso. Similarly, Souza et al. [43] noted that an EVI time series could be compared to official data to identify soybean croplands. For Victoria et al. [44], estimates based on MODIS vegetation indices presented good agreement with large cultivated areas, but poor agreement with municipalities with small cropland areas. These studies used 16-day composites derived only from MOD13Q1 images. In contrast, we used eight-day composites derived through combining the MOD13Q1 and MYD13Q1 images. Furthermore, the aforementioned studies did not consider large-scale interannual variations resulting from the agricultural dynamics in Mato Grosso. We address this issue with the proposed SEI_{Pixel}, which covers the phenological cycle of the crop through the 2000/2001 to 2010/2011 crop seasons. The analysis of eight-day rather than 16-day composites, as well as the consideration of large-scale interannual variations, allows the detection of subtle spectral–temporal differences between crop types, improving the method's crop area identification.

Although IBGE data were used as a reference, the high correlation between the estimates generated by the SEI_{Pixel} and the IBGE data for some municipalities indicates that the sampling may not accurately represent the IBGE estimate. In this respect, adopting a more reliable reference, with satellite image mapping at greater spatial detail, allowed us to determine the level of uncertainty in the objective and subjective methods of crop forecasting. We were able to ascertain the planted soybean area at the end of the crop development cycle but before harvest, unlike the IBGE survey, which extends beyond the end of the harvesting period.

The intensification of soybean cropland was described by the iSEI_{Pixel}. The minimum and maximum iSEI_{Pixel} values were 0 and 12, respectively. Based on the chosen threshold values, SEI_{Pixel} overestimated the total soybean harvested areas when compared with the IBGE data. Due to the impossibility of determining the real soybean harvested area in Mato Grosso, the use of binomial kriging was necessary to measure the uncertainty in the estimates.

4.1.1. Binomial Areal Kriging Model

The binomial areal kriging results show the soybean expansion in Mato Grosso. Kastens et al. [9] evaluated soybean cultivation in Mato Grosso during 2001 and 2014 in relation to the Soy Moratorium initiative, finding the deforestation rate in the pre-Soy Moratorium period to be more than five times the post-Soy Moratorium rate, while simultaneously observing the pre-Soy Moratorium forest-to-soy conversion rate to be more than twice the post-Soy Moratorium rate. Gibbs et al. [45] observed that, in the two years prior of the agreement, almost 30% of the soybean expansion had occurred through deforestation. After the Soy Moratorium, this value decreased to 3% in the Amazon biome until 2011. These observations indicate that the Soy Moratorium played a role in reducing both deforestation and the subsequent use of the land for soybean production. However, the authors considered that, in the Cerrado biome, where the Soy Moratorium does not apply, the annual rate of soybean expansion over native vegetation areas has remained high, varying from 11% to 23% between 2007 and 2013, demonstrating that the Cerrado is less protected by environmental laws than the Amazon rainforest.

This information corroborates the results obtained by the SEI_{Pixel} index. An analysis of the maps of each crop season shows that the soybean crop areas expanded mostly in the Cerrado biome. Figure 7 illustrates that the expansion of the soybean crop areas occurred in the four main agricultural regions cited above in item 2.5. The same scenario was found by Gusso et al. [11], Arvor et al. [46] and Dubreuil et al. [47]. In these four agricultural areas, located in the Cerrado biome, the crop identification showed that soybean is the main crop.

Regarding the Amazon biome, region (1) presented soybean expansion in forest areas. This deforestation can be explained by the direct influence exerted by the BR-163 (the main route between the north and south regions of Mato Grosso and principal access road to the Santarém port) on the regional landscape change [47]. According to Morton et al. [48], the improvements of the roads network and the construction of grain storage facilities in northern Mato Grosso have encouraged the direct conversion of forest to soybean crops. To circumvent the Soy Moratorium, the deforested areas in this region were first converted to pasture before being used for soybean production [16,37,47].

Considering that the spatial structure of the soybean areas in Mato Grosso shows that soybean crops are spatially correlated, and that the correlation of the underlying probability of soybean occurrence extends by about 180 km on average, the binomial kriging results mean that soybean crop areas and the uncertainty in their prediction can be mapped geostatistically. The maps show that sparse rural areas present lower percentages of soybean occurrence than the highly aggregated soybean crop areas. Using binomial kriging, we combined different types of geographical units and various sources of information to create maps illustrating where the percentage of soybean harvested areas varied continuously in the space, thereby reducing the visual bias associated with the interpretation of the classified vector maps. By measuring the variance of the prediction errors, we identified large and sparse areas where the estimates were less reliable.

Binomial kriging can be considered a noise filter for generating maps of regional patterns, such as information about the concentration of soybean-harvested areas. This result can be observed in the map of the averaged years, where sparse areas present a low probability of soybean occurrence, due to less agricultural intensification over the years, compared to consolidated agroecosystems (Figure 7). Galford et al. [19] used a wavelet analysis of eight-day MODIS time series data as a filtering approach to detect crops in Mato Grosso, but they left a gap in relation to the determination of uncertainties. The binomial kriging results are promising for the determination of spatial prediction uncertainties as well as for soybean mapping.

4.2. Accuracy Assessment of Soybean Crop Area Results

The comparison between the reference data and the classification of soybean crop areas for the 2010/2011 crop season indicated high alignment. The global accuracy and the Kappa index obtained from the error matrix were 92.1% and 0.84 for the soybean crop, respectively. According to Foody [49], a global accuracy higher than 85% is desirable. However, according to the levels of agreement of Kappa index values proposed by Landis; Koch [50], a classification above 0.80 is considered excellent and highly aligned with the *in situ* data. On the basis of these indices, the classification could be considered excellent, which indicates the potential for the identification of soybean crop areas in Mato Grosso. These results confirm the potential for using eight-day MODIS EVI composites with geostatistical techniques to map soybean crops.

The values we obtained were even more significant when compared with studies that used data from the same sensor. Arvor et al. [8], using only the MOD13Q1 (16-day time series) to evaluate soybean crop expansion in the vicinity of protected areas in Mato Grosso, obtained 74% global accuracy and a Kappa value of 0.675. Lamparelli et al. [51] obtained Kappa values between 0.60 and 0.80 when estimating soybean crop areas with MODIS data. The values we obtained for user and producer accuracy were also considered excellent. The omission errors were 0.05 for the "Soybean" class, and 0.11 for the "Non-soybean" class. Nevertheless, the commission errors were 0.10 for "Soybean" and 0.05 for "Non-soybean". Epiphanio et al. [52], using the spectral–temporal response surface (STRS) method for soybean classification in Mato Grosso, obtained omission errors of 0.06 and commission errors of 0.17 for the soybean class, as well as 80% global accuracy and a Kappa value of 0.26, with 343 checkpoints.

The detected errors can be explained by both *in situ* and orbital data. The *in situ* data reveals that the distribution of some crop fields is not representative in terms of yield in edge areas bordering other land uses. Therefore, their temporal values of maximum and minimum EVI, amplitude, and mean and standard deviation, are different to the medium and high-yield crop fields. With the mixed response resulting from spectral mixing between different land uses, these crop fields tended to fall below the threshold that is used to define the "Soybean" class, and were thus considered by the classification algorithm as within the "Non-soybean" class. Another important factor is the spatial variability of agricultural calendars. Due to the large extent of Mato Grosso and the differences in edaphoclimatic conditions, soybeans are not sowed uniformly on all farms, and this non-uniformity of phenological cycles caused spectral confusions. This problem was also detected by Zhu et al. [20] and Arvor et al. [8], who evaluated the spatial distribution of soybean crops in Mato Grosso. The soybean crop is sown from late September to early November (the onset of the rainy season), and harvested from January to March. The soybean cotton crop can be confused with the cotton crop, because the sowing period of cotton is in December, immediately following the September–November sowing period of soybean [16].

Other factors that can affect the sowing date are the phenological cycle of the crop variety, edaphoclimatic differences, and economic and social factors related to the global food supply. Uncertainty about the exact location of soybean crops, which can be caused by the satellite image spatial resolution, sowing period, crop variety, agricultural intensification, and census data from large municipalities, can result in commission and omission errors in soybean crop detection.

Regarding the analysis of orbital data, two factors can cause errors: spectral similarity and pixel size. The spectral similarity between soybean crops and other summer crops such as corn and cotton may confuse the classification algorithm. Lu et al. [53] observed that the errors in agriculture classifications using MODIS are directly linked to the size of the pixel in proportion to the size of the crop field. Each MODIS pixel has a spatial detection potential of 6.25 hectares. Therefore, crop fields smaller than 6.25 ha may be subject to interference from other land uses, which obscure their exact temporal pattern; consequently, they may not be considered as "Soybean". From the *in situ* data, we identified crop fields smaller than 6.25 hectares in all of the agglomerates.

In terms of area, the classification made by the SEI_{Pixel} algorithm detected 6,460,429 ha cultivated with soybean crops in the 2010/2011 crop season. In comparison, the IBGE estimate for the same period was 6,454,331 ha, showing a coefficient of determination (R^2) of 0.98. These results are in line with those obtained by Gusso et al. [42], who also analyzed the temporal variation of the soybean EVI profile, finding scores between 0.93 and 0.98 for the period between 2001 and 2005.

4.3. Soybean Crop Yield Results

4.3.1. Gaussian Areal Kriging Model

The results for the agglomerates reveal a high alignment between the *in situ* data and the classification achieved by the SEI_{Pixel}. The classification presented a probability of soybean occurrence of 50–75% for the Vale do Rio Verde, Santa Luzia, and Colorado agglomerates, and 25–50% for Malu and Colibri. Geographic location may explain this variation: Malu is in the eastern part of the State, and Colibri is in the south. In addition to their agricultural areas, both of these regions have extensive pasturelands, which show seasonal variation in satellite images. This factor may have caused spectral confusion by reducing the algorithm’s accuracy in capturing the areas of soybean crop fields. The Vale do Rio Verde, Santa Luzia, and Colorado agglomerates are located in the mid–west portion, which is known as the agribusiness region [54], with most of these areas devoted to soybean production [8]. The resulting uniformity of development limits the spectral confusion from mixed pixels, and allows better soybean crop detection.

Some adjustments may increase the reliability of soybean identification. One of these consists of using the phenological metrics of the crop cycle. Recently, the use of these parameters as the main input data has gained recognition for directly representing crop characteristics throughout their phenological cycles, such as dates of sowing, emergence, peak maturation, and harvest [23].

The Gaussian areal kriging results were predicted for the soybean-harvested vector areas that were detected by remote sensing using block kriging. The highest yields occurred not only in the aggregated soybean crop areas, but also in sparse rural areas. However, major producer areas represented a higher proportion of the total yield, and those with smaller areas represented a lower proportion of the total yield. The main aggregated producer areas presented high yields, particularly in the 2010/2011 crop season (Figure 11).

4.4. Accuracy Assessment of Soybean Crop Yield Results

Considering that international trade decisions based on inadequate information about the global food supply can have severe economic and social effects, the Gaussian areal kriging results indicate that the downscaling methodology is promising for soybean crop yield visualization with uncertain spatial prediction results. When we have information about the planted soybean crop area per municipality in Mato Grosso, it is possible to make predictions for soybean crop areas using Gaussian areal kriging and remote sensing, generating prediction and uncertainty results.

The high average standard error observed for all of the harvest periods was expected, considering the large size of the Mato Grosso municipalities and the characteristics of the IBGE data. The high prediction standard error values in the soybean crop yield prediction demonstrated ecological fallacy effects. According to Gotway and Young [5], ecological fallacy occurs when analysis based on grouped data leads to conclusions that differ from those based on individual data. The resulting bias is often called ecological bias. The aggregation bias is analogous to the scale and zoning effects in the modifiable areal unit problem. Openshaw [55] observed that ecological fallacy effects are endemic to areal census data, although their magnitude is perhaps not as large as might be expected.

The yield results show that increasing yield was not directly correlated with area expansion. Arvor et al. [16] explained that agricultural intensification has evolved rapidly in Mato Grosso, particularly the proportion of the net planted area cultivated with double-cropping systems and producing two successive harvests of commercial crops. Rudorff et al. [21] categorized the expansion into previously deforested areas in Mato Grosso as a process that was encouraged by the Soy Moratorium established by producers, non-governmental organizations (NGOs), State administrations, and private enterprises. Sentelhas et al. [37], Arvor et al. [16] and Dubreuil et al. [47] explained that the impact of soybean expansion on deforestation in Mato Grosso was indirect, as the deforested areas were first converted to pasture before being used for soybean production.

In relation to the yield prediction, many factors could explain the errors that were detected. In addition to the natural lack of uniformity in cultivation, due to the territorial extent, five other variables complicate the estimation of crop yield at the State level: different crop varieties, crop diseases, climatic components at the regional scale, soil type, and management. According to Prasad et al. [56], factors such as pests, plant diseases, the mass adoption of new hybrid crop varieties with high yields, and the large variations in climatic conditions and agricultural practices can cause local or regional deviations in the predicted crop yield, seriously limiting any forecasting method.

Battisti and Sentelhas [57] has noted that the soybean cultivars that are most tolerant to water stress presented the lowest potential yield, and the less tolerant group presented a higher potential yield. Observing the *in situ* data, we note that the soybean varieties are selected for sowing in Mato Grosso according to higher productivity and higher resistance to environmental stresses. In the agglomerates used in the validation phase, the three most planted varieties were TMG 123 RR, TMG 132 RR, and GB 874 RR, which are considered superprecocial, tolerant to rain at harvest, and resistant to nematodes. Other disease-resistant varieties were also planted on a large scale between 2000 and 2011, such as FMT Tabarana, M 7639 RR, M-SOY 8866, and BRSMT Pintado. While this diversity decreased the uniformity of the soybean variation, diseases still affected production throughout the State. Sentelhas et al. [37] showed that the use of crop rotation could improve yield and mitigate the effects of diseases. According to Franchini et al. [58], there are many medium and long-term benefits to crop rotation, such as improvements in soil quality and the disruption of the dynamics of diseases, pests, and weeds. The use of different crops in the same area improves soil biology and increases the efficiency of nutrient cycling and nitrogen fixation. However, crop rotation is not practiced uniformly in Mato Grosso, which is evidenced by the EVI profiles of the soybean crop fields in the agglomerates.

Regarding climatic conditions, the 2010/2011 crop season was only slightly affected by the effects of La Niña or El Niño. However, the natural variability causes water deficit and large-scale effects on soybean crop yield. According to Sinclair et al. [59], the principal effect relates to in-season weather stresses due to water deficit. Although irrigation during the critical crop phases, and in years with reduced rain, could minimize the impact of water deficit, the practice in not widespread in Mato Grosso.

The soil variability also complicates yield prediction, because the soils in Mato Grosso have low natural fertility and impeding layers, limiting root growth. When the soil exhibits good fertility and favorable physical properties, soybean roots develop well [60], reducing the water deficit and increasing the availability of nutrients. For southern Brazil, Franchini et al. [61] highlighted that crop rotation increased the soil microbial biomass, reduced soil compaction, and improved root growth, increasing water availability for the soybean crop during growth and reducing drought and the water deficit-related yield gap.

Based on the validation phase results, 206 crop fields were classified according to the established yield classes, with an SEI_{Pixel} accuracy of 56.13%. This value can be explained by the regional variations presented in section 4.3.1, and by the characteristics of the data. IBGE's Automatic Recovery System (SIDRA) database, which is available at (http://www.sidra.ibge.gov.br/), provides time-series data from 1990 to the present that is aggregated at the national, State, and municipal levels. This dataset does not distinguish between yields obtained in different cropping cycles, providing only an estimate of the annual gross yield for each crop type [62].

The IBGE yield survey is not completely consistent. The “yield” is defined in the IBGE dataset as the average ratio of production (kg) per harvested area (ha), rather than sowed area, for each crop. Since the data pertain to municipal administrative limits, IBGE's information does not include the occurrence of crops that are planted in one municipality but commercialized in another, as reported by Arvor et al. [63] and Dubreuil et al. [54]. These authors explain that many farms are distributed in more than one municipality, and that final yields are taken to the municipality where the farm has its headquarters, changing the municipal statistics. Many multinationals are located in these municipalities, and some agricultural negotiations are carried out by producers who grow crops in neighboring communities. This scenario can affect the extrapolation of the yield model to the total State count.

The yield prediction presents a 95.09% accuracy considering the standard deviation and the probable error, demonstrating the potential of the method based on census and remote sensing data. Considering the temporal variation of this study, the use of historical crop production data can optimize the yield predictions, or at least reduce the error.

Analyzing the spatial distribution of the soybean crops, Vale do Rio Verde and Santa Luzia agglomerates confirmed the high productive potential of the agribusiness region [54], presenting a yield of over 55 bags of 60 kg per ha^{-1}. The Malu agglomerate also presented a yield of over 55 bags of 60 kg per ha^{-1}, contributing significantly to the soybean production observed in eastern Mato Grosso. This region experienced the largest expansion of soybean area over the observed crop seasons. The Colorado and Colibri agglomerates were classified with 45 to 55 bags of 60 kg per ha^{-1}.

5. Conclusions

The methodology presented here links geographically aggregated yield data from different sources and different spatial supports, downscaling crop yields from one set of spatial units to another set of overlapping spatial units. Soybean crop detection and yield monitoring can be improved by this geostatistical approach. It can also be used to predict environmental variables related to soybean agroecosystems, such as disease incidence, pest infestation, chemical application, and other variables of interest. The advantages of this approach over those developed specifically for crop detection are the availability of *in situ* data and the ability to adjust predictions using the covariate values and the measures of uncertainty obtained for each prediction.

Binomial areal kriging was used to generate maps of soybean areas in Mato Grosso, presenting the probability of soybean occurrence over the years. Gaussian areal kriging was used to predict the crop yields of soybean areas detected by remote sensing, having a downscaling effect on the results. The global accuracy and the Kappa index values for soybean crop detection were 92.1% and 0.84, respectively. The yield prediction presented 95.09% accuracy considering the standard deviation and the probable error. The validation phase confirmed this accuracy rate and revealed the uncertainties of the SEI_{Pixel} classification. The higher accuracy values show that the MODIS eight-day composite time series data performs best for the discrimination of crops with a similar phenology, because the higher temporal resolution allows subtle spectral–temporal differences between crop types to be detected, improving identification.

Acknowledgements: The authors wish to thank (i) the farmers for providing the *in situ* data, (ii) the Coordination for the Improvement of Higher Education Personnel (CAPES) for support, (iii) the Federal University of Lavras (UFLA) for providing office space and infrastructure to achieve this article and (iiii) the Reviewers and the Editor for their constructive comments and helpful suggestions, which have greatly improved this article.

Author Contributions: Marcelo de C. Alves led in the design of the experiment. Michel E. D. Chaves performed data analysis. Michel E. D. Chaves, Marcelo de C. Alves, Marcelo S. de Oliveira and Thelma Sáfadi interpreted the results. Michel E. D. Chaves wrote the paper with significant contributions from all authors.

Conflict of Interest: The authors declare no conflict of interest.

References

1. Gotway, C.A.; Young, L.J. Combining incompatible spatial data. *J. Am. Stat. Assoc.* **2002**, *97*, 632-648. doi:10.1198/016214502760047140.

2. Atkinson, P.M. Downscaling in remote sensing. *Int. J. Appl. Earth Obs. Geoinform.* **2013**, *22*, 106-114. doi:10.1016/j.jag.2012.04.012.
3. Krivoruchko, K.; Gribov, A.; Krause, E. Multivariate areal interpolation for continuous and count data. *Proc. Environ. Sciences* **2011**, *3*, 14-19. doi:10.1016/j.proenv.2011.02.004.
4. Atkinson, P.M.; Tate, N.J. Spatial scale problems and geostatistical solutions: A review. *Prof. Geograph.* **2000**, *52*, 607-623.
5. Gotway, C.A.; Young, L.J. A geostatistical approach to linking geographically aggregated data from different sources. *J. Comput. Graph. Stat.* **2007**, *16*, 15-35.
6. Openshaw, S. *The Modifiable Areal Unit Problem*; Concepts and Techniques in Modern Geography, No. 38; GeoBooks: Norwich, UK, 1984.
7. Goovaerts, P. Combining areal and point data in geostatistical interpolation: Applications to soil science and medical geography. *Math. Geosci.* **2010**, *42*, 535-554. doi:10.1007/s11004-010-9286-5.
8. Arvor, D.; Jonathan, M.; Meirelles, M.S.P.; Dubreuil, V.; Durieux, L. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *Int. J. Remote Sens.* **2011**, *32*, 7847-7871. doi:10.1080/01431161.2010.531783.
9. Kastens, J.H.; Brown, J.C.; Coutinho, A.C.; Bishop, C.R.; Esquerdo, J.C.D.M. Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. *PLoS One* **2017**, *12*, 1-21. doi:10.1371/journal.pone.0176168.
10. Brown, J.C.; Kastens, J.H.; Coutinho, A.C.; Victoria, D. de C.; Bishop, C.R. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sens. Environ.* **2013**, *130*, 39-50. doi:10.1016/j.rse.2012.11.009.
11. Gusso, A.; Arvor, D.; Ducati, J.R.; Veronez, M.R.; Silveira Junior, L.G. Assessing the MODIS Crop Detection Algorithm for Soybean Crop Area Mapping and Expansion in the Mato Grosso State, Brazil. *The Scientific World Journal* **2014**, doi:10.1155/2014/863141.
12. Davidson, E.A.; Araújo, A.C.; Artaxo, P.; Balch, J.K.; Brown, I.F.; Bustamante, M.M.C.; Coe, M.T.; DeFries, R.S.; Keller, M.; Longo, M.; Munger, J.W.; Schroeder, W.; Soares-Filho, B.S.; Souza, C.M.; Wofsy, S.C. The Amazon basin in transition. *Nature* **2012**, *481*, 321-328. doi:10.1038/nature10717.
13. Raucci, G.S.; Moreira, C.S.; Alves, P.A.; Mello, F.F.C.; Frazão, L.A.; Cerri, C.E.P.; Cerri, C.C. Greenhouse gas assessment of Brazilian soybean production: A case study of Mato Grosso State. *J. Clean. Prod.* **2015**, *96*, 419-425. doi:10.1016/j.jclepro.2014.02.064.
14. Brazilian Institute of Geography and Statistics (IBGE), 2017. Geociências: Produtos, IBGE, Available online: http://downloads.ibge.gov.br/downloads_geociencias.htm (accessed on 18 January 2017).
15. Arvor, D.; Dubreuil, V.; Mendez Del Villar, P.; Ferreira, C.M.; Meirelles, M.S.P. Développement, crises et adaptation des territoires du soja au Mato Grosso: l'exemple de Sorriso. *Confins*, **2009**, *6*, Available online: http://confins.revues.org/ index5934.html (accessed on 27 February 2017).

16. Arvor, D.; Meirelles, M.S.P.; Dubreuil, V.; Bégué, A.; Shimabukuro, Y.E. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Appl. Geogr.* **2012**, *32*, 702-713. doi:10.1016/j.apgeog.2011.08.007.
17. Solano, R., Didan, K., Jacobson, A., Huete, A. *MODIS Vegetation Index (MOD 13) C5 User's Guide*; The University of Arizona, Tucson, AZ, USA, 2010; p. 38.
18. Risso, J.; Rizzi, R.; Rudorff, B.F.T.; Adami, M.; Shimabukuro, Y.E.; Formaggio, A.R.; Epiphanio, R.D.V. Modis vegetation indices applied to soybean area discrimination. *Pesq. Agropec. Bras.* **2012**, *47*, 1317-1326. doi:10.1590/S0100-204X2012000900017.
19. Galford, G.L.; Mustard, J.F.; Melillo, J.; Gendrin, A.; Cerri, C.C.; Cerri, C.E.P. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sens. Environ.* **2008**, *112*, 576-587. doi:10.1016/j.rse.2007.05.017.
20. Zhu, C.; Lu, D.; Victoria, D. de C.; Dutra, L.V. Mapping Fractional Cropland Distribution in Mato Grosso, Brazil using time series MODIS Enhanced Vegetation Index and Landsat Thematic Mapper data. *Remote Sens.* **2016**, *8*, 1-14. doi:10.3390/rs8010022.
21. Rudorff, B.F.T.; Adami, M.; Aguiar, D.A.; Moreira, M.A.; Mello, M.P.; Fabiani, L.; Amaral, D.F.; Pires, B.M. The soy moratorium in the Amazon biome monitored by remote sensing images. *Remote Sens.* **2011**, *3*, 185-202; doi:10.3390/rs3010185.
22. Huete, A.R.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195-213. doi:10.1016/S0034-4257(02)00096-2.
23. Zhong, L.; Hu, L.; Yu, L.; Gong, P.; Biging, G. S. Automated mapping of soybean and corn using phenology. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 151-164. doi:10.1016/j.isprsjprs.2016.05.014.
24. Yan, H.; Xiao, X.; Huang, H.; Liu, J.; Chen, J.; Bai, X. Multiple cropping intensity in China derived from agro-meteorological observations and MODIS data. *Chin. Geogra. Sci.* **2014**, *24*, 205-219. doi:10.1007/s11769-013-0637-2.
25. Huete, A.R.; Liu, H.Q.; Batchily, K.; van Leewen, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sens. Environ.* **1997**, *59*, 440-451.
26. Roy, D.P.; Borak, J.S.; Devadiga, S.; Wolfe, R.E.; Zheng, M.; Descloitres, J. The MODIS Land product quality assessment approach. *Remote Sens. Environ.* **2002**, *83*, 62-76. doi:10.1016/S0034-4257(02)00087-1.
27. National Aeronautics and Space Administration (NASA). *Land Processes Distributed Active Archive Center (LP DAAC), MOD13Q1*; USGS/Earth Resources Observation and Science (EROS) Center: Sioux Falls, SD, USA, 2001.
28. Brazilian Institute of Geography and Statistics (IBGE), 2011. Sistema IBGE de Recuperação Automática, Produção Agrícola Municipal. Instituto Brasileiro de Geografia e Estatística. Available online: https://sidra.ibge.gov.br/home/lspa/brasil (accessed on 18 January 2017).
29. Lajaunie, C. *Local Risk Estimation for a Rare Non Contagious Disease Based on Observed Frequencies*; Centre de Geostatistique de l'Ecole des Mines de Paris: Fontainebleau, France, 1991, Note N-36/91/G.

30. Waller L.A.; Gotway C.A. *Applied Spatial Statistics for Public Health Data*; John Wiley and Sons: Hoboken, NJ, USA, 2004; p. 494.
31. Wackernagel, H., 2003. *Multivariate Geostatistics: An Introduction with Applications,* 3rd ed.; Springer-Verlag, Berlin, Germany, 2003; p. 388.
32. Webster, R.; Oliver, M.; Muir, K.; Mann, J. Kriging the Local Risk of a Rare Disease from a Register of Diagnoses. *Geogr. Anal.* **1994**, *26*, 168-185. doi:10.1111/j.1538-4632.1994.tb00318.x.
33. Cressie, N.A.C. *Statistics for Spatial Data*, 1st ed.; Wiley Interscience, New York, NY, USA; 1993; p. 903.
34. Chilès, J.P.; Delfiner, P. *Geostatistics: Modeling spatial uncertainty,* 1st ed.; Wiley Interscience, New York, USA; 1999; p. 695.
35. Goovaerts, P. Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Math. Geosci.* **2008**, *40*,101-128.
36. Kerry, R.; Goovaerts, P., Rawlins, B.G., Marchant, B.P. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma* **2012**, *170*, 347-358. doi:10.1016/j.geoderma.2011.10.007.
37. Sentelhas, P.C.; Battisti, R.; Câmara, G.M.S.; Farias, J.R.B.; Hampf, A.C.; Nendel, C. The soybean yield gap in Brazil – magnitude, causes and possible solutions for sustainable production. *J. Agric. Sci.* **2015**, *65*, 1-18. doi:10.1017/S0021859615000313.
38. Ronchail, J.; Cochonneau, G.; Molinier, M.; Guyot, J. L.; Chaves, A. G. de M.; Guimaraes, W.; Oliveira, E. de. Rainfall variability in the Amazon Basin and SSTs in the tropical Pacific and Atlantic oceans. *Int. J. Climatol.* **2002**, *22*, 1663-1686. http://dx.doi.org/10.1002/joc.815.
39. Almeida, C.A.; Coutinho, A.C.; Esquerdo, J.C.D.M.; Adami, M.; Venturieri, A.; Diniz, C.G.; Dessay, N.; Durieux, L.; Gomes, A.R. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. *Acta Amazon.* **2016**, *46*, 291-302. doi:10.1590/1809-4392201505504.
40. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remote Sensing Data: Principles and Practices*, 2nd ed.; CRC Press: Boca Raton, FL, USA; 2009; p. 200.
41. Ghilani, C.D., Wolf, P.R. *Elementary Surveying: An Introduction to Geomatics*, 13th ed.; Prentice Hall: Boston, MA, USA, 2012; p. 958.
42. Gusso, A.; Formaggio, A.R.; Rizzi, R.; Adami, M.; Rudorff B.F.T. Soybean crop area estimation by MODIS/EVI data. *Pesq. Agropec. Bras.* **2012**, *47*, 425-435. doi:/10.1590/s0100-204x2012000300015.
43. Souza, C.H.W.; Mercante, E.; Johann, J.A.; Lamparelli, R.A.C.; Uribe-Opazo, M.A. Mapping and discrimination of soya bean and corn crops using spectro-temporal profiles of vegetation indices. *Int. J. Remote Sens.* **2015**, *36*, 1809-1824. doi:10.1080/01431161.2015.1026956.
44. Victoria, D. de C.; Paz, A.R. da; Coutinho, A.C.; Kastens, J.; Brown, J.C. Cropland area estimates using Modis NDVI time series in the state of Mato Grosso, Brazil. *Pesq. Agropec. Bras.* **2012**, *47*,1270-1278.

45. Gibbs, H.K.; Rausch, L.; Munger, J.; Schelly, I.; Morton, D.C.; Noojipady, P.; Soares-Filho, B.; Barreto, P.; Micol, L.; Walker, N.F. Brazil's Soy Moratorium. *Science* **2015**, *347*, 377-378. doi:10.1126/science.aaa0181.
46. Arvor, D.; Dubreuil, V.; Meirelles, M.S.P.; Bégué, A. Mapping and spatial analysis of the soybean agricultural frontier in Mato Grosso, Brazil, using remote sensing data. *GeoJournal* **2013**, *78*, 833-850. doi:10.1007/s10708-012-9469-3.
47. Dubreuil, V.; Nédélec, V.; Arvor, D.; Le Derout, M.; Laques, A. E.; Mendez, P. Colonisation agricole et déforestation en Amazonie Brésilienne: exemple du front pionnier du Mato Grosso. *Enquêtes Rurales* **2009**, *12*, 107-135.
48. Morton, D.; DeFries, R.S.; Shimabukuro, Y.E.; Anderson, L.O.; Del Bon Espírito-Santo, F.; Hansen, M.; Carroll, M. Rapid assessment of annual deforestation in the Brazilian Amazon using MODIS data. *Earth Interact.* **2005**, *9*, 1-22. doi:10.1175/EI139.1.
49. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*,185-201. doi:10.1016/S0034-4257(01)00295-4.
50. Landis, J.; Koch, G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159-174.
51. Lamparelli, R.A.C.; Carvalho, W.M.O. de; Mercante, E. Mapeamento de semeaduras de soja (*Glycinemax (L.) Merr.*) mediante dados MODIS/Terra e TM/Landsat 5: Um comparativo. *Eng. Agr.* **2008**, *28*, 334-344. doi:10.1590/S0100-69162008000200014.
52. Epiphanio, R.D.V.; Formaggio, A.R., Rudorff, B.F.T., Maeda, E.E., Luiz, A.J.B. Estimating soybean crop areas using spectral-temporal surfaces derived from MODIS images in Mato Grosso, Brazil. *Pesq. Agropec. Bras.*, **2010**, *45*, 72-80.
53. Lu, D.; Batistella, M.; Moran, E.; Hetrick, S.; Alves, D.; Brondizio, E. Fractional forest cover mapping in the Brazilian Amazon with a combination of MODIS and TM images. *Int. J. Remote Sens.* **2011**, *32*, 1-19. doi:10.1080/01431161.2010.519004.
54. Dubreuil, V.; Laques, A-E.; Nédélec, V.; Arvor, D.; Gurgel, H. Paysages et fronts pionniers amazoniens sous le regard des satellites: L'exemple du Mato Grosso. *Espace Geographique*, **2008**, *37*, 57-74.
55. Openshaw, S. Ecological fallacies and the analysis of areal census data. *Environment & Planning A* **1984**, *16*, 17-31. doi:10.1068/a160017.
56. Prasad, A.K.; Singh, R.P.; Tare, V.; Kafatos, M. Use of vegetation index and meteorological parameters for the prediction of crop yield in India. *Int. J. Remote Sens.* **2007**, *28*, 5207–5235. doi:10.1080/01431160601105843.
57. Battisti, R.; Sentelhas, P.C. Drought tolerance of Brazilian soybean cultivars simulated by a simple agrometeorological yield model. *Exp. Agric.* **2015**, *51*, 285-298. doi:10.1017/S0014479714000283.
58. Franchini, J.C.; Costa, J.M.; Debiasi, H.; Torres, E. *Importância da rotação de culturas para a produção agrícola sustentável no Paraná*; Documentos 327; Embrapa Soja: Londrina, Brazil, 2011; p. 52.

59. Sinclair, T.R.; Messina, C.D.; Beatty, A.; Samples, M. Assessment across the united states of the benefits of altered soybean drought traits. *Agron. J.* **2010**, *102*, 475-482. doi:10.2134/agronj2009.0195.
60. Torrion, J.A.; Setiyono, T.D.; Cassman, K.G.; Ferguson, R.B.; Irmak, S.; Specht, J.E. Soybean root development relative to vegetative and reproductive phenology. *Agron. J.* **2012**, *104*, 1702-1709. doi:10.2134/agronj2012.0199.
61. Franchini, J.C.; Debiasi, H.; Sacoman, A.; Nepomuceno, A.L.; Farias, J.R.B. *Manejo do solo para redução das perdas de produtividade pela seca*; Documentos 314; Embrapa Soja: Londrina, Brazil, 2009; p. 39.
62. Anderson, M.C.; Zolin, C.A.; Sentelhas, P.C.; Hain, C.R.; Semmens, K.; Yilmaz, M.T.; Gao, F.; Otkin, J.A.; Tetrault, R. The Evaporative Stress Index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. *Remote Sens. Environ.* **2016**, *174*, 82-99. doi:10.1016/j.rse.2015.11.034.
63. Arvor, D.; Dubreuil, V. Etude par Télédétection de la Dynamique du soja et de L'impact des Précipitations sur les Productions au Mato Grosso (Brésil). PhD Thesis, Université Rennes 2, Rennes, France, 2009.

ARTIGO 3 - AGRICULTURAL DYNAMIC DETECTED BY PATTERN RECOGNITION IN MODIS SATELLITE IMAGE TIME SERIES OF MATO GROSSO

Normas do periódico *Land Use Policy*, ISSN: 0264-8377.
Versão preliminar.

Michel Eustáquio Dantas Chaves[a*], Marcelo de Carvalho Alves[b], Thelma Sáfadi[c], Marcelo Silva de Oliveira[d]

[a] Engineering Department, Federal University of Lavras. Campus Universitário, PO Box 3037, ZIP code 37200-000, Lavras, Brazil. E-mail: medchaves@posgrad.ufla.br; Phone: +55 35 99968 2078

[b*] Engineering Department, Federal University of Lavras. Campus Universitário, PO Box 3037, ZIP code 37200-000, Lavras, Brazil. E-mail: marcelo.alves@deg.ufla.br; Phone: + 55 35 3829 1027

[c] Statistics Department, Federal University of Lavras. Campus Universitário, PO Box 3037, ZIP code 37200-000, Lavras, Brazil. E-mail: safadi@des.ufla.br; Phone: +55 35 3829 1370

[d] Statistics Department, Federal University of Lavras. Campus Universitário, PO Box 3037, ZIP code 37200-000, Lavras, Brazil. E-mail: marcelo.oliveira@des.ufla.br; Phone: +55 35 3829 1373

***** Correspondence: medchaves@posgrad.ufla.br; Phone: + 55 35 99968 2078

Abstract

Techniques of time series analysis of agricultural phenological dynamics have shown efficacy when applied to the identification of crop areas, providing more accurate information in a shorter processing time. Given the dynamics of agriculture in Mato Grosso, with alternating plantings, successions and rotations different to each crop, these techniques are useful because they detect subtle changes in land use and land cover. The objective of this study was to identify the dynamics of agricultural crops between 2008 and 2012 in the Santa Luzia agglomerate of farms, located in Sapezal municipality, State of Mato Grosso, Brazil, to identify soybean, cotton and maize cultivation crop fields, as well as the planted crops in *safrinha*, through the spectral-temporal pattern presented by each crop, and the data support obtained *in situ*. The pattern recognition in time series and the classifications were done using the Time-Weighted Dynamic Time Warping - TWDTW algorithm. The Cotton-fallow, Soybean-cotton, Soybean-maize and Soybean-millet sequences were identified by the TWDTW method with Overall accuracy and Kappa index of 89.5% and 80%, respectively, which indicates that the TWDTW algorithm was able to detect crop fields and their seasonal variations, caused by the interannual succession and rotation.

Keywords: Remote sensing; Vegetation indices; Phenology; Crop monitoring; Time-Weighted Dynamic Time Warping (TWDTW).

1. Introduction

With the technologies application advancement in agriculture, the decision-making of agricultural commodities globalized market began to demand greater dynamism and precision in the data collection. The absence of these characteristics can lead to speculations and uncertainties that affect the entire economic-productive and social chain, from the prices definition to the establishment of public policies for agriculture.

For a long time, the fixed agricultural calendar was considered to interpret results obtained by orbital data. However, given the substantial variability in planting and maturation of a broad range of planted crops (Brown et al. 2013, Coutinho et al. 2013), this information is no longer trusted to delimit the areas (Gusso; Arvor; Ducati 2017, Bolton; Friedl 2013) when it is possible to identify the patterns of the crops in time series. Another important concern for effectively change detection is the appropriate determination of analysis periods. A long period may conceal real changes, while a short period may not capture subtle changes (Li et al. 2018).

In this way, the scientific community has been developing applications for agricultural dynamics detection and gradual changes indication in phenological development, using time series to detect land use changes (Brown et al. 2013) and to monitor the agricultural advance (Zhu et al. 2016, Gusso et al. 2014, Rudorff et al. 2011). The result is a constant elaboration of crop monitoring tools, using mainly vegetation indices temporal analysis. Some algorithms have been developed in last decade to extract spectral-temporal trajectories from terrestrial surface and to eliminate noises that cause confusions in results interpretation, such as the Breaks For Additive Seasonal and Trend - BFAST (Verbesselt et al. 2010), Spectral-Temporal Analysis by Response Surface - STARS (Mello et al. 2012) and Time-Weighted Dynamic Time Warping - TWDTW (Maus et al. 2016a).

Providing terrestrial surface data is the MODerate resolution Imaging Spectroradiometer (MODIS) sensor, with almost daily temporal resolution at a 250 meters spatial resolution, very used in agricultural studies by having high imaging frequency, images with high geometric quality, a sophisticated procedure for atmospheric correction and a larger number of bands and specific algorithms to generates composite products (Soares; Batista; Shimabukuro 2007, Justice et al. 1998), which the vegetation indices Enhanced Vegetation Index - EVI (Huete et al. 1997) and Normalized Difference Vegetation Index - NDVI (Rouse et al. 1973), capable of assess the vegetative vigor during crop phenological cycles. In the cycle beginning, when the amount of phytomass is scarce and the spectral response is influenced by soil, the indices values are low. As the crop develops and phytomass production increases, the values rise to vegetative peak. With the senescence and harvest, they decrease until reach values found in beginning (Coutinho et al. 2013).

The possibility of chronological images ordering allowed the use of data provided by the MODIS sensor to assess crop fields (Zhu et al. 2016, Gusso et al. 2014, Brown et al. 2013). The time series analysis allows to verify trends and seasonality in the data. Specifically, in relation to agriculture, MODIS vegetation indices time series favor the patterns identification of different agricultural uses. In addition, the use of phenological metrics (Johann et al. 2016, Gibbs et al. 2015) and pattern recognition (Belgiu; Csillik 2018, Maus et al. 2016a) increase subtle differences detection and the degree of separability between different land use classes, improving the crop identification, and the crop forecast models that have this information as input data.

The alternations of practices in each harvest period and the succession and rotation of crops are factors that prevent the adequate contextualization of agriculture and the formulation of more accurate estimates, being necessary to approach them in the classification of areas (Gusso; Arvor; Ducati 2017). In addition to the annual changes in phenological cycles caused by climate or variations in agricultural practices, the mapping agriculture based in time series analysis is also hampered by the lack of samples from different crops and land uses to train the supervised algorithm, and the obscuration by clouds (Petitjean; Inglada; Gançarski 2012). However, Dynamic Time Warping - DTW (Sakoe; Chiba 1978) proved to be an efficient solution to minimize these problems (Baumann et al. 2017) by comparing an unknown time series with a known time signature and dealing with temporal distortions (Petitjean; Inglada; Gançarski 2012) that occur from one phenological cycle to other.

Maus et al. (2016a) improved the DTW for the agriculture dynamism by proposing a time-weighted version, the Time-Weighted Dynamic Time Warping - TWDTW, capable of improve the crop classification with different vegetation dynamics through application of dissimilarity measures. The TWDTW had satisfactory performance in the identification of single crop, double crop, forest and pasture in recent studies (Manabe; Melo; Rocha 2018, Belgiu; Csillik 2018, Maus et al. 2016a), characterizing the sensitivity to the vegetation seasonal changes, considering climatic and seasonal variability and obtaining high accuracy for the heterogeneous fields mapping, improving the classification of areas with high agricultural dynamics.

The objective of this work was to use the patterns recognition of different crops in vegetation indices time series to identify and classify the land use and land cover in the crop fields, as well as the successions and rotations, in the Santa Luzia agglomerate of farms, located at Sapezal, Mato Grosso, Brazil, in 2008/2009, 2009/2010, 2010/2011 and 2011/2012 harvest periods, through the TWDTW algorithm, with support of crop fields *in situ* data. This work is based on the hypothesis that, through the time series analysis of MODIS sensor composite products and the patterns recognition in temporal trajectories of each crop, it is possible to identify different crops, both in first and in second crop season of the harvest period (*safrinha*), in crop field level.

2. Material and methods

The methodological procedure was performed in four steps: (1) data input; (2) time series processing for recognition of phenological patterns; (3) time series and images classification; (4) reliability assess of the generated classifications (Figure 1).

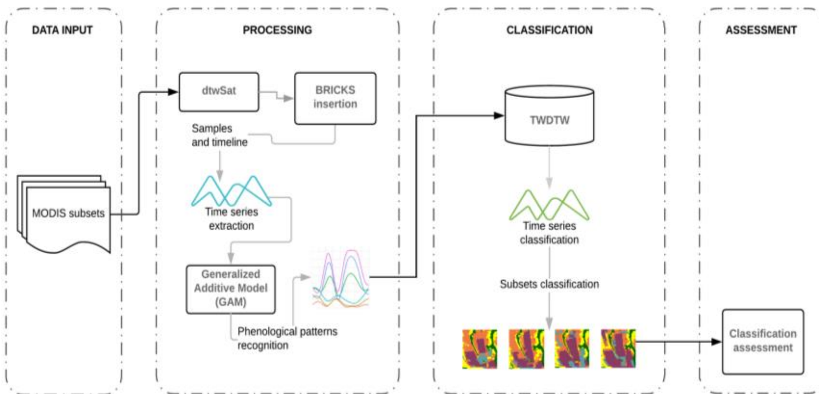


Figure 1. Representative scheme of the methodological procedures adopted to interpret the agricultural dynamics in Santa Luzia agglomerate, in Sapezal-MT, through the TWDTW algorithm.

2.1 Study area characterization

The experimental field corresponded to soybean (*Glycine max (L.) Merril*), maize (*Zea mays L.*), cotton (*Gossypium hirsutum L.*) and millet (*Pennisetum glaucum (L.) R. Br.*) crop fields from Santa Luzia agglomerate (13,682.59 ha), located in the municipality of Sapezal, northern mesoregion of Mato Grosso State (Figure 2).

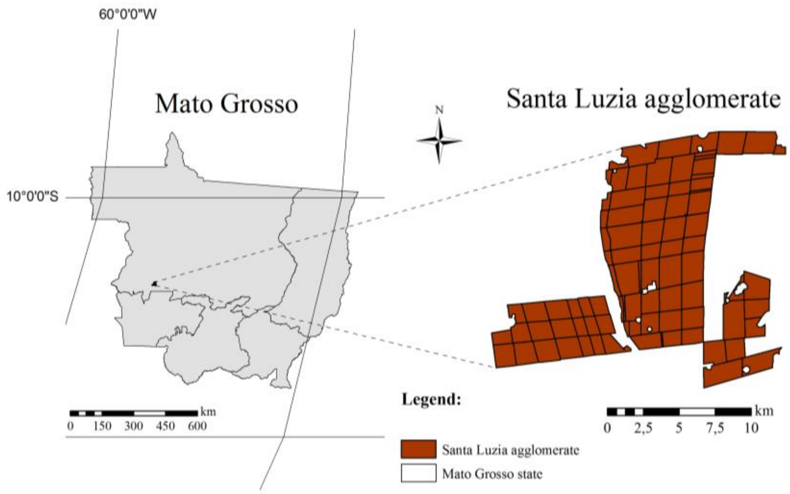


Figure 2. Geographical location of the municipality of Sapezal, Mato Grosso State, and geographical delimitation of the Santa Luzia agglomerate of farms.

Located at the Chapada dos Parecis region, Sapezal was one of the first agricultural poles in Mato Grosso (Arvor et al. 2014), and underwent a continuous process of land cover conversion, due to the advance of mechanized agriculture and livestock. By 2000, the deforested total was 83.2 km². In 2011, it was 184.9 km² (INPE 2017). The municipality is part of a region known for practicing a large scale Agribusiness (Dubreuil et al. 2008), characterized by obtaining two productions per year. According to Arvor et al. (2011) under these conditions, succession of crops occurs, with soybeans being the most planted crop in the first crop season and cotton and maize in *safrinha*, which occurs soon after soybean harvest. This pattern was also observed in field campaigns. This succession aims to avoid erosion and improve soil quality, to break down pest cycles, to maintain soil moisture and to establish conditions for high-quality no-till (Arvor et al. 2012).

According to Arvor et al. (2013), agriculture in Sapezal has evolved from the intensification phase, where small farms are replaced by large farms and commodities are sold nationally or internationally for the intensive phase, where more than 50% of cropped areas were cultivated with double cropping systems, and urbanization and environmental concern grow. With MODIS time series, Galford et al. (2008) identified the incorporation of new areas for agriculture and the double cropping systems consolidation in Sapezal region between 2001 and 2006.

Regionally, soybean sowing is determined by the onset of rainy season (Gusso et al. 2014), starting from mid-September and ending on the end of October, which may be related to the late onset of rainfall in Chapada dos Parecis region in relation to the rest of State, as reported by Arvor et al. (2014), Arvor; Dubreuil; Meirelles (2008). According to Arvor et al. (2017), in Mato Grosso agricultural poles, such as Sapezal, Sorriso and Lucas do Rio Verde, about 90% of the area planted with soybean is sown in no-tillage.

2.2 Data

2.2.1 *In situ* data

The study area was chosen due to the existence of a useful *in situ* data set for the validation phase. The *in situ* data were kindly provided by the farmers of the visited agglomerate. This data referenced the cultivation practices applied in the crop field level from 2008/2009 to 2011/2012 crop seasons, and included: variety, sowing, germination, and harvesting dates, soil texture, total planted area, and the number of 60-kg bags of harvested crop per hectare (ha) of each crop field.

2.2.2 Orbital data

The orbital data used were the MODIS composite vegetation indices EVI and NDVI and the red, blue, near infrared (nir) and medium infrared (mir), derived from MOD13Q1 and MYD13Q1 products (Collection 6), with 16-day temporal frequency, spatial resolution of 250 meters, Sinusoidal projection and TIFF format. These compositions corresponded to the maximum-value composition every 16 days. Although both data were available every 16 days, one starts in the middle of the compositing period of the other. This situation effectively creates an 8-day product, which improves temporal change detection (Solano et al. 2010).

The orbital data were obtained from the Oak Ridge National Laboratory Distributed Active Archive Center, through the MODIS Collection 6 Land Products Global Subsetting and Visualization Tool (https://modis.ornl.gov/) (ORNL DAAC 2018), which allows subsets of a MODIS tile for a particular area of interest, maintaining the spatial resolution of 250 meters, the 8-day temporal frequency and the atmospheric corrections (ORNL DAAC 2017, Santhana-Vannan et al. 2011). Subsets obtained from this portal were used to assess agriculture in Mato Grosso by Biudes et al. (2015) and Zeilhofer et al. (2015), demonstrating utility to reduce the time spent with downloading and big data processing.

Was downloaded a subset of 30.25 km wide x 30.25 km high around the point determined by latitude: -13.72264 and longitude: -58.85835, corresponding to the Santa Luzia agglomerate center, contemplating all farms (Figure 3). The images were obtained from August-29-2008 to March-15-2013.

Figure 3. Representative scheme of the subset delimited in the ORNL portal to obtain the MODIS images of the area that covers the Santa Luzia agglomerate, in Sapezal-MT.

2.3 Time-Weighted Dynamic Time Warping method

In order to generate land use and land cover classifications from 2008/2009 to 2011/2012 crop seasons, was applied the TWDTW algorithm (Maus et al. 2016a), a time series satellite imagery implemented in an open source R (R Core Team 2015) package, dtwSat (Maus 2016b).

This method, an adaptation of the DTW method, is more sensitive to seasonal variations of vegetation by adding time constraints, logistic weights (α) and midpoint (β), to find the ideal time alignment between the time series pattern obtained from each crop samples pixels and the series of pixels to be classified, aiming to assign land use classes in each moment of the pixel trajectory during the entire period of analysis. For this, for each time pattern, TWDTW finds all corresponding subintervals in complete time series, providing a measure of dissimilarity, a TWDTW distance between them, in which values closer to zero indicate greater similarity in temporal trajectory of the time series (Maus et al. 2016a).

The use of TWDTW classification method for time series analysis and identification was more effective when compared to other methods, such as Random Forest (Belgiu; Csilik 2018). This method considers the typical temporal variation of the agriculture dynamism as a more important factor than the spatial variation (Manabe et al. 2018), since the crops phenological cycles can vary from year to year, depending of the climatic conditions and the land management, thus causing the displacement of the stages (Maus et al. 2016a); changing the time series. Consider this variation is suitable with the Santa Luzia agglomerate scenario, in which the temporal crops dynamics is intense, with two annual productions, but the crop fields area and size dimensions does not change.

Thus the method doesn't require that two time series have the same length, which allows finding all possible matches of a pattern within a longer time series. This avoids a possible inconsistent matching of phenological cycles caused by time series division (Maus et al. 2016a).

The application of the method was divided into three main steps: (1) obtaining patterns in the time series of the field samples; (2) application of the class separability analysis for the classification of time series and (3) classification of the MODIS subsets from which time series were extracted, generating land use and land cover maps on the end of each harvest period.

2.3.1 Detection of temporal patterns in samples

For the detection of time series patterns of each class, which were compared with the profiles obtained from the images, a set of 603 field samples, obtained by Maus et al. (2016a), and used in the dtwSat package default, was inserted, being 68 for Cotton-fallow, 138 for Forest, 79 for Soybean-cotton, 134 for Soybean-maize and 184 for Soybean-millet. These samples were chosen because are from Porto dos Gaúchos municipality, also located in Mato Grosso, inside of the Amazon Biome, distant 437 km from Sapezal and with similar agricultural land use and land cover pattern and dynamic.

As a support to this step, we adopted as input: (1) the images in chronological order, forming temporal series of images, that is, the raster files related to each variable time series stacking in chronological order (2008-2012), (2) a file containing the geographical coordinates of each point in the sample, and (3) a file containing the start and end dates of the cycle, in Julian days. The mean time profile of each sample in each variable was extracted and analyzed, generating a mean time profile representative of each class pattern in each variable. For each class, the time series of the samples were grouped and smoothed by the Generalized Additive Model - GAM (Wood 2011, Hastie; Tibshirani 1986), a filter adopted to fit better the orbital data than purely parametric models (Maus et al. 2016a), observing the temporal frequency of 8 days.

2.3.2 Time-series classification

The patterns obtained from the samples of each agricultural crop and crop sequence were used as profile example to classify the trajectories over time and to identify the land use and land cover also in *safrinha*, thus classifying the time series of the assessed period in full.

The identification and classification of the time series was accomplished through the generation of TWDTW distances from the comparison between the temporal patterns of each crop and the time series for each pixel. Each pixel was classified using the identified temporal pattern with the shortest TWDTW distance, k-nearest neighbor, where k = 1.

2.3.3 Land Use Land Cover classification

The TWDTW was tested by Maus et al. (2016a) in linear and logistic forms, and the logistic form obtained better performance for the generation of land use and land cover maps in Mato Grosso. The logistic TWDTW uses a low penalty for small time and high for greater time distortions, which increases sensitivity to land use and land cover classifications (Maus et al. 2016a). In this way, the logistic TWDTW was used (Equation 1):

$$\omega_{i,j} = \frac{1}{1+e^{-\alpha(g(t_i,t_j)-\beta)}} \tag{1}$$

where $g(t_i, t_j)$ is the elapsed time in days between the dates t_i in the pattern and t_j in the time series, β and α represent slope and midpoint.

As recommended by Maus et al. (2016a), we considered the logistic weights (α) and midpoint (β) of the series in analysis, important factors when analyzing time series of different years and when the phenological cycles of the crops are different from one season to the next. These authors adopted α = - 0.1 and β = 50. In this work, as the dates of planting and harvesting in each crop field were known, α = - 0.05 and β = 35 were adopted, aiming to capture subtle differences in phenological dynamics. These values mean that a weight-time was added to the DTW, with a low penalty for times less than 35 days. Other values of α (- 0.1, - 0.02 or 0.2) and β (50, 15 or 75) were tested, but the classification results were lower.

By these parameters, the time series of each pixel of the study area were analyzed in comparison with the time patterns identified in the training samples. In this way, it was possible to generate maps of land use and land cover in the end of each harvest period.

2.4 Validation phase

In order to validate the results obtained and verify interannual variations, were used information collected *in situ* in the crop field level, discussed in section 2.2.1.

The accuracy assessment involved the temporal analysis of each crop field, considering the land use and land cover registered in the crop fields *in situ* data. For this, the shapefile of each harvest periods, with the spatial arrangement and crop fields location, were superimposed on the final maps. The classification accuracy measures of each harvest period were obtained based on confusion matrices, the most widely accepted approach to perform an accuracy assessment is through the use of a confusion matrix (Congalton 1991). In the confusion matrices, the overall accuracy, Kappa index, omission and commission errors, and producer and user accuracy were assessed.

3. Results and discussion

3.1 Agricultural crops temporal patterns

Firstly, spectral-temporal patterns of the three main commercial agricultural crops planted in the Santa Luzia agglomerate (soybean, cotton and maize) were obtained from the samples distributed by Maus et al. (2016), detailed in the 2.3.1 section, in NDVI and EVI indices and in the red, blue, nir and mir bands (Figure 4).

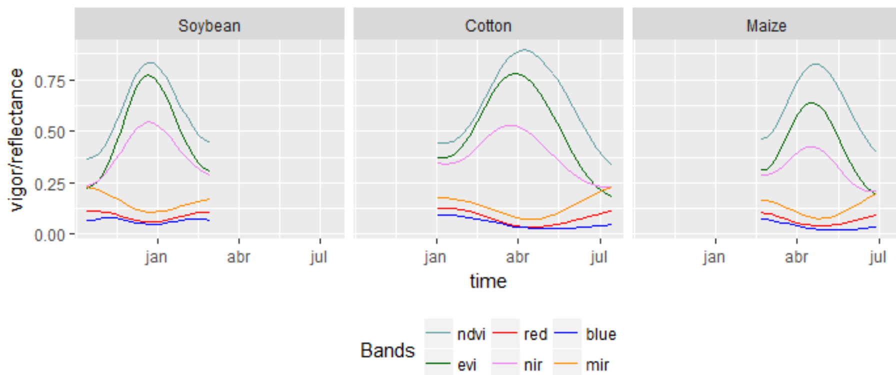


Figure 4. Mean patterns of specific time series of 'Soybean', 'Cotton' and 'Maize' classes obtained by Maus et al. (2016) used to compare with Santa Luzia agglomerate crop profiles over the phenological cycles between 2008 and 2012.

These were the patterns of the ‘Soybean', 'Cotton' and 'Maize' classes trajectories. Evaluating these patterns, it was possible to identify important differences. As reported by Souza et al. (2015) and Wardlow; Egbert (2008), these summer crops have differences caused by their distinct phenological characteristics, variations in cultural tracts, and their different responses to edaphoclimatic conditions.

Regarding the planting season, soybean sowing occurred in the beginning of rainy season. Cotton and maize succeeded soybeans and were planted as *safrinha*, from the beginning of January for cotton, and from the second half of February for maize. These observations are in line with that observed on the planting dates reported in the *in situ* data.

Regarding the period of maximum vegetative vigor, it was found that soybean was different from maize and cotton, being shorter (less than 30 days). Risso et al. (2012) consider that during the high biomass period a developing soybean crop without serious phytosanitary problems exceeds 0.8 in EVI. Values close to this were found on this stage, especially in December. In the same way, it was possible to detect the abrupt fall of values from January, natural, due to phytomass decrease along the cycle. Adami (2010) observed that, after the vegetative vigor peak, associated with flowering and grain filling, the plant translocates nutrients to the soybean grain, causing the derivative of the index curve to become negative, initiating the decrease of values in MODIS vegetation indices.

There was a difference in the cotton trajectory in relation to maize and soybean, presenting different phenological characterization in the indices and bands assessments. As detected by Couto Junior (2013), there was greater amplitude and variation, with dry periods marked by values lower than 0.2 and rainfall, close to 0.8, with higher seasonal variation.

The maize trajectory resembled those of soybeans, with an abrupt peak, preceded by a period with great vegetative vigor after the onset of rains. This peak was followed by a period of decrease in values, in months of low precipitation, which evidenced the close relationship between rainfall and the crop phenological development. The lowest values occurrence in August and September is explained as being the immediate period to winter, the interim between two cycles, where the vegetation is drier and the soil is discovered. The lowest EVI values recorded were 0.18 and 0.16 in the first observations of September 2010 and 2011, respectively, and 0.20 in the second observation of September 2012.

In common, it was verified the relation between phenological development and precipitation accumulation. In dry period and characterized by low vegetation cover, with high soil exposure, lower values were obtained. However, during the rainy season, the higher water availability in that time and a better use of this resource by vegetation showed the predominance of higher values.

The individual patterns obtained for each agricultural crop were used to classify the trajectories over time. The temporal profiles analysis involving an entire harvest period allowed the detection of trajectory sequences, identifying the crops planted in *safrinha*, as observed by the follow classes: Cotton-fallow, Soybean-cotton, Soybean-maize and Soybean-millet (Figure 5), the patterns obtained by Maus et al. (2016) used as parameter.

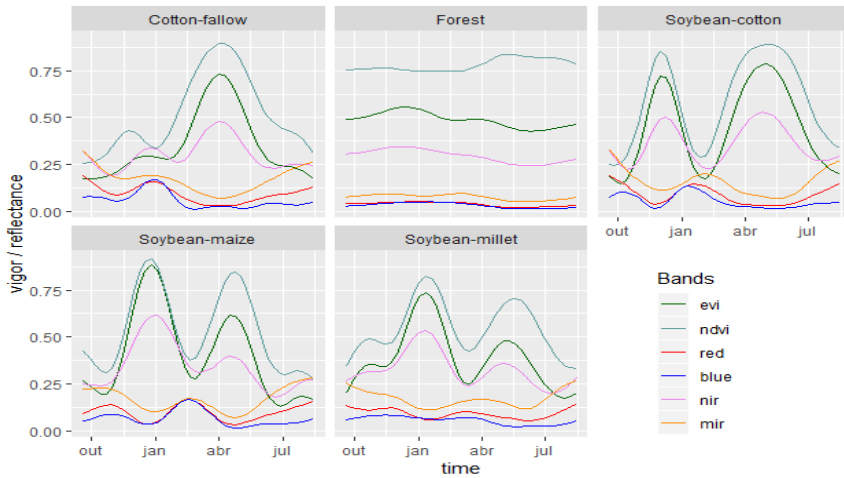


Figure 5. Trajectories identified in the time series analysis for 'Cotton-fallow', 'Forest', 'Soybean-cotton', 'Soybean-maize' and 'Soybean-millet' classes obtained by Maus et al. (2016) used to comparison with Santa Luzia agglomerate crop profiles over the phenological cycles between 2008 and 2012.

3.2 Time-series classification

Although it was possible to detect the successional stage, the similar spectral responses of the crops caused confusion to the classification algorithm. This problem was also reported by Souza et al. (2015) and Grzegozewski et al. (2016), also using MODIS vegetation indices to discriminate soybean and maize. In view of the similarities, it was necessary to observe dissimilarity measures originated from comparisons between the series so that subtle differences could be emphasized that could favor the crops separability, aiming to refine the classification (Figure 6).

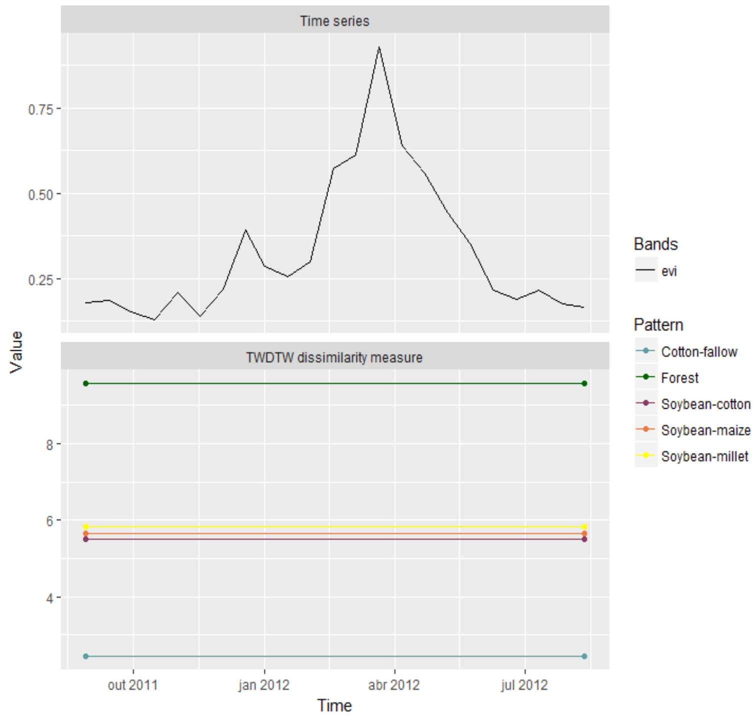


Figure 6: Example of how the result of the application of the dissimilarity measure assists in the identification of the class to which the assessed pixel belongs at a certain moment in the time series.

On the top, the graph corresponds to a sample period of the EVI variable corresponding to one year of the analysis period. On the bottom, it refers to the TWDTW distance of each class (colors) for the segment of the time series shown at the top. The smaller this distance, the more similar the segment of the class pattern. TWDTW is a "closest prototype" classifier. In the example case, the period would be classified as "Cotton-Fallow".

Amplifying this analysis to verify the changes occurred from one crop to other, it was possible to interpret each trajectory and to identify the land uses according to the similarities detected. The adoption of dissimilarity measures favored the land use identification during the period. The temporal profiles trajectory showed variations that, according to the detected patterns, denote continuous changes in phenology. Considering the intense dynamics of agriculture in Mato Grosso (Chaves et al. 2018, Arvor et al. 2014, Brown et al. 2013), the occurrence of these changes is natural, due the succession and rotation practices.

Concerning these terms of crop sequence, the crop succession indicates just the succession of the crop types in a defined time span. In contrast, the term crop rotation refers to certain multiannual crop pattern, which is repeated multiple times, is connected to a cropping system and is usually intended to have positive effects on the agro-environment or the crop yield (Waldhoff; Lussem; Bareth 2017, Leteinturier et al. 2006).

In this context, the TWDTW algorithm was able to capture the crop sequence and, considering each crop time series pattern detection, NDVI and EVI vegetation indices and the red, blue, nir and mir bands, it was able to classify the time series of each pixel.

3.3 Land Use Land Cover classification

Throughout the logistic TWDTW with α = - 0.05 and β = 35 logistic weights, it was possible to assess the land use and land cover and observe the dynamics of phenological changes within Santa Luzia agglomerate, which made it possible to identify changes in the crop fields use and coverage (Figure 7). Other values of α (- 0.1, - 0.02 or 0.2) and β (50, 15 or 75) were tested, but the classification results were lower.

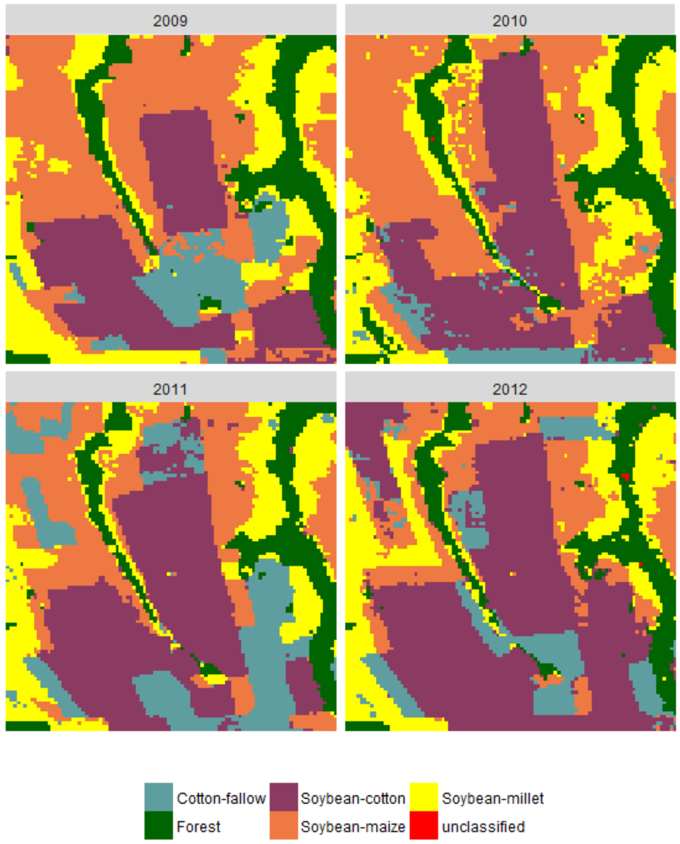


Figure 7. Land use and land cover of land in Santa Luzia agglomerate between harvest periods from 2008/2009 to 2011/2012. Each Classification map represents the agricultural use detected in the end of the crop season.

The interannual variations indicate the dynamism of agriculture (Figure 8). Since this is an agricultural agglomerate in continuous production process, this variation is natural, since the dynamism of agricultural processes, the fast rhythm of production and the care for the land demand actions such as succession and rotation of crops, which, in and of themselves, represent processes that alter the use and interannual land use and land cover, leaving them in transition for each harvest period. It was also possible to identify forest fragments.

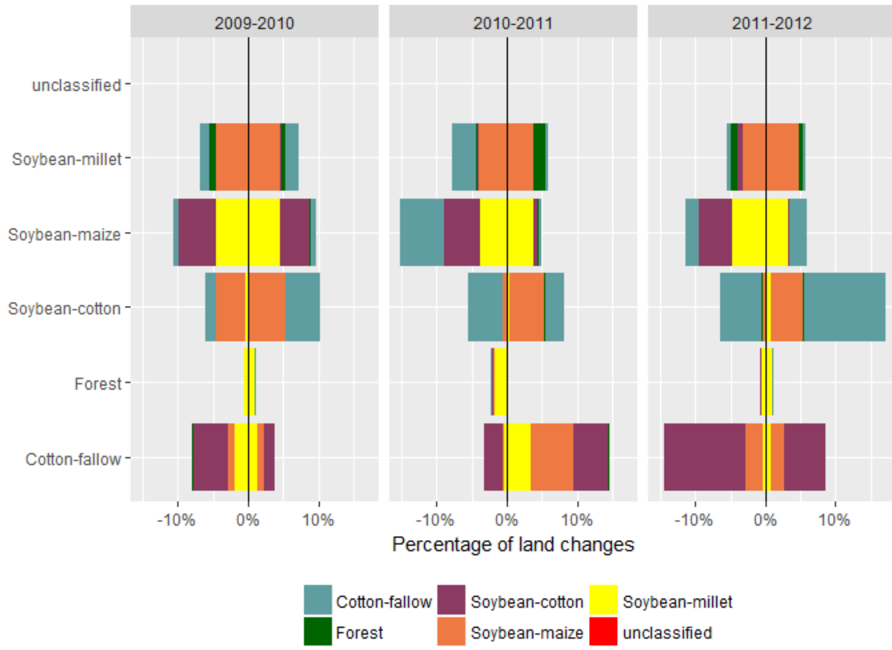


Figure 8. Degree of change detected. Axis y: current class. Positive direction on the x-axis: area gains of each class. Negative direction: area losses of each class. The colors represent which class the gain and loss of area belonged to.

In general, class area gains and losses balanced. For example, from 2008-2009 to 2009-2010 harvest periods, there was a greater variation in the Soybean-maize class, which lost more than 10% of its area to Soybean-millet, Soybean-cotton and Cotton-fallow, but, occupied other areas of the same classes, almost in the same proportion. Already, in loss of area, the Cotton-fallow class lost almost 10% of area for Soybean-cotton, Soybean-millet and Soybean-maize. In gain, the Soybean-cotton class began to occupy areas of Soybean-maize and Soybean-cotton.

In 2010-2011, the largest variations in the Cotton-fallow class, which now occupy areas of Soybean-millet, Soybean-maize and Soybean-cotton, representing an area gain of almost 15%, and in the Soybean-maize class, which lost a little more of this same percentage for Cotton-fallow, Soybean-cotton and Soybean-millet. In 2011-2012, the Soybean-maize class lost 12% of the Soybean-millet class (5%), Soybean-cotton (5%) and Cotton-fallow (2%) occupy 5% of the area previously occupied by Soybean-maize and 13% of the area that belonged to Cotton-fallow.

We noted that the variation of forest area remained low during the period. For several authors, the expansion of double-cropping system also dissociated deforestation from production at the agricultural poles (Arvor et al. 2013, Macedo et al. 2012), since it is more profitable and environmentally appropriate for producers to invest in new technologies to be applied in areas already deforested than to finance the opening of new fronts. In double-cropping systems, yields of individual crops are often lower than in single cropping systems, but the overall productivity and revenue tends to be higher due to the increased cropping intensity (Hampf et al. 2018).

These changes demonstrated the variations during the period, mainly expound the sow of soybeans in the first crop season and the growth of cotton in *safrinha* (Figure 9).

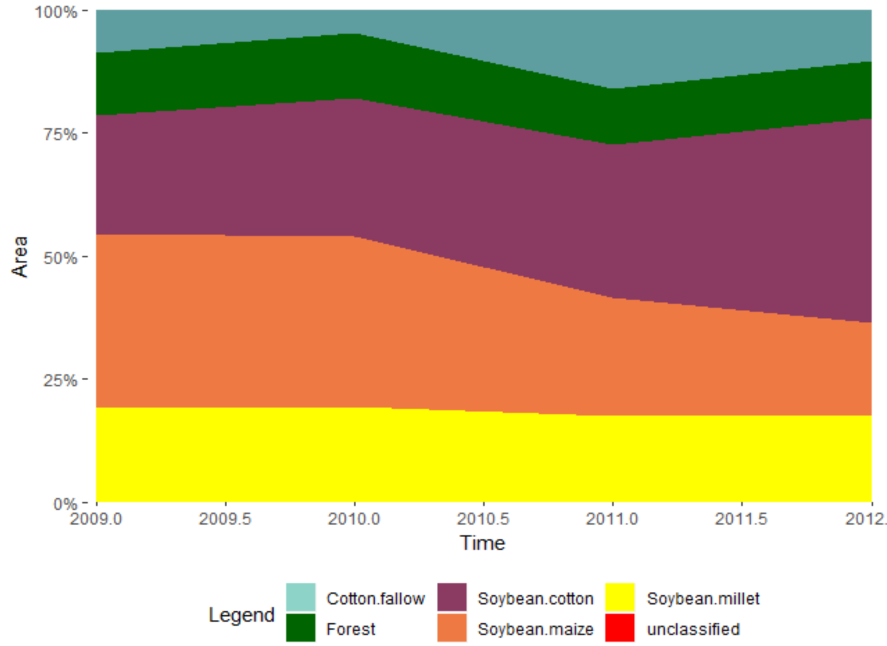


Figure 9. Percentage area variation of each land use and land cover class from 2008/2009 to 2011/2012.

The variations presented indicates that the crop succession occurred on a large scale than the crop rotation. In general, the farmers opting for a specific crop based on the forecast of better opportunities to market the final product. Considering the economic factor as a main driver of the agricultural practices, the continuous cultivation of soybeans in the first crop season and maize and cotton crops in *safrinha* predominate. About the cotton growth, Hampf et al. (2018) detect that large-scale agent farms (> 500 ha) with unconstrained access to machinery, inclusive in Sapezal, shifted their production from soybean-maize to soybean-cotton rotation due the major increase of the cotton yield and profit between 1999-2015.

Alternatively, cotton can be grown in rotation with a cover crop such as millet, where latter mainly serves to prevent soil erosion and enhance nitrogen cycling, but the most common practice is the soybean sow at the onset of the rainy season, while maize and cotton are sown between December and March (IMEA 2016). This crop system is traditional and rentable in Mato Grosso.

However, these adopted continuous system can have negative consequences, such as the intensification of pests, diseases, weeds and organic matter reduction in long term. The crop rotation can mitigate these problems, ensuring fertile soil, disease control and increased productivity, invigorating the soil, improving farming conditions where soils were too sandy (and, thus, not perfectly suitable for soybean cultivation) because the deep root system of some grasses helped structuring the soil (Gil et al. 2015).

Further the crop rotation, the variations of rotation, which the crop rotation with pasture, the integrated crop–livestock and integrated crop–livestock–forestry systems, bring benefits to the soil and the farmers, such as the increase in carbon stocks in the soils (Asai et al. 2018, Moraine et al. 2016, Silva et al., 2014), the improvement of nutrient cycling by re-coupling nitrogen and carbon cycles (Ryschawy et al. 2017, Martin et al. 2016, Piva et al. 2014) and the increase of economic returns (Hampf et al. 2018, Gil et al. 2015, Franzluebbers; Sawchik; Taboada 2014).

It is noteworthy that the evaluation occurred between 2008 and 2012, and these practices were not as widespread as today. The adoption of integrated production systems is gradually growing, but the rotation practice still as a challenge in Mato Grosso.

3.4 Classifications assessment

The accuracy analysis involved the land use and land cover analysis of each crop field in each harvest period, allowing to identify differences between the classifications and the *in situ* data. The total number of crop fields in the Santa Luzia agglomerate was 72 in the 2008/2009 harvest period and 81 in the following three harvest periods, maintaining its spatial location throughout the period.

Were analyzed the land use and land cover registered *in situ* and obtained in classification. After, were generated confusion matrices to assess the classifications. Because the rarely occurrence, some land uses that Soybean-millet and Cotton-fallow are undersampled each harvest period. However, these land uses were not retired of the confusion matrices in function to explore the capability of the TWDTW method to identify all crop fields land uses present in each harvest period, individually. The same case was presented by Congalton; Green (2009). Belgiu; Csillik (2018) presents that the TWDTW proved to be less sensitive in relation to the training samples. This is an important asset in areas where inputs for training samples are limited.

Based on the example presented by these authors, we believe that, also in this case, it is more important to ensure correct crop identification, the principal purpose of the TWDTW method, than it is to ensure that enough samples are collected in rarely occurring crop types.

In relation to the 2008/2009 harvest period, there was variation only in Soybean-cotton classes, with one crop field less and Soybean-maize, with one more crop field (Table 1).

Table 1. Comparison between the land use and land cover identified in the crop fields and the land use and land cover registered in the *in situ* dataset for the 2008/2009 harvest period.

Crop Fields	Cotton-fallow	Soybean-cotton	Soybean-maize	Soybean-millet	Crop fields total
In situ	4	33	32	3	72
Classification	**4**	**32**	**33**	**3**	**72**
Total	**=**	**-1**	**+1**	**=**	

By the confusion matrix referring to the 2008/2009 harvest period classification (Table 2), containing the Global accuracy and Kappa index values, omission and commission errors and the user’s and producer’s accuracy, it was possible to observe the quality of classification.

Table 2. Confusion matrix of the 2008/2009 harvest period classification.

Classification	*In situ*						
	Cotton-fallow	Soybean-cotton	Soybean-maize	Soybean-millet	Total	Comission errors	User's accuracy
Cotton-fallow	**2**	0	1	1	4	0,50	0,50
Soybean-cotton	2	**29**	1	0	32	0,09	0,91
Soybean-maize	0	4	**29**	0	33	0,12	0,88
Soybean-millet	0	**0**	1	**2**	3	0,34	0,66
Total	4	33	32	3	72		
Omission errors	0,50	0,12	0,09	0,34			
Producer's accuracy	0,50	0,88	0,91	0,66			
Global accuracy	***0,86***						
Kappa index	***0,76***						

In the following harvest period, 2009/2010, more errors of classification were identified in relation to the *in situ* data (Table 3), with a higher variation for Soybean-cotton (less 5 crop fields) and Soybean-maize (5 more).

Table 3. Comparison between the land use and land cover identified in the crop fields and the land use and land cover registered in the *in situ* dataset for the 2009/2010 harvest period.

Crop fields	Cotton-fallow	Soybean-cotton	Soybean-maize	Soybean-millet	Crop fields total
In situ	0	46	28	7	81
Classification	**0**	**41**	**33**	**7**	**81**
Total	=	**-5**	**+5**	=	

By the confusion matrix referring to the 2009/2010 harvest classification (Table 4), containing the Global accuracy and Kappa index values, omission and commission errors and the user's and producer's accuracy, it was possible to observe the quality of classification.

Table 4. Confusion matrix of the 2009/2010 harvest period classification.

Classification	*In situ*						
	Cotton-fallow	Soybean-cotton	Soybean-maize	Soybean-millet	Total	Comission errors	User's accuracy
Cotton-fallow	**0**	0	0	0	0	0	1
Soybean-cotton	0	**41**	0	0	41	0	1
Soybean-maize	0	5	**26**	2	33	0,21	0,79
Soybean-millet	0	0	2	**5**	7	0,29	0,71
Total	0	46	28	7	81		
Omission errors	0	0,11	0,07	0,29			
Producer's accuracy	1	0,89	0,93	0,71			
Global accuracy	***0,89***						
Kappa index	***0,80***						

In the 2010/2011 harvest period, the classification present little difference in relation to the *in situ* data (Table 5), with the highest variation in the Soybean-maize class: 3 more crop fields.

Table 5. Comparison between the land use and land cover identified in the crop fields and the land use and land cover registered in the *in situ* dataset for the 2010/2011 harvest period.

Crop fields	Cotton-fallow	Soybean-cotton	Soybean-maize	Soybean-millet	Crop fields total
In situ	1	57	8	15	81
Classification	**0**	**57**	**11**	**13**	**81**
Total	**-1**	**=**	**+3**	**-2**	

By the confusion matrix referring to the 2010/2011 harvest period classification (Table 6), containing the Global accuracy and Kappa index values, omission and commission errors and the user's and producer's accuracy, it was possible to observe the quality of classification.

Table 6. Confusion matrix of the 2010/2011 harvest period classification.

Classification	*In situ*						
	Cotton-fallow	Soybean-cotton	Soybean-maize	Soybean-millet	Total	Comission errors	User's accuracy
Cotton-fallow	**0**	0	0	0	0	0	1
Soybean-cotton	1	**54**	0	2	57	0,05	0,95
Soybean-maize	0	3	**8**	0	11	0,27	0,73
Soybean-millet	0	**0**	0	**13**	13	0,07	0,93
Total	1	57	8	15	81		
Omission errors	0	0,05	0	0,13			
Producer's accuracy	1	0,95	1	0,87			
Global accuracy	***0,93***						
Kappa index	***0,84***						

Following the pattern, there was little variation in the 2011/2012 crop. The highest variation occurred for the Soybean-cotton class, being 5 more lofts (Table 7).

Table 7. Comparison between the land use and land cover identified in the crop fields and the land use and land cover registered in the *in situ* dataset for the 2011/2012 harvest period.

Crop fields	Cotton-fallow	Soybean-cotton	Soybean-maize	Soybean-millet	Crop fields total
In situ	12	52	17	0	81
Classification	**10**	**57**	**14**	**0**	**81**
Total	**-2**	**+5**	**-3**	=	

By the confusion matrix referring to the 2011/2012 harvest period classification (Table 8), containing the Global accuracy and Kappa index values, omission and commission errors and the user's and producer's accuracy, it was possible to observe the quality of classification.

Table 8. Confusion matrix of the 2011/2012 harvest period classification.

Classification	*In situ*						
	Cotton-fallow	Soybean-cotton	Soybean-maize	Soybean-millet	Total	Comission errors	User's accuracy
Cotton-fallow	**9**	1	0	0	10	0,10	0,90
Soybean-cotton	3	**51**	3	0	57	0,11	0,89
Soybean-maize	0	0	**14**	0	14	0	1
Soybean-millet	0	0	0	**0**	0	0	1
Total	12	52	17	0	81		
Omission errors	0,25	0,02	0,18	0			
Producer's accuracy	0,75	0,98	0,82	1			
Global accuracy	***0,91***						
Kappa index	***0,82***						

The results obtained in the validation phase demonstrate that the TWDTW method favored the identification of interannual variations in the agglomerate, making it possible to assess the patterns of use and occupation of the agricultural lands. According to Foody et al. (2002), an overall accuracy greater than 85% (0.85) is desirable. However, according to the accuracy levels of the Kappa index proposed by Landis; Koch (1977), a Kappa classification above 0.80 is considered excellent and highly consistent with the *in situ* data. Based on these indices, the classification generated by TWDTW, which obtained a global accuracy and average Kappa index of 89.5% and 0.80, respectively, can be considered excellent for all crops, indicating the potential for the crop identification in the Santa Luzia agglomerate.

As decisions in agriculture are taken on crop field level, research in this plan of analysis becomes necessary, because it allows a better comprehension of the agricultural dynamics (Petitjean; Inglada; Gançarski 2012). Therefore, the TWDTW method, which detected the dynamics on crop field level, favored the understanding of the vegetative dynamics intensity in the Santa Luzia agglomerate, in Sapezal-MT.

Other methods obtained good results, but did not identify interannual variations and the crop dynamics. This is the case of MCDA method, proposed by Gusso et al. (2012), which identified soybean areas in Mato Grosso, but did not consider the management practices and interannual dynamics of the agricultural year within the period of analysis (2001-2013). This factor caused confusions between soybean and typical natural landscapes of Cerrado, such as pastures, on the rainy season beginning, a problem contextualized by Arvor et al. (2011). In general, the planting period varies between days 225 and 337 of the year, beginning in September. However, the beginning of the rainy season, a determining factor for planting, is variable. This was confirmed by Wardlow;

Egbert; Kastens (2007), who identified the increment of EVI MODIS values due to the fast and intense vegetative growth. The TWDTW considered these variations.

Gusso et al. (2014) found that the MCDA identified soybean areas better than 50,000 hectares in Rio Grande do Sul. This was also observed in Mato Grosso, in previous analyzes (ABIOVE 2010, Lobell; Asner 2004). The TWDTW did not present this problem, being able to identify crops in the largest and smallest crop fields without significant variation.

Another problem was in double crop dynamics. Arvor et al. (2011) assessed temporal profiles of different crops and double cropping system in Mato Grosso, in 2006 and 2007, and identified that this practice can reduce the classification accuracy. The double crop dynamic factor was not an obstacle to the TWDTW analysis, due the capability of detect alignments of short patterns in a long time series, considering the agricultural calendar.

One of TWDTW virtues in relation to other methods is the application of a more robust model for filtering noises in vegetation indices time series. Specifically for Mato Grosso, Arvor; Dubreuil; Meirelles (2008) pointed out that, even if the data derived from MODIS are already processed, the temporal profiles of vegetation indices obtained still present noises due to cloudiness (considered to be highest in a region tropical as the south of Amazon Basin), as well as sensor problems and possible errors of atmospheric corrections or action of Bidirectional Reflectance Distribution Function (BRDF) effect, factors that make it necessary to use algorithms to smooth the multitemporal profiles of vegetation indices, in order to further improve the ability to identify and differentiate land use and land cover patterns.

Since TWDTW and other methods directly analyze variations in vegetative vigor in time series, the application of filters to eliminate or minimize noise interference that masks the phenology of agricultural crops is important

(Jönsson; Eklundh 2002). As there are several filters, one should apply the one that minimizes anomalous values without characterizing the vegetation index temporal profile (Adami 2010). In TWDTW, the elimination of these noises consists of the GAM application (Hastie; Tibshirani 1986, Wood 2011), which presents robustness and provides better adjustment to the orbital data of areas with intense agricultural and climatic dynamics than others models, especially the purely parametric (Maus et al. 2016a).

Another virtue was the use of 8-day time series. Guindin-Garcia et al. (2012) concluded that the MOD09Q1 product, with 250 m of spatial resolution and eight days of temporal resolution, showed a greater capacity to monitor agricultural crops due to its higher temporal resolution compared to MOD13Q1, which is 16 days. However, MOD09Q1 does not provide the EVI and NDVI indices. The MOD13Q1 and MYD13Q1 union improve the detection of differences, and the 8-day compositions were better suited to local conditions and agricultural calendars, favoring the identification of crops. In addition, they allowed to capture subtle differences that compositions of 16 days, generally, do not capture. The same was reported by Chaves et al. (2018), Solano et al. (2010), Galford et al. (2008), Wardlow; Egbert (2008).

4. Conclusions

The use of TWDTW method to analyze MODIS time series improved the spectral separability between land use classes and allowed to analyze gradually the changes occurred in Santa Luzia agglomerate between the 2008/2009 and 2011/2012 harvest periods. This level of observation was obtained in other studies, but with less capacity to identify the phenological change point and which crop is planted in the second crop (*safrinha*). The TWDTW obtained the identification and improved the detail level of remote analysis in an area with intense variation in land use and land cover.

The consideration of time weight constraints and the generation of dissimilarity measures favored a comprehension of the intensity of agricultural dynamics in the Santa Luzia agglomerate. Also, the use of multiband time series, as the original spectral bands and transformed ones EVI and NDVI improved the classifications, as observed in the processing step.

As decisions in agriculture are taken on crop field level, the automatic change detection obtained by the TWDTW in this level of analysis, with high classification accuracy and reduced computational efforts and time, make the TWDTW a method that could be integrated into operational programs dedicated to cropland mapping and monitoring based on satellite image time series.

Acknowledgements

This work was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), project entitled: Remote sensing of soybean grain yield in Mato Grosso for precision management and harvest forecasting. Also, the authors wish to thank the farmers for providing the *in situ* data and the Federal University of Lavras (UFLA) for providing office space and infrastructure to achieve this article.

References

ABIOVE. 2010. Soy Moratorium Report: mapping & monitoring of soy plantings in the amazon biome in the third year (http://www.abiove.com.br/english/sustent/relatorio09/moratoria09_relatorio_jul10_us.pdf). Acesso em: 12/07/2018.

Adami, M. 2010. Estimativa da data de plantio da soja por meio de séries temporais de imagens MODIS. 2010. 163p. Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

Arvor et al. 2017. Land use sustainability on the South-Eastern Amazon agricultural frontier: Recent progress and the challenges ahead. *Applied Geography*, 80, 86–97. doi:10.1016/j.apgeog.2017.02.003.

Arvor, D. et al. 2014. Spatial patterns of rainfall regimes related to levels of double cropping agriculture systems in Mato Grosso (Brazil). *International Journal of Climatology*, 34, 2622–2633. doi:10.1002/joc.3863.

Arvor, D. et al. 2013. Mapping and spatial analysis of the soybean agricultural frontier in Mato Grosso, Brazil, using remote sensing data. *GeoJournal*, 78, 833–850. doi:10.1007/s10708-012-9469-3.

Arvor, D. et al. 2012. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Applied Geography*, 32, 702-713. doi:10.1016/j.apgeog.2011.08.007.

Arvor, D. et al. 2011. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing*, 32, 7847-7871. doi:10.1080/01431161.2010.531783.

Arvor, D., Dubreuil, V., Meirelles, M.S.P. 2008. Detection de situations à risques pour la culture du soja a partir de données satellitaires TRMM et MODIS. XXI^{ème} colloque de l'Association Internationale de Climatologie, Montpellier, 99-104.

Asai, M. et al. 2018. Critical factors for crop-livestock integration beyond the farm level: A cross-analysis of worldwide case studies. Land use policy 73, 184–194. doi:10.1016/j.landusepol.2017.12.010.

Baumann, M. 2017. Phenology from Landsat when data is scarce: using MODIS and dynamic time-warping to combine multi-year Landsat imagery to derive annual phenology curves. *International Journal of Applied Earth Observation*, 54, 72–83.doi: 10.1016/j.jag.2016.09.005.

Belgiu, M., Csillik, O. 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204, 509–523. doi:10.1016/j.rse.2017.10.005.

Biudes, M. S. et al. (2015). Patterns of energy exchange for tropical ecosystems across a climate gradient in Mato Grosso, Brazil. *Agricultural and Forest Meteorology*, 202, 112–124. doi:10.1016/j.agrformet.2014.12.008.

Bolton, D.K.; Friedl, M.A. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, 17, 74–84. doi: 10.1016/j.agrformet.2013.01.007.

Brazilian Institute of Geography and Statistics (IBGE), 2011. Sistema IBGE de Recuperação Automática, Produção Agrícola Municipal. Instituto Brasileiro de Geografia e Estatística. Available at: < https://sidra.ibge.gov.br/home/lspa/brasil>. (Last accessed June 18, 2017).

Brown, J.C. et al. 2013. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sensing of Environment*, 130, 39-50. doi:10.1016/j.rse.2012.11.009.

Chaves, M.E.D. et al. 2018. A geostatistical approach for modeling soybean crop area and yield based on census and remote sensing data. *Remote Sensing*, 10, 1–29. doi:10.3390/rs10050680.

Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*. 1991, 37, 35-46.

Congalton, R.G., Green, K. 2009. Reference Data Collection. In: Assessing the Accuracy of Remotely Sensed Data Principles and Practices. 2nd ed. CRC Press, Boca Raton. 85-103. doi:10.1201/9781420055139.ch6.

Coutinho, A.C. et al. 2013. Methodology for systematical mapping of annual crops in Mato Grosso do Sul state (Brazil). *Geografia*, 38, 45-54.

Couto Junior, A.F. et al. 2013. Characterization of the agriculture occupation in the Cerrado biome using MODIS time-series. *Revista Brasileira de GeofÍsica*, 31, 393-402. doi: 10.22564/rbgf.v31i3.312.

Dubreuil, V. et al. 2008. Paysages et fronts pionniers amazoniens sous le regard des satellites: l’exemple du Mato Grosso. *L’Espace Géographique*, 37, 57-74. doi:10.3917/eg.371.0057.

Foody, G.M. 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, 185–201. doi: 10.1016/S0034-4257(01)00295-4.

Franzluebbers, A. J., Sawchik, J., Taboada, M. A. 2014. Agronomic and environmental impacts of pasture–crop rotations in temperate North and South America. *Agriculture, Ecosystems & Environment*, 190, 18–26. doi:10.1016/j.agee.2013.09.017.

Galford, G. et al. 2008. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sensing of Environment*, 112(2), 576–587. doi: 10.1016/j.rse.2007.05.017.

Gibbs, H.K. et al. 2015. Brazil's Soy Moratorium. *Science*, 347, 377-378. doi:10.1126/science.aaa0181.

Gil, J., Siebold, M., Berger, T. 2015. Adoption and development of integrated crop-livestock-forestry systems in Mato Grosso, Brazil. *Agriculture, Ecosystems and Environment*. 199, 394–406. doi:10.1016/j.agee.2014.10.008.

Grzegozewski, D.M. et al. 2016. Mapping soya bean and corn crops in the State of Paraná, Brazil, using EVI images from the MODIS sensor. *International Journal of Remote Sensing*, 37, 1257–1275. http://dx.doi.org/10.1080/01431161.2016.1148285.

Guindin-Garcia, N. et al. 2012. An Evaluation of MODIS 8- and 16-Day Composite Products for Monitoring Maize Green Leaf Area Index. *Agricultural and Forest Meteorology*, 161, 15–25. doi:10.1016/j. agrformet.2012.03.012.

Gusso, A., Arvor, D., Ducati, J.R. 2017. Model for soybean production forecast based on prevailing physical conditions. *Pesquisa Agropecuária Brasileira,* 52, 95–103. doi:10.1590/S0100-204X2017000200003.

Gusso, A. et al. 2014. Assessing the MODIS Crop Detection Algorithm for Soybean Crop Area Mapping and Expansion in the Mato Grosso State, Brazil. *The Scientific World Journal*, 2014, 863141. doi:10.1155/2014/863141.

Gusso, A. et al. 2012. Soybean crop area estimation by MODIS/EVI data. *Pesquisa Agropecuária Brasileira*, 47, 425-435. doi: 10.1590/S0100-204X2012000300015.

Hampf, A.C. et al. 2018. The biophysical and socio-economic dimension of yield gaps in the southern Amazon – A bio-economic modelling approach. Agric. Syst. 165, 1–13. doi:10.1016/j.agsy.2018.05.009.

Hastie T, Tibshirani R. 1986. Generalized Additive Models. *Statistical Science*, 1(3), 297–310.

Huete, A. R. et al. 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, 59, 3, 440-451.

IMEA - Instituto Mato Grossense de Economia Agropecuária, 2016. Captura e análise de dados micro do Agronegócio em Mato Grosso. Relatórios de Mercado. Retrieved from. http://www.imea.com.br.

Instituto Brasileiro de Geografia e Estatística (IBGE). 2011. Sistema IBGE de Recuperação Automática, Produção Agrícola Municipal. Instituto Brasileiro de Geografia e Estatística. Available online: https: //sidra.ibge.gov.br/home/lspa/brasil (accessed on 18 January 2017).

Instituto Nacional de Pesquisas Espaciais (INPE). 2017. Projeto de Monitoramento do Desflorestamento na Amazônia Legal - PRODES. Available at: <www.obt.inpe.br/prodes>. (Last accessed June 18, 2017).

Johann, J.A. et al. 2016. Uso de imagens do sensor orbital MODIS na estimação de datas do ciclo de desenvolvimento da cultura da soja para o Estado do Paraná - Brasil. *Eng. Agrícola*, 36, 126-142. doi:10.1590/1809-4430-Eng.Agric.v36n1p126-142/2016.

Jönsson, P., Eklundh, L. 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40, 1824-1832. doi: 10.1109/TGRS.2002.802519.

Justice, C.O. et al. 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1228-1249. doi: 10.1109/36.701075.

Landis, J., Koch, G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.doi: 10.2307/2529310.

Leteinturier, B. et al. 2006. Adaptation of a crop sequence indicator based on a land parcel management system. Agriculture, Ecosystems and Environment. 112, 324–334. doi:10.1016/j.agee.2005.07.011.

Li, G. et al. 2018. Examining deforestation and agropasture dynamics along the Brazilian TransAmazon Highway using multitemporal Landsat imagery. *GIScience & Remote Sensing*, 2018, 1–23. doi:10.1080/15481603.2018.1497438.

Lobell, D.B., Asner, G.P. 2004. Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, 93, 412-422. doi: 10.1016/j.rse.2004.08.002.

Macedo, M. et al. 2012. Decoupling of deforestation and soy production in the southern Amazon during the late 2000s. Proceedings of the National Academy of Sciences of the United States of America.

Manabe, V., Melo, M., Rocha, J. 2018. Framework for Mapping Integrated Crop-Livestock Systems in Mato Grosso, Brazil. *Remote Sensing*, 10, 1322. doi:10.3390/rs10091322.

Martin, G. et al. 2016. Crop–livestock integration beyond the farm level: a review. Agronomy for Sustainable Development, 36(3). doi:10.1007/s13593-016-0390-x.

Maus, V. et al. 2016a. A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8), 3729-3739. doi: 10.1109/JSTARS.2016.2517118.

Maus, V. et al. 2016b. dtwSat: Time-weighted dynamic time warping for satellite image time series analysis in r. https://cran.r-project.org/web/packages/dtwSat/index.html.

Mello, M.P. et al. 2012. STARS: a new method for multitemporal remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 51, 1-17. doi: 10.1109/TGRS.2012.2215332.

Moraine, M. et al. 2016. Co-design and assessment of cropping systems for developing crop-livestock integration at the territory level. Agricultural Systems, 147, 87–97. doi:10.1016/j.agsy.2016.06.002.

ORNL DAAC. 2018. MODIS Collection 6 Land Product Subsets Web Service. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1600.

ORNL DAAC. 2017. MODIS Collection 6 Land Products Fixed Sites Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1567.

Petitjean, F., Inglada, J., Gançarski, P. 2012. Satellite image time series analysis under time warping. *IEEE Transactions on Geoscience and Remote Sensing,* 50, 3081–3095. doi: 10.1109/TGRS.2011.2179050.

Piva, J. T. et al. 2014. Soil gaseous N2O and CH4 emissions and carbon pool due to integrated crop-livestock in a subtropical Ferralsol. *Agriculture, Ecosystems & Environment*, 190, 87–93. doi:10.1016/j.agee.2013.09.008.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria (URL: ⟨http://www.R-project.org/⟩).

Risso, J. et al. 2012. Modis vegetation indices applied to soybean area discrimination. *Pesquisa Agropecuária Brasileira*, 47, 1317-1326. doi: 10.1590/S0100-204X2012000900017.

Rouse, J. W. et al. 1973. Monitoring vegetation systems in the great plains with ERTS. Earth Resources Technology Satellite-1 Symposium. 3, Washington. Proceedings…Washington: NASA: 309-317. doi: citeulike-article-id:7234782.

Rudorff, B.F.T. et al. 2011. The soy moratorium in the Amazon biome monitored by remote sensing images. *Remote Sensing*, 3, 185-202; doi:10.3390/rs3010185.

Ryschawy, J. et al. 2017. Designing crop–livestock integration at different levels: Toward new agroecological models? Nutrient Cycling in Agroecosystems, 108(1), 5–20. doi:10.1007/s10705-016-9815-9.

Sakoe, H., Chiba S. 1978. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43–49. doi:10.1109/TASSP.1978.1163055.

Santhana-Vannan, S. K. et al. 2010. A Web-Based Subsetting Service for Regional Scale MODIS Land Products. *IEEE JSTARS*, 2(4): 319–328.

Silva, F. D. da et al. 2014. Soil carbon indices as affected by 10 years of integrated crop–livestock production with different pasture grazing intensities in Southern Brazil. Agriculture, Ecosystems & Environment, 190, 60–69. doi:10.1016/j.agee.2013.12.005.

Soares, J.V, Batista, G.T., Shimabukuro, Y.E. 2007. Sensor MODIS: Histórico e Descrição. In: O Sensor Modis e suas aplicações ambientais no Brasil - Shimabukuro. Y. E.; Rudorff, B. F. T.; Ceballos, J. C. (Coords). São José dos Campos: Editora Parêntese, SP, Brasil, 2007.

Solano, R. et al. 2010. MODIS Vegetation Index (MOD 13) C5 User's Guide. The University of Arizona. 38p.

Souza, C.H.W. et al. 2015. Mapping and discrimination of soya bean and corn crops using spectro-temporal profiles of vegetation indices. *International Journal of Remote Sensing*, 36, 1809-1824. doi:10.1080/01431161.2015.1026956.

Verbesselt, J. et al. 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114, 106-115. doi: 10.1016/j.rse.2009.08.014.

Zeilhofer, P. et al. 2012. Seasonal variations in litter production and its relation with MODIS vegetation indices in a semi-deciduous forest of Mato Grosso. *Remote Sensing Letters*, 3(1), 1–9. doi:10.1080/01431161.2010.523025.

Zhu, C. et al. 2016. Mapping Fractional Cropland Distribution in Mato Grosso, Brazil using time series MODIS Enhanced Vegetation Index and Landsat Thematic Mapper data. *Remote Sensing*, 8, 1-14. doi:10.3390/rs8010022.

Waldhoff, G., Lussem, U., Bareth, G., 2017. Multi-Data Approach for remote sensing-based regional crop rotation mapping: A case study for the Rur catchment, Germany. *International Journal of Applied Earth Observation and Geoinformation*. 61, 55–69. doi:10.1016/j.jag.2017.04.009.

Wardlow, B. D., Egbert, S. L. 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the U.S. Central Great plains. *Remote Sensing of Environment*, 112, 1096–1116. doi: 10.1016/j.rse.2007.07.019.

Wardlow, B.D., Egbert, S.L., Kastens, J.H. 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment,* 108, 290–310. http://dx.doi.org/10.1016/j.rse.2006.11.021.

Wood, S.N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3-36.

ARTIGO 4 - SOJASAT: WEB PLATFORM TO MONITORING SOYBEAN CROPS IN MATO GROSSO, BRAZIL

Normas do periódico *Applied Geomatics*, ISSN: 1866-9298
Versão preliminar

Michel Eustáquio Dantas Chaves[a], Marcelo de Carvalho Alves[b*]

[a*] Engineering Department, Federal University of Lavras. Campus Universitário, PO Box 3037, ZIP code 37200-000, Lavras, Brazil. E-mail: medchaves@posgrad.ufla.br; Phone: +55 35 99968 2078

[b*] Engineering Department, Federal University of Lavras. Campus Universitário, PO Box 3037, ZIP code 37200-000, Lavras, Brazil. E-mail: marcelo.alves@deg.ufla.br; Phone: +55 35 3829 1027

***** Correspondence: medchaves@posgrad.ufla.br; Phone: +55 35 99968 2078

Abstract

The virtual web platform development to disseminate data about land use land cover have proved useful for the interactive presentation of detected changes, avoiding computational efforts and processing time. In the specific case of agriculture in Mato Grosso, with alternate plantations, successions and different rotations for each crop, the interactive presentation of the variations detected between the harvest periods is a key part for the study and the understanding of its dynamics. Therefore, the objective of this work was to present the process of elaborating a virtual platform for the visualization and download of information on the areas of soybean plantation and its yield in the State of Mato Grosso between 2000 and 2011. The methodological procedure involved the elaboration of a virtual environment containing area and yield maps from 2001 to 2011, transformed into web maps. The results show the platform interactivity and the possibility of observing and downloading the changes between harvest crops. The contribution of the SojaSAT virtual platform is to make it possible to follow the dynamics of sojicultura in the territory, its processes of occupation and its transformations, through a spatial and quantitative monitoring of the vegetation cover and the land use in Mato Grosso.

Keywords: Internet Technologies; Scientific diffusion; Agriculture data.

1. Introduction

Due the possibility of extract parameters about agricultural vegetation, the scientific community has been developing applications to indicate gradual changes in phenological development using time series to detect changes in agricultural land use (Brown et al. 2013, Maus et al. 2015), and to monitor the agricultural advance (Zhu et al. 2016, Gusso et al. 2014). The result has been the emergence of tools to crop monitoring using the application of remote sensing techniques of terrestrial surface and temporal analysis of vegetation indices.

Most of these tools are the web platforms available on Internet to disseminate the data for the general public. The development of geobrowser tools, based on virtual globes, has provided free access to high spatial resolution images and geographical maps derived from remote sensing satellites. The development of these virtual globes allowed researchers and general public to visualize geospatial data, to understand multi-scale geography, to process data and to publish information (Butler 2006., Ballagh et al. 2011, Chiang et al. 2011).

The availability of maps and Geoinformation on the Internet arises as a consequence of the advance of its use in the present times. In particular, the dissemination of content on the Internet represents the most current way of transmitting information that can now be useful to any space actor. Internet technology could change the way of development and distribution of expert systems and GIS, since knowledge about any subject can be directly available to users, by cloud computing storage and processing, on servers shared and interconnected through the internet (Duan et al. 2005).

The rapid advances in Internet technologies have opened new opportunities for enhancing traditional Expert Systems (ES). Internet technology can change the way that an ES is developed and distributed. For the first time, knowledge on any subject can directly be delivered to users through a web-based

ES. Since its main function is to mimic expertise and distribute expert knowledge to non-experts, such benefits can be greatly enhanced by using the Internet.

With the presence of people in the virtual world, the availability of information about diverse themes, that the agriculture, for example, allows the diffusion of technology and the generation of subsidies for interventions in the sector. In the agricultural case, the generation and availability of crop maps every year allows the effective crop areas and yield advance monitoring. In the other hand, the availability of historical maps allows the knowledge about the formation of productive sectors and demonstrates the gradual crop expansion.

In view of the above, this paper aimed to present the methodology used for development of the SojaSAT web platform, a virtual platform created to disseminate the results obtained by Chaves et al. (2018) about the soybean crop area and yield expansion in Mato Grosso along to 2000 years first decade, the decade of major soybean crop expansion in Brazil. The contribution of this platform is allow the historical review and investigate the gradual crop expansion in the Mato Grosso territory.

2. Material and methods

2.1 Study area

The area selected for the implementation of SojaSAT web platform is the State of Mato Grosso (MT), which is located in the Central–west region of Brazil (Figure 1) and covers an area of approximately 905,000 km^2. Mato Grosso was chosen because due to the heterogeneous landscape, a result of an active pioneer frontier shaped by different populations (public and private colonists, logging industries, indigenous societies), land uses, practices, and varying natural conditions (climate, soil, vegetation).

Figure 1. Geographic location of the State of Mato Grosso, Brazil, with its mesoregions. Source: Brazilian Institute of Geography and Statistics (IBGE) (2017).

The southern region of the State is a tropical wetland known as the Pantanal (61,000 km^2). In the north are the moist forests of Amazonia (484,000 km^2). The central region is dominated by vast tropical savannas known as Cerrado (360,000 km^2), where agribusiness is concentrated (Kastens et al. 2017). According to Brown et al. (2013), Mato Grosso's climate (Köppen Aw) is hot-semi-humid to humid - with pronounced seasonality marked by a dry season from May to October. The rainy season occurs from October to May (Gusso et al. 2014), and the climatic gradient is largely coincident with a gradient in land-use change, indicating the interconnectedness of biophysical and socio-economic processes (Davidson et al. 2012).

Richards et al. (2015a) explained that Mato Grosso has benefitted from a number of geographic and institutional conditions that have increased the capacity of the state's agriculture sector to serve as a growth engine, constituting in a separate case, with three peculiarities: first, Mato Grosso is a relatively recent agricultural frontier. Even today, properties are still being cleared or converted to agriculture as investment capital is made available. Second, Mato Grosso's agricultural sector is tightly coupled with local, urban-based supply chains (Garrett el al. 2013). Due the strength of the local supply chain, and the presence of downstream processing facilities in Mato Grosso, the local economy captures and circulates a larger proportion of the potential value of each harvest. Third, and finally, in Mato Grosso many farm owners live and spend locally (Richards et al. 2015b).

This context converted Mato Grosso into a globally important center of agricultural production (Cohn et al. 2016, Spera et al. 2014). The total land-use shift into soybean in Mato Grosso from 2001 to 2011 was almost 8.7 million ha, of which almost 3.5 million ha belonged to the Brazilian Amazon biome (Gusso et al. 2014). According to Brazilian Institute of Geography and Statistics (IBGE) (Garrett et al. 2013), the total planted area in Mato Grosso increased by 297% (from 4.74 million to 14.10 million ha) between 2000 and 2015.

These scenario represents difficulty to establish data with no overlap land use land cover categories, a perfect challenge to the mapping elaborated.

2.2 Soybean crop area and yield identification

Advances in satellite imagery and remote sensing have enabled the acquisition of spatial data at several different resolutions. Geographic information systems (GIS) can be used to link geographic data from different sources. The chronological ordering of images allowed the use of data provided by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor to evaluate crop fields (Chen et al. 2018; Zhu et al. 2016, Gusso et al. 2014, Brown et al. 2013). Specifically, in relation to Mato Grosso's agriculture, the time series derived from MODIS vegetation indexes favors the typical patterns observation of different agricultural uses, having spatial and temporal resolutions compatible with the crops size and agricultural dynamics.

The SojaSAT web platform was constructed to disseminate the results obtained by Chaves et al. (2018), who presented the soybean crop detection and yield prediction by linking census data, GIS, remote sensing, and geostatistics of the 2000 first decade. The adopted approach combines Brazilian Institute of Geography and Statistics (IBGE) census data with an eight-day enhanced vegetation index (EVI) time series derived from MODIS Collection 5 data to monitor soybean areas and yields in Mato Grosso.

A dataset of *in situ* data from farms were used to validate the obtained results. Binomial areal kriging was used to generate maps of soybean occurrence over the years, and Gaussian areal kriging was used to predict soybean crop yield census data inside detected soybean areas, which had a downscaling effect on the results. The global accuracy and the Kappa index for the soybean crop detection were 92.1% and 0.84, respectively. The yield prediction presented 95.09% accuracy considering the standard deviation and probable error. Soybean crop detection and yield monitoring can be improved by this approach.

2.3 Web platform architecture – Web maps creation in QGIS

In order to disseminate the SojaSAT results, a GIS Web application was developed by the Web Map Service (WMS) to enable interactive display and instant access to the results. The web platform was developed using Leaflet and OpenLayers tools and hosted in a GIS server using the soybean area and yield outputs generated, in order to integrate the database and the internet. The data, server directories, and configuration store reside locally on the temporal GIS server architecture, developed to organize the database.

The main steps (Figure 2) involved the insertion of the area and yield maps generated by the method in QGIS Desktop 2.16 Nødebo, where was configured the scales and categories of area and yield. After, the configured maps were transformed and prepared to be exported to web maps. Server and directories were defined for export the results to web browser.

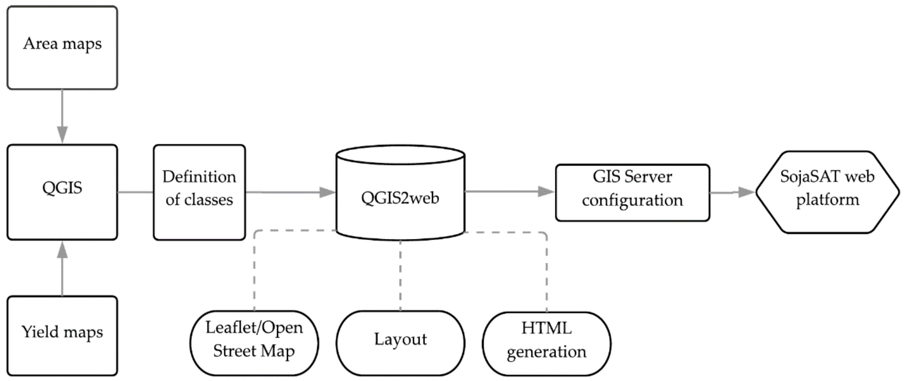


Figure 2. Flow chart of the methodological procedures.

2.4 From QGIS to web

To share the maps in the internet we build a highly-interactive map, which is based on web services. This recipe will show you how to build a simple web map using Leaflet - a popular JavaScript library that is used to create web maps.

The qgis2web plugin converts the map into something that is compatible with the web. Generally, this means converting vector data to the GeoJSON format and generating an HTML page (web page) with some JavaScript to create and populate the map. In this plugin, the process basically involved the load of interest layers, the definition of appearance style, including the addition of a Web Map Services (WMS) layer to the maps, and the configuration of labels. This was used to transform the area and yield maps (raster and shape files) in web maps. The map final arts were prepared in QGIS. This process involved the chronological ordering of the maps and definition of scale, legend and projection. After, was choose the background basemap for spatialize the soybean crops in Mato Grosso.

The basemap selected was Open Street Map (OSM), a crowdsourcing project aimed at the development of an open map of the world. Worldwide, the OSM community uses high-resolution remote sensing imagery, GPS surveying, and local knowledge to make the map as accurate and detailed as possible (Bruy; Svidzinska 2015). Thus, this basemap contains the municipalities, roads and Indigenous lands of the Mato Grosso State, important references for the work purpose, and can provide a more detailed representation of Land Use Land Cover (LULC) than that achieved solely through remote sensing.

After the basemap definition, were determined the layers and web map aspects, as extent, geolocation, scale/zoom and appearance, highlighting features on mouse over and popups for finish the template. These configurations were applied to all layers generated (Figure 3).

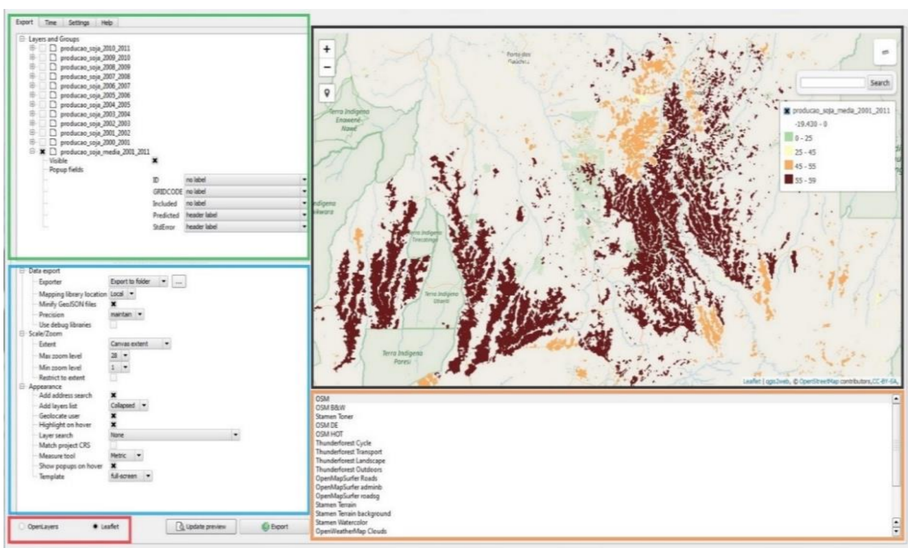


Figure 3. Web maps layout configuration on the OSM basemap. In green box, the set of imported layers. In blue box, the configurations to export to folder. In red box, the choose between OpenLayers or Leaflet modules. In orange box, the basemap options. In black box, the instantaneous preview of the web platform layout.

Subsequently, were defined folder directories to storage the web maps created. The projects containing all configurations and results were exported to the respective folder directories. After this, the maps were opened in web browser. The result was a new folder named with this sequence: export_year_month_day_hour_minute_seconds (for example, export_2018_09_18_12_23_29). Inside this folder is index.html. The open of this file with a web browser allow to see the map.

The last step involved the configuration of server that maintain the platform on the web. The mechanism used was Adobe Dreamweaver, with HTML and CSS routines. This mechanism also was adopted to create a responsive website, in order to guaranties the web platform interactivity.

3. Results and discussion

3.1 Interactive visualization of SojaSAT data on the web

The SojaSAT web platform results are accessible through the URL (http://www.sergeo.deg.ufla.br/sergeo/Projetos/SojaSat/SojaSat.html). In this website, we present a visualization tool of soybean area and yield intensification between 2000/2001 and 2010/2011 crop seasons (Figure 4). Browsers with IE (Internet Explorer) core, or other browsers such as FireFox and Chrome with Adobe Flash Player plugin are recommended.

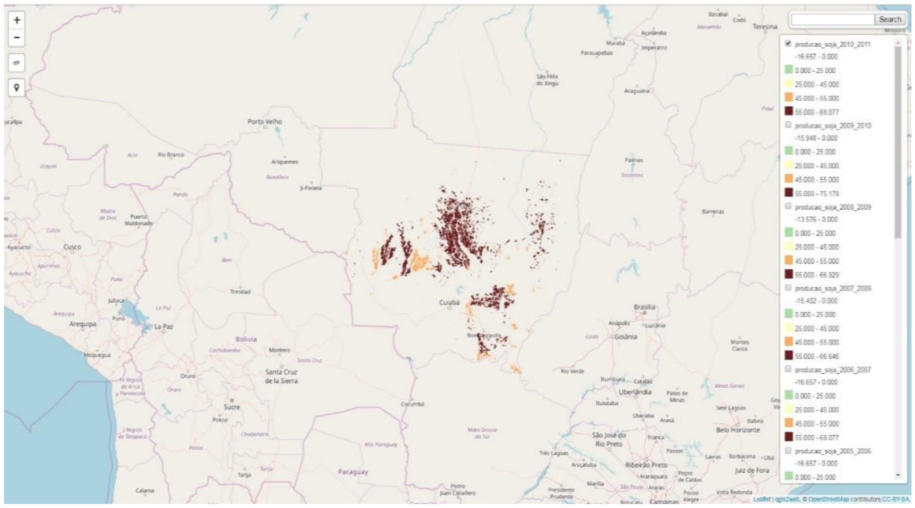


Figure 4. The user interface of SojaSAT web site on the internet. In this example, were accessed the 2010/2011 soybean crop yield results.

The user interface (UI) is a key element in any web application, because it is the communication point between the user and the system. In general terms, our aim was to develop a web application that allows the user to easily observe and analyze the soybean crop expansion in Mato Grosso between 2000 and 2011, as seen throughout this paper. For this, a user friendly interface was implemented to connect the maps module and the maps module facilities with virtual laboratory users. Note that all the vectors are clickable, and the popup will display the attribute table information. If you turned on labels and hover, then hovering over a point will display the name.

The SojaSAT web site presents the spatio-temporal distribution of the soybean crop area and yield in Mato Grosso (Figure 5).

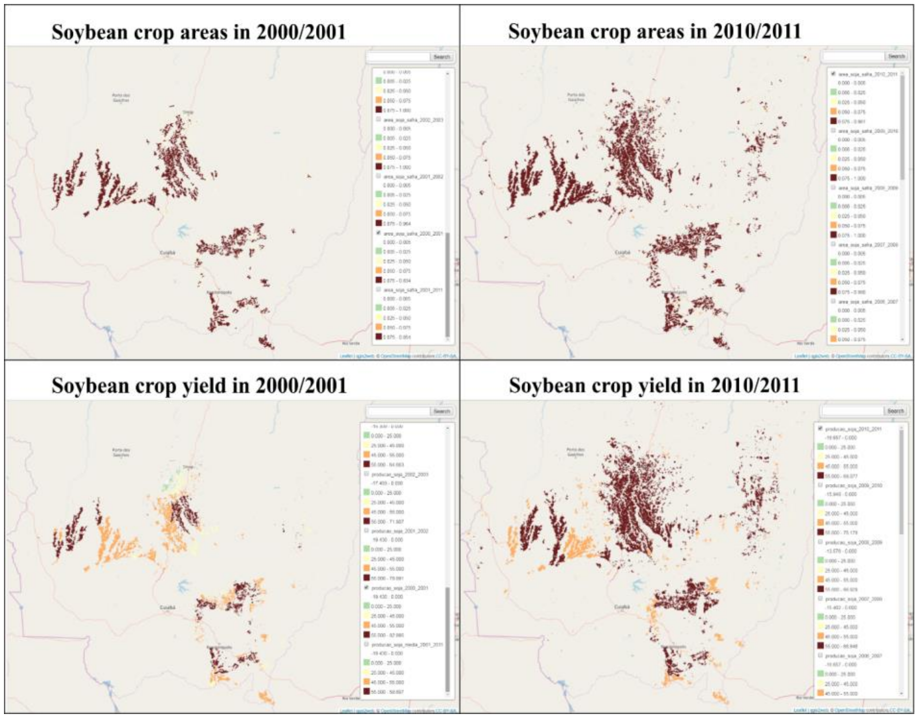


Figure 5. Examples of the data presented in SojaSAT. Soybean crops area and yield between 2000/2001 and 2010/2011.

In the legend bar are presented the categories of area and yield obtained in each harvest period interactively. By the zoom tool it is possible to observe the soybean crops distribution and the land continuous of Mato Grosso, with roads, municipalities, indigenous lands and other protected areas. The interactivity presented by the platform allows to users as instantaneous comprehension about the advance of the soybean crops in Mato Grosso.

3.2 The data catalog

All the generated shapefile (.shp) data of the cultivated areas and the yield obtained in each crop season are available for download. These data serves as a historical repository about the soybean crop expansion in Mato Grosso, in the first decade of the actual century. The maps can explain the dynamic of soybean crop boom, occurred in this years. For the download is not necessary anyone register.

3.3 Future directions

The next phase of the SojaSAT project is the implementation of a virtual library of the crop patterns detected in Mato Grosso. These patterns are from soybean, corn, cotton and millet only and in the successional stage from different regions (Figure 6). These steps are in development, and the generated temporal profiles will be inserted in the SojaSAT web platform.

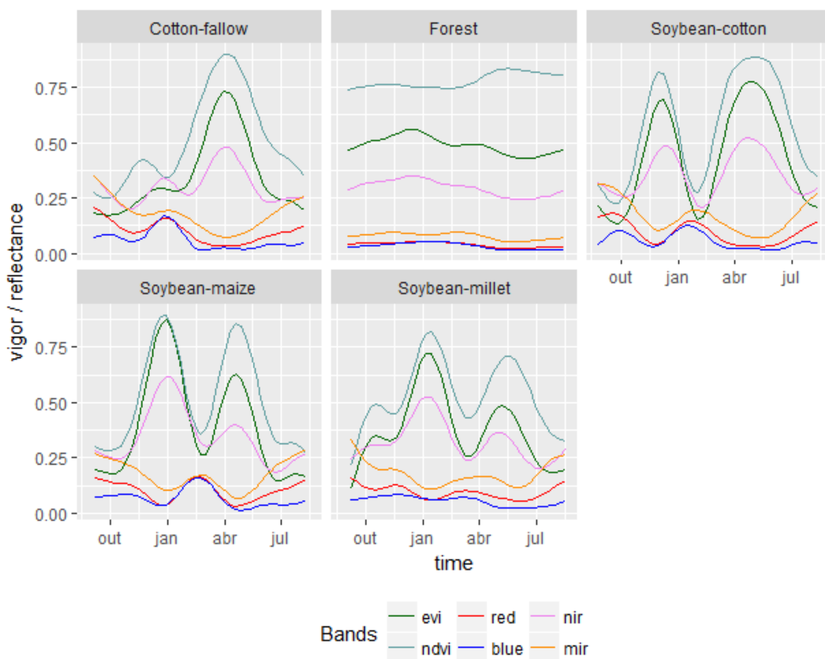


Figure 6. Example of virtual library of crop temporal patterns detected in Mato Grosso. Adapted from Maus et al. (2016).

The library will be important to disseminate the crop patterns data detected in Mato Grosso, favoring the summer crops identification. This information is directly related to the agricultural calendar, that represents the phenological cycle of the crops. Regarding to the integration with virtual globes such as OSM and GoogleMaps, this work showed an innovation because it allows the public access and instantaneous visualization. This work can be extended for any geographical region since MODIS data are available for the entire globe.

Another future step is the actualization of web maps, providing continuous monitoring of the soybean crop area and yield expansion in Mato Grosso by the MODIS sensor data.

4. Conclusions

The Internet technology able innumerous possibilities to disseminate contents. The Brazilian agriculture need more intelligence and nowadays is vast the forms of complete this task. For the specific case of Mato Grosso, the historical monitoring is very important to orient guidelines and discover tendencies in the crop distribution.

This paper presents the elaboration of the web platform for computer modeling of soybean crop production phenomena. The platform provides access to database of MODIS soybean crop detected maps, capability to execute computational tasks in highlight of the area and yield obtained in each local. These data can be used as ancillary to local and regional analysis about the soybean crops expansion in the first 2000's decade in Mato Grosso, being compared with current data and avoiding a new classification of this years for temporal analysis. This description indicated that with a minimum of technical knowledge about soybean agricultural practices it is possible to the analyst recovering the 10 years' history of specific plots and fields. This can be of great interest to certifiers that need to know the LULC change history.

We have discussed the general scheme of SojaSAT and the model of interactions between the platform modules. Such a resource will allow researchers to access high-performance computing resources that will significantly reduce the time and cost of the research and development process of soybean crops area and yield expansion in Mato Grosso.

Acknowledgements

The present work was carried out with the support of the National Council of Scientific and Technological Development (CNPq) for the financing of the project entitled: Remote sensing of soybean grain yield in Mato Grosso for precision management and harvest forecasting. The present work also was carried out with the support of the Coordination of Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001.

References

Ballagh LM, Raup BH, Duerr RE, Khalsa SJS, Helm C, Fowler D, Gupte A. Representing scientific data sets in KML: Methods and challenges, *Comp Geosci*. 2011;37(1):57–64.

Brown JC, Kastens JH, Coutinho AC, Victoria D de C, Bishop CR. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sens Environ.* 2013;130:39–50.

Bruy A, Svidzinska D. QGIS by example: Leverage the power of QGIS in real-world applications to become a powerful user in cartography and GIS analysis. *The Physical Science Basis*. 2015;1-299.

Butler D. Virtual globes: the web-wide world. *Nature*. 2006;439(7078):776–778.

Chaves MED, Alves M de C, de Oliveira MS, Sáfadi T. A geostatistical approach for modeling soybean crop area and yield based on census and remote sensing data. *Remote Sens.* 2018;10(5):680.

Chen Y, Lu D, Moran E, Batistella M, Dutra LV, Sanches ID, et al. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int J Appl Earth Obs Geoinf.* 2018;69(March):133–47.

Chiang G, Toby OH, Dove MT, Bovolo CI, Ewen J. Geo-visualization Fortran library, *Comp Geosci*. 2011;37(1):65–74.

Cohn AS, VanWey LK, Spera SA, Mustard JF. Cropping frequency and area response to climate variability can exceed yield response. *Nat Clim Chang.* 2016;6(6):601-4.

Davidson EA, Araújo AC, Artaxo P, Balch JK, Brown IF, Bustamante MMC, Coe MT, DeFries RS, Keller M, Longo M et al. The Amazon basin in transition. *Nature.* 2012;481:321-328.

Duan, Y, Edwards, JS, Xu, MX. Web-based expert systems: Benefits and challenges. *Inf. Manag.* 2005. 42:799-811.

Garrett RD, Lambin EF, Naylor RL. Land institutions and supply chain configurations as determinants of soybean planted area and yields in Brazil. *Land Use Pol*. 2013;31:385-96.

Gusso A, Arvor D, Ricardo Ducati J, Veronez MR, Da Silveira LG. Assessing the modis crop detection algorithm for soybean crop area mapping and expansion in the Mato Grosso state, Brazil. *Sci World J.* 2014;2014:863141.

Kastens JH, Brown JC, Coutinho AC, Bishop CR, Esquerdo JCDM. Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. *PLoS ONE*. 2017;12(4):e0176168.

Maus V, Câmara G, Cartaxo R, Sanchez A, Ramos FM, Queiroz GR. A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2016;9(8), 3729-3739.

Richards P, Pellegrina H, VanWey L, Spera S. Soybean development the impact of a decade of agricultural change on urban and economic growth in Mato Grosso, Brazil. *PLoS One.* 2015;10(4): e0122510.

Richards P, VanWey L. Where Deforestation Leads to Urbanization: How Resource Extraction Is Leading to Urban Growth in the Brazilian Amazon. *Ann Assoc Am Geogr*. 2015b;105(4):806–23.

Spera SA, Cohn AS, VanWey LK, Mustard JF, Rudorff BF, Risso J, et al. Recent cropping frequency, expansion, and abandonment in Mato Grosso, Brazil had selective land characteristics. *Environ Res Lett.* 2014;9(6):064010.

Zhu C, Lu D, Victoria D, Dutra LV. Mapping fractional cropland distribution in Mato Grosso, Brazil using time series MODIS enhanced vegetation index and Landsat Thematic Mapper data. *Remote Sens.* 2016;8(1):22-3.