UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS RURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

Luan Pierre Pott

MAPEAMENTO AGRÍCOLA UTILIZANDO SENSORIAMENTO REMOTO, MODELAGEM DE CULTURAS E APRENDIZADO DE MÁQUINA NO RIO GRANDE DO SUL

Santa Maria, RS
2023

Luan Pierre Pott

MAPEAMENTO AGRÍCOLA UTILIZANDO SENSORIAMENTO REMOTO, MODELAGEM DE CULTURAS E APRENDIZADO DE MÁQUINA NO RIO GRANDE DO SUL

Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Doutor em Engenharia Agrícola**.

Orientador: Prof. Dr. Telmo Jorge Carneiro Amado

Santa Maria, RS
2023

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001

Pott, Luan Pierre
MAPEAMENTO AGRÍCOLA UTILIZANDO SENSORIAMENTO REMOTO, MODELAGEM DE CULTURAS E APRENDIZADO DE MÁQUINA NO RIO GRANDE DO SUL / Luan Pierre Pott.- 2023.
137 p.; 30 cm

Orientador: Telmo Jorge Carneiro Amado
Coorientador: Ignacio Antonio Ciampitti
Tese (doutorado) - Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós Graduação em Engenharia Agrícola, RS, 2023

1. Agricultura digital 2. Mapeamento de culturas agrícolas 3. Modelos de culturas agrícolas 4. Mapeamento da rotação de culturas 5. Ciência de dados na agricultura I. Amado, Telmo Jorge Carneiro II. Ciampitti, Ignacio Antonio III. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

Declaro, LUAN PIERRE POTT, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Tese) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

Luan Pierre Pott

MAPEAMENTO AGRÍCOLA UTILIZANDO SENSORIAMENTO REMOTO, MODELAGEM DE CULTURAS E APRENDIZADO DE MÁQUINA NO RIO GRANDE DO SUL

Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do título de **Doutor em Engenharia Agrícola**.

Aprovado em 23 de agosto de 2023:

Telmo Jorge Carneiro Amado, Dr. (UFSM)
(Presidente/Orientador)

Ignacio Antonio Ciampitti, PhD (KSU)
Co-orientador

Raí Augusto Schwalbert, Dr. (GDM)

Geomar Mateus Corassa, Dr. (CCGL)

Christian Bredemeier, PhD (UFRGS)

Santa Maria, RS
2023

Dedico todo o trabalho desenvolvido aos meus pais que fundamentalmente me concederam ambiente para construção do estudo.

AGRADECIMENTOS

Gostaria de agradecer à Deus por iluminar minha trajetória para conquista deste passo em minha vida.

À minha mãe Teresa Elena Pott, pai Eldemar Pott, irmã Luciana Pott Prediger, e toda minha família pelo apoio incondicional, amor e carinho sempre prestados. Sem vocês nada seria possível.

Também agradeço aos meus avós *in memoriam* que me motivaram sempre para minhas conquistas, muito obrigado, me orgulho muito de vocês.

À minha esposa Priscila Luisa Schuster por todo amor e carinho de sempre, compreendendo todos os momentos da minha vida.

À todos meus professores que me proporcionaram a educação, principalmente aos professores Telmo Jorge Carneiro Amado e Ignacio Antonio Ciampitti que foram meus conselheiros e grande educadores, não apenas cientificamente, mas também pessoal e profissional.

À todos meus amigos pelo apoio e convívio, irei levar para toda a vida.

Aos amigos do laboratório do Projeto Aquarius e KSU-Crops por toda ajuda, convívio diário, viagens, coleta de dados, parceria, amizade de sempre.

Aos amigos do Grupo Don Mario (GDM), agradeço pelo companheirismo e crescimento profissional.

Aos avaliadores da banca de qualificação e por fim da tese final Christian Bredemeier, Raí Augusto Schwalbert, Geomar Mateus Corassa, Leonardo Bastos e Lúcio de Paula Amaral, meu muito obrigado.

“There is another world, but it is in this one.”

(Paul Éluard)

RESUMO

MAPEAMENTO AGRÍCOLA UTILIZANDO SENSORIAMENTO REMOTO, MODELAGEM DE CULTURAS E APRENDIZADO DE MÁQUINA NO RIO GRANDE DO SUL

AUTOR: Luan Pierre Pott
ORIENTADOR: Telmo Jorge Carneiro Amado

A agricultura está sob intensa revolução, inúmeros dados são gerados a cada instante, seja pelo produtor, sensores, ou por geração de novos produtos para o setor agrícola, favorecendo a agricultura digital. Assim, os objetivos desse trabalho foram coletar dados de campo, bem como utilizar dados disponíveis como de sensoriamento remoto e dados geoespaciais de uso público para gerar conhecimento para a agricultura do Rio Grande do Sul (RS), Brasil. Os objetivos do primeiro trabalho foram i) avaliação da variabilidade espacial de dados de campo para gerar o modelo de classificação de culturas; ii) avaliar o modelo de transferência de aprendizagem com os dados da estação de cultivo subsequente; iii) avaliar a precisão do modelo de previsão para previsões antecipadas, e iv) desenvolver um modelo de classificação e mapeamento das culturas agrícolas para o RS. Os objetivos do segundo trabalho foram: i) comparar dados gerados através de simulações de desenvolvimento de cultivos agrícolas e dados de campo; ii) avaliar máscaras de campos de produção geradas pelo Cadastro Ambiental Rural, MapBiomas e modelo de *random forest*; e iii) avaliar modelos de classificação não-supervisionada, classificação supervisionada com dados de simulações de desenvolvimento de cultivos agrícolas, e classificação supervisionada com dados de campo, bem como a combinação dos mesmos. Os objetivos do terceiro trabalho foram: i) mapear padrões de monocultivos e rotação de culturas nas diferentes mesoregiões do estado do RS; ii) identificar variáveis de solo e de clima que coincide com maiores percentuais de área de monocultivo; iii) avaliar o efeito da rotação de culturas na produtividade de grãos das culturas. Como resultados dos trabalhos, foram gerados modelo de classificação e mapeamento de culturas agrícolas do RS, com a possibilidade de transferência de aprendizagem para safras subsequentes, obtendo predições a partir de 1º de janeiro da safra agrícola, aumentando a acurácia na medida que são capturadas mais imagens de sensoriamento remoto da safra. Também, no segundo trabalho foi possível gerar modelos de classificação de culturas agrícolas com diferentes modelos, classificação não-supervisionada, classificação supervisionada com dados de campo, com simulações de modelos de desenvolvimento de culturas, e agregando dados de campo e de simulações para aumentar a acurácia do modelo. O mapeamento de rotação de culturas e dos padrões de rotação de culturas para o estado do RS foram gerados possibilitando um olhar mais holístico para a adoção de estratégias de rotação de culturas, intensificação, e sustentabilidade da agricultura para o estado. Os resultados apresentados nesse estudo têm potencial para contribuir com a digitalização na agricultura, podendo auxiliar agricultores, e agentes formuladores de políticas durante o processo de tomada de decisão.

Palavras-chave: Agricultura digital; Mapeamento de culturas agrícolas; Modelos de culturas agrícolas; Mapeamento da rotação de culturas.

ABSTRACT

CROP MAPPING USING REMOTE SENSING, CROP MODELING, AND MACHINE LEARNING IN RIO GRANDE DO SUL

AUTHOR: Luan Pierre Pott
ADVISOR: Telmo Jorge Carneiro Amado

Agriculture is under intense revolution, numerous data is generated every moment, either by the farmers, sensors, or by generating new products for the agricultural sector, favoring digital agriculture. Thus, the objectives of this work were to collect field data, as well as use available data such as remote sensing and geospatial public data to generate knowledge for agriculture of Rio Grande do Sul (RS), Brazil. The objectives of the first study were to i) evaluate of the spatial variability of field data to generate the crop classification model; ii) evaluate the transfer learning model with the subsequent growing season data; iii) evaluate the accuracy of the forecast model for early forecasts, and IV) develop a classification and mapping model of agricultural crops for RS. The objectives of the second study were to: i) compare data generated through simulations of development of agricultural crops and field data; II) evaluate production fields masks generated by the Rural Environmental Registry, MapBiomas and random forest model; and iii) evaluate non-supervised classification models, supervised classification with data from agricultural crop development simulations, and supervised classification with field data, as well as their combination. The objectives of the third study were to: i) map monoculture patterns and crop rotation in the different mesoregions of the state of RS; ii) identify soil and climate variables that coincides with the highest percentages of monoculture area; iii) evaluate the effect of crop rotation on crop grain yields. As a result of the first study, the model of classification and mapping of agricultural crops of RS were generated, with the possibility of transfer learning to subsequent growing seasons, obtaining predictions from January 1 of the agricultural crop, increasing accuracy as more remote sensing images of the crops are captured. Also, in the second study it was possible to generate models of classification of crop types with different models, non-supervisioned classification, supervised classification with field data, simulations of crop development models, and adding field data and simulations data to increase the accuracy of the model. Crop rotation mapping and crop rotation patterns for the state of RS were generated by enabling a more holistic look at the adoption of crop rotation strategies, intensification, and sustainability of agriculture to the state. The results presented in this study have the potential to contribute to digitization in agriculture, and may assist farmers, and policymakers during the decision-making process.

Keywords: Digital agriculture; Crop type mapping; Crop models; Crop rotation mapping.

SUMÁRIO

1 APRESENTAÇÃO

A agricultura está sob intensa revolução digital, onde há imensa geração de dados de diversos formatos que necessita de tratamento, processamento até a geração de novos conhecimentos ou produtos para melhores tomadas de decisões, seja para maior economia, produtividade ou sustentabilidade do setor agrícola.

Diversas fontes de dados são utilizados rotineiramente para geração de informações e novos conhecimentos para o setor agrícola, no entanto o sensoriamento remoto (LOBELL et al., 2019; WEISS et al., 2020) e mais recentemente os modelos de crescimento de culturas que possibilitam a geração de inúmeras simulações de crescimento e desenvolvimento de culturas agrícolas em diferentes cenários tanto de clima, solo e manejo, auxilia e agrega para geração de conhecimento (KASAMPALIS et al., 2018; SAIZ-RUBIO & ROVIRA-MÁS, 2020).

O primeiro passo para o monitoramento digital da agricultura é a classificação e mapeamento das culturas agrícolas. Classificação de culturas agrícolas vem sendo realizada utilizando sensoriamento remoto e classificação não-supervisionada – sem dados de campo (WANG et al., 2019; KLUGER et al., 2021), com a utilização de sensoriamento remoto com classificação supervisionada – com dados de campo (BELGIU & CSILLIK, 2018; HAO et al., 2020; DADO et al., 2020), bem como a utilização de dados de campo associados à simulações geradas com modelos de culturas (JIN et al., 2019; DEINES et al., 2020).

Com a classificação e mapeamento das culturas agrícolas, além de obter informações importantes intrínsecas à esse produto, como atualização de área cultivada nas diferentes regiões e diferenciação da área de cultivo ao longo dos anos, a geração do mapeamento de culturas agrícolas possibilita também o monitoramento dos cultivos agrícolas ao longo do ciclo em plataformas digitais, previsão de safra (SCHWALBERT et al., 2020; DADO et al., 2020; DEINES et al., 2020), previsão de produção agrícola (SCHWALBERT et al., 2020), e mapeamento de rotação de culturas (SEIFERT et al., 2017; COHEN et al., 2019).

Em tempos onde busca-se a sustentabilidade da agricultura, o mapeamento da rotação de culturas agrícolas busca identificar sistemas agrícolas mais sustentáveis (WALDHOFF et al., 2017), bem como a possibilidade de adoção de novos cultivos agrícolas, assim como a intensificação da agricultura das regiões (BELTRAN-PEÑA et al., 2020).

Entre os principais fatores que contribuíram para a agricultura digital e novas gerações de conhecimento, e produtos são o sensoriamento remoto por satélite, digitalização de dados da propriedade, dados de solo, clima, muitas vezes utilizados na pesquisa para geração de

novos produtos, modelos de crescimento e desenvolvimento de culturas agrícolas agregando em geração de dados de simulações para incorporar informação para modelos agrícolas, bem como plataformas de processamento de dados em nuvem, novos algoritmos, entre outros (SAIZ-RUBIO & ROVIRA-MÁS, 2020).

1.1 REFERENCIAL TEÓRICO

1.1.1 Dados para agricultura – *Big Data*

Dados são registros de atributos de um ente, objeto ou fenômeno que quando agregados geram informação e conhecimento para os sistemas que o pertence. Para a era atual da agricultura, dados vêm sendo importante para geração de novos conhecimentos, tecnologias agregando produtividade, economia e sustentabilidade para a atividade agrícola (WOODARD et al., 2018; SAIZ-RUBIO & ROVIRA-MÁS, 2020).

Os dados na agricultura são, portanto, uma estratégia de valor para os sistemas de produção. A quantidade, velocidade e variedade cada vez maior de dados disponíveis, o que caracteriza o termo *big data*, para gerenciamento de campo torna necessária a implementação de algum tipo de processo automático para extrair informações operacionais de dados em massa.

De acordo com Laney (2001) o termo *big data* possui três dimensões (“3Vs”): volume, velocidade e variedade. Kunisch (2016) adicionou um quarto “V” para veracidade. Finalmente, um quinto “V” foi adicionado por Chi et al. (2016) para a valorização dos dados. Outro “V” relevante poderia ser para visualização (KARMAS et al., 2016). No geral, os “Vs” (dimensões) de *big data* representam:

- Volume refere-se para tamanho de banco de dados que ultrapassa a capacidade típica de *softwares* de bancos de dados tradicionais para capturar, armazenar, manejar e analisar as informações;
- Velocidade refere-se à capacidade de adquirir, entender e interpretar os dados momentaneamente. Também se refere ao tempo que estes dados possuem valor, sem a obsolescência dos mesmos;
- Variedade refere-se a diferentes formas de dados e da diversidade de complexidade;
- Veracidade refere-se à qualidade, realidade, e confidenciabilidade dos dados;
- Valor refere-se à habilidade de propagar conhecimento, aplicação e inovação;

- Visualização refere-se à necessidade de estruturação complexa das informações para melhor entendimento.

Embora essas dimensões descrevem o termo *big data*, a análise de *big data* não precisa satisfazer todas as dimensões (RODRIGUEZ et al., 2017).

1.1.2 Internet das coisas – *Internet of Things*

O aumento contínuo no volume e detalhes dos dados capturados seja por produtores, organizações, sensores, entre outros, produziram um fluxo esmagador de dados em formato estruturado ou não estruturado impulsionado pela internet das coisas – *Internet of Things* (IoT).

A IoT em um contexto agrícola refere-se ao uso de sensores e outros dispositivos para transformar cada elemento e ação envolvida na agricultura em dados. A IoT impulsiona a digitalização da agricultura (TZOUNIS et al., 2017). De fato, as tecnologias IoT são uma das razões pelas quais a agricultura pode gerar uma grande quantidade de informações valiosas, e espera-se que o setor agrícola seja altamente beneficiado pelos avanços dessas tecnologias (TZOUNIS et al., 2017). Atualmente dados de sensoriamento remoto são a grande quantidade de dados gerados por sensores acoplados em satélites com diferentes resoluções espectrais, temporais e espaciais.

Estima-se que, com novas técnicas, a IoT tem potencial para aumentar a produtividade agrícola em 70% até 2050 (MYKLEVY et al., 2016), o que é positivo, pois segundo Myklevy et al. (2016), o mundo precisa aumentar a produção global de alimentos em 60% em 2050 devido a um crescimento populacional de mais de nove bilhões.

1.1.3 Computação em nuvem – *Cloud Computing*

A computação em nuvem – *cloud computing* é uma tecnologia poderosa para realizar computação de dados de forma complexa e em grande escala. A computação em nuvem elimina a necessidade de manter *hardware* de computação caro, espaço dedicado e *software* na organização - *on premise*. Foi observado um crescimento maciço na escala de dados ou *big data* gerados por meio da computação em nuvem (HASHEM et al., 2015). Lidar com *big data* é uma tarefa desafiadora e demorada que requer uma grande infraestrutura computacional para garantir o processamento e a análise de dados bem-sucedidos. Algumas ferramentas como Google Earth Engine (GEE) possibilita o acesso, processamento e análise em nuvem.

O GEE consiste em um catálogo de dados, muitos de sensoriamento remoto pronto para análise de vários petabytes co-localizado com um serviço de computação intrinsecamente paralelo de alto desempenho. Ele é acessado e controlado por meio de uma interface de programação de aplicativo - *Application Programming Interface* (API) acessível pela internet e um ambiente de desenvolvimento interativo - *Integrated Development Environment* (IDE), que permite o processamento e visualização dos resultados (GORELICK et al., 2017).

1.1.4 Agricultura de precisão – *Precision Agriculture*

A agricultura orientada a dados usa técnicas de *big data* para tomar melhores decisões usando os dados agrícolas para gerar recomendações de utilização de insumos no local, tempo e quantidades corretas, agregando para a filosofia da agricultura de precisão.

A agricultura de precisão tornou-se a terceira onda da revolução da agricultura moderna, a primeira foi a mecanização e a segunda a revolução verde com sua modificação genética (ZHANG, 2019), e atualmente, está sendo aprimorado com o aumento dos sistemas de conhecimento agrícola devido à disponibilidade de maiores quantidades de dados (SAIZ-RUBIO & ROVIRA-MÁS, 2020).

Hoje em dia, as práticas agrícolas estão sendo apoiadas por biotecnologia (Rahman et al., 2013), manejo de sistemas agrícolas baseado em dados (CORASSA et al., 2018; POTT et al., 2019; BASTOS et al., 2020; SCHWALBERT et al., 2020), e tecnologias digitais emergentes, como sensoriamento remoto (LOBELL et al., 2019), computação em nuvem – *cloud computing* (HASHEM et al., 2015) e IoT (WEBER & WEBER, 2010), levando à noção de "agricultura inteligente" (KAMILARIS et al., 2017).

A implantação de novas tecnologias de informação, comunicação e digitalização para cultivo em nível de campo e manejo agrícola estende o conceito de agricultura de precisão (LOKERS et al., 2016), aprimorando as tarefas existentes de manejo e tomada de decisão da agricultura (KAMILARIS et al., 2017).

1.1.5 Agricultura digital – *Digital Agriculture*

A nova filosofia centrada na aquisição, tratamento e análise dos dados agrícolas, unindo a telemática e gestão combinando ao conceito de agricultura de precisão é expressa pela agricultura 4.0, uma clara referência à indústria 4.0, também podendo ser chamada de agricultura inteligente, ou agricultura digital (SAIZ-RUBIO & ROVIRA-MÁS, 2020).

Tradicionalmente, os agricultores vão aos campos para verificar a situação de suas safras e tomar decisões com base em sua experiência acumulada. Esta abordagem pode ser onerosa em função do tamanho da área, vinculação com a atividade agrícola, entre outras razões, podendo afetar a eficiência, sustentabilidade e produtividade do sistema agrícola.

Como resultado, a agricultura digital é baseada nos princípios da agricultura de precisão com produtores utilizando sistemas e plataformas que geram dados em suas fazendas, que serão processados de forma a tomar decisões estratégicas e operacionais adequadas.

Os sistemas de gerenciamento avançados dentro do contexto da agricultura digital estão fornecendo soluções práticas. Além disso, apesar de alguns agricultores terem uma longa experiência adquirida após muitos anos de trabalho no campo, a tecnologia pode fornecer uma ferramenta sistemática para detectar problemas imprevistos difíceis de perceber por inspeção visual em verificações ocasionais (SAIZ-RUBIO & ROVIRA-MÁS, 2020).

A agricultura digital utiliza-se de *big data*, IoT, *cloud computing*, métodos e soluções analíticas – *data science* para construir sistemas de suporte à tomada de decisões de manejo. Além disso, contribui para elevar os índices de eficiência, produtividade, racionalização do uso de insumos, da redução de custos com mão de obra, melhorar a qualidade do trabalho e a segurança dos trabalhadores, e sustentabilidade do meio ambiente (MASSRUHÁ & LEITE, 2017).

1.1.6 Sensoriamento remoto e dados geoespaciais – *Remote Ssensing and Geospatial Data*

O sensoriamento remoto desempenha um papel fundamental no progresso da agricultura digital, uma vez que qualquer campo agrícola pode ser acessado gerando índices de vegetação que ressaltam diferenças e destacam percepções capazes de gerar informações para alimentar a tomada de decisão – *data-driven*.

Sensoriamento remoto a partir de sensores acoplados em satélites artificiais são a grande maioria dos dados gerados de sensoriamento remoto. Satélites importantes que fornecem informações agrícolas de forma pública são os sensores ópticos como *Moderate Resolution Imaging Spectroradiometer* (MODIS) disponibilizado pela *National Aeronautics and Space Administration* (NASA) com 250 m de resolução espacial e tempo de revisita de 1-2 dias, Landsat disponibilizado pela *United States Geological Survey* (USGS) com 30 m de resolução espacial e tempo de revisita de 16 dias, e Sentinel-2 disponibilizado pela *European*

Space Agency (ESA-Copernicus) com resolução espacial de 10 m e tempo de revisita de 5 dias. Outra importante classe são os radares como Sentinel-1 *Synthetic Aperture Radar* (SAR) disponibilizado pela *European Space Agency* (ESA-Copernicus) com resolução espacial de 10 m e tempo de revisita de 5 dias.

Também há vários produtos originados de sensoriamento remoto, combinando com dados levantados no mundo inteiro formando dados geoespaciais – *geospatial data* onde são gerados os modelos e disponibilização das camadas de dados – *layers*. Produtos de informações de vegetação, uso da terra, terreno, clima como precipitação, temperatura, entre outras são gerados e disponibilizados em plataformas de *cloud computing* como o GEE.

O modelo de elevação digital da *Shuttle Radar Topography Mission* (SRTM) é um esforço de pesquisa internacional que obteve modelos de elevação digital em uma escala quase global. É fornecido pela NASA com uma resolução de aproximadamente 30 m (FARR et al. 2007).

Os *layers* da NASA *Prediction of Worldwide Energy Resources* (NASA-POWER) fornece conjuntos de dados solares e meteorológicos da pesquisa da NASA para suporte de energia renovável, eficiência energética de edifícios e necessidades agrícolas. Sua resolução espacial é de 1° de latitude por 1° de longitude para os conjuntos de dados de radiação e ½° de latitude por ⅝° de longitude para os conjuntos de dados meteorológicos. Usualmente esses dados são fornecidos diariamente.

O *layer* Land Surface Temperature (LST), produzido com a composição de 8 dias de imagens termais do satélite Aqua MODIS (produto MYD11A2) tem relação com temperatura do ar e vem sendo utilizado para predições na agricultura como no trabalho de Schwalbert et al. (2020) com predição de produtividade de soja. Outros dados geoespaciais a ser destacado são dados de gride - *grid* de precipitação do *Climate Hazards Group Infrared Precipitation with Stations* (CHIRPS), o qual disponibiliza dados de precipitação agregando informações de satélite e estações meteorológicas com resolução espacial 0.05° de latitude e de 0.05º de longitude.

1.1.7 Modelagem de culturas agrícolas – *Crop Modeling*

Na era da agricultura baseada em dados, a modelagem de culturas agrícolas com intuito de gerar simulações de crescimento e desenvolvimento das culturas tem sido amplamente difundida. Nesse contexto, os modelos de cultura podem auxiliar na avaliação

das melhores estratégias para atingir melhores manejos agrícolas, uma vez que são capazes de evidenciar desempenho agronômico nas melhores datas de semeadura (BATTISTI & SENTELHAS, 2014), os efeitos das mudanças climáticas na produtividade (ASSENG et al., 2013), identificam traços de tolerância à seca (SINCLAIR et al., 2010; BATTISTI & SENTELHAS, 2015), e predição de produtividade (JIN et al., 2019; DEINES et al., 2020).

Ao longo dos anos muitos modelos de crescimento de culturas foram propostos de forma generalizada ou específica para cultura e região, cabe aos interessados modificar e calibrar para utilização no sistema agrícola de interesse. Dentre os modelos mais populares são: *Food and Agriculture Organization of the United Nations* (FAO) - AQUACROP (STEDUTO et al. 2009); *Simulation Model for Nitrogen and Carbon Dynamics in Agro-Ecosystems* (MONICA) (NENDEL et al. 2011); *Decision Support System for Agrotechnology Transfer* (DSSAT) (BOOTE et al. 1998), e *Agricultural Production Systems Simulator* (APSIM) (KEATING et al. 2003), e *Simple Generic Crop Model* (SIMPLE) (ZHAO et al., 2019).

Os modelos de culturas diferem na forma que simulam os processos dinâmicos, como balanço hídrico, variáveis meteorológicas, e parâmetros de crescimento e desenvolvimento (WHITE et al., 2011). O modelo da FAO é um modelo de cultura considerado simples utilizado para avaliar a relação entre o rendimento da cultura e as condições climáticas (DOORENBOS & KASSAM, 1994). Por outro lado, os modelos de cultura mais complexos consideram mais variáveis e parâmetros nos processos de cultura, o que aumenta seus possíveis usos para avaliar crescimento, desenvolvimento e a produtividade da cultura, como os modelos DSSAT e APSIM.

1.1.8 Inteligência artificial – *Artificial Intelligence*

A inteligência artificial é um campo de estudo abrangente, que envolve robótica, técnicas de aprendizado de máquina – *machine learning*, aprendizado profundo – *deep learning*, o qual teve seu início na década de 1940, quando foi descrita a primeira rede neural artificial – *artificial neural network*. O campo de estudo incorpora diversas áreas de estudo, mas principalmente a programação, matemática e estatística.

Machine learning é um ramo da inteligência artificial e da ciência da computação que se concentra no uso de dados e algoritmos para relacionar a maneira como os humanos aprendem, melhorando gradualmente sua precisão. *Machine learning* surgiu junto com

tecnologias de *big data* e computação de alto desempenho para criar novas oportunidades para desvendar, quantificar e compreender processos intensivos de dados em ambientes operacionais. Modelos de *machine learning* amplamente utilizados para descrever fenômenos e análises na agricultura são: regressão linear e logística, agrupamento, modelos de Bayesian, modelos baseados em instância – mais comum *k-nearest neighbor*, árvores de decisão (*decision trees, random forest, bagging, boostig*), *support vector machines* (SVM), e aprendizado por conjunto – *emsemble learning* (LIAKOS et al., 2018). *Deep learning* é uma subárea de *machine learning* que utiliza vários *layers* para extrair e processar informação. Os modelos de *deep learning* são basicamente topologias diferentes de redes neurais as quais baseiam-se no aprendizado conectivo. Diversos estudos vêm ampliando conhecimento agronômico utilizando análises mais robustas de *machine* e *deep learning* para inferir, modelar e predizer processos nos sistemas agrícolas (CORASSA et al., 2018; BASTOS et al., 2020; SCHWALBERT et al., 2020; NIETO et al., 2021).

1.2 PROPOSIÇÃO

Apesar dos importantes avanços na agricultura digital no Brasil, a classificação e mapeamento de culturas agrícolas ainda é incipiente, e, portanto, fazem parte dos objetivos desse trabalho.

A classificação e mapeamento de culturas agrícolas nas diversas escalas necessita de entendimento dos sistemas agrícolas como o ciclo de desenvolvimento dos cultivos, e demais informações sobre outros alvos como área urbana, florestas, corpos de água, banhados, pastagem, etc. Para tanto, dados de campos de produção das principais culturas cultivadas, e dados de outros alvos são necessários para relacionar com variáveis - *features* de sensoriamento remoto como bandas espectrais e índices de vegetação. Para tal, dentro de um modelo de classificação de culturas agrícolas é importante identificar possíveis ruídos para transferir a predição do modelo – *transfer learning*, para diferentes regiões e para os próximos anos, bem como avaliação do modelo através de novos bancos de dados – *dataset*, para validação do mesmo. Os objetivos do primeiro trabalho foram: i) avaliação da variabilidade espacial de dados de campo para gerar o modelo de classificação de culturas agrícolas; ii) avaliar o modelo de *transfer learning* com os dados da estação de cultivo subsequente; iii) avaliar a precisão do modelo de previsão para previsões antecipadas, e iv) desenvolver um

modelo de classificação e mapeamento das culturas agrícolas para o Rio Grande do Sul, Brasil.

Outro trabalho explorado dentro da classificação e mapeamento de culturas agrícolas é a classificação não-supervisionada (WANG et al., 2019). Para tal, pode-se criar uma máscara de campos de cultivos, para o Brasil tem-se a possibilidade de utilização de dados do Cadastro Ambiental Rural (CAR), MapBiomas, ou realizar a própria máscara de campos de cultivos previamente com algoritmos de classificação. Posteriormente, algoritmos como *k-means* tem produzido classificação em grupos, onde a única entrada do modelo é o número de classes a serem geradas (WANG et al., 2019). Também dados de simulações gerados por modelos de culturas agrícolas têm sido utilizados isoladamente ou somadas à dados de campos para geração de bancos de dados com maior variabilidade aumentando a acurácia dos modelos de previsão de produtividade de grãos (JIN et al., 2019; DEINES et al., 2020). No entanto utilização de modelos de desenvolvimento de culturas para geração de dados e utilização de classificação supervisionada ainda não está bem reportado na literatura. Os objetivos do trabalho foram: i) comparar dados gerados através de simulações de desenvolvimento de cultivos agrícolas e dados de campo; ii) avaliar máscaras de campos de produção geradas pelo CAR, MapBiomas e modelo de *random forest*; e iii) avaliar modelos de classificação não-supervisionada, classificação supervisionada com dados de simulações de desenvolvimento de cultivos agrícolas, e classificação supervisionada com dados de campo, bem como a combinação dos mesmos.

Um terceiro trabalho, o qual é ainda pouco reportado na literatura é o mapeamento da rotação de culturas agrícolas em nível de campo para escalas regionais, utilizando de *layers* do mapeamento de cultivos agrícolas de anos anteriores para realizar a inferência sobre monocultivos e rotação de culturas. Aspectos importantes no mapeamento de rotação de culturas é a regionalização desta caracterização em diferentes sistemas agrícolas, bem como os padrões de rotação de culturas ao longo dos anos. Trabalhos com o efeito de rotação de culturas agrícolas e monocultivo utilizando dados de solos, e clima, bem como produtividade destacam o benefício da rotação de culturas (SEIFERT et al., 2017; COHEN et al., 2019). Os objetivos do trabalho foram: i) mapear padrões de monocultivos e rotação de culturas nas diferentes mesoregiões do estado do Rio Grande do Sul, Brasil; ii) identificar variáveis de solo e de clima que coincide com maiores percentuais de área de monocultivo; iii) avaliar o efeito da rotação de culturas na produtividade de grãos das culturas.

1.3 MATERIAIS E MÉTODOS

Os materiais e métodos utilizados nos trabalhos desenvolvidos serão descritos por trabalho de forma sucinta nesta seção. Para mais informações dos materiais e métodos dos trabalhos, consulte as seções de materiais e métodos dos artigos dispostos na sequência.

1.3.1 Materiais e métodos do artigo 1

A área de estudo consistiu no estado do Rio Grande do Sul, Brasil. Foram coletados dados de campos de produção de soja (*Glycine max* (L.) Merr.), milho (*Zea mays* L.) e arroz (*Oryza sativa* L.) de dois diferentes métodos. Primeiramente foi utilizado o banco de dados do Google Street View, o qual possui um compêndio de imagens panorâmicas ao redor do globo reunindo informação de data e respectivas coordenadas geográficas. Também foram coletados dados das culturas agrícolas utilizando um receptor portátil *Global Navigation Satellite System* (GNSS) para coletar informações de safra ao longo de estradas em todo o estado do Rio Grande do Sul. Posteriormente foram agregadas informações de outras classes como florestas, vegetação nativa, outras culturas agrícolas, corpos de água, setor urbano, etc, através de imagens de alta resolução do Google Earth. Ambos bancos de dados foram reunidos para confecção de bases de treinamento e de teste para o modelo de classificação de culturas.

A partir dos dados de campos foram combinados *features* de sensoriamento e remoto e modelo digital de elevação para compor as variáveis preditoras do modelo. Foram utilizadas as bandas espectrais, bem como confecção de índices de vegetação do sensor do satélite Sentinel-2 para o período do ciclo de cultivo das culturas agrícolas (1º de julho à 1º de maio) para as safras agrícolas 2018-2019 e 2019-2020. Como transformação dos dados – *feature engineering*, foi realizada regressões harmônicas já reportadas em estudos anteriores (WANG et al., 2019; DEINES et al., 2020; DADO et al., 2020). Também foram utilizados dados do sensor do satélite Sentinel-1, o qual fornece dados de radar com polarização dupla como VV e VH, para tal foram utilizadas composições mensais de VV e VH de julho à abril de ambos anos de dados. Por fim, foram utilizados dados de terreno como, elevação, declividade e aspecto do modelo digital de elevação SRTM. Fontes de dados do Sentinel-2, Sentinel-1 e SRTM foram combinados para construir o modelo de classificação de cultura na plataforma do GEE.

Foram realizadas rotinas de seleção de variáveis – *feature selection*, como correlação de Spearman para exclusão de variáveis autocorrelacionadas, bem como rotinas de seleção de variávies do modelo *random forest* como *mean decrease accuracy*. O algoritmo utilizado para gerar a classificação de culturas agrícolas foi *random forest* devido seu alto desempenho e disponibilidade de desenvolvimento e hospedagem na plataforma GEE (JIN et al., 2019; WANG et al., 2019; WANG et al., 2020).

Foram analisados a variabilidade espacial do cultivo das culturas agrícolas através de sensoriamento remoto, a possibilidade de predizer a classificação das culturas para subsequentes, previsão antecipada de 1º de maio para 1º de dezembro e por fim foram gerados o mapa de culturas agrícolas de soja, milho e arroz do Rio Grande do Sul, bem como estatísticas de área cultivada. As análises estatísticas foram geradas na linguagem R (R CORE TEAM, 2023) com auxílio da Google Colaboratory (Colab) (BISONG, 2019). O modelo proposto foi implementado na plataforma GEE (GORELICK et al., 2017).

1.3.2 Materiais e métodos do artigo 2

A área de estudo selecionada foi a mesorregião Noroeste do estado do Rio Grande do Sul, Brasil, o qual possui em sua maior área agricultável de verão as culturas de soja e milho do estado, sendo as culturas que foram utilizadas no estudo. Foram utilizados dados de campo coletados de duas diferentes formas. A primeira com a utilização do catálogo de imagens do Google Street View, armazenando informação georreferenciada deslocada para o campo em questão, ao longo das vias rodoviárias de seu catálogo com sua respectiva cultura agrícola (soja ou milho). A segunda forma de coleta de dados foi realizada com GNSS portátil através de marcação de pontos com os deslocamentos através de rodovias da área do estudo. Posteriormente, todas as informações foram organizadas para a criação do *dataset* para confecção dos modelos de treinamento e teste de classificação de culturas.

Imagens de satélites Sentinel-2 Level 1-C (10m resolução) de júlio a março (correspondente à estação de crescimento das culturas de soja e milho) foram utilizadas com intuito da extração de regressões harmônicas como estratégia de *feature engineeering* e utilização destas *features* como input dos modelos de *machine learning*. Como máscaras de cultivos agrícolas - *agricultural masks* foram utilizados três diferentes recursos. O primeiro com a confecção de um modelo de *random forest* com *features* das regressões harmônicas de culturas agrícolas e de outras classes (florestas, água, vegetação nativa, região urbana, etc). O

segundo recurso foi utilizando a camada de máscara de áreas agricultáveis do MapBiomas Brasil (MAPBIOMAS, 2023) o qual possui informações anuais de mapeamento de cobertura e uso da terra. O terceiro recurso foi a utilização do Cadastro Ambiental Rural (CAR) (SICAR, 2023) onde possui dados de áreas consolidadas com agricultura delimitado de forma georreferenciada, digital e pública.

Com intuito de testar a utilização de simulações de crescimento de cultivos agrícolas como dados para associar com dados de satélite, modelos de crescimento de culturas do *framework* APSIM Next Generation (APSIM-NG) (HOLZWORTH et al., 2018) foram utilizados com *inputs* de clima de NASA/POWER (SPARKS et al., 2018), solo de SoilGrids (HENGL et al., 2017), e de manejos da CONAB (2020) para serem geradas simulações de crescimento de soja e milho para municípios da região, obtendo como *output* valores de índice de área foliar os quais foram associados com equações empíricas para converte-los em índice de vegetação e poder utilizar a escala regional das imagens de satélite para classificação e mapeamento das culturas.

Modelos de *machine learning* como de *supervised* e *unsupervised learning* com os algoritmos de *random forest* e *k-means*, respectivamente foram utilizados como modelos de classificação e agrupamento para comparar número de anos, número de dados necessários, e da espacialização necessária dos dados para atingir acurácias aceitáveis nos modelos bem como associações de dados de campo, dados de simulações e agregações de ambos para compor um modelo mais robusto de classificação e mapeamento de culturas agrícolas foram testadas. Ademais, foram comparadas as *features* de dados de campo com dados simulados, e comparação das performances das máscaras de cultivos agrícolas. As análises estatísticas foram geradas com a linguagem R (R CORE TEAM, 2023). O modelo proposto foi implementado na plataforma GEE (GORELICK et al., 2017).

1.3.3 Materiais e métodos do artigo 3

Á área de estudo compreende toda a extensão do estado do Rio Grande do Sul na região Sul do Brasil. Rio Grande do Sul se destaca em culturas de grãos como a soja, arroz e milho, com ~5.7Mha, ~1 Mha, e ~0.76 Mha como média de área cultivada das últimas safras agrícolas, respectivamente (IBGE 2022). Devido à grande extensão de área com distintas características climáticas, de solo e de manejo agronômico, mesorregiões do estado foram consideradas para realizar o estudo de rotação de culturas agrícolas de verão. As mesorregiões

são: Noroeste Rio-Grandense, Nordeste Rio-Grandense, Centro Ocidental Rio-Grandense, Centro Oriental Rio-Grandense, Região Metropolitana de Porto Alegre, Sudoeste Rio-Grandense, e Sudeste Rio-Grandense.

Modelos de classificação e mapeamento das culturas agrícolas de estudo anteriores de POTT et al. (2021) foram utilizados como identificação da cultura cultivada das safras agrícolas de 2017-2018 à 2020-2021 utilizando dados geoespaciais de áreas consolidadas do banco de dados do Cadastro Ambiental Rural (CAR) (SICAR, 2023) para obter a área de cada campo, com sua respectiva cultura agrícola. POTT et al., (2021) desenvolveu modelo de classificação e mapeamento de culturas agrícolas para o estado do Rio Grande do Sul utilizando imagens de satélite de Sentinel-2, Sentinel-1 e dados de elevação digital de *Shuttle Radar Topographic Mission* (SRTM), para extrair *features* como *input* de modelo de *random forest*, obtendo acurácia de 0.95.

Os dados de cultura cultivada temporal (2017-2018 à 2020-2021) associada às áreas agrícolas foram utilizadas para compor o banco de dados de rotação de culturas. Classes de rotação de culturas como: sem rotação, um ano de rotação, dois anos de rotação, e rotação de culturas plena foram geradas para classificar o comportamento de rotação para cada área de campo considerando a espacialização de mesorregiões do estado obtendo características de rotação de culturas de verão espacialmente.

Dados de solo e de clima de catálgos geospaciais foram relacionados com os dados de rotação de culturas com intuito de caracterizar as condições edafoclimáticas com a utilização ou não de rotação de culturas. Como dados de solos foram utilizados os atributos: argila, areia, silte, água disponível no solo, densidade do solo, carbono orgânico do solo, pH, e capacidade de troca de cátions do SoilGrids (HENGL et al., 2017). No entanto os dados climáticos como produto altamente relacionado com temperatura do ar foram acessados do produto MYD11A2 do satélite MODIS Aqua (ASTER, 2018). Para precipitação foi utilizado dados de *Climate Hazards Group Infrared Precipitation with Stations* (CHIRPS) (FUNK et al., 2015). Todos esses dados foram sumarizados à nível de município para relacionar com dados de rotação de culturas à nível municipal.

Finalmente, o banco de dados de rotação de culturas relacionado com dados ambientais de clima e solo foram utilizados juntamente com dados de produtividade das culturas agrícolas à nível de município (IBGE 2022) para estimar através de regressão linear o efeito da utilização de rotação de cultura na produtividade de grãos à nível de estado e de

regiões pré-definidas com base em características de solo e clima, objetivando obter dados do benefício da utilização da prática de manejo de rotação de culturas em larga escala.

2 ARTIGO 1 - SATELLITE-BASED DATA FUSION CROP TYPE CLASSIFICATION AND MAPPING IN RIO GRANDE DO SUL, BRAZIL

Artigo originalmente publicado e seguindo as normas da Revista ISPRS Journal of Photogrammetry and Remote Sensing.

Pott, L. P., Amado, T. J. C., Schwalbert, R. A., Corassa, G. M., & Ciampitti, I. A. (2021). Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing*, *176*, 196-210.

ABSTRACT

Field-scale crop monitoring is essential for agricultural management and policy making for food security and sustainability. Automating crop classification process while elaborating a workflow is a key step for reliable and precise crop mapping. This study aims to develop an approach for crop classification in the state of Rio Grande do Sul, Brazil, following the specific goals of i) evaluating spatial satellite-based features to guide crop data collection; ii) testing transfer learning model with subsequent growing season data; iii) examining accuracy in early-season prediction model; and lastly, iv) developing a crop classification model for estimating large scale crop area. As main data inputs, Sentinel-2, Sentinel-1, and Shuttle Radar Topographic Mission (SRTM) Digital Elevation data were used to extract features to input in the Random Forest classifier. Spatial variability of satellite features was evaluated using Moran's I Index and cluster k-means. Crop area prediction data were obtained at municipality level to compare with census data (standard method). A crop summer map layer was generated for three major crops: soybeans (*Glycine max* L.), corn (*Zea mays* L.), and rice (*Oryza sativa* L.) in the state of Rio Grande do Sul, Brazil. The crop classification model achieved an overall accuracy of 0.95. Model performance was influenced by sample size and spatial variability of the samples. The random forest model was transferred to the next growing season with 0.89 and 0.91 overall accuracy for 250 and 750 samples, respectively. However, overall accuracy increased from 0.93 to 0.95 when 50 to 250 samples of same-year data was aggregated to the model. Similar accuracy was obtained for predictions done with data until March relative to when the entire season was considered, until May. When data for

more growing seasons were aggregated, the model produced more accurate early season predictions (January and February). Soybean prediction area obtained the highest performance (R^2 = 0.94), relative to rice (R^2 = 0.90) and corn (R^2 = 0.37). The rice prediction area presented a high precision, but the crop area was overestimated due to errors with wetland target relative to other class. Lastly, this study presents the first crop map layer of the three major field crops for the state of Rio Grande do Sul, Brazil, serving as a foundation for the creation of crop type maps for other states in the country and around the globe.

Keywords: Agricultural monitoring, Remote sensing, Machine learning, Sentinel-1, Sentinel-2, SRTM Digital Elevation.

INTRODUCTION

Crop monitoring of agricultural systems can provide information for policymakers related to food production and sustainability, identify research priorities, and evaluate commodity markets (Van Ittersum et al., 2013; Carletto et al., 2015a). The first step to monitor agricultural crops is obtaining crop mapping information at the field-scale with high accuracy of land crop regions to improve agricultural policies.

Field surveys and census are the standard method for obtaining data about crop area for most agricultural regions (Wang et al., 2019). In Brazil, data collection is achieved by both the Brazilian Institute of Geography and Statistics (IBGE) and the National Supply Company (CONAB). While survey and census provide valuable crop data, limitations related to high cost, spatial inference level, frequency of updates, and subjectivity to bias are main constraints linked to these methods (Carletto et al., 2015b; Gourlay et al., 2017).

In Brazil, crop type classification has been studied using satellite data MODIS with maximum likelihood estimation (Rudorff et al., 2007; Arvor et al., 2011); unsupervised classification, multi-spectral and -temporal visual interpretation (Lemos, 2016); and decision tree classifier (Zhong et al., 2016). Mengue and Fontana (2005) proposed a parallelepiped classifier from MODIS and Shuttle Radar Topographic Mission (SRTM) data. Schultz et al. (2015) utilized random forests with Landsat 8 images. Silva Junior et al. (2020) proposed Perpendicular Crop Enhancement Index (PCEI) as a threshold decision from Sentinel-2, Landsat and MODIS data. Furthermore, the CONAB has been proposing satellite-based crop monitoring for major crops such as soybean (*Glycine max* (L.) Merr.), corn (*Zea mays* L.) and rice (*Oryza sativa* L.) in order to overcome challenges due to the high spatial variability, such

as large range of latitude, terrain, multiple crops and high variability of management of Brazilian production (CONAB, 2020a).

The advent of more accessible satellite data and new statistical methods in machine learning will assist a fast implementation of crop classification approaches to enable agricultural monitoring at high spatio-temporal resolutions (Balzter et al., 2015; Bargiel, 2017; Cai et al., 2018; Jin et al., 2019; Wang et al., 2020; Weiss et al., 2020). The Sentinel-2 and Sentinel-1 satellites offer near-real time images with high spatial (10 to 60 m) and temporal (1 to 5 days) resolution. Recent studies have reported high performance of optical data for crop mapping over different cropping systems and climates (Ozdogan and Gutman, 2008; Belgiu and Csillik, 2018; Wang et al., 2019). Additionally, recent investigations have portrayed benefits of combining data from optical (e.g., Sentinel-2 or Landsat archives) and radar (e.g., SAR as the Sentinel-1) for improving model performance for crop mapping (Abubakar et al., 2010, Inglada et al., 2016; Sonobe et al., 2017; Jin et al., 2019; Orynbaikyzy et al., 2020). Furthermore, features from SRTM aggregated with different input sources have significant relevancy for improving land cover (Balzter et al., 2015; Li et al., 2020) and for in-season crop mapping (Demarez et al., 2019).

Crop classification models need proper classification algorithms and quality ground truth data to train and test models. Data beyond the years included in the model are also required to test spatio-temporal transferability and to foster better in-season predictions. From the many classification techniques, machine learning algorithms and supervised classifications have been globally used. Among those methods, random forest models are highlighted (Demarez et al., 2019; Jin et al., 2019; Wang et al., 2020; Deines et al., 2020). Random forests are an ensemble machine learning method comprised of many decision trees in aggregate (Breiman, 2001), which usually outperform other learning methods (Kayad et al., 2019). From the ground truthing perspective, spatial distribution of crop data should be collected to fairly represent gradients of environment and management for each crop class in the region of study (Waldner et al., 2019; Fowler et al., 2020).

Brazil is the largest country in South America and the fifth largest in the world, presenting a high range of latitude and variability in terrain, soils, crops, and agricultural management. The state of Rio Grande do Sul is a key agricultural producer for the country, but still presents a high spatial variability creating challenges for large-scale crop monitoring. Thus, high quality survey data is essential for developing a stable satellite-based model

transferable to other environments (regions) and growing seasons for crop classification. Identification and proper characterization of crop phenology patterns for various crops, regions, and seasons can lead to successful transfer learning performance in classification models (Cai et al., 2018; Wang et al., 2019; Hao et al., 2020a). Improving near-real time crop mapping prediction will have large impacts in agricultural and socioeconomic decision-making procedures such as crop insurance, supply-chain logistics, and financial market forecasting. For in-season crop classification, instead of using satellite data from the entire season, many studies have evaluated the effect of time series length on crop classification performance (Cai et al., 2018; Hao et al., 2020a; Hao et al., 2020b). Demarez et al. (2019) reported an improvement in random forest classifier for mapping irrigated crops with SRTM data mainly for early season classification.

Thus, the aims of this study were to i) evaluate spatial satellite-based features to guide crop data collection; ii) testing transfer learning model with subsequent growing season data; iii) examine accuracy in early-season prediction model; and lastly, iv) develop a crop classification model for mapping and estimating the crop area for soybean, corn, and rice field crops in Rio Grande do Sul state, Brazil.

MATERIAL AND METHODS

Study area

The study area consisted of Rio Grande do Sul state (27º to 34º S and 50º to 58º W) in Southern Brazil (Fig. 1A). Rio Grande do Sul state has approximately 281.707 km² with 497 municipalities (IBGE, 2019). According to Alvares et al. (2013), the dominant climate types are Humid subtropical without dry season with -hot summer (Cfa), and with -temperate summer (Cfb). Rio Grande do Sul has a wide variety of soil types with predominance of Oxisols, Alfisols, Mollisols, Ultisols, Entisols and Inceptisols (Santos et al., 2018; Soil Survey Staff, 2014). The state is ranked first nationally in irrigated rice production, third in soybean, and sixth in corn. The projection for soybean cropping area is 6.06 Mha with a production of 19.8 Mt, irrigated rice area is 0.969 Mha with a production of 7.6 Mt, the corn area is 0.808 Mha with a production of 3.6 Mt in the 2020-21 growing season (CONAB, 2021). These three grain crops comprise more than 81% of the crop area in the state (IBGE, 2020). Soybean is planted from October to December and harvested from February to May (CONAB, 2020b). Corn is planted in a wide window from August to January, and harvested from January to July

(CONAB, 2020b). Rice, concentrated in the southern and central regions of the state, is planted from September to January and harvested from February to May (CONAB, 2020b).

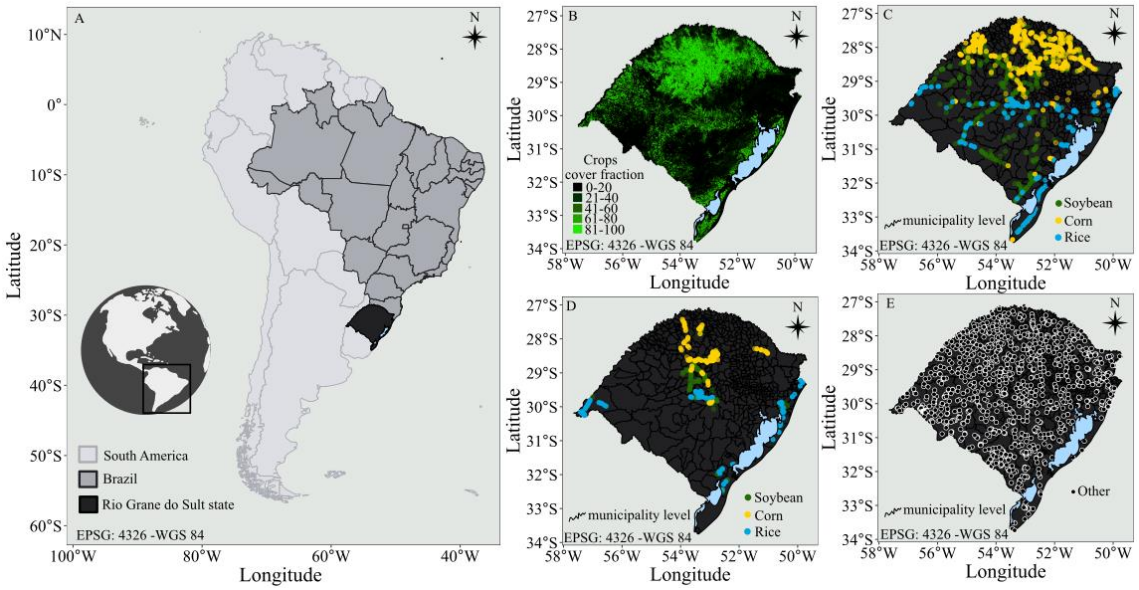


Fig. 1. (A) South America map, Brazil (shaded in grey), and Rio Grande do Sul state (shaded in black). (B) Copernicus Global Land Cover Layers – crops cover fraction 2019 (Buchhorn et al., 2020). (C) Crop type data points accessed by Google Street View for 2018-2019 growing season. (D) Crop type data points collected manually utilizing GNSS system for 2019-2020 growing season. (E) Other classes (pasture/native vegetation, forest, urban, water, and other crops data points) retrieved from Google Street View and Google Earth Engine.

Data acquisition

The ground truth crop data were collected following two approaches. The first, via Google Street View, presents a compendium of panoramic images around the globe gathering geographic coordinates and time of the collected data. Google Street View has been recently used in several scientific studies (Ringland et al., 2019; Yan and Ryu, 2021). Yan and Ryu (2021) developed an automatic ground truthing via Google Street View images utilizing convolutional neural network model for ground referencing and crop mapping. Data from Google Street View has been manually searched specifying latitude and longitude coordinates at intervals of around 500–1000 m considering crop cover fractions (Fig. 1B) along roads in the state of Rio Grande do Sul. The Google Street View database in Rio Grande do Sul

contains images from October 2018 to April 2019. Imagery including one of the three crops (soybean, corn, or rice), geographic coordinates, and time were retrieved from this database (Fig. 1C). The second collection method utilized a Global Navigation Satellite System (GNSS) portable receiver, GPSmap 62®, to gather crop information along roads throughout the state of Rio Grande do Sul for 2019-2020 growing season (Fig. 1D). The dataset for the 2018-2019 season contained 1,000 data points (Google Street View) for each crop, while the dataset for 2019-2020 season (field survey-based) generated 500 data points for each crop.

Other class were collected by visually assessing high-resolution imagery from Google Earth, including water, urban, forest, pasture, and native vegetation (Fig. 1E). Species of native vegetation for the Southern Brazilian grasslands are commonly *Andropogon* spp., *Axonopus* spp., *Aristida* spp., and *Paspalum* spp. (Hasenack et al., 2010; Andrade et al., 2019). Other summer crops such as horticulture crops, tobacco (*Nicotiana tabacum* L.), potato (*Solanum tuberosum* L.), yerba mate (*Ilex paraguariensis* A. St. -Hil.), grape (*Vitis* spp.), apple (*Malus domestica* Borkh.), walnut (*Carya illinoinensis* (Wangenh) K. Koch), olive (*Olea europaea* L.) and orange (*Citrus* spp.) were covered assessing Google Street View and Google Earth and belong to other class.

Remote sensing data and feature engineering

Optical, radar satellite, and digital elevation data sources were combined to build a crop classification model using Google Earth Engine (GEE) (Fig. 2 I).

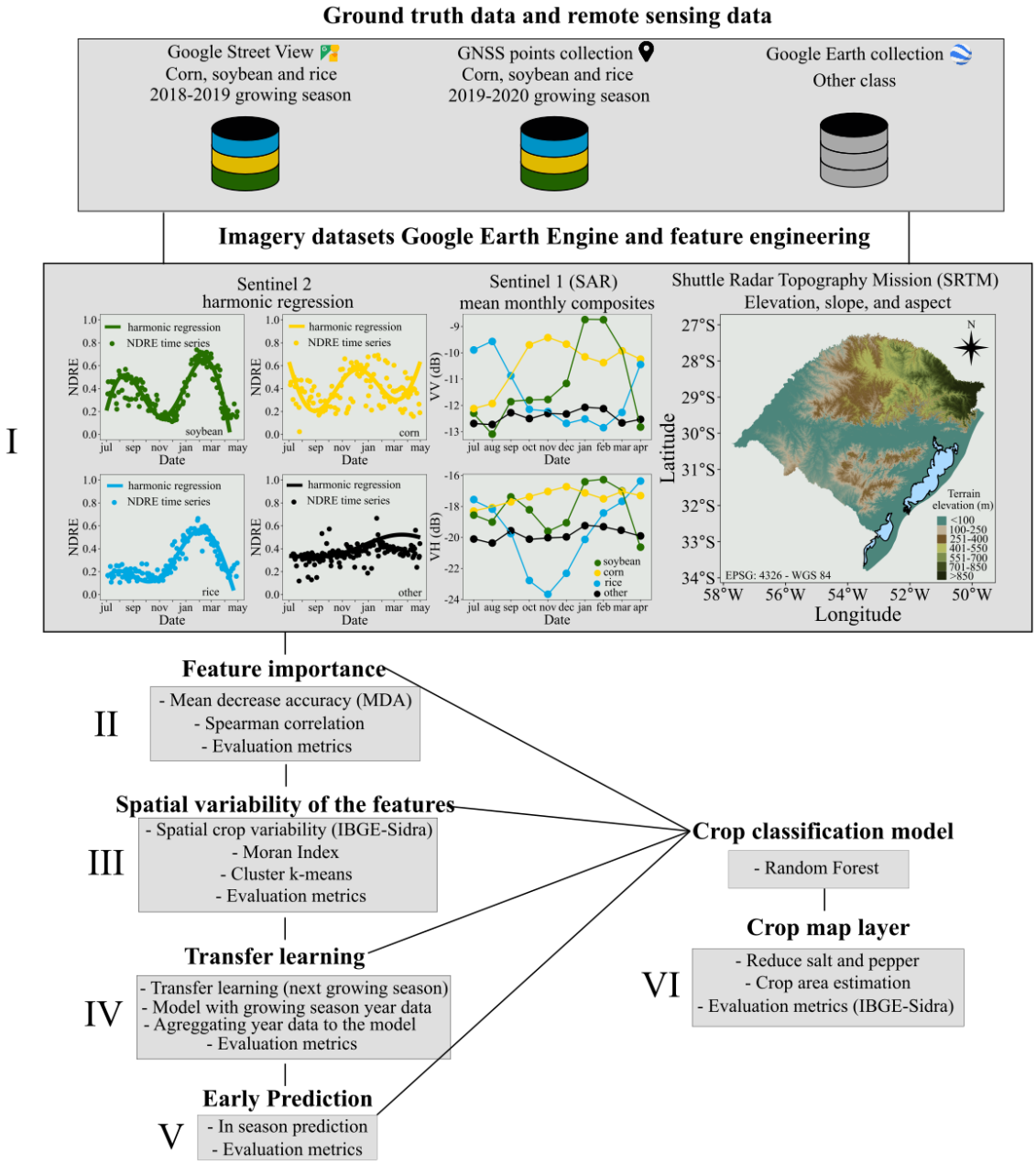


Fig. 2. Framework for model development. Data acquisition – crop type points based on Google Street View for 2018-2019 growing season; crop type points collected utilizing GNSS manually for 2019-2020 growing season; other class collected by visually assessing high-resolution imagery from Google Earth collection. I- Imagery datasets and feature engineering – Google Earth Engine - Sentinel-2 (harmonic regression), Sentinel-1 SAR (mean monthly composition), and SRTM Digital Elevation (elevation, slope, aspect). II- Feature importance of the model. III- Spatial variability of the features. IV- Transfer learning. V- Early prediction. VI- Crop map layer and crop area estimation.

Sentinel-2

Sentinel-2 satellites carry an optical sensor Multi-Spectral Instrument (MSI), which acquires data with a wide-swath and high-resolution (10 m). Images are gathered every 5 days supporting Copernicus Land Monitoring studies, including vegetation, soil, and water cover. The Sentinel-2 product Level-1C consists of top of atmosphere (TOA) reflectance observations in cartographic geometry. The higher-level product, Level-2A, consists of a process including scene classification and atmospheric corrections applied to Level-1C. Jin et al. (2019) reported Sentinel-2 Level-2A as non-essential for crop classification but important to crop yield estimates. We have selected Sentinel-2 Level-1C due the availability of this data for Rio Grande do Sul state in GEE. Sentinel-2 Level-2A presented a collection of imagery from December 2018 onwards in the region of study, while Sentinel-2 Level-1C has imagery data for the interest period (July 1, 2018 to May 1, 2019, and July 1, 2019 to May 1, 2020). We used the following reflectance bands: blue, green, red, red edge 1, near infrared (NIR), shortwave infrared 1 (SWIR1), and shortwave infrared 2 (SWIR2). In addition, we have derived four vegetation indices (VIs) commonly used for agricultural purposes: NDVI (Tucker, 1979), NDRE (Barnes et al., 2000), GCVI (Gitelson et al., 2005), and EVI2 (Jiang et al., 2008). The time frame for image collection ranged from July 1, 2018 to May 1, 2019, and July 1, 2019 to May 1, 2020 in order to perform harmonic regressions. The imagery data were filtered for less than 20% pixel percentage of clouds, and masked with a Quality Assessment (QA) band provided by European Space Agency (ESA) that flags clouds and cirrus pixels. Furthermore, we used a decision tree as proposed by Hollstein et al. (2016) to detect contaminated pixels, including clouds, cirrus, shadows, and snow. Decision tree classifiers, such as Hollstein's trees, were completely serialized allowing us to generate scalable engineering through GEE (Hollstein et al. 2016; Jin et al., 2019). The harmonic regression was selected as the method to extract features from the image time series to apply machine learning models. The harmonic regression of optical data has been used as a potential feature for crop classification (Jakubauskas et al., 2002; Ghazaryan et al., 2018), with the advantage of being directly deployable in GEE using the built-in linear regression function (Wang et al., 2019).

Sentinel-2 bands and VIs can be represented by the following harmonic regression function:

$$f_i(t) = c_i + \sum_{k=1}^{n} [a_{ik} cos(2\pi\omega kt) + b_{ik} sin\,(2\pi\omega kt)]$$

where a_{ik} are cosine coefficients, b_{ik} are sine coefficients, and c_i is the intercept term. The independent variable t represents the time of the images taken within a year expressed as a fraction between 0 (July 1) and 1 (May 1).

Increasing the number of harmonic terms (n) results in a closer fit to the function, requiring a balance of fit and overfitting in the function. For the ω, it was necessary find the value which best fit the data. Coefficient analyses were performed using the mean-squared error for the three crop classes (soybean, corn and rice) and other classes (Fig 2 II), resulting in the best fit utilizing the same coefficients of Wang et al. (2019). Likewise, the coefficients were utilized in Wang et al. (2020), Deines et al. (2020) and Dado et al. (2020) studies for crop classification in different states of the United Sates. The final regression at each Sentinel-2 point for band or VIs is accordingly

$$f_i(t) = c_i + a_{i1} cos(3\pi t) + b_{i1} sin(3\pi t) + a_{i2} cos(6\pi t) + b_{i2} sin(6\pi t).$$

We have extracted the coefficients c_{i1}, a_{i1}, b_{i1}, a_{i2} and b_{i2} for 7 bands and to 4 VIs, resulting in 55 features for the Sentinel-2 data.

Sentinel-1

The Sentinel-1 mission provides data from a dual-polarization C-band Synthetic Aperture Radar (SAR) instrument. The Sentinel-1 SAR GRD is processed using the Sentinel-1 Toolbox to generate a calibrated, ortho-corrected product with 10 m and 6-day revisit-time resolution available in GEE. We have used the dual polarization (VV and VH) and the ratio (VV/VH), resulting in three backscatter bands. Finally, we have computed mean monthly composition from July to April in both years resulting in 30 features for the SAR data.

Shuttle Radar Topography Mission (SRTM)

The SRTM digital elevation data is an international research effort for digital elevation models on a near-global scale (Farr et al. 2007). For the purpose of this study, we have used the SRTM V3 product provided by NASA JPL at a resolution of 30-m available in GEE. From this dataset, we have derived elevation, slope, and aspect, resulting in three main features from SRTM digital elevation data.

Feature selection

Due to the high number of features, we have performed feature selection based on Spearman correlation and mean decrease accuracy (MDA) of the random forest model to infer the complexity of the model and the tradeoff between accuracy and computational cost (Fig. 2 II). Feature selection based on Spearman correlation and MDA retains the properties of the feature engineering, as opposed to the dimensionality reduction using a transformation algorithm such as the principal component analysis (PCA) (Wang et al., 2019).

Random Forest model

The features were fused to train the random forest models. Random forest is a machine learning method encompassing of many decision trees in aggregate (Breiman, 2001), and offering a high performance for crop classification, with the algorithm available to perform training in GEE platform with a large possibility of executing transfer learning for an entire area or other years as a transfer year learning (Jin et al., 2019; Wang et al., 2019; Wang et al., 2020). To train the random forest model, we used the *randomForest* (Breiman, 2001) package in the R environment. We tuned the parameter to reduce model variance, and the final parameters resulted in 100 trees and variables per split equal to 8, with all other parameter as default. We used 75% of the data to train and 25% to test the model with overall accuracy, out-of-bag (OOB) error and balanced accuracy using *caret* (Kuhn, 2008) package in R. To create the final Rio Grande do Sul state map, we used Google Earth Engine´s *ee.Classifer.smileRandomForest* algorithm.

Data analyses

We have analyzed the features for spatial crop variability, transfer learning, early prediction, and generated the crop map and statistics. Statistical analyses were completed in R language (R Core Team, 2020) in Google Colaboratory (Colab) (Bisong, 2019). The proposed model was implemented in GEE platform (Gorelick et al., 2017).

Spatial variability of crops features

Spatial variability of the features for each crop throughout the state of Rio Grande do Sul state were evaluated due to the high variability in soils and climate, in addition to the large range of sowing dates and farming management (Fig. 2 III).

First, we have assessed the selected features derived from Sentinel-2, Sentinel-1, and SRTM with geographic coordinates to perform Moran's I index analyses to verify the spatial dependence of each crop. A good spatial balance when collecting training data is important for classification of remotely-sensed crops, especially when no prior information is available about spatial pattern in crop growth (Fowler et al., 2020). In our study, we have analyzed the features separately for each crop, using the Moran's I index.

Finally, after the Moran's I index analyses we have performed the cluster k-means algorithm. The k-means clustering algorithm has been used in unsupervised crop classification (Wang et al. 2019; Xu et al., 2019). In our study, we have utilized k-means to associate satellite features in cluster regions within the state. As the features presented different magnitude values, we have normalized them on same scale. The number of clusters was defined according to the gap statistic. Gap statistic compares the total within intra-cluster variation for different values of k with their expected values under null reference distribution of the data (Tibshirani et al., 2001).

Sample size and transfer learning

In this study, we have evaluated various sample allocations in the 2018-2019 growing season, testing 50 to 750 samples to train the model while utilizing 250 samples to test the model. Additionally, we have tested the transfer learning approach to evaluate the model for prediction of the subsequent growing season (2019-2020) with 250 and 750 samples from 2018-2019 growing season for training and 250 samples from 2019-2020 growing season for validation. Furthermore, we aggregated data from the two growing seasons to train the model, validating with 2019-2020 growing season (Fig. 2 IV). More details about sample size and transfer learning are presented in Table 1.

Table 1. Summary of sample sizes utilized to train the model and transfer learning. All the models evaluated were validated with 250 samples.

Model	Trained samples
Trained model, 2018-2019	50; 150; 250; 350; 450; 550; 650; 750
Transfer year learning	250; 750
Year model, 2019-2020	50; 100; 150; 200; 250

Aggregating year data to trained model	700_50; 650_100; 600_150; 550_200; 500_250

Early prediction

In-season crop classification was evaluated by modifying the date for Sentinel-2 and Sentinel-1 (Fig. 2 V). The complete model has been set utilizing dates ranging from July 1 to May 1. For early predictions, we have trained and tested the model for the period ranging from July 1 until i) April 1 (herein termed as 'April'), ii) March 1 (termed as 'March'), iii) February 1 (termed as 'February'), and iv) January 1 (termed as 'January').

Crop map and statistics

Before producing the final crop map, we filtered a majority vote of size 5×5 pixels to reduce the salt and pepper effect. Pixels outside the mask were left unchanged, and using a modal filter results in an improvement of overall accuracy (Booth and Oldfield, 1989; Waldner et al., 2015; Belgiu and Csillik, 2017).

Rio Grande do Sul crop maps for 2017-2018, 2018-2019, and 2019-2020 growing seasons were created, and the area composed by each crop was extracted at a municipality-level. To evaluate the area prediction performance, we have compared the predicted crop area for 497 municipalities from IBGE estimates (IBGE, 2020) for each season (Fig. 2 VI).

The following statistical parameters were used to evaluate the prediction crop area model: coefficient of determination (R^2), mean absolute percentage error (MAPE), and mean absolute error (MAE). The R^2 examines how well differences in one variable can be explained by the difference in a second variable. MAPE measures the accuracy as a percentage, and can be calculated as the average absolute percent error minus actual values, divided by actual values. The mean absolute error (MAE) represents the average magnitude of the errors. As MAPE has the significant disadvantage that it produces undefined values for zero or close-to-zero actual values, we considered municipalities with more than 0.02% of Rio Grande do Sul crop area in the census to avoid affecting the statistic evaluation.

RESULTS

Feature selection

Based on the combination of Spearman correlation and MDA of the variables, we have selected and discarded features when comparing the model performance (Fig. A1 and A2). The discarded features included the blue and green bands, EVI2 and NDVI VIs; mean monthly composition of the ratio (VV/VH) from July to April; and aspect. The selected features were the red, NIR, SWIR1, and SWIR2 bands, NDRE and GCVI VIs; mean monthly composition of VV and VH from July to April; elevation, and slope. The random forest model with all features obtained an OOB error of 6.61%, and overall accuracy of 0.93, while the model for selected features presented an OOB error of 5.63% and overall accuracy of 0.95.

The models utilizing only selected features from Sentinel-2, Sentinel-1, and SRTM obtained overall accuracy of 0.92, 0.88, and 0.56, respectively. The models utilizing two data sources with Sentinel-2 + Sentinel-1, Sentinel-2 + SRTM, and Sentinel-1 + SRTM presented overall accuracy of 0.94, 0.93, 0.90, respectively; while the data fusion merging the three data sources achieved an overall accuracy of 0.95 (Fig. 3).

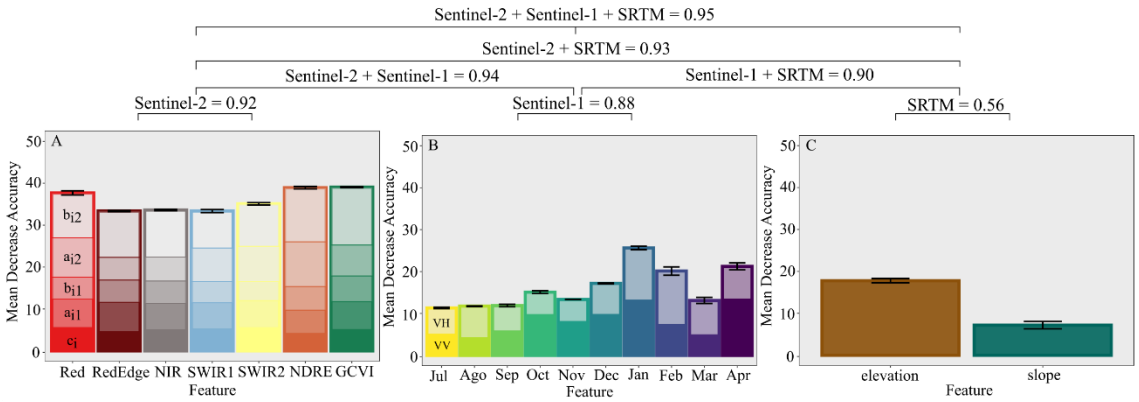


Fig. 3. Mean Decrease Accuracy of features from Sentinel-2 (A), Sentinel-1 (B), and SRTM Digital Elevation (C). c_i, a_{i1}, b_{i1}, a_{i2}, b_{i2} are the harmonic coefficients extracted from the Sentinel-2 time-series. VV and VH are the monthly mean features from Sentinel-1. Elevation and slope from SRTM Digital Elevation. Random forests overall accuracy of the feature's sources (Sentinel-2, Sentinel-1 and SRTM Digital Elevation) and data fusion were presented in the upper with the horizontal lines.

Spatial variability of crops features

The collected data points k-means cluster reflected the spatial variability of each cultivated crop area (Fig. 4 A, B, C). The selected features from Sentinel-2, Sentinel-1, and

SRTM digital elevation displayed spatial dependence, mainly for corn and rice, with minimal spatial dependence for soybeans, with Moran's I index equal to 0.37, 0.34, and 0.12, respectively.

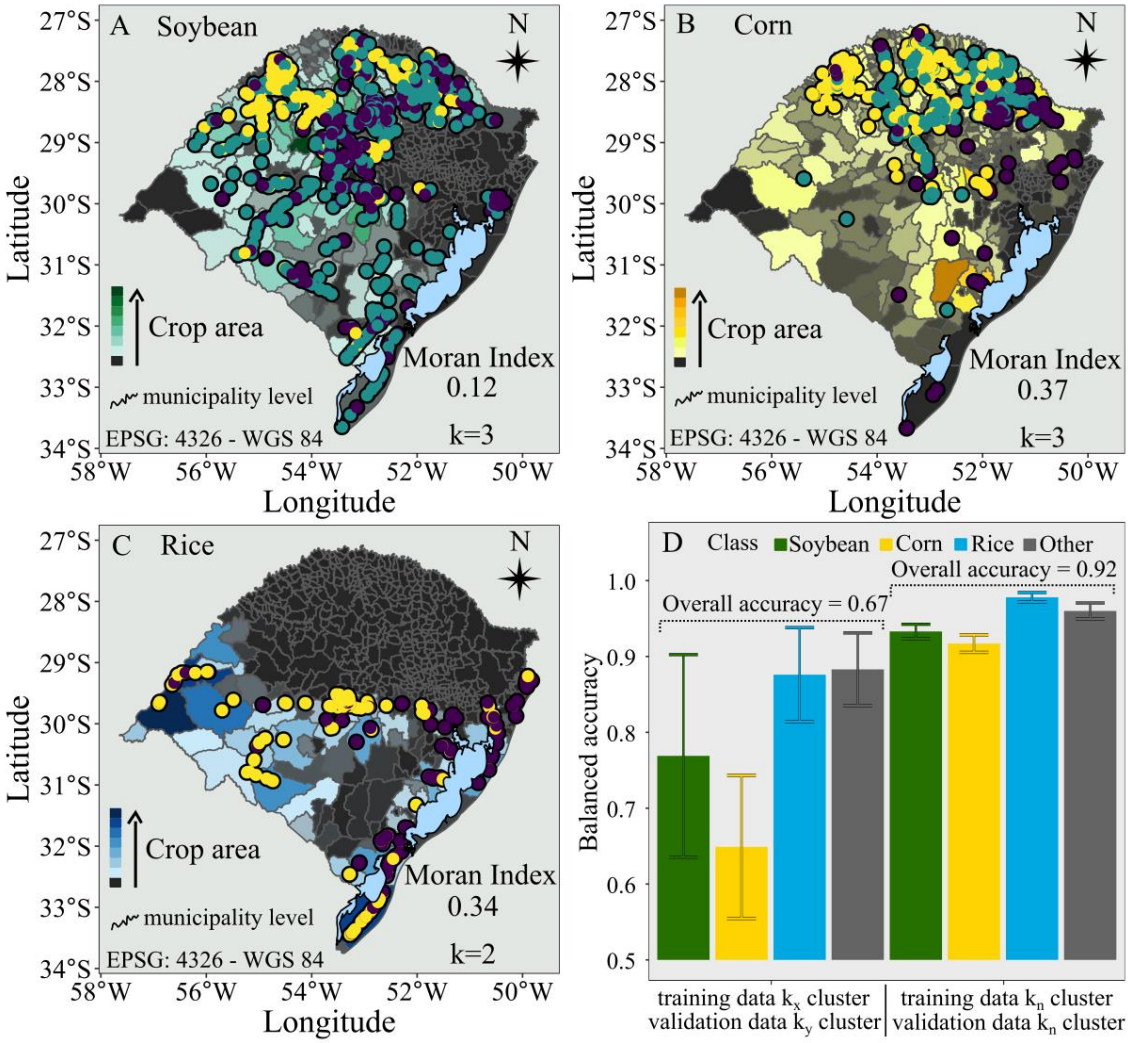


Fig. 4. (A) Spatial variabilities of the features of the model (Moran's Index and Cluster k-means analyses for soybean; (B) corn; (C) rice. Circles with different colors in the maps refer to the clusters of the features from Sentinel-2, Sentinel-1, and SRTM data. Background colors in the maps refer to cultivated area for each crop, mean from 2000-2001 to 2019-2020 growing seasons, IBGE statistics. (D) Balanced accuracy for each class regarding different clusters sampling data, model utilizing k_x training data and k_y validation data, model utilizing k_n training data and k_n validation data. Both training and validation with 100 samples for each cluster.

The cluster k-means analysis had k=3 for soybeans and corn, but k=2 for rice (Fig. 4 A, B, C). The clusters denoted the variability of the features from satellite data, representing an indirect spatial variability of soils, climate, and crop management.

A sensitivity analysis of the cluster selection, exploring all possible combinations, reported an overall decrease in accuracy from 0.92 to 0.67 due to the lack of consideration of the spatial variability in the ground truth data from Rio Grande do Sul state. The most affected crop in the classification was corn followed by soybean, with balanced accuracy equal to 0.64 and 0.74, respectively. Rice crop and other classes did not present a significant reduction in accuracy (Fig. 4D).

Sample size and transfer learning

Increases in sample size to train the model improved OOB error from 10.5 to 5.56%, and overall accuracy from 0.87 to 0.94, for 50 samples and 750 samples, respectively. Training the model with 750 samples achieved an accuracy of 0.94, no different from the model with 250 samples for the 2018-2019 growing season (Fig. 5A).

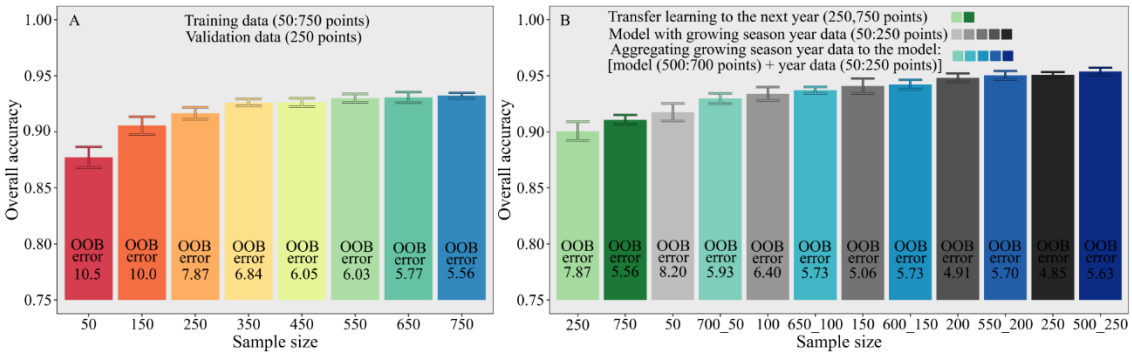


Fig. 5. (A) Comparison of overall accuracy of sample sizes for training the model (50:250 data samples). (B) Comparison of overall accuracy for: subsequent year growing season transfer learning (250 and 750 data samples of the 2018-2019 growing season model - greens), model with same-year growing season data (50:250 data samples - grays), and aggregating same-year growing season data to the model, training with 750 data samples, [model(500:700 data samples) + same-year data (50:250) - blues]. All the models evaluated were validated with 250 samples.

In addition, when the model was transferred to the next growing season with 250 and 750 data samples, respectively, the greater number of samples increased overall accuracy from 0.89 to 0.91. The model with same-year growing season data, even with low samples (50 data samples), presented greater overall accuracy relative to transfer year learning, with a change of 0.92 to 0.94 as the sample year data increased. The OOB error decreased as the sample size increased in the model with same-year growing season data (Fig. 5B).

Although the transfer learning with single year data obtained lower accuracy relative to fewer same-year samples, when more years of data were aggregated to the model with greater variation in climate, soil, and crop management, the overall accuracy increased. The model accuracy aggregating all data for the growing season to the model ranged from 0.93 (with 50 samples for 2019-2020, + 700 samples for 2018-2019 season) to 0.95 (with 250 samples for 2019-2020, + 500 samples for 2018-2019 season) (Fig. 5B).

In-season prediction

Relative to the entire season, early season predictions in crop classification with single growing season data decreased overall accuracy from 0.94 to 0.86, and increased OOB error from 5.56 to 11.8% for May and January, respectively (Fig. 6A). However, predictions with data through March obtained an overall accuracy similar to the May predictions (Fig. 6). Moreover, when the model was built with two growing seasons, 2018-2019 and 2019-2020, March predictions presented the highest overall accuracy relative to the early and late season predictions (Fig. 6B). Balanced accuracy of the classes with single year data had a small penalty for soybean and corn until March, with ~0.95 balanced accuracy from April onwards. However, with two growing seasons of data, corn was able to obtain a high balanced accuracy as early as January, and a balanced accuracy >0.95 was reached for soybean from March onwards (Table A1).

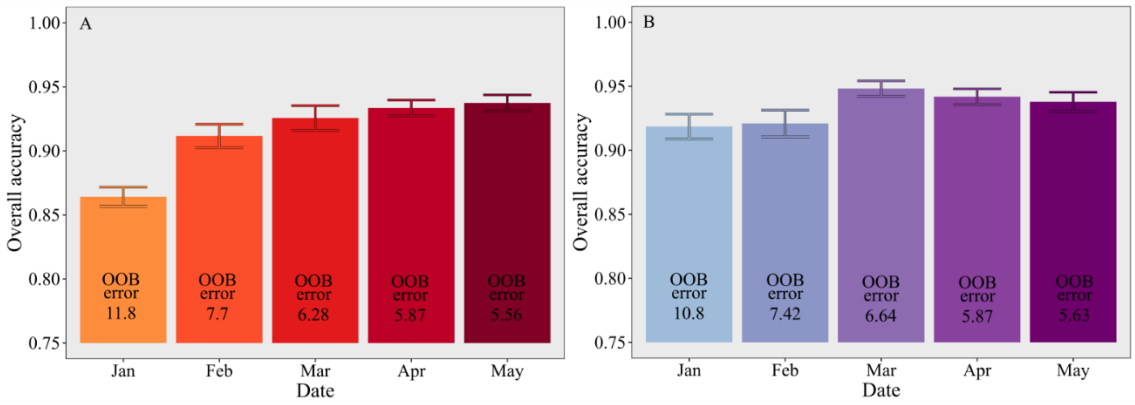


Fig. 6. (A) Comparison of overall accuracy for early crop classification to 2018-2019 growing season (2018-2019 data samples); and (B) 2019-2020 growing season (aggregating same-year data to the 2018-2019 model).

Crop map product and statistics

The statistics of the crop classification model reveal a high performance of soybean crop area predictions (R^2 = 0.94 and MAPE = 0.21) related to IBGE area estimation. The soybean area MAE for the three growing seasons was ~2,046 ha at the municipality level (Fig. 7A). For corn, the model performance achieved was lower relative to soybeans (R^2 = 0.37 and MAPE = 0.81) with a MAE for the three growing seasons of ~1,323 ha (Fig. 7B). While the rice prediction presented high performance (R^2 = 0.90), with an over estimation due to the confounding effect of other class (targets of wetland with rice targets throughout the state), with MAPE = 0.98 and MAE of ~1,702 ha (Fig. 7C).

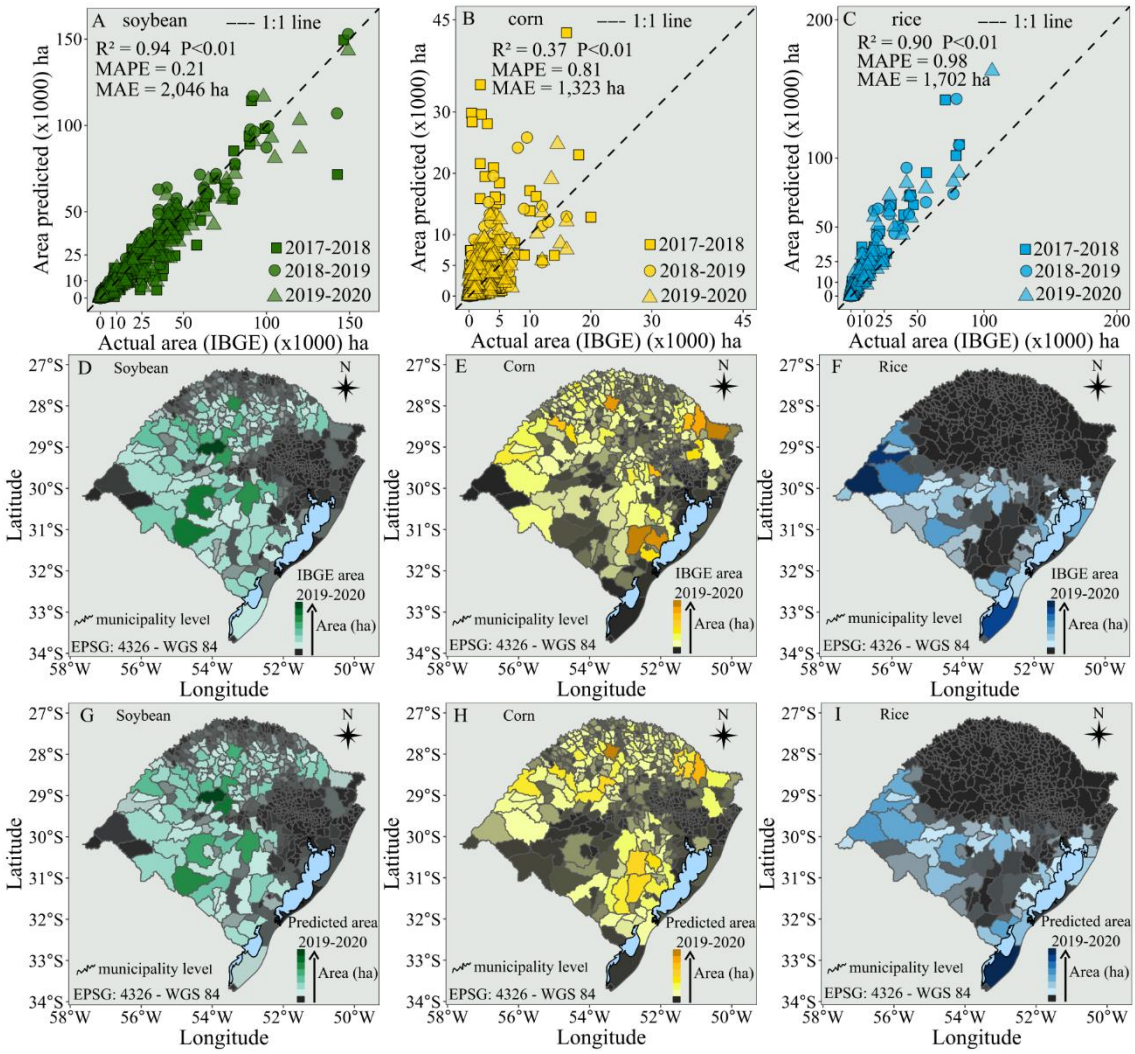


Fig. 7. (A) Relation 1:1 area predicted by the model and actual area IBGE statistics for soybean, (B) corn, and (C) rice. Data from 2017-2018, 2018-2019, 2019-2020 growing seasons. (D) Crop municipality area map of IBGE statistics from 2019-2020 growing season for soybean, (E) corn, and (F) rice. (G) Crop municipality area map predicted by the model from 2019-2020 growing season for soybean, (H) corn, and (I) rice.

The crop classification model at municipality level presented a normal distribution error, with no specific spatial associated error in the crop area (Fig. 7D, E, F, G, H, and I).

Lastly, a final crop map layer for 2019-2020 season (Fig. 8A), and one example of a municipality with two crop classes (Fig. 8B), and a municipality detailed with three crop classes (Fig. 8C) were mapped. In regions mainly cultivated with two crops (soybean and corn), such as in Northern Rio Grande do Sul, the crop classification map layer produced

more defined field boundaries in the model with less pixel noises compared to municipalities with three crop classes like Southern Rio Grande do Sul, where soybean, corn, and rice are the major three crops during the summer. In Table A2, examples are presented of crop area prediction based on satellite-based data fusion and IBGE census (standard method) for main crop cultivating municipalities in Rio Grande do Sul, Brazil.

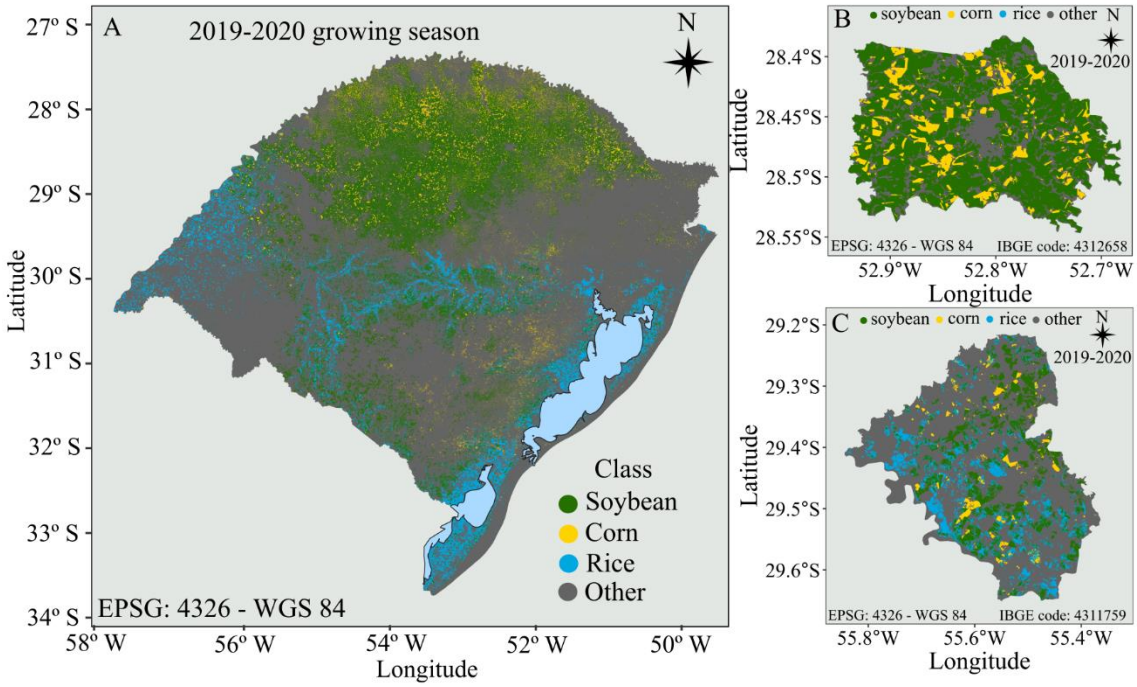


Fig. 8. (A) Rio Grande do Sul state, Brazil, crop map of 2019-2020 growing season. (B) Municipality detailed with two crop classes. (C) Municipality detailed with three crop classes.

DISCUSSION

The data fusion merging Sentinel-2, Sentinel-1, and Digital Elevation (SRTM) was relevant for developing this complex regional classification with an overall accuracy of 0.95. Similar to past studies, the combination of both optical and SAR data have improved crop classification (Abubakar et al., 2010, Inglada et al., 2016; Jin et al., 2019; Orynbaikyzy et al., 2020), with overall accuracy improving when SRTM data was included in the current study.

For crop classification in Brazil, Rudorff et al. (2007) presented a kappa of ~0.50 for soybean crop land in the north region of Rio Grande do Sul state with MODIS. In the state of Mato Grosso (Brazil), a study of crop classification presented an overall accuracy of 0.74 with EVI using MODIS (Arvor et al., 2011). Lemos (2016) developed a soybean mapping

classification utilizing EVI from MODIS in Southern Rio Grande do Sul with accuracy ~0.95, while Mengue and Fontana (2015) utilized MODIS and SRTM data for mapping soybean and rice crops in Rio Grande do Sul obtaining a kappa of 0.61 and 0.66, respectively for each crop. Zhong et al. (2016) obtained 87% overall accuracy utilizing MODIS for mapping soybean and corn cropland in Paraná (Brazil). Schultz et al. (2015) achieved 80% overall accuracy for classification of soybean, sugarcane (*Saccharum officinarum* L), cassava (*Manihot esculenta* Crantz), peanut (*Arachis hypogaea* L.), and other crops. Furthermore, Silva Junior et al. (2020) used a new index developed by Silva Junior et al. (2017) to map planted soybean area in the state of Mato Grosso comparing MODIS, Landsat 8, and Sentinel 2 sensors with greater accuracy for MODIS ($R^2 = 0.90$).

One of the main problems associated with optical sensors in satellite platforms such as Sentinel-2 is the issue of missing pixels due to cloud, cirrus, and shadow. For this reason, sensors having more temporal with high to medium spatial resolutions have the advantage to reduce time series data losses (Jin et al., 2019; Silva Junior et al., 2020). To handle missing pixels, feature engineering techniques such as harmonic regression coefficient extraction have been broadly used for crop classification (Ghazaryan et al., 2018; Wang et al. 2019; 2020; Deines et al., 2020; Dado et al., 2020). Multitemporal Sentinel-1 data have been used as a powerful tool for crop classification and agricultural monitoring in several studies (Skriver et al., 2011; Bargiel 2017; Bazzi et al., 2019). Furthermore, SRTM Digital Elevation data play an important role in land cover classification (Balzter et al., 2015; Sadeghi et al., 2018). The state of Rio Grande do Sul has a wide variation in terrain features such as elevation and slope, and the input of these features in the model improved crop classification performance in our study.

The spatial distribution of ground truth crop data exerted a strong impact on model performance, as reflected by the Moran's Index and k-mean cluster analyses. The quality of the data collected for the supervised classification model also influences model performance (Campbell and Wynne, 2011; Fowler et al., 2020). A representative subset of the population for each crop class with varying environmental and management conditions is important in the data collection process (Waldner et al., 2017; Waldner et al., 2019). The Moran's index and cluster k-mean provided more efficient guidance for crop type collection. Surveying data points/polygons in all the cluster regions of the Rio Grande do Sul state substantially improved the crop classification model.

Transfer learning for the subsequent growing season presented high accuracy when the model was built with more data input (large sample size). Random forests trained on regions and seasons with similar growing degree days (GDD) presented accuracies greater than 80% when transferred to target region and season with a comparable crop season (Wang et al., 2019). The similarity of crop phenology patterns, along with location and years lead to higher performance of transfer learning in random forest classification models (Hao et al., 2016; Wang et al., 2019; Hao et al., 2020a). Although transfer model learning showed good potential in crop classification for the subsequent growing season, when some samples data from same-year growing season were aggregated to the model, the model performance presented higher accuracy. A random forest classification utilizing local training samples increased overall accuracy compared to transfer learning for another region in United States (US) (Hao et al., 2020a). Improving model performance of transfer learning by merging data from a variety of growing conditions enlarges the potential that the training samples represent the region or growing season.

In-season crop classification is valuable information for decision-makers. We obtained similar prediction accuracy for early March (DOY 60) relative to the entire time-series (with all season data until May), with the potential use of January and February predictions when aggregating same-season data to the model. For the present study, aggregating data across years increased accuracy for in-season crop classification. Likewise, Hao et al. (2020a) reported an increase in early season predictions when local observations were utilized. Cai et al. (2018) reported high accuracy for late-June (DOY 180) in-season crop classification for both soybean and corn in the US, which represents approximately late December in the Rio Grande do Sul growing season.

Future studies should focus on high-resolution satellite sources, integrating crop modeling, and including more soils and climate features with detected agricultural field boundaries in order to improve model performance across regions and growing seasons. Moreover, further study should focus on crop classification including other minor crops, additional land features, and specific crop rotation patterns.

CONCLUSION

Satellite-based crop classification and mapping for soybean, corn, and rice crops was developed for the state of Rio Grande do Sul, Brazil. Features derived from Sentinel-2, Sentinel-1, and SRTM Digital Elevation provided high accuracy for crop classification.

From the ground truthing perspective, sample size and spatial distribution are key aspects, with more than 250 samples resulting in high overall accuracy. Random forest presented high accuracy in transferring the model to the next season, with the performance of the model increasing when full-season data was included.

Predictions of crop classification for March obtained similar results to when the entire season was taken into account (until May). Moreover, when data from additional growing seasons were aggregated, earlier predictions (January and February) reached higher overall accuracy.

Lastly, crop area prediction was most accurate for soybeans and least accurate for corn, with rice area overestimated due to errors with wetland target for other class.

In summary, to the extent of our knowledge, this is the first crop map layer for soybean, corn and rice crops for Rio Grande do Sul state (Brazil), establishing a foundation for creating maps for other crops, seasons, and regions around the globe.

CRediT authorship contribution statement

Luan Pierre Pott: Conceptualization of this study, Methodology, Software, Formal analysis, Investigation, Data Curation Visualization, Validation, Writing - Original Draft. Telmo Jorge Carneiro Amado: Validation, Supervision, Writing - Review & Editing. Raí Augusto Schwalbert: Investigation, Software, Formal analysis, Writing - Review & Editing. Geomar Mateus Corassa: Validation, Writing - Review & Editing. Ignacio Antonio Ciampitti: Conceptualization of this study, Investigation, Methodology, Supervision, Writing - Review & Editing.

Acknowledgements

We thank A.K.B., C.R., E.M.G., E.R., G.H., and R.D. for helping the data collection process. The first author thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The second author thanks to CNPq for research scholarship n 305622/2017-0. Contribution no. 21-244-J from the Kansas Agricultural Experiment Station.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Abubakar, G. A., Wang, K., Shahtahamssebi, A., Xue, X., Belete, M., Gudo, A. J. A., … Gan, M. (2020). Mapping Maize Fields by Using Multi-Temporal Sentinel-1A and Sentinel-2A Images in Makarfi, Northern Nigeria, Africa. *Sustainability*, *12*(6), 2539. https://doi.org/10.3390/su12062539

Alvares, C. A., Stape, J. L., Sentelhas, P. C., de Moraes Gonçalves, J. L., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, *22*(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Andrade, B. O., Bonilha, C. L., Overbeck, G. E., Vélez-Martin, E., Rolim, R. G., Bordignon, S. A. L., … Boldrini, I. I. (2019). Classification of South Brazilian grasslands: Implications for conservation. *Applied Vegetation Science*, *22*(1), 168–184. https://doi.org/10.1111/avsc.12413

Arvor, D., Jonathan, M., Meirelles, M.S.P., Dubreuil, V., & Durieux, L. (2011). Classification of MODIS EVI temporal series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing. 32 (22),* 7847–7871. https://doi.org/*10.1080/01431161.2010.531783*

Balzter, H., Cole, B., Thiel, C., & Schmullius, C. (2015). Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM Digital Elevation Model Data using Random Forests. *Remote Sensing*, *7*(11), 14876–14898. https://doi.org/10.3390/rs71114876

Bargiel, D. (2017). A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment*, *198*, 369–383. https://doi.org/10.1016/j.rse.2017.06.022

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., … Moran, M. S. (2000) Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In 'Proceedings of the Fifth International Conference on Precision Agriculture'. (Eds PC Robert, RH Rust, WE Larson) (American Society of Agronomy: Madison, WI) (CD-ROM)

Bazzi, H., Baghdadi, N., Ienco, D., El Hajj, M., Zribi, M., Belhouchette, H., … Demarez, V. (2019). Mapping Irrigated Areas Using Sentinel-1 Time Series in Catalonia, Spain. *Remote Sensing*, *11*(15), 1836. https://doi.org/10.3390/rs11151836

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. https://doi.org/10.1016/j.rse.2017.10.005

Bisong, E. (2019). Google Colaboratory. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, 59–64. https://doi.org/10.1007/978-1-4842-4470-8_7

Booth, D. J., & Oldfield, R. B. (1989). A comparison of classification algorithms in terms of speed and accuracy after the application of a post-classification modal filter. *International Journal of Remote Sensing,10*, 1271-1276. https://doi.org/10.1080/01431168908903965

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., & Smets, B. (2020). Copernicus Global Land Cover Layers—Collection 2. *Remote Sensing*, *12*(6), 1044. https://doi.org/10.3390/rs12061044

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., & Li, Z. (2018). A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, 210, 35–47. https://doi.org/10.1016/j.rse.2018.02.045

Campbell, J.B., Wynne, R.H. (2011). Introduction to Remote Sensing. Guilford Press.

Carletto, C., Jolliffe, D., Banerjee, R. (2015a). From tragedy to renaissance: improving agricultural data for better policies. The Journal of Development Studies, *51*, 133-148, *10.1080/00220388.2014.968140*

Carletto, C., Gourlay, S., & Winters, P. (2015b). From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. *Journal of African Economies*, *24*(5), 593–628. https://doi.org/10.1093/jae/ejv011

CONAB, Companhia Nacional de Abastecimento (2020a). - Mapeamentos Agrícolas. Retrieved November 19, 2020, from Conab.gov.br website: https://www.conab.gov.br/info-agro/safras/mapeamentos-agricolas

CONAB, Companhia Nacional de Abastecimento. (2020b). Acompanhamento da safra brasileira de grãos, v. 7 - Safra 2019/20 - Décimo segundo levantamento, Brasília, p. 1-68, setembro 2020. Retrieved November 4, 2020, from Conab.gov.br website: https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos

CONAB, Companhia Nacional de Abastecimento. (2020b). Boletim da Safra de Grãos. Estimativas - Produção e Balanço de Oferta e Demanda. Quarto Levantamento - Safra 2020/21. Retrieved January 15, 2021, from Conab.gov.br website: https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos

Dado, W. T., Deines, J. M., Patel, R., Liang, S.-Z., & Lobell, D. B. (2020). High-Resolution Soybean Yield Mapping Across the US Midwest Using Subfield Harvester Data. *Remote Sensing*, *12*(21), 3471. https://doi.org/10.3390/rs12213471

Deines, J. M., Patel, R., Liang, S.-Z., Dado, W., & Lobell, D. B. (2020). A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sensing of Environment*, 112174. https://doi.org/10.1016/j.rse.2020.112174

Demarez, V., Helen, F., Marais-Sicre, C., & Baup, F. (2019). In-Season Mapping of Irrigated Crops Using Landsat 8 and Sentinel-1 Time Series. *Remote Sensing*, *11*(2), 118. https://doi.org/10.3390/rs11020118

EMATER, Emater/RS-Ascar, (2020). Safra de verão 2020-2021 – Estimativas inicias de área, produtividade e produção. Setembro 2020. Retrieved December 5, 2020, from Emater.tche.br website: http://www.emater.tche.br/site/info-agro/acompanhamento_safra.php#.X_NdFdhKjIU

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., … Alsdorf, D. (2007), The Shuttle Radar Topography Mission, *Reviews of Geophysics, 45,* RG2004. https://doi.org/10.1029/2005RG000183.

Fowler, J., Waldner, F., & Hochman, Z. (2020). All pixels are useful, but some are more useful: Efficient in situ data collection for crop-type mapping using sequential exploration methods. *International Journal of Applied Earth Observation and Geoinformation*, *91*, 102114. https://doi.org/10.1016/j.jag.2020.102114

Ghazaryan, G., Dubovyk, O., Löw, F., Lavreniuk, M., Kolotii, A., Schellberg, J., Kussul, N. (2018). A rule-based approach for crop identification using multi-temporal and multi-

sensor phenological metrics. *European Journal of Remote Sensing, 51(1),* 511–524. https://doi.org/10.1080/22797254.2018.1455540

Gitelson, A.A., Vina, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, *32*(8). https://doi.org/10.1029/2005gl022688

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gourlay, S., Kilic, T., Lobell, D. (2017). Could the Debate be Over? Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda. The World Bank.

Hao, P., Di, L., Zhang, C., & Guo, L. (2020a). Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. *Science of The Total Environment*, *733*, 138869. https://doi.org/10.1016/j.scitotenv.2020.138869

Hao, P., Tang, H., Chen, Z., Meng, Q., & Kang, Y. (2020b). Early-season crop type mapping using 30-m reference time series. *Journal of Integrative Agriculture*, *19*(7), 1897–1911. https://doi.org/10.1016/s2095-3119(19)62812-1

Hao, P., Wang, L., Zhan, Y., & Niu, Z. (2016). Using Moderate-Resolution Temporal NDVI Profiles for High-Resolution Crop Mapping in Years of Absent Ground Reference Data: A Case Study of Bole and Manas Counties in Xinjiang, China. *ISPRS International Journal of Geo-Information*, *5*(5), 67. https://doi.org/10.3390/ijgi5050067

Hollstein, A., Segl, K., Guanter, L., Brell, M., & Enesco, M. (2016). Ready-to-Use Methods for the Detection of Clouds, Cirrus, Snow, Shadow, Water and Clear Sky Pixels in Sentinel-2 MSI Images. *Remote Sensing*, *8*(8), 666. https://doi.org/10.3390/rs8080666

IBGE, Instituto Brasileiro de Geografia e Estatística. (2019). Retrieved November 4, 2020, from Ibge.gov.br website: https://www.ibge.gov.br/cidades-e-estados/rs/

IBGE, Instituto Brasileiro de Geografia e Estatística. Área plantada, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras temporárias. (2020). Retrieved October 5, 2020, from Ibge.gov.br website: https://sidra.ibge.gov.br/tabela/1612#

IBGE, Instituto Brasileiro de Geografia e Estatística. Série histórica da estimativa anual da área plantada, área colhida, produção e rendimento médio dos produtos das lavouras. (2020). Retrieved November 4, 2020, from Ibge.gov.br website: https://sidra.ibge.gov.br/tabela/6588

Inglada, J., Vincent, A., Arias, M., & Marais-Sicre, C. (2016). Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. *Remote Sensing*, *8*(5), 362. https://doi.org/10.3390/rs8050362

Jakubauskas, M. E., Legates, D. R., & Kastens, J. H. (2002). Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture*, *37*(1–3), 127–139. https://doi.org/10.1016/s0168-1699(02)00116-3

Jiang, Z., Huete, A., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, *112*(10), 3833–3845. https://doi.org/10.1016/j.rse.2008.06.006

Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., & Lobell, D. B. (2019). Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment*, 228, 115–128. https://doi.org/10.1016/j.rse.2019.04.016

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., & Pirotti, F. (2019). Monitoring Within-Field Variability of Corn Yield using Sentinel-2 and Machine Learning Techniques. *Remote Sensing*, *11*(23), 2873. https://doi.org/10.3390/rs11232873

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1-26. https://doi.org/10.18637/jss.v028.i05

Lemos, G.S. (2016). Mapeamento de áreas de soja em municípios da metade sul do estado do Rio Grande Do Sul a partir de imagens de satélite. (Mestrado em Ciências) Programa de Pós-graduação em Manejo e Conservação do solo e da Água, Universidade Federal de Pelotas, Pelotas, pp. 82.

Li, Q., Qiu, C., Ma, L., Schmitt, M., & Zhu, X. (2020). Mapping the Land Cover of Africa at 10 m Resolution from Multi-Source Remote Sensing Data with Google Earth Engine. *Remote Sensing*, *12*(4), 602. https://doi.org/10.3390/rs12040602

Mengue, V. P., & Fontana, D. C. (2015). Avaliação da dinâmica espectro-temporal visando o mapeamento dos principais cultivos de verão no Rio Grande do Sul. *Bragantia*, *74*(3), 331–340. https://doi.org/10.1590/1678-4499.0452

Orynbaikyzy, A., Gessner, U., Mack, B., & Conrad, C. (2020). Crop Type Classification Using Fusion of Sentinel-1 and Sentinel-2 Data: Assessing the Impact of Feature Selection, Optical Data Availability, and Parcel Sizes on the Accuracies. *Remote Sensing*, *12*(17), 2779. https://doi.org/10.3390/rs12172779

Ozdogan, M., & Gutman, G. (2008). A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: An application example in the continental US. *Remote Sensing of Environment*, *112*(9), 3520–3537. https://doi.org/10.1016/j.rse.2008.04.010

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ringland, J., Bohm, M., & Baek, S.-R. (2019). Characterization of food cultivation along roadside transects with Google Street View imagery and deep learning. *Computers and Electronics in Agriculture*, *158*, 36–50. https://doi.org/10.1016/j.compag.2019.01.014

Rudorff, C. de M., Rizzi, R., Rudorff, B. F. T., Sugawara, L. M., & Vieira, C. A. O. (2007). Superfícies de resposta espectro-temporal de imagens do sensor MODIS para classificação de área de soja no Estado do Rio Grande do Sul. *Ciência Rural*, *37*(1), 118–125. https://doi.org/10.1590/s0103-84782007000100019

Sadeghi, Y., St-Onge, B., Leblon, B., Prieur, J.-F., & Simard, M. (2018). Mapping boreal forest biomass from a SRTM and TanDEM-X based on canopy height model and Landsat spectral indices. *International Journal of Applied Earth Observation and Geoinformation*, *68*, 202–213. https://doi.org/10.1016/j.jag.2017.12.004

Santos, H. G., Jacomine, P. K. T., Dos Anjos, L. H. C., De Oliveira, V. A., Lumbreras, J. F., Coelho, M. R., ... & Cunha, T. J. F. (2018). Sistema brasileiro de classificação de solos. Brasília, DF: Embrapa, 2018.

Schultz, B., Immitzer, M., Formaggio, A., Sanches, I., Luiz, A., & Atzberger, C. (2015). Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in Southeastern Brazil. *Remote Sensing*, *7*(11), 14482–14508. https://doi.org/10.3390/rs71114482

Silva Junior, C. A., Leonel-Junior, A. H. S., Rossi, F. S., Correia Filho, W. L. F., Santiago, D. de B., Oliveira-Júnior, J. F. de, … Capristo-Silva, G. F. (2020). Mapping soybean planting area in midwest Brazil with remotely sensed images and phenology-based

algorithm using the Google Earth Engine platform. *Computers and Electronics in Agriculture*, *169*, 105194. https://doi.org/10.1016/j.compag.2019.105194

Silva Junior, C. A., Nanni, M. R., Teodoro, P. E., & Silva, G. F. C. (2017). Vegetation Indices for Discrimination of Soybean Areas: A New Approach. *Agronomy Journal*, *109*(4), 1331–1343. https://doi.org/10.2134/agronj2017.01.0003

Skriver, H., Mattia, F., Satalino, G., Balenzano, A., Pauwels, V. R. N., Verhoest, N. E. C., & Davidson, M. (2011). Crop Classification Using Short-Revisit Multitemporal SAR Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *4*(2), 423–431. https://doi.org/10.1109/jstars.2011.2106198

Soil Survey Staff. (2014). Keys to soil taxonomy (12th ed.). Washington, DC: USDA-Natural Resources Conservation Service.

Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N.; Mochizuki, K-I. (2017). Assessing the suitability of data from Sentinel-1A and 2A for crop classification. *GIScience & Remote Sensing, 54,* 918–938. *https://doi.org/10.1080/15481603.2017.1351149*

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423. https://doi.org/10.1111/1467-9868.00293

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., & Hochman, Z. (2013). Yield gap analysis with local to global relevance—A review. *Field Crops Research*, *143*, 4–17. https://doi.org/10.1016/j.fcr.2012.09.009

Waldner, F., Bellemans, N., Hochman, Z., Newby, T., de Abelleyra, D., Verón, S. R., … Defourny, P. (2019). Roadside collection of training data for cropland mapping is viable when environmental and management gradients are surveyed. *International Journal of Applied Earth Observation and Geoinformation*, *80*, 82–93. https://doi.org/10.1016/j.jag.2019.01.002

Waldner, F., Canto, G. S., & Defourny, P. (2015). Automated annual cropland mapping using knowledge-based temporal features. *ISPRS Journal of Photogrammetry and Remote Sensing*, *110*, 1–13. https://doi.org/10.1016/j.isprsjprs.2015.09.013

Waldner, F., Jacques, D. C., & Löw, F. (2017). The impact of training class proportions on binary cropland classification. *Remote Sensing Letters*, *8*(12), 1122-1131. https://doi.org/10.1080/2150704X.2017.1362124

Wang, S., Azzari, G., & Lobell, D. B. (2019). Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, *222*, 303–317. https://doi.org/10.1016/j.rse.2018.12.026

Wang, S., Di Tommaso, S., Deines, J. M., & Lobell, D. B. (2020). Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. *Scientific Data*, *7*(1). https://doi.org/10.1038/s41597-020-00646-4

Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, *236*, 111402. https://doi.org/10.1016/j.rse.2019.111402

Xu, L., Zhang, H., Wang, C., Zhang, B., & Liu, M. (2018). Crop Classification Based on Temporal Information Using Sentinel-1 SAR Time-Series Data. *Remote Sensing*, *11*(1), 53. https://doi.org/10.3390/rs11010053

Yan, Y., & Ryu, Y. (2021). Exploring Google Street View with deep learning for crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *171*, 278–296. https://doi.org/10.1016/j.isprsjprs.2020.11.022

Zhong, L., Hu, L., Yu, L., Gong, P., & Biging, G. S. (2016). Automated mapping of soybean and corn using phenology. *ISPRS Journal of Photogrammetry and Remote Sensing*, *119*, 151–164. https://doi.org/10.1016/j.isprsjprs.2016.05.014

Appendix A

Table A1. Balanced accuracy of the classes in early predictions with one (2018-2019 data samples) and two (2018-2019 + 2019-2020 data samples) growing seasons.

Class	Balanced accuracy									
	2018-2019 data samples					2018-2019 + 2019-2020 data samples				
	Jan	Feb	Mar	Apr	May	Jan	Feb	Mar	Apr	May
Corn	0.91	0.92	0.91	0.95	0.95	0.97	0.97	0.96	0.96	0.96
Soybean	0.90	0.94	0.94	0.95	0.96	0.94	0.93	0.96	0.96	0.96
Rice	0.97	0.98	0.99	0.97	0.98	0.99	0.98	0.99	0.98	0.98
Other	0.91	0.93	0.92	0.96	0.97	0.97	0.97	0.98	0.98	0.97

Table A2. Municipality crop area of select cities in Rio Grande do Sul. Census from IBGE and crop area predicted by model.

City	IBGE			Predicted		
	Corn	Rice	Soybean	Corn	Rice	Soybean
Agudo	3500	9100	950	825	10112	1642
Alegrete	3800	49641	38000	3700	77923	38985
Cachoeira do Sul	3500	25522	103000	6489	61037	92775
Carazinho	2300	0	40420	6564	1	42006
Cruz Alta	5100	0	92000	12469	2	90716
Dom Pedrito	550	38923	120000	882	56685	102975
Erechim	1320	0	9700	6189	1	13345
Ijuí	1900	0	40000	2114	2	42122
Jóia	3700	0	81000	5575	17	71941
Lagoa Vermelha	4500	0	46000	13039	4	53331
Não-Me-Toque	3000	0	22500	4220	0	22843
Palmares do Sul	150	12845	4090	639	19728	4566
Palmeira das Missões	14500	0	90400	24767	1	81077
Passo Fundo	1400	0	41000	3924	2	43070
Pelotas	7000	7177	22000	6658	22356	15165
Santa Bárbara do Sul	3200	0	76000	8345	0	68648
Santa Maria	190	6250	49345	231	17874	41701
Santa Rosa	2500	0	19500	5705	1	19200
Santo Ângelo	3400	0	36300	4685	7	33155
São Borja	5000	38800	65000	9571	72483	68469
Soledade	800	0	42000	2206	3	52754
Tupanciretã	2530	0	149100	9469	18	143320
Uruguaiana	0	76319	2980	2443	89168	2564
Vacaria	13500	0	55000	24767	1	81077

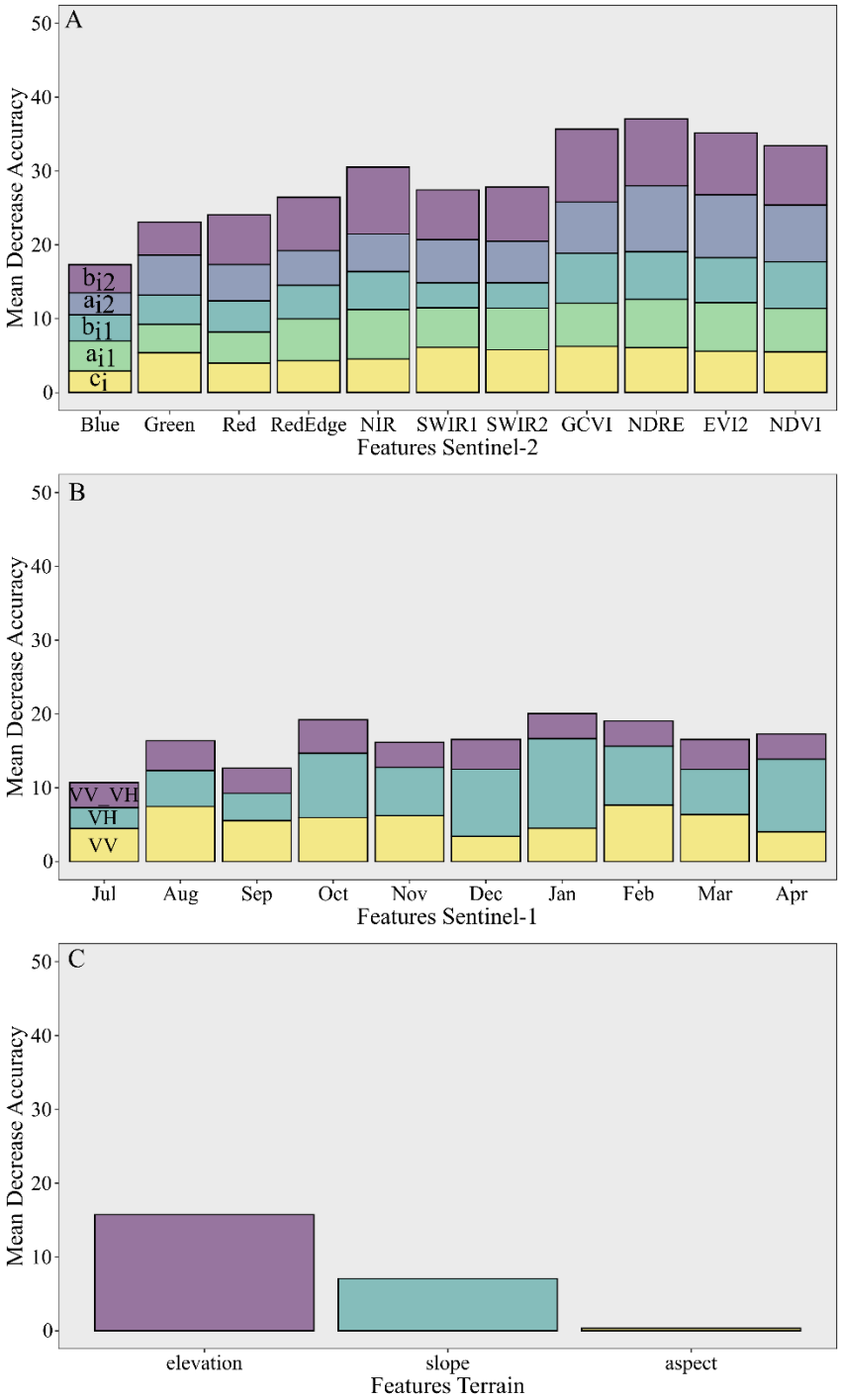
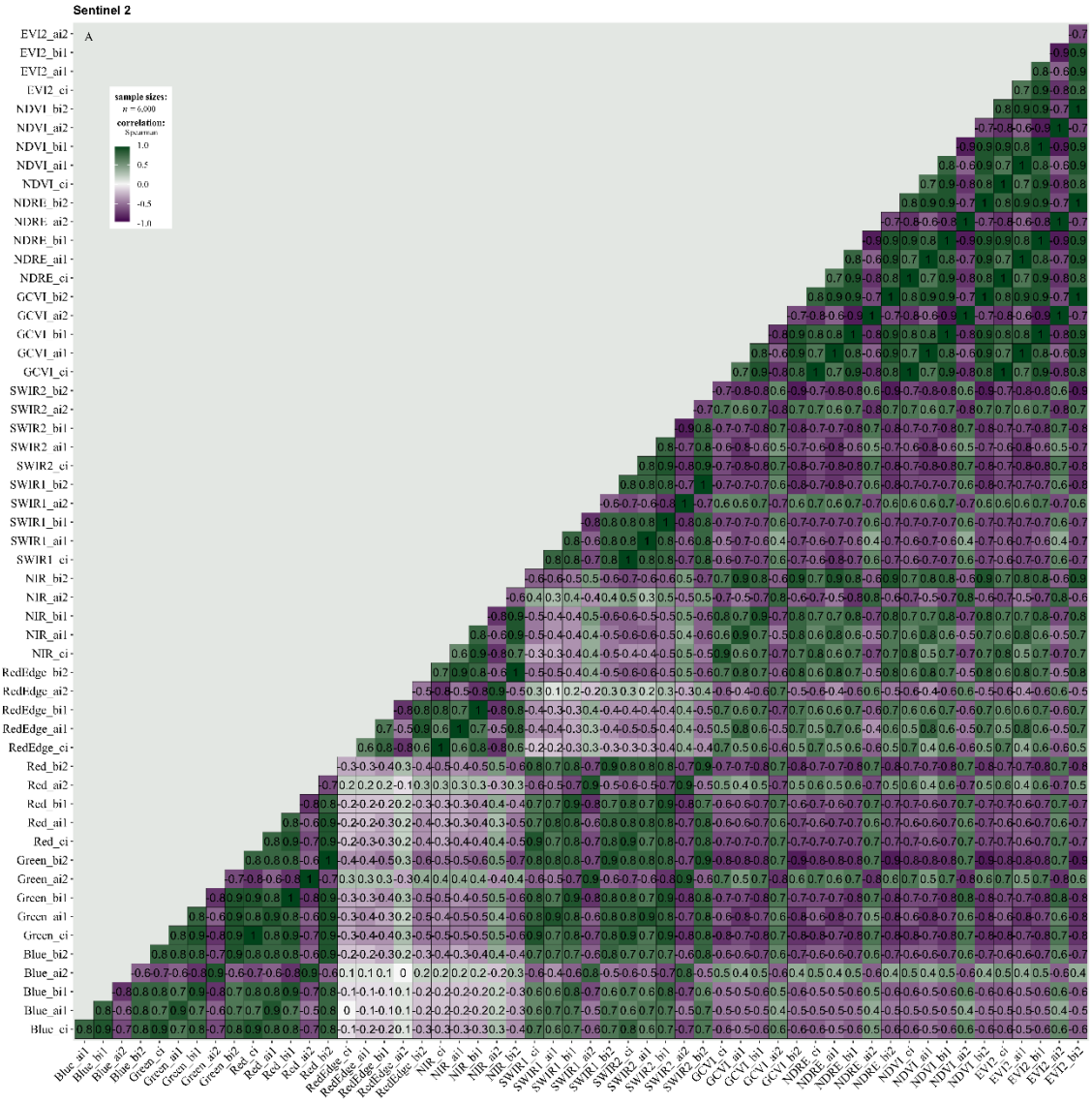
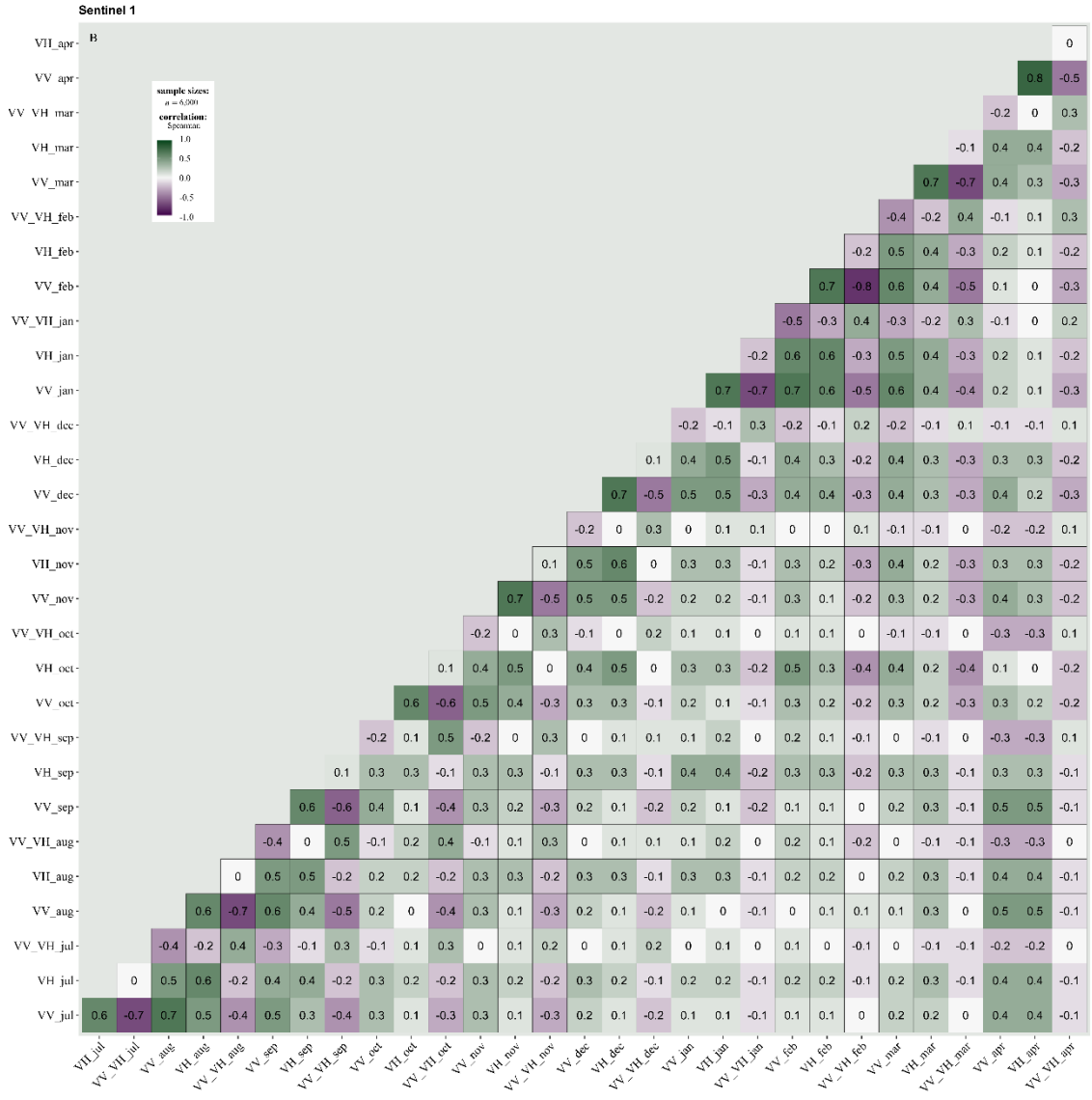


Fig. A1. Mean Decrease Accuracy of all the features from Sentinel-2 (A), Sentinel-1 (B), and SRTM Digital Elevation (C). c_i, a_{i1}, b_{i1}, a_{i2}, b_{i2} are the harmonic coefficients extracted from the Sentinel-2 time-series. VV, VH, and VV_VH are the monthly mean features from Sentinel-1. Elevation, slope, and aspect from SRTM Digital Elevation.


Sentinel 2
A
sample sizes:
n = 6,000
correlation:
Spearman
1.0
0.5
0.0
-0.5
-1.0
EVI2_ai2
EVI2_bi1
EVI2_ai1
EVI2_ci
NDVI_bi2
NDVI_ai2
NDVI_bi1
NDVI_ai1
NDVI_ci
NDRE_bi2
NDRE_ai2
NDRE_bi1
NDRE_ai1
NDRE_ci
GCVI_bi2
GCVI_ai2
GCVI_bi1
GCVI_ai1
GCVI_ci
SWIR2_bi2
SWIR2_ai2
SWIR2_bi1
SWIR2_ai1
SWIR2_ci
SWIR1_bi2
SWIR1_ai2
SWIR1_bi1
SWIR1_ai1
SWIR1_ci
NIR_bi2
NIR_ai2
NIR_bi1
NIR_ai1
NIR_ci
RedEdge_bi2
RedEdge_ai2
RedEdge_bi1
RedEdge_ai1
RedEdge_ci
Red_bi2
Red_ai2
Red_bi1
Red_ai1
Red_ci
Green_bi2
Green_ai2
Green_bi1
Green_ai1
Green_ci
Blue_bi2
Blue_ai2
Blue_bi1
Blue_ai1
Blue_ci


Sentinel 1
B
sample sizes:
n = 6,000
correlation:
Spearman
1.0
0.5
0.0
-0.5
-1.0
VH_apr
VV_apr
VV_VH_mar
VH_mar
VV_mar
VV_VH_feb
VH_feb
VV_feb
VV_VH_jan
VH_jan
VV_jan
VV_VH_dec
VH_dec
VV_dec
VV_VH_nov
VH_nov
VV_nov
VV_VH_oct
VH_oct
VV_oct
VV_VH_sep
VH_sep
VV_sep
VV_VH_aug
VH_aug
VV_aug
VV_VH_jul
VH_jul
VV_jul

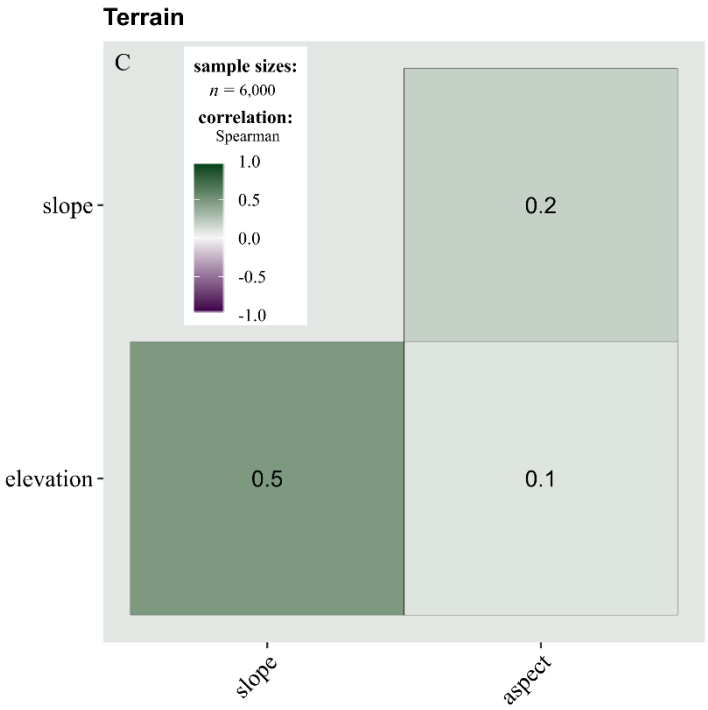


Fig. A2. Spearman correlation for all the features from Sentinel-2 (A), Sentinel-1 (B), and SRTM Digital Elevation (C).

3 ARTIGO 2 - CROP TYPE CLASSIFICATION IN SOUTHERN BRAZIL: INTEGRATING REMOTE SENSING, CROP MODELING AND MACHINE LEARNING

Artigo originalmente publicado e seguindo as normas da Revista Computers and Electronics in Agriculture.

Pott, L. P., Amado, T. J. C., Schwalbert, R. A., Corassa, G. M., & Ciampitti, I. A. (2022). Crop type classification in Southern Brazil: Integrating remote sensing, crop modeling and machine learning. *Computers and Electronics in Agriculture*, 201, 107320.

ABSTRACT

Crop type mapping is essential for agricultural monitoring, but lack of terrestrial labels (herein termed as ground truth data) limit these models around the globe. In this study, we tested different methods for creating crop type maps for soybean (*Glycine max* L.) and corn (*Zea mays* L.), with aims of i) compare ground truth data related to crop modeling; ii) evaluate agricultural field masks generated by Environmental Rural Register, MapBiomas product, and random forest model; iii) evaluate models of a) unsupervised classification, b) supervised classification with ground truth data of 1 year, c) with ground truth data of 2 years, d) supervised classification with crop modeling only, e) with the combination of crop modeling and ground truth data of 1 year, and f) with ground truth data of 2 years; iv) test sample size for training the model utilizing ground truth data and crop modeling in transfer learning approach; and v) examine spatial remote sensing features to guide crop data collection. The APSIM-NG crop model was utilized for generate crop modeling simulations to compare with satellite images harmonic regression of ground truth data. We found similarity of harmonic regression coefficients derived from crop modeling and satellite imagery of Sentinel-2 of the labeled ground truth data for both soybean and corn crops. Agricultural masks showed efficiency for crop area estimation for soybean, with highest accuracy with random forest model. Crop modeling aggregated with growing season data as input for supervised learning presented the greater model performance with overall accuracy of 0.94. Crop area prediction was most accurate for soybean [R^2 = 0.93, and mean absolute

error (MAE) = 2,052 ha] for the model with the combination of crop modeling and ground truth data of 2 years, and least for corn, [R^2 = 0.18, and MAE = 1,146 ha] when only using crop modeling. Model performance was influenced by sample size, with greater accuracy (0.93) for aggregating crop modeling (150 samples) and year data (100 samples). In addition, considering spatial field data variability as model input increased overall accuracy from 0.84 to 0.93, with higher impact when only ground truth data was utilized for input in the model. Our results suggest that these methods offer options for crop type classification when less adequate ground truth data is available and with unsupervised learning. On the other hand, supervised learning utilizing crop modeling with presence of field data did improve model performance overall.

Keywords: Agricultural monitoring; Crop mapping; Land cover; Sentinel-2; APSIM-NG, Transfer learning.

INTRODUCTION

Crop type classification and mapping can provide information for food production, farm management, and agricultural intensification. With the increasing abundance and availability of satellite data, many studies have reported crop type maps based on remote sensing (Bargiel, 2017; Cai et al., 2018; Kluger et al., 2021; Yan et al., 2021). Mostly, machine learning methods with supervised learning have been widely utilized for crop type classification (Wang et al., 2019; 2020; Fowler et al., 2020; Pott et al., 2021). However, supervised algorithms request surveys and censuses for terrestrial labels (herein termed as ground truth data) which are labor and time intensive, and expensive to be conducted with accuracy (Fowler et al., 2020; Kluger et al., 2021).

Agricultural masks can be obtained with land use and land cover (LULC) layers, primary models with crop and non-crop classes, and another sources. In Brazil, MapBiomas project reconstructed decades of LULC changes with several classes including agricultural layers (Souza et al., 2020; MapBiomas, 2021). The Environmental Rural Register (Cadastro Ambiental Rural - CAR) (Sicar, 2021) in Brazil have formed a geographical database aiming to environmental planning for tracking deforestation rate, but also including consolidated areas of agriculture (Sicar, 2021). Furthermore, various studies have being conducted to build high-resolution models with more robust and efficient approaches to generate LULC maps for Brazilian agriculture (Picoli et al., 2018; Souza et al., 2020).

The efficacy of machine learning algorithms is partially determined by the choice of feature engineering method. Harmonic regression coefficient extraction has been widely utilized for crop type classification with unsupervised learning (Wang et al., 2019), and supervised learning (Wang et al., 2019; 2020; Pott et al., 2021). Moreover, harmonic regression has being utilized as feature engineering for crop yield mapping integrating crop modeling with remote sensing in the United States (US) for soybean (*Glycine max* L.) (Dado et al., 2020) and corn (*Zea mays* L.) (Deines et al., 2021).

Large amounts of periodically ground truth data are needed to label using supervised machine learning algorithms to generate crop type classification maps across different regions and growing seasons with high model performance (Fowler et al., 2020; Wang et al., 2019). Unlike the progress on satellite resolution, the development of regularly-updated crop type maps in Brazil have been limited in geographic due to the scarcity of terrestrial data. In regions where ground truth data is not abundant, machine learning methods obtained some success in crop type mapping (Defourny et al., 2019; Kluger et al., 2021). Furthermore, due to the narrow geographical concentration or limited growing seasons for crop label data, scaling-up or transfer-learning of models (outside of the limits offered by the training dataset) can severely impact model performance by heterogeneity of agricultural settings (Wang et al., 2019; Pott et al., 2021; Kluger et al., 2021).

To deal with a scarcity of field-level crop labels, Wang et al. (2019) utilized unsupervised learning aggregated with crop statistics for crop type mapping in US states. However, the use of unsupervised methods for crop type classification is sparse, often vulnerable to outliers, high dimensionality and require the user input such as the number of clusters, but with the advantage of generating crop type maps when no ground truth data is available (Hastie et al., 2009; Wang et al., 2019). Transfer learning approach with supervised learning utilizing ground truth data have been reported with success for many studies around the globe (Wang et al., 2019; 2020; Pott et al. 2021; Lin et al., 2022). However, the use of crop modeling, as a generation of crop development simulations, with their output serving as data features for input in supervised models is not well addressed in the literature for crop type classification purposes. Furthermore, crop classification models need both quantity and quality of labeled data to train and test models. Data beyond the years included in the model are also required to test sample sizes and spatio-temporal transferability to improve crop

prediction and in-season crop mapping (Wang et al., 2019; Pott et al., 2021; Zhang et al., 2021; Lin et al., 2022).

This study develops and tests different methods for creating crop type maps in settings without ground truth data, utilizing crop modeling to label satellite imagery and when there is presence of field ground data. Therefore, the aims of this study were to: i) compare the ground truth data related to simulations via crop modeling; ii) evaluate agricultural field masks generated by Environmental Rural Register, MapBiomas and random forest model; iii) evaluate models of a) unsupervised classification, b) supervised classification with 1 year ground truth data, c) with 2 years ground truth data, d) supervised classification with crop modeling, e) supervised with the combination of crop modeling and ground truth data from 1 year and, f) 2 years; iv) test sample size for model training utilizing ground truth data augmented with simulated crop data (from APSIM-NG) in a transfer learning approach; and v) examine spatial remote sensing features to guide crop data collection.

MATERIAL AND METHODS

Study area

The selected area to this study consists of Northwest region (27° to 29° S and 51° to 56° W) in Rio Grande do Sul state, Southern Brazil (Fig. 1A, and B). The Northwest region has 216 municipalities with approximately 64,931 km^2 (IBGE, 2020) (Fig. 1C) representing the most important grain production region of the state. The climate types of this selected region are Humid subtropical without dry season with -hot summer (Cfa), and with -temperate summer (Cfb) according to Alvares et al. (2013). The predominant soil type is Oxisols, with presence of Mollisols, Entisols, and Ultisols (Santos et al., 2018; Soil Survey Staff, 2014). In the study area, soybean and corn crops comprise more than 78% of the grain crop area in the region (IBGE, 2021). Soybean cropping area of the last four years was on average 3.1 Mha, representing around 50% of state and 8% national area, and corn area was on average 0.4 Mha representing 48% of state and 2.3% national area (IBGE, 2021). Soybean planting time is preferably around October extending up to January and harvested from February to May, while corn is usually planted in a large window from July to February and harvested from January to June (CONAB, 2020).

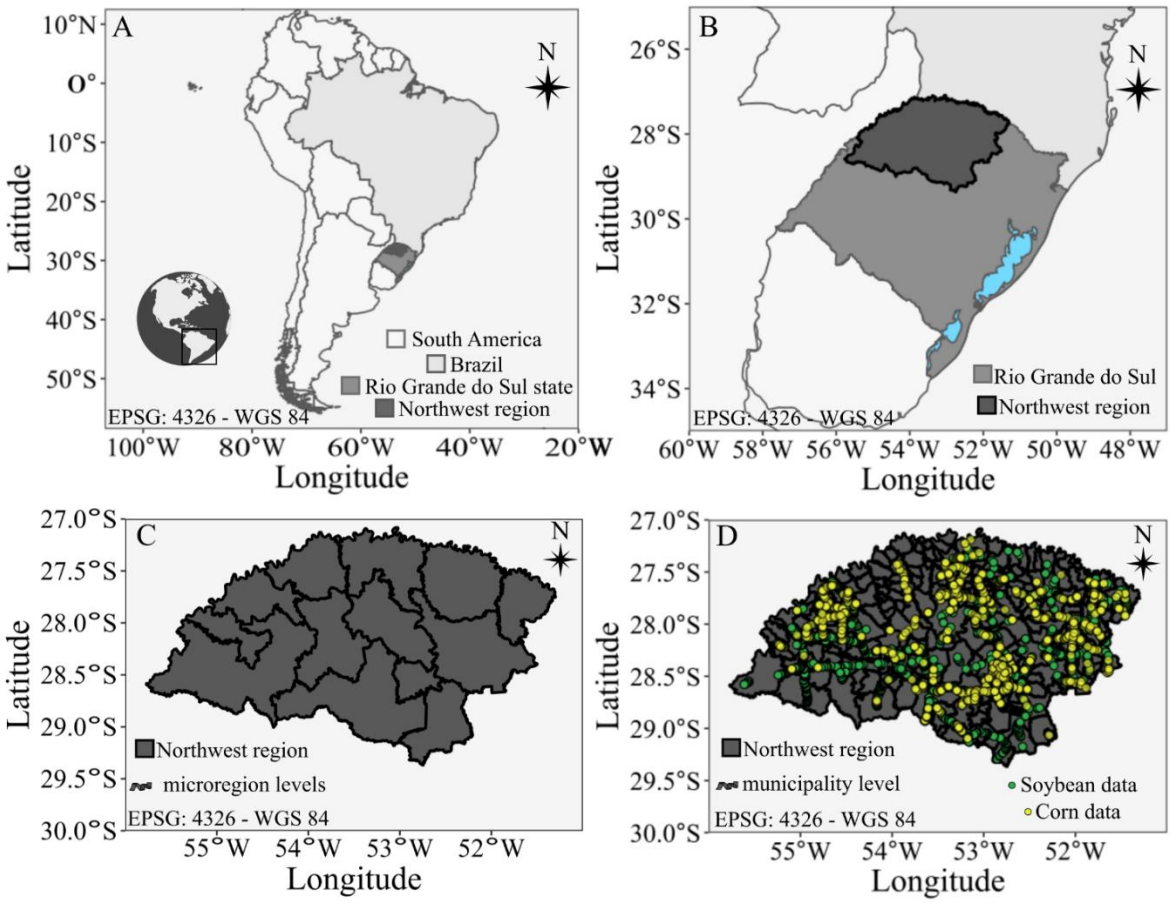


Fig. 1. (A) South America map, Brazil country, Rio Grande do Sul state, and Northwest region

. (B) Rio Grande do Sul state and Northwest region (shaded in black). (C) Microregions in Northwest region. (D) Crop type data points collected during 2018-2019, 2019-2020, and 2020-2021 growing seasons.

Ground truth data

The ground truth data were collected with two different approaches. The first approach was via Google Street View, which presents a compendium of panoramic images along roads around the globe gathering geographic coordinates and time of the collected data. Google Street View databases has been recently used as data source for crop classification in scientific studies (Yan and Ryu, 2021; Pott et al., 2021). For the Rio Grande do Sul state, the database contains images from October 2018 to April 2019. We have accessed this database manually informing latitude and longitude coordinates and identifying the crop type along the roads in the Northwest region. The dataset was formed by identification of soybean and corn

crops gathered with latitude and longitude coordinates for the 2018-2019 growing season. The second collection process utilized a Global Navigation Satellite System (GNSS) portable receiver, GPSmap 62®, to collect crop type information gathering latitude and longitude coordinates along Northwest region for 2019-2020, and 2020-2021 growing seasons. For all the three growing seasons we have selected 500 data points for each crop (Fig. 1D).

Remote sensing data

We have utilized remote sensing Sentinel-2 Level-1C imageries for the growing season interest period (July 2017 to May 2018, July 2018 to May 2019, July 2019 to May 2020) for the 2017-18, 2018-19 and 2019-20 growing seasons, respectively. We have selected Sentinel-2 Level-1C due the availability of this data for Rio Grande do Sul state in Google Earth Engine (GEE) along the years of study. The Sentinel-2 Level-1C product obtained similar results compared with Sentinel-2 Level-2A for crop classification tasks (Jin et al., 2019). The imagery data were filtered for clouds using a modified version of the Fmask algorithm (Frantz et al., 2018).

We have utilized the feature engineering harmonic regression previous reported as potential for crop classification around the globe (Jakubauskas et al., 2002; Wang et al., 2019; Pott et al., 2021) with the green chlorophyll vegetation index (GCVI) (Gitelson et al., 2005) to extract features from the image time series to apply in machine learning models. GCVI have been utilized for crop modeling leaf area index (LAI) linkage with remote sensing vegetation indices (Lobell et al., 2015; Dado et al., 2020; Deines et al., 2021). In our analyses, GCVI obtained the highest performance compared with others vegetation indices. We have utilized the same numbers of harmonic terms and coefficients reported in Wang et al. (2019; 2020), Deines et al. (2020) and Dado et al. (2020) in the United States (US) and Pott et al. (2021) in Brazil. The final regression at each Sentinel-2 point for the GCVI is accordingly Eq. (1).

$$f_i(t) = c_i + a_{i1}\,cos(3\pi t) + b_{i1}\,sin(3\pi t) + a_{i2}\,cos(6\pi t) + b_{i2}\,sin(6\pi t) \quad (1)$$

where $a_{ik,}$ and b_{ik} are cosine, and sine coefficients respectively, while c_i is the intercept. *t* is an independent variable that represents the time of the Sentinel-2 images taken within a growing season expressed as a fraction between 0 (July 1) and 1 (May 1). We have extracted the coefficient: c_{i1}, a_{i1}, b_{i1}, a_{i2} and b_{i2} for GCVI using the built-in linear regression function in GEE, resulting in 5 features to input in the machine learning methods.

Agricultural masks

We have evaluated three different sources of agricultural masks: i) random forest model, ii) MapBiomas product (MapBiomas, 2021), and iii) Environmental Rural Register (Cadastro Ambiental Rural - CAR) (Sicar, 2021) (Fig. 2).

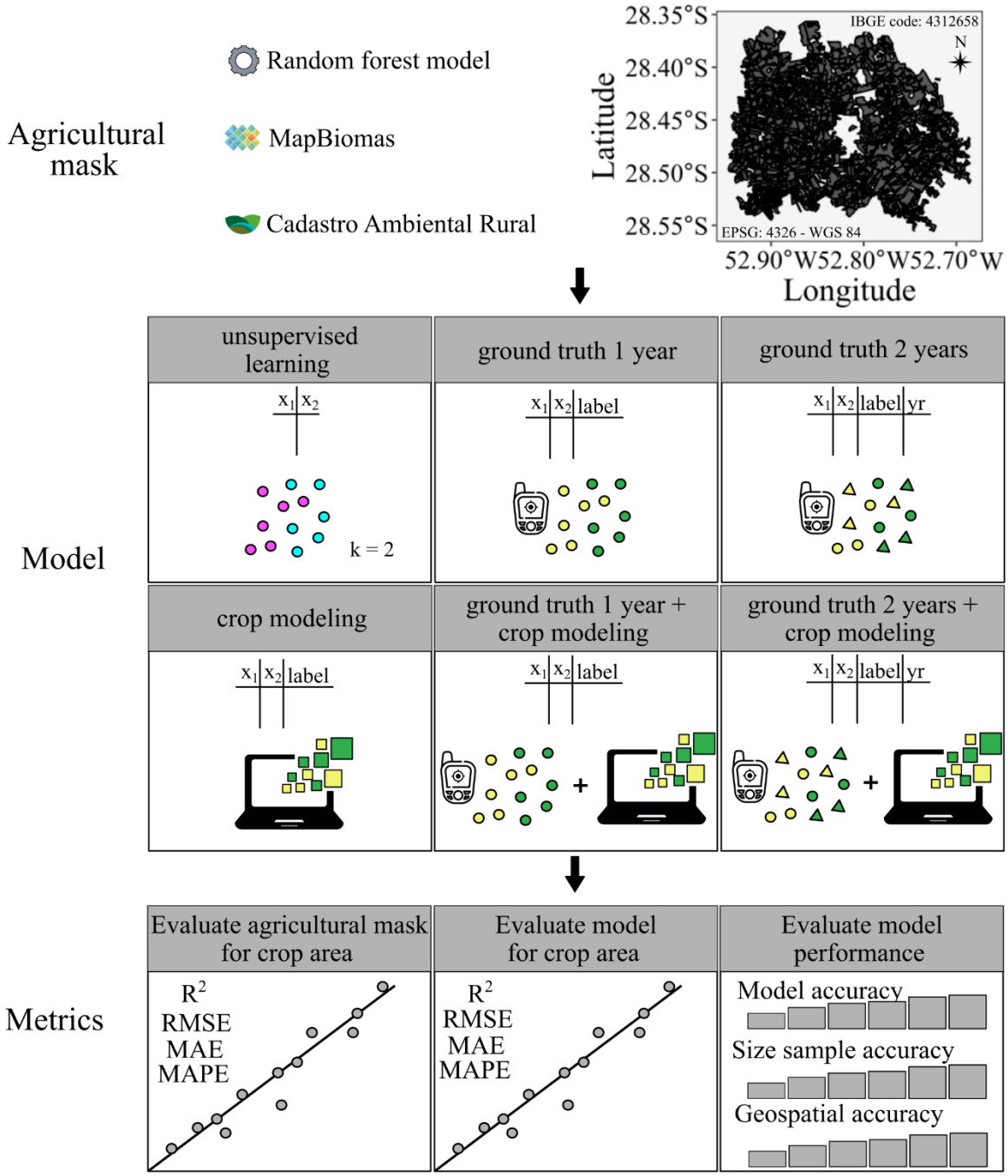


Fig. 2. Framework for model development. Agricultural masks utilized - Random forest model, MapBiomas and Environmental Rural Register (Cadastro Ambiental Rural); Datasets

utilized to run the model – unsupervised learning, supervised learning with: ground truth data of 1 year, ground truth of 2 years, crop modeling, ground truth of 1 year + crop modeling, and full model with ground truth of 2 years + crop modeling; Metrics – comparison of statistics for agricultural masks, for the different models, and overall accuracy of the model evaluating sample size and geospatial model accuracies.

Random forest model

The agricultural random forest masks were built with Sentinel-2 Level-1C dataset extracting coefficients of harmonic regression from GCVI images time-series as already reported in section 2.3.

Random forest classification was performed for two classes: agricultural crops (data merged from soybean and corn ground truth for three growing seasons – class: agricultural crops), and other class (collected by visually assessing high-resolution imagery from Google Earth and Google Street View, including water, urban, forest, pasture, and native vegetation, and other minor crop – class: other class). The dataset contained 3,000 data points for both classes (crop and non-crop).

To train the random forest model, we used the *tidymodels* package (Kuhn et al., 2020) in R (R Core Team, 2021). We tuned the parameter to reduce model variance, and the final parameters resulted in 250 trees and variables per split equal to 3, with all other parameter as default. We used 75% of the data to train and 25% to test the model. To create the final agricultural mask raster in Northwest Rio Grande do Sul, we used GEE *ee.Classifier.smileRandomForest()* algorithm. The raster masks for each year were vectorized in GEE to compose the agricultural masks.

MapBiomas product

MapBiomas is a Brazilian project that works with remote sensing, computer science and geoespatial techniques with different institution and experts collaboration (MapBiomas, 2021). MapBiomas products use remote sesning images from Landsat archive with machine learning algorithm within GEE platform to make available maps of LULC in Brazil since 1985 (Souza et al., 2020).

In this study, we have accessed MapBiomas collection 6, that covers the period of 1985-2020, via GEE ID: projects/mapbiomas-

workspace/public/collection6/mapbiomas_collection60_integration_v1. It was selected the ID classes numbers 18, 19, 39, and 41, representing agriculture, temporary crop, soybean, and other temporary crops classes, respectively (MapBiomas, 2021). We have generated the raster masks of these ID classes for 2017, 2018 and 2019 years. The raster masks were vectorized in GEE to compose the agricultural masks for each year investigated.

Environmental Rural Register

The Environmental Rural Register (Cadastro Ambiental Rural – CAR) is a digital public registry, including all the rural farms information such as environmental area for preservation, forests, and consolidated area boundaries for retrieving data for controlling and monitoring against deforestation (Sicar, 2021). For this study, we have accessed the field boundaries of the consolidated agricultural areas for all the 216 municipalities in the Northwest Rio Grande do Sul. The field boundaries archieve was downloaded individualy as shapefile format and merged via *sf* package (Pebesma, 2018) in R, and uploaded into the GEE platform.

Crop modeling

In order to build crop modeling simulations, we have employed the Agricultural Production Systems sIMulator Next Generation (APSIM-NG) (Holzworth et al., 2018), taking LAI as output of the simulations. We converted LAI from APSIM-NG simulations of soybean and corn crops to GCVI via linear regression based on Nguy-Robertson et al. (2012) equations, Eqs. (2)-(3), that already been reported as the linkage for crop modeling and remote sensing (Lobell et al., 2015; Dado et al., 2020; Deines et al., 2021).

$$GCVI = 1.1 + 1.4 \times LAI^{1.3}, \text{ for soybean} \quad (2)$$

$$GCVI = 0.93 + 1.4 \times LAI^{1.03}, \text{ for corn} \quad (3)$$

For generating the simulations, we have selected one municipality per microregion in the Northwest Rio Grande do Sul region to compose the sites of the simulations (Fig. 1C). For the soybean cultivar choice, we have employed generic cultivars with appropriated maturity groups (MG) (Archontoulis et al., 2014) for the geographic interest region. Furthermore, we have combined soybean cultivar parameters available after calibration by Battisti et al. (2017) in Brazil. For corn, we have utilized the parameters of a generic hybrid calibrated by Duarte

and Sentelhas (2019) in Brazil. Sowing dates were selected based on the entire range of sowing dates for the study region reported in CONAB (2020). More details of the parameters for the APSIM-NG crop simulations are presented in Table 1.

Table 1. Summary of Agricultural Production Systems sIMulator (APSIM) settings for crop simulations of soybean and corn.

Factor	Values used for soybean	Values used for corn	Units
Year	1999/2000-2019/2020	Same	
Site	Cruz Alta (28.6º S, 53.6º W) Erechim (27.6º S, 52.3º W) Guarani das Missões (28.1º S, 54.5º W) Marau (28.4º S, 52.2º W) Não-Me-Toque (28.4º S, 52.8º W) Palmeira das Missões (27.9º S, 53.3º W) Panambi (28.2º S, 53.5º W) Santo Cristo (27.8º S, 54.7º W) São José do Ouro (27.8º S, 51.6º W) São Luiz Gonzaga (28.4º S, 54.9º W) Soledade (28.8º S, 52.5º W) Tenete Portela (27.4º S, 53.7º W) Trindade do Sul (27.5º S, 52.9º W)	Same	
Cultivar choice	BRS_284 Battisti et al. (2017) Generic_MG5 Generic_MG6 Generic_MG7 Archontoulis et al. (2014)	Generic Duarte and Sentelhas (2019)	

Sowing date	25-sep, 5-oct, 15-oct, 25-oct, 5-nov, 15-nov, 25-nov, 5-dec, 15-dec, 25-dec	5-aug, 25-aug, 15-sep, 5-oct, 25-oct, 15-nov, 5-dec, 25-dec, 15-jan, 30-jan	
Sowing density	1, 2, 3, 4, 5	6, 8, 10	plants per m^2
Fertilizer	-	200, 400	kg of urea N per ha
Row spacing	500	same	mm
Total simulations	54,600 (200 per site-year)	16,380 (60 per site-year)	

The climate variables were retrieved from NASA Prediction of Worldwide Energy Resources (NASA/POWER-NP), which provides daily data of all variables required by crop simulation models in a 0.5 to 1° resolution (Stackhouse et al., 2015). For the dataset acquisition, we have utilized the *nasapower* package (Sparks, 2018) in R, that provides functionality to generate weather input files for agricultural crop modelling. We have utilized the SoilGrids 2.0 for the soil profiles data to input in the crop model process. SoilGrids provides a digital global maps of interpolated soil properties important for crop modelling such as bulk density, soil texture, organic matter content and pH at finer scales (Hengl et al., 2017). The preparation of all the soils profiles are done utilizing the APSoil (Version 7.20). We have utilized pedotransfer functions to estimate water content on the soil following Reichert et al. (2009) equations for volumetric water content at 0.33 bars level (VWC33) and 15 bars (VWC1500), Eqs. (4)-(5).

$$\text{VWC33 (drained upper limit - DUL)} = 0.366 - 0.34 \times \text{SAND} \quad (4)$$

$$\text{VWC1500 (lower limit - LL15)} = 0.236 + 0.045 \times \text{CLAY} - 0.21 \times \text{SAND} \quad (5)$$

Following the APSoil parameter estimation protocol (Dalgliesh et al., 2016), air dry (AD) can be calculated as 50% of the LL15 value in the top layers (0–0.05 and 0.05–0.15m), 80% in the 0.15-0.30m layer and 100% for all other profiles. Saturation (SAT) was estimated using porosity (PO) as follows Eqs. (6)-(7).

$$PO = 1 - BD/TD, \text{ where BD is bulk density, and TD is true density (2.65 g cm}^{-3}) \quad (6)$$

$$SAT = PO - e, \text{ we have considered } e \text{ for loams soil, } e = 0.05 \quad (7)$$

As the pH from SoilGrids is based on a 1:2.5 soil water ratio, and the modern crop modeling uses the ratio 1:5, the pH values were converted following the methods from Kabała et al. (2016), Eq. (8).

$$pH\ (1{:}5W) = 0.14 + 0.99 \times pH\ (1{:}2.5W) \quad (8)$$

We have compared the climate data from NASA POWER with National Institute of Meteorology (INMET) database, which provides access to maximum and minimum temperature, solar radiation or sunshine hours and rainfall. Solar radiation was estimated with Ångström–Prescott equation (Ångström, 1924) with coefficient for Rio Grande do Sul reported by Bexaira et al. (2018). Furthermore, SoilGrids were compared to RADAM Brazil Project (1974), providing information on clay, silt and sand contents, pH, and soil organic carbon (SOC). We found high agreement in NASA POWER and INMET, R^2 from 0.76 to 0.85, and index of agreement (d) from 0.90 to 0.96 for the climate variables, but rainfall in daily values (Fig. 3). SoilGrids and RADAM Brazil Project obtained satisfactory relationship, R^2 from 0.39 to 0.53, and d index from 0.72 to 0.81 for the soil variables, except for SOC (Fig. 4), however we need to consider the oldness of the RADAM Brazil database.

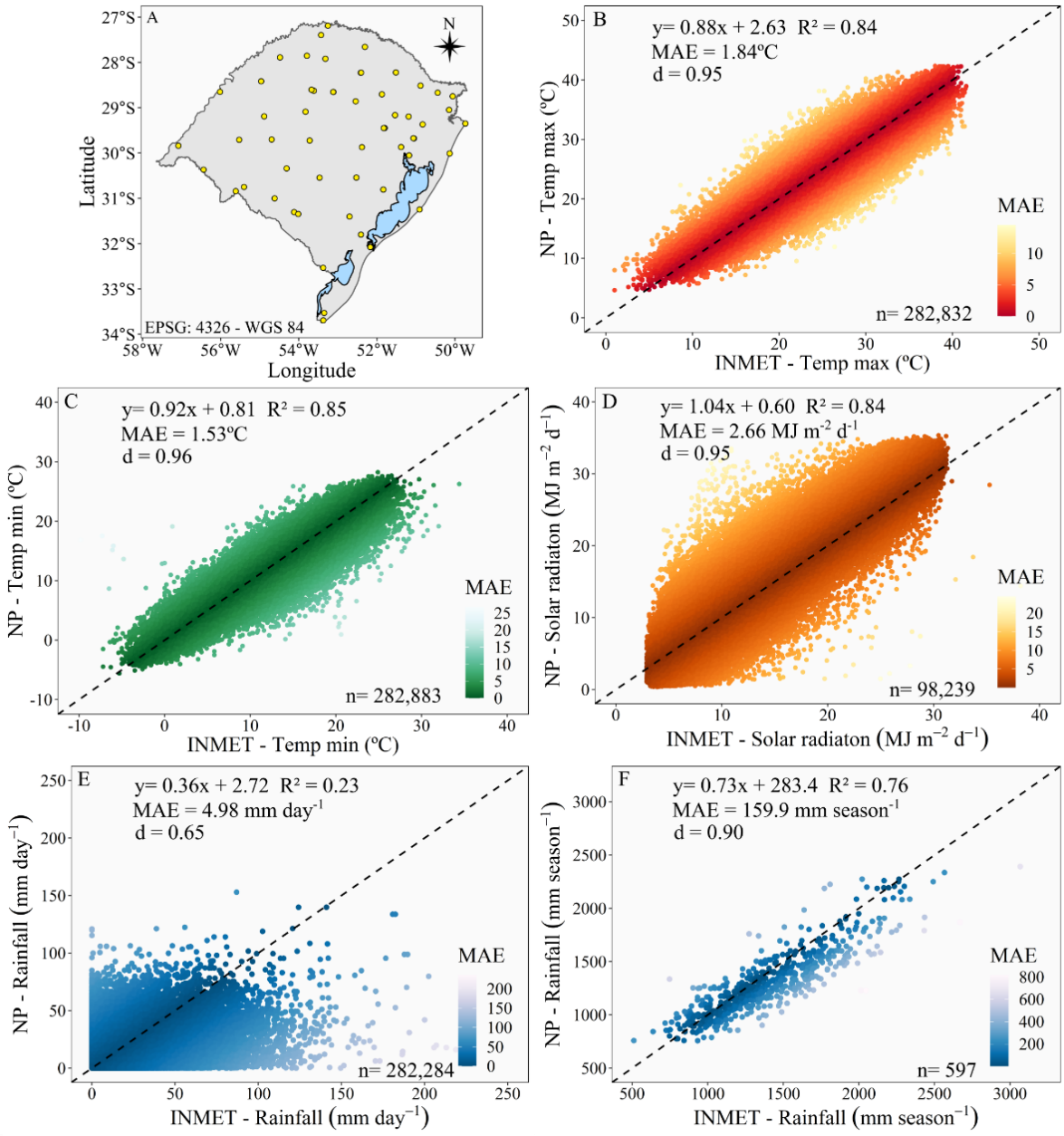


Fig. 3. (A) Meteorological station data collected in Rio Grande do Sul, Brazil. (B) Relationship between maximum temperature, (C) minimum temperature, (D) solar radiation, (E) daily rainfall, and (F) season rainfall obtained from INMET and NASA/POWER database.

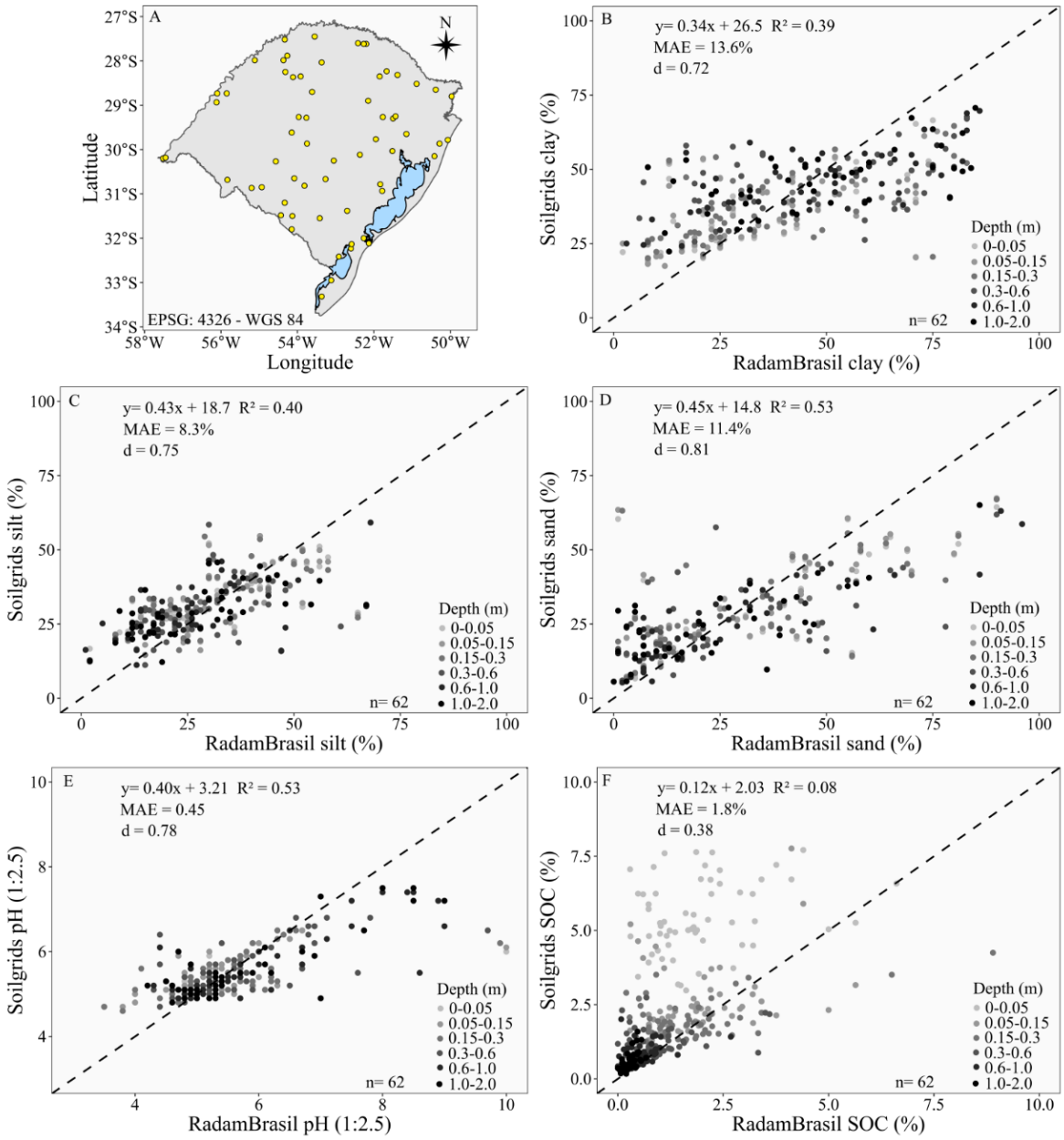


Fig. 4. (A) Data points from RadamBrasil collected in Rio Grande do Sul, Brazil. (B) Relationship between clay, (C) silt, (D) sand, (E) pH, and (F) soil organic matter (SOC) obtained from RadamBrasil and SoilGrids database.

Machine learning methods

We have evaluated six different models for crop type classification in the study area (Fig. 2). Unsupervised and supervised learning were implemented for the crop type classification. In addition, we evaluated different sources of data for training the supervised machine learning method random forest algorithm with ground truth data of one year,

combining two years of ground truth data, data from simulations of crop modeling, and combining ground truth data of one and two years with crop modeling. The summary of the models is presented in Table 2.

Table 2. Summary of the crop type classification models implemented in the study area.

Model	Machine learning algorithm
Unsupervised	K-means algorithm
Ground truth data – 1 year	Random forest
Ground truth data – 2 years	Random forest
Crop modeling	Random forest
Ground truth data – 1 year + crop modeling	Random forest
Ground truth data – 2 years + crop modeling	Random forest

To train the random forest models, we used the *tidymodels* package (Kuhn et al., 2020) in R. We have utilized grid search for tunning the parameters with the best choice of: 250 trees and variables per split equal to 3, with all other parameter as default. For training the models and estimating the new growing season as transfer learning approach, we used the entire dataset of the year or years of the ground truth, and for the crop modeling model all the simulations, except the year to predict, as the validation process of leave one out cross validation. To create the final crop type classification maps in Northwest Rio Grande do Sul, we used GEE *ee.Classifier.smileRandomForest()* algorithm.

To generate the clusters in the unsupervised model, we have utilized the k-means clustering with tidy data principles with *tidymodels* package (Kuhn et al., 2020) in R. In the pre-processing preceding the clustering, we have normalized the features on the same scale. We have utilized the number of clusters (k) = 2, for generating the classes, since the clusters are the crops soybean and corn. Since IBGE (2021) reports soybean as the crop with major area for all the municipalities in the study region, we have considered the cluster with more area for soybean, and less area for corn crop. To create the final crop type classification maps in Northwest Rio Grande do Sul, we used the GEE *ee.Clusterer.wekaKMeans()* algorithm.

Data analyses

The comparison between the observations of the features from ground truth data and the crop modeling simulations were evaluated with the Kullback-Leibler (KL) divergence using *FNN* package (Beygelzimer et al., 2019) in R. The KL divergence is a measure of the difference between two probability distributions (Kullback and Leibler, 1951).

To evaluate the crop area prediction performance for the agricultural masks and for the six models, we have compared the predicted crop area for 216 municipalities from IBGE estimates (IBGE, 2021) for each growing season (Fig. 2). The following statistical parameters were used to evaluate crop area predictions: coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Furthermore, we have utilized the index of agreement (d) (Willmott, 1981) for the comparisons between NASA/POWER x INMET, and SoilGrids x RADAM Brazil Project. Related to the model comparison, we have computed the confusions matrix and calculated the overall accuracy using *tidymodels* workflow with *yardstick* package (Kuhn and Vaughan, 2021) in R. Model performance was evaluated for sample size, combining the factor sample with different source of data, and the balanced of historical ground truth of 1- and 2-years data, and crop modeling with year data.

The features of the harmonic regression of Sentinel-2 data were assessed to perform Moran's I index analyses to verify the spatial dependence of each crop. After Moran's I index analyses, we have performed the cluster k-means algorithm with the normalized features. The number of clusters were defined according to gap statistics. The gap statistic is a standard method for determining the best number of clusters "k", in a dataset (Tibshirani et al., 2001).

RESULTS

Comparison of ground truth data and crop modeling simulations

The coefficients of the harmonic regression with n=2 presented distinguishable of crop types for corn and soybean for both ground truth data and crop modeling showed in an illustration of two-dimensional feature space (Fig. 5A). The models we have implemented have utilized the five coefficients being the model more complex.

The density distribution of the features from ground truth data and crop modeling presented similarity for soybean and corn (Fig. 5B, and C). The KL divergence was greater for corn (KL = 0.46), compared with soybean (KL = 0.16). These differences can indicate the high range of possibilities of different years for the ground truth data and the crop modeling

(Fig. 5D, E). Furthermore, management such as sowing date generated by crop simulations for corn (sowing dates from August to January), can indicate that the ground truth data did not cover these periods of representatively data collection.

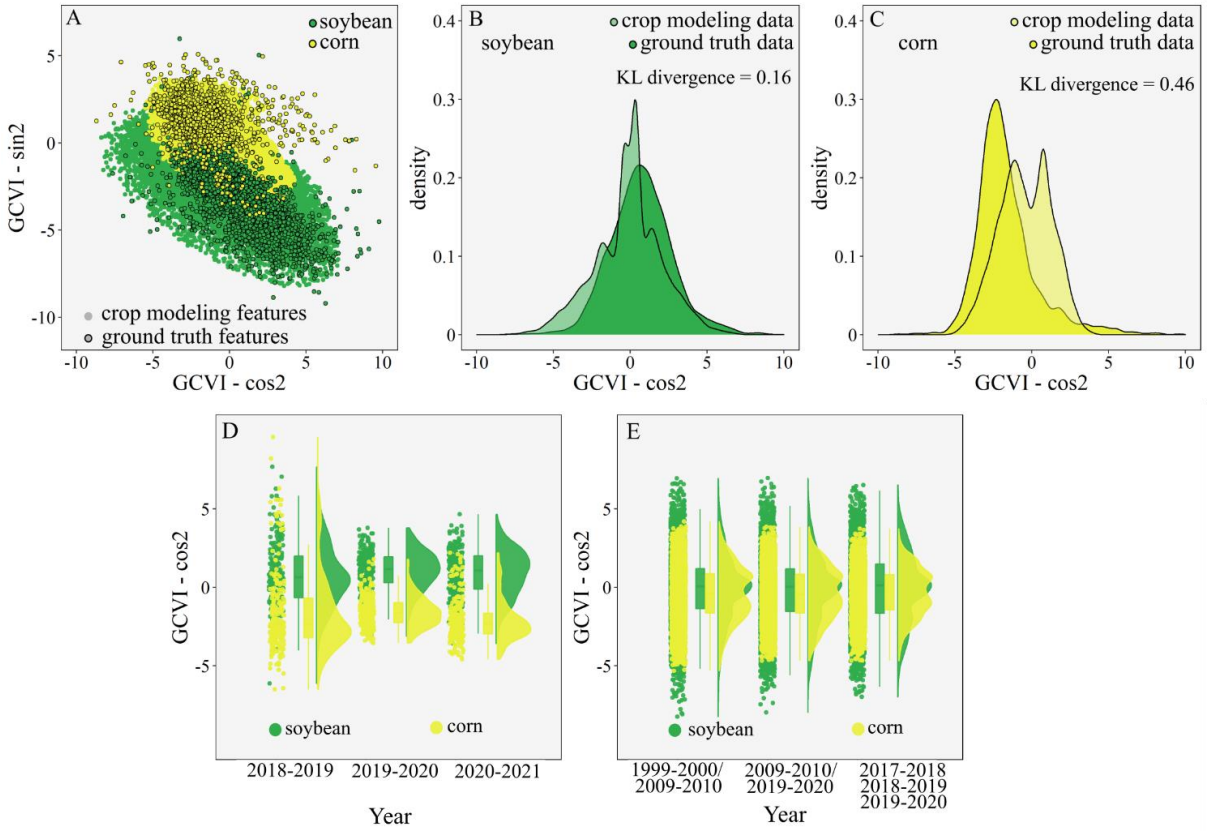


Fig. 5. (A) Harmonic regression features – second order cosine and sine terms of GCVI, of soybean and corn for the ground truth data and the crop modeling simulations. (B) density distribution of harmonic regression feature – second order cosine term, of data truth and crop modeling simulations for soybean, (C) and corn. (D) Raincloud plot of the harmonic regression feature - second order cosine term, of the ground truth data in the different growing seasons, (E) and of the crop modeling data in the different growing seasons simulations.

Municipality-level comparison with IBGE crop area for agricultural masks

The three different agricultural masks presented high accuracy and precision for soybean crop area when compared with IBGE statistics, R^2 from 0.87 to 0.92, MAE from 2,535 to 2,946 ha, while for corn the models with all agricultural masks showed low precision and accuracy, R^2 from 0.10 to 0.23, MAE from 1,208 to 4,099 ha (Fig. 6). The predicted and IBGE statistics comparison of soybean crop area with the agricultural masks utilizing the

random forest model with Sentinel-2 images to generate the mask obtained the highest performance compared with MapBiomas and CAR agricultural masks, with R^2 = 0.92, and MAE = 2,591 ha (Fig. 6A, B, and C).

The predicted and IBGE statistics of corn crop area did not show good relationship, with higher accuracy for MapBiomas agricultural mask, R^2 = 0.10, and MAE = 1,208 ha, followed by random forest model, and last CAR with the lower accuracy (Fig. 6 D, E, and F). Details of the field boundaries presented by the agricultural masks can be found in Figure 7.

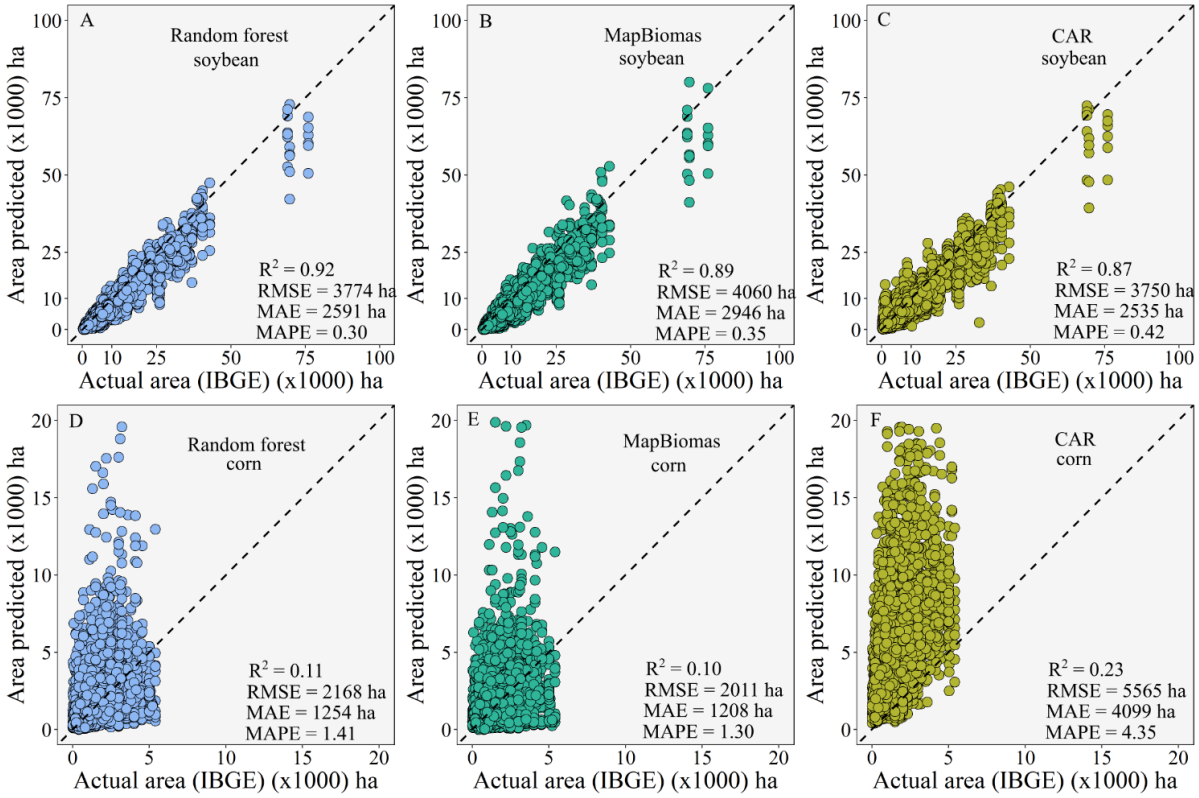


Fig. 6. (A) Relation 1:1 area predicted by the model and actual area IBGE statistics of soybean for random forest model, (B) MapBiomas, and (C) Environmental Rural Register (CAR) agricultural masks. (D) Relation 1:1 area predicted by the model and actual area IBGE statistics of corn for random forest model, (E) MapBiomas, and (F) Environmental Rural Register agricultural masks.

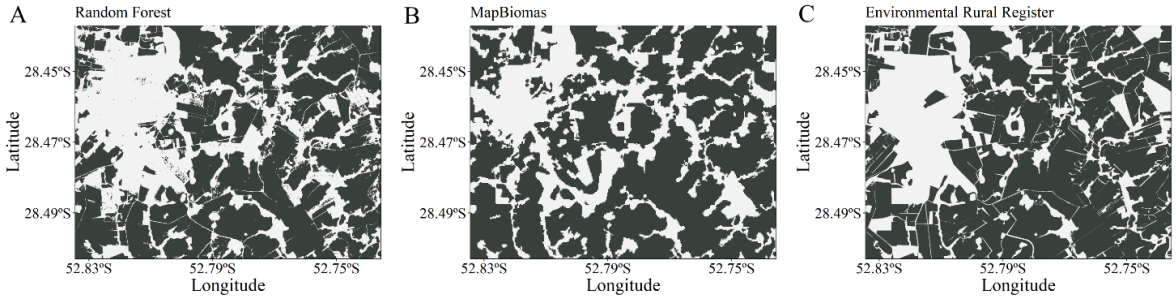


Fig. 7. Agricultural masks from random forest model with Sentinel-2 images (A), MapBiomas product (B), and Environmental Rural Register consolidated areas database (C).

Municipality-level comparison with IBGE crop area for the different models

To understand whether unsupervised and supervised models can be used to classify soybean and corn crops in a new growing season, transfer learning approach and crop area prediction were evaluated for three growing seasons (Fig. 8).

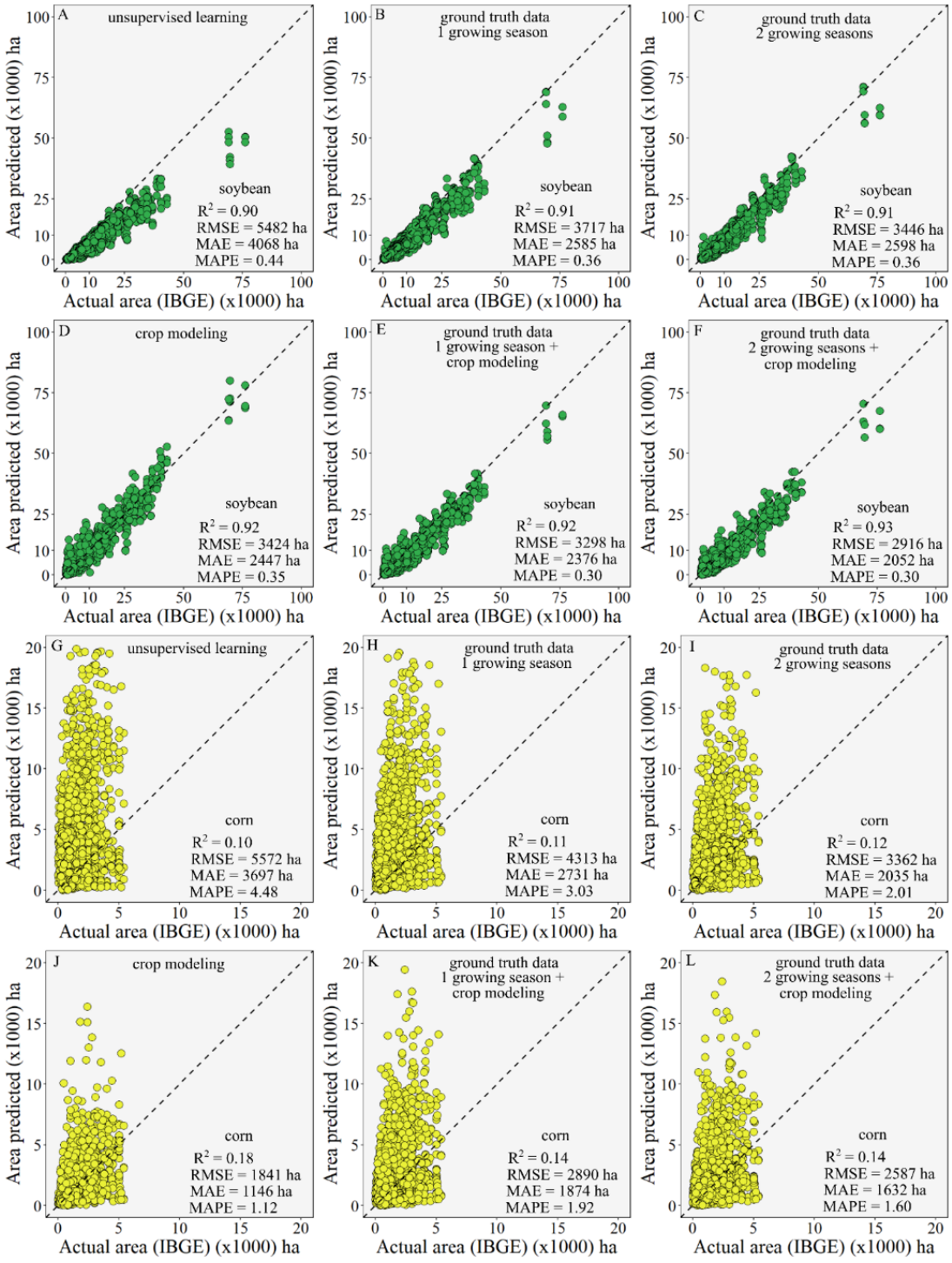


Fig. 8. (A) Relation 1:1 area predicted by the model and actual area IBGE statistics of soybean for unsupervised model, (B) utilizing ground truth data of one growing season, (C) and two growing seasons, (D) utilizing crop modeling data, (E) ground truth data of one growing season + crop modeling, and (F) and two growing seasons + crop modeling. (G) Relation 1:1 area predicted by the model and actual area IBGE statistics of corn for unsupervised model, (H) utilizing ground truth data of one growing season, (I) and two

growing seasons, (J) utilizing crop modeling data, (K) ground truth data of one growing season + crop modeling, and (L) and two growing seasons + crop modeling.

Soybean crop area predictions compared with IBGE statistics generally showed high precision and accuracy (Fig. 8 A, B, C, D, E and F). Unsupervised learning method for soybean area predictions presented high precision, but low accuracy compared with the supervised learning method with the different databases, R^2 = 0.90, MAE = 4,068 ha (Fig. 8A). For the supervised learning methods utilizing ground truth data in the random forest model, the models with one and two growing season year data presented similar performance for crop area predictions, increasing performance, with reduction of MAPE from 0.44 to 0.36, related to unsupervised model, R^2 = 0.91, MAE = 2,585 ha, and R^2 = 0.91, MAE = 2,598 ha, for ground truth data of 1-, and 2-years data, respectively (Fig. 8B, and C). The model trained only with simulations improved performance relative to the supervised learning utilizing ground truth data, R^2 = 0.92, MAE = 2,447 ha (Fig. 8D). Next, aggregating ground truth data of one year and crop modeling simulations as data for train the model, the performance improved (mainly in a reduction of MAE), R^2 = 0.92, MAE = 2,376 ha (Fig. 8E). Moreover, when ground truth data of two growing seasons were added to the crop simulations data, the model achieved the greatest performance, R^2 = 0.93, MAE = 2,052 ha (Fig. 8F).

However, corn crop area predictions did not show significantly relationship with IBGE statistics (Fig 8. G, H, I, J, K, and L).

More statistics of predicted and actual area from IBGE of the models and agricultural masks for soybean and corn are presented in Table 3.

Table 3. Statistics of predicted model and actual area from IBGE of the different model and agricultural masks for soybean and corn.

Model	Agricultural mask											
	Random forest				MapBiomas				CAR			
	R^2	MAE (ha)	MAPE	RMSE (ha)	R^2	MAE (ha)	MAPE	RMSE (ha)	R^2	MAE (ha)	MAPE	RMSE (ha)
soybean												
Unsupervised learning	0.91	3784	0.42	5139	0.91	4304	0.49	5639	0.92	4115	0.42	5652
Truth	0.93	2252	0.26	3271	0.91	2756	0.33	3870	0.92	2747	0.32	3973
Truth 2 years	0.92	2568	0.29	3566	0.92	2747	0.32	3723	0.94	2027	0.28	2928
Crop	0.94	1838	0.23	2611	0.91	2085	0.26	2978	0.90	2231	0.62	3134

modeling												
Truth + crop modeling	0.95	2278	0.27	3149	0.94	2766	0.34	3705	0.91	2084	0.44	2996
Truth 2 years + crop modeling	0.95	2827	0.32	3797	0.95	3016	0.33	3960	0.92	1953	0.42	2765
						corn						
Unsupervised learning	0.08	2271	3.12	3660	0.06	1983	2.62	3269	0.27	6848	7.68	8311
Truth	0.10	1548	1.88	2624	0.08	1361	1.59	2338	0.30	5293	5.61	6593
Truth 2 years	0.23	918	0.88	1336	0.19	943	0.84	1357	0.30	4256	4.32	5506
Crop modeling	0.32	711	0.57	1020	0.25	921	0.91	1250	0.28	1810	1.87	2751
Truth + crop modeling	0.14	1195	1.22	1859	0.12	1110	1.06	1714	0.31	3325	3.48	4321
Truth 2 years + crop modeling	0.22	888	0.81	1278	0.18	993	0.80	1331	0.30	3083	3.18	4084

Comparison of overall accuracy of the models

The transfer learning approach, crop type classification models for next growing seasons, training without year data and testing with year data, presented different accuracies for the unsupervised and supervised learning. Unsupervised learning obtained the lower performance with overall accuracy equals to 0.86 (Fig. 9A). For the ground truth data to train the supervised learning, when the data collection of two growing seasons were aggregated, the overall accuracy increased from 0.89 to 0.91, and decreased standard deviation in the overall accuracy when three versus one growing season were tested (Fig. 9A). Crop modeling obtained greater performance in transfer learning crop classification models, achieving overall accuracies greater than 0.92 and the lowest standard deviations. Furthermore, when two growing seasons years data were aggregated to historical crop modeling, the model presented the greater overall accuracy of 0.93 (Fig. 9A). The confusion matrix of the test data for each model showed balanced error between the two crops. Furthermore, the confusion matrix showed lower values of false positives and false negatives for truth 2 years + crop modeling, resulting in the high accuracy achieved in the study (Fig. 9B).

Sample size and transfer learning

Sample size have affected the model performance for transfer learning approach differently for the source of the data. Generally, training the model with 50 samples achieved

an accuracy of ~0.85 for transfer learning for next growing season, while increasing the sample size until 1000 samples increased accuracy up to ~0.94 (Fig. 9C). However, when the source of data utilized to train and transfer learning to next growing season were evaluated, crop modeling showed greater accuracy compared with ground truth data of 2 years, and 1 year when the sample size was reduced (Fig. 9C).

Training the model with year data for crop classification improved model performance when sample size has increased from 50 to 250, with overall accuracy from 0.86 to 0.92, respectively (Fig. 9D). Aggregating historical ground truth data or historical crop modeling data presented the most stable strategy in overall accuracy for crop type classification. Balanced data with 50 samples of historical data aggregated with 200 samples of year data, (50_200), presented similar results related to 200 samples of historical data with 50 samples of year data, (200_50) (Fig. 9E). Related to the source of data to aggregate with year data, crop modeling featured prominently in overall accuracy related to historical of 1- and 2-years data, with overall accuracy, with ~0.91 overall accuracy (Fig. 9E).

Spatial variability of Sentinel-2 harmonic regression coefficients

The features of Sentinel-2 reflected the variability of the collected ground truth data points for soybean and corn crops (Fig. 9F, and H). The geospatial analyses with Moran´s I index showed spatial dependence for the two crops, with Moran´s I index equal to 0.20 and 0.23, for soybean and corn, respectively. The cluster k-means analysis resulted in k = 2 for both, soybean and corn (Fig. 9G, and I). The clusters indicate the variability from the features from Sentinel-2 satellite data, denoting an indirect spatial variability in climate, soils and crop management such as sowing date. The model performance decreased overall accuracy from 0.91 to 0.84 due to the lack of consideration of the spatial variability of the Northwest region in Rio Grande do Sul, Brazil (Fig. 9J). However, the model consisting in both crop modeling (150 samples) with year data (100 samples) improved from 0.91 to 0.93 in overall accuracy when the spatial variability component was included (Fig. 9K).

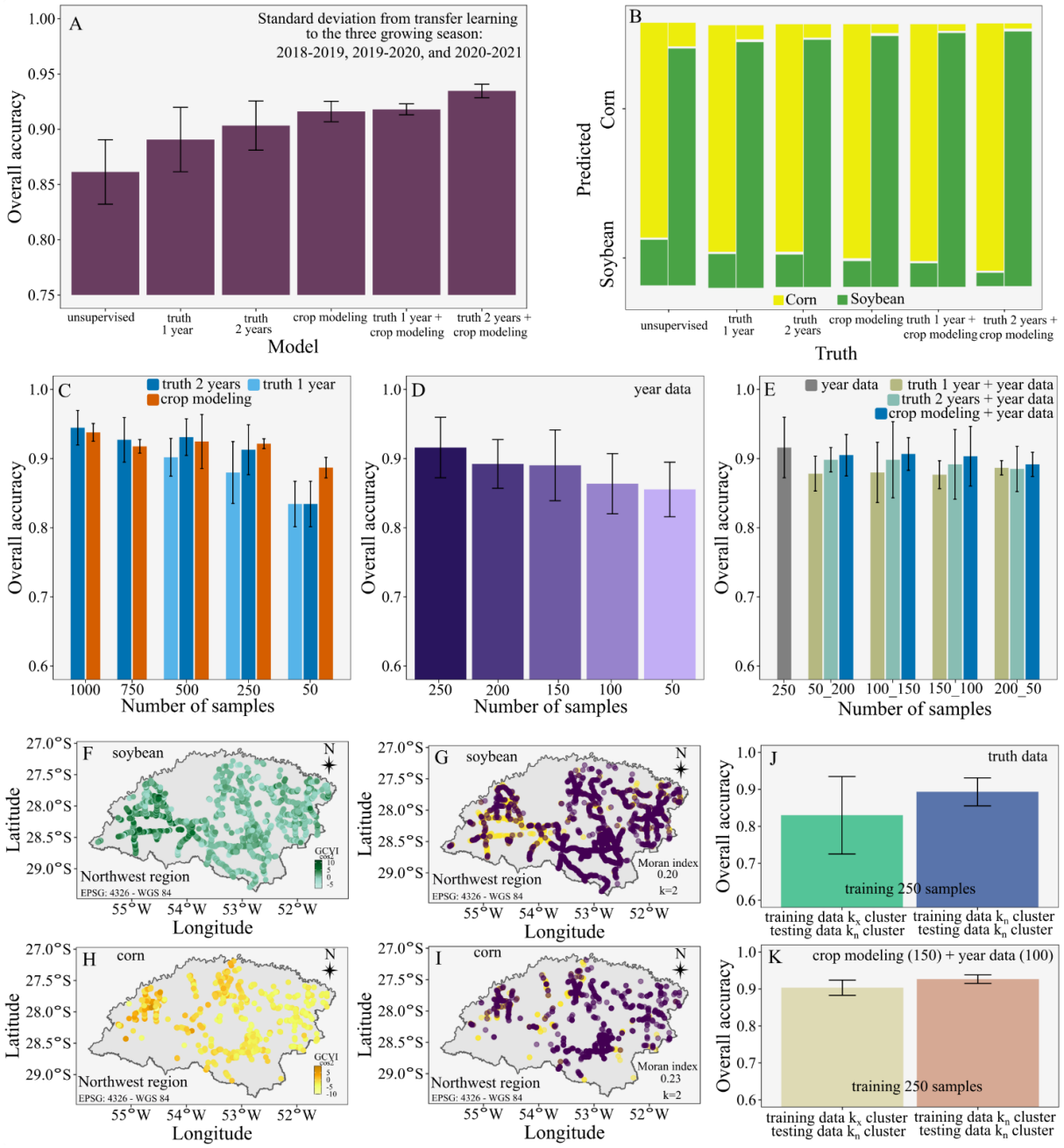


Fig. 9. (A) Comparison of overall accuracy for transfer learning models of the three growing seasons -2018-2019, 2019-2020, and 2020-2021. (B) Confusion matrices for the test data of the six models. (C) Comparison of overall accuracy of sample sizes for transfer learning approach with ground truth with 1-, and 2-years data, and crop modeling data. (D) Comparison of overall accuracy of sample sizes for year data (ground truth data). (E) Comparison of overall accuracy of balanced ground truth with 1-, and 2-years data, and crop modeling data aggregating year data to the training model. All the models with 250 sample size. The first number (x_) in the x axis correspond the truth of 1-, and 2-years data and crop

modeling, while the second one is related to the year data (_x). Spatial variabilities of the features from harmonic regression of Sentinel-2 GCVI for soybean (F) and corn (H). Moran's Index and Cluster k-means analyses for soybean (G) and corn (I). (J) Overall accuracy of the ground truth data regarding the different clusters sampling data, and (K) overall accuracy of crop modeling (150 samples) + year data (100 samples) regarding the different clusters sampling data. Models utilizing k_x training data and k_n testing data (first bars), and models utilizing k_n training data and k_n validation data (second bars). Whiskers in all the panels denotes the standard deviation.

DISCUSSION

Main outcomes

This study provides new insights on the integration and testing of unsupervised learning method with regional statistics, supervised learning model with ground truth data and with crop modeling and testing their combinations achieving high performance, overall accuracy of 0.94 – 1000 samples. Crop type classification with crop area estimation and crop mapping were generated for the entire interested region (Fig. 10).

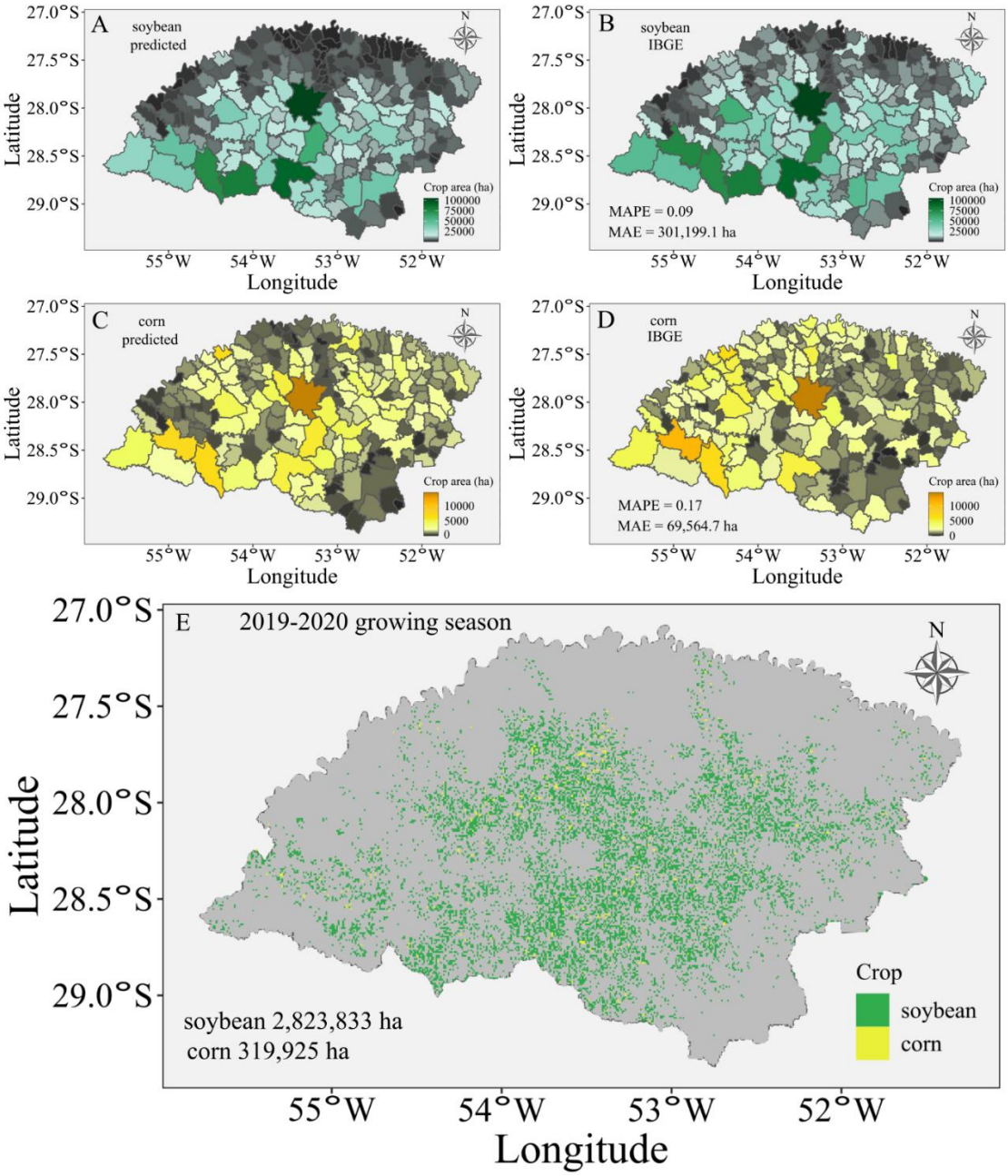


Fig 10. (A) Crop municipality area map predicted by the crop modeling model from 2019-2020 growing season for soybean, and map with IBGE statistics (B). (C) Crop municipality area map predicted by the crop modeling model from 2019-2020 growing season for corn, and map with IBGE statistics (D). (E) Rio Grande do Sul state, Northwest region crop map of 2019–2020 growing season. Mean absolute percentage error (MAPE) and mean absolute error (MAE) values are for the sum of crops area in the region.

Feature engineering

The harmonic regressions as input for the machine learning methods obtained high performance for crop type classification as well as for the agricultural crop masks with the random forest model. Likewise, the feature engineering with the coefficients of the harmonic regressions provided high performance for crop type classification in recent past studies in US (Jakubauskas et al., 2002; Wang et al., 2019; 2020; Dado et al., 2020; Deines et al., 2021) and Brazil (Pott et al., 2021).

Agricultural masks

High resolution of LULC for all biomes is still a gap information for Brazil, that is the fifth biggest country in the world. In our study, the random forest model using harmonic regression from Sentinel-2 presented the highest accuracy for the agricultural masks compared with MapBiomas and Environmental Rural Register aiming crop area estimation. The MapBiomas utilizes time-series information in the freely available Landsat data based on GEE cloud-computing to generate temporal changes in the LULC Brazilian Biomes (Souza et al., 2020). The Environmental Rural Register database relates a more environmental and economic planning aiming to track deforestation (Sicar, 2021), including manual and individual delineation of consolidated areas for all rural farms, however with possility of human errors in the delineation. Recently, the Dynamic World V1 product was made available in GEE, which is a near real-time 10 m spatial resolution associated with Sentinel-2 imagery for global land use and cover monitoring dataset that includes class probabilities and label information for nine classes, including agricultural crops. The composites of Dynamic World for user-specified data ranges and continuous products can benefit LULC classification around the globe (Brown et al., 2022).

Advantages of crop type classification models

Related to the different crop type classification models, when field ground truth data is not publicly available or is difficult to acquire, the unsupervised learning with regional statistics or supervised learning with crop modeling simulations may develop crop type classification to generate maps for crop area predictions. Furthermore, this current study reported that when growing seasons ground field data are available and aggregated with crop modeling yielded in more consistently accurate classification. In addition, when historical crop modeling data (150 samples) was integrated with few year data (100 samples)

considering the spatial variability of the region, the model presented the highest performance, reducing labor cost, time and improving efficiency for crop type classification to next growing seasons.

Crop type classification utilizing remote sensing data have been proposed in several regions around the world (Wang et al., 2019; Pott et al., 2021; Deines et al., 2021). Wang et al. (2019) have studied crop type mapping with unsupervised and supervised learning utilizing images from Landsat, to transfer learning for various states and across years in US, reporting similar results matched we found in our study. Most of the studies have developed supervised learning with ground truth data to implement machine learning with images from remote sensing (Cai et al., 2018; Wang et al., 2019; Fowler et al., 2020; Wang et al., 2020; Pott et al., 2021). Crop type classification with supervised learning utilizing crop modeling is not well addressed in the scientific literature yet. However, crop modeling approaches have been successfully utilized for yield predictions with remote sensing (Lobell et al., 2015; Jin et al., 2017; Dado et al., 2020; Deines et al., 2021).

Limitations and future studies

Some limitations of this study are related to the scale-up this methodology of aggregate crop modeling simulations for new regions, where there are other major crops with uncalibrated parameters in crop modeling frameworks to generate the simulations.

There are ways to further improve the current crop modeling simulations in terms of management factors such as selecting specifics sowing dates according to the growing season of interest (incorporate data along the sowing period).

Future studies should focus on field research studies with the goal of assisting crop modeling in the development of more regional focused calibration of Brazilian genetic cultivars and testing new spectral bands/vegetation indices for prediction of crop phenology and monitoring changes in crop vegetation. Moreover, studies should focus on including ancillary data statistics about sowing dates and other relevant crop management factors, and evaluating different crop models in an ensemble learning approach. Besides, studies should focus on develop empirical relationships of different crops development such as LAI, and biomass with vegetation indices for specific regions in Brazil to improve the linkage of crop modeling with remote sensing. Future work could also extend these approaches to aiming in-season forecasting of crop area and predictions of total crop production.

CONCLUSIONS

This study introduced different methods for crop type classification with different resources for remote sensing-based field-level crop mapping. Our results showed that crop type classification maps improved when ground truth data was available and aggregated to crop modeling, even more when taking into account data quantity and spatial distribution for crop type classification to the next growing seasons.

In addition, coefficients of harmonic regressions derived from crop modeling presented similarity to those retrieved from Sentinel-2 data for both soybean and corn. Agricultural masks showed high efficiency for crop area estimation for soybean crop, with higher accuracy for the random forest model utilizing GCVI harmonic regression coefficient extraction from Sentinel-2 images, while for corn did not present good accuracy. The supervised learning model with crop modeling aggregating ground truth data of growing seasons obtained the greater overall accuracy. Lastly, crop area prediction was most accurate for soybean and least for corn.

In summary, we have generated an end-to-end approach to build crop type maps utilizing unsupervised models, crop modeling to provide simulations of labeled crop data, and even a combination of crop modeling with ground truth data to increase supervised model performance. The methods in this study may be applied to other regions, and countries around the globe, as well as for other crops.

CRediT authorship contribution statement

Luan Pierre Pott: Conceptualization of this study, Methodology, Software, Formal analysis, Investigation, Data Curation Visualization, Validation, Writing - Original Draft. Telmo Jorge Carneiro Amado: Validation, Writing - Review & Editing. Raí Augusto Schwalbert: Investigation, Writing - Review & Editing. Geomar Mateus Corassa: Investigation, Writing - Review & Editing. Ignacio Antonio Ciampitti: Conceptualization of this study, Investigation, Methodology, Supervision, Writing - Review & Editing.

Acknowledgements

We thank E.M.G and G.H. for helping in the data collection process. The first author thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance

Code 001. The second author thanks to CNPq for research scholarship n 305622/2017-0. Contribution n. X from the Kansas Agricultural Experiment Station.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift, 22*(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507.

Ångström, A. (1924). Solar and Terrestrial Radiation. *Quarterly Journal of the Royal Meteorological Society, 50,* 121-126. http://dx.doi.org/10.1002/qj.49705021008

Archontoulis, S. V., Miguez, F. E., Moore, K. J. (2014). A methodology and an optimization tool to calibrate phenology of short-day species included in the APSIM PLANT model: Application to soybean. *Environmental Modelling & Software*, *62*, 465–477. https://doi.org/10.1016/j.envsoft.2014.04.009

Bargiel, D. (2017). A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment*, *198*, 369–383. https://doi.org/10.1016/j.rse.2017.06.022

Battisti, R., Sentelhas, P. C., Boote, K. J. (2017). Inter-comparison of performance of soybean crop simulation models and their ensemble in southern Brazil. *Field Crops Research, 200,* 28–37. https://doi.org/10.1016/j.fcr.2016.10.004

Bexaira, K. P., Streck, N. A., Cera, J. C., Prestes, S. D. (2018). Coeficientes de Angström-Prescott para Estimar a Radiação Solar no Rio Grande do Sul. *Revista Brasileira de Meteorologia, 33(3),* 401–411. https://doi.org/10.1590/0102-7786333001

Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., Li, S. (2019). FNN: Fast Nearest Neighbor search algorithms and applications. *R package version 1.1.3*. https://CRAN.R-project.org/package=FNN

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., … Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, *9*(1). https://doi.org/10.1038/s41597-022-01307-4

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z. (2018). A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, *210*, 35–47. https://doi.org/10.1016/j.rse.2018.02.045

CONAB, Companhia Nacional de Abastecimento. (2020). Calendário de Plantio e Colheita de Grãos no Brasil. Retrieved November 3, 2021, from Conab.gov.br website: https://www.conab.gov.br/institucional/publicacoes/outras-publicacoes/item/download/36427_9534db174ba2bcddb8bad4be22818839

Dado, W. T., Deines, J. M., Patel, R., Liang, S.-Z., Lobell, D. B. (2020). High-Resolution Soybean Yield Mapping Across the US Midwest Using Subfield Harvester Data. *Remote Sensing*, *12*(21), 3471. https://doi.org/10.3390/rs12213471

Dalgliesh, N., Hochman, Z., Huth, N., Holzworth, D. (2016). A protocol for the development of APSoil parameter values for use in APSIM. Version 4; CSIRO: Black Mountain, Australia.

Deines, J. M., Patel, R., Liang, S.-Z., Dado, W., Lobell, D. B. (2021). A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sensing of Environment*, *253*, 112174. https://doi.org/10.1016/j.rse.2020.112174

Duarte, Y. C. N., Sentelhas, P. C. (2019). Intercomparison and Performance of Maize Crop Models and Their Ensemble for Yield Simulations in Brazil. *International Journal of Plant Production*, *14*(1), 127–139. https://doi.org/10.1007/s42106-019-00073-5

Fowler, J., Waldner, F., Hochman, Z. (2020). All pixels are useful, but some are more useful: Efficient in situ data collection for crop-type mapping using sequential exploration methods. *International Journal of Applied Earth Observation and Geoinformation*, *91*, 102114. https://doi.org/10.1016/j.jag.2020.102114

Frantz, D., Haß, E., Uhl, A., Stoffels, J., Hill, J. (2018). Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. *Remote Sensing of Environment*, *215*, 471–481. https://doi.org/10.1016/j.rse.2018.04.046

Gitelson, A. A. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, *32*(8). https://doi.org/10.1029/2005gl022688

Hastie, T., Tibshirani, R., Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., … Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE, 12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Holzworth, D., Huth, N. I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., … Snow, V. (2018). APSIM Next Generation: Overcoming challenges in modernising a farming systems model. *Environmental Modelling & Software*, *103*, 43–51. https://doi.org/10.1016/j.envsoft.2018.02.002

IBGE, Instituto Brasileiro de Geografia e Estatística. (2020). Cidades e Estados. Retrieved November 3, 2021, from Ibge.gov.br website: https://www.ibge.gov.br/cidades-e-estados/rs/.

IBGE, Instituto Brasileiro de Geografia e Estatística. (2021). Área plantada, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras temporárias. Retrieved November 3, 2021, from Ibge.gov.br website: https://sidra.ibge.gov.br/tabela/1612#.

Jakubauskas, M. E., Legates, D. R., & Kastens, J. H. (2002). Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture, 37*(1-3), 127–139. https://doi.org/10.1016/s0168-1699(02)00116-3

Jin, Z., Azzari, G., Lobell, D. B. (2017). Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. *Agricultural and Forest Meteorology, 247*, 207–220. https://doi.org/10.1016/j.agrformet.2017.08.001

Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D. B. (2019). Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment, 228*, 115–128. https://doi.org/10.1016/j.rse.2019.04.016

Kabała, C., Musztyfaga, E., Gałka, B., Łabuńska, D., Mańczyńska, P. (2016). Conversion of Soil pH 1:2.5 KCl and 1:2.5 H2O to 1:5 H2O: Conclusions for Soil Management, Environmental Monitoring, and International Soil Databases. *Polish Journal of Environmental Studies, 25*(2), 647–653. https://doi.org/10.15244/pjoes/61549

Kluger, D. M., Wang, S., Lobell, D. B. (2021). Two shifts for crop mapping: Leveraging aggregate crop statistics to improve satellite-based maps in new regions. *Remote Sensing of Environment*, *262*, 112488. https://doi.org/10.1016/j.rse.2021.112488

Kuhn et al., (2020). Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. https://www.tidymodels.org

Kuhn, M., Vaughan, D. (2021). yardstick: Tidy characterizations of model performance. R package version 0.0.8. https://CRAN.R-project.org/package=yardstick

Kullback, S., Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694

Lin, C., Zhong, L., Song, X.-P., Dong, J., Lobell, D. B., & Jin, Z. (2022). Early- and in-season crop type mapping without current-year ground truth: Generating labels from historical information via a topology-based approach. *Remote Sensing of Environment*, *274*, 112994. https://doi.org/10.1016/j.rse.2022.112994

Lobell, D. B., Thau, D., Seifert, C., Engle, E., Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, *164*, 324–333. https://doi.org/10.1016/j.rse.2015.04.021

MapBiomas, Brasil. (2021). Retrieved November 8, 2021, from Mapbiomas.org website: https://mapbiomas.org/en

Nguy-Robertson, A., Gitelson, A., Peng, Y., Viña, A., Arkebauer, T., & Rundquist, D. (2012). Green Leaf Area Index Estimation in Maize and Soybean: Combining Vegetation Indices to Achieve Maximal Sensitivity. *Agronomy Journal*, *104*(5), 1336–1347. https://doi.org/10.2134/agronj2012.0065

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439. https://doi.org/10.32614/rj-2018-009

Picoli, M. C. A., Camara, G., Sanches, I., Simões, R., Carvalho, A., Maciel, A., … Almeida, C. (2018). Big earth observation time series analysis for monitoring Brazilian agriculture. *ISPRS Journal of Photogrammetry and Remote Sensing*, *145*, 328–339. https://doi.org/10.1016/j.isprsjprs.2018.08.007

Pott, L. P., Amado, T. J. C., Schwalbert, R. A., Corassa, G. M., Ciampitti, I. A. (2021). Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing*, *176*, 196–210. https://doi.org/10.1016/j.isprsjprs.2021.04.015

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RADAM Brazil Project, 1974. Levantamento de recursos naturais. Rio de Janeiro, 4.

Reichert, J. M., Albuquerque, J. A., Kaiser, D. R., Reinert, D. J., Urach, F. L., & Carlesso, R. (2009). Estimation of water retention and availability in soils of Rio Grande do Sul. *Revista Brasileira de Ciência Do Solo*, *33*(6), 1547–1560. https://doi.org/10.1590/s0100-06832009000600004

Santos, H.G., Jacomine, P.K.T., Dos Anjos, L.H.C., De Oliveira, V.A., Lumbreras, J.F., Coelho, M.R., Cunha, T.J.F. (2018). Sistema brasileiro de classificação de solos. Brasília, DF: Embrapa, 2018.

Sparks, A. (2018). nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. *Journal of Open Source Software*, *3*(30), 1035. https://doi.org/10.21105/joss.01035

Sicar, Sistema Nacional de Cadastro Ambiental Rural (2021). Retrieved September 16, 2021, from Car.gov.br website: https://www.car.gov.br/publico/imoveis/index

Soil Survey Staff (2014). Keys to soil taxonomy, 12th ed. USDA-Natural Resources Conservation Service, Washington, DC.

Souza, C. M., Z. Shimbo, J., Rosa, M. R., Parente, L. L., A. Alencar, A., Rudorff, B. F. T., … Azevedo, T. (2020). Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sensing*, *12*(17), 2735. https://doi.org/10.3390/rs12172735

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. Roy. Statist. Soc.: Series B (Statistical Methodology) 63 (2), 411–423. https://doi.org/10.1111/1467-9868.00293.

Wang, S., Azzari, G., Lobell, D. B. (2019). Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, *222*, 303–317. https://doi.org/10.1016/j.rse.2018.12.026

Wang, S., Di Tommaso, S., Deines, J.M., Lobell, D.B. (2020). Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. *Scientific Data*, *7*(1). https://doi.org/10.1038/s41597-020-00646-4

Willmott, C.J. (1981). On the validation of models. *Physical Geography, 2,* 184-194. https://doi.org/10.1080/02723646.1981.10642213

Yan, S., Yao, X., Zhu, D., Liu, D., Zhang, L., Yu, G., … Yun, W. (2021). Large-scale crop mapping from multi-source optical satellite imageries using machine learning with discrete grids. *International Journal of Applied Earth Observation and Geoinformation, 103*, 102485. https://doi.org/10.1016/j.jag.2021.102485

Yan, Y., Ryu, Y. (2021). Exploring Google Street View with deep learning for crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing, 171*, 278–296. https://doi.org/10.1016/j.isprsjprs.2020.11.022

Zhang, C., Di, L., Hao, P., Yang, Z., Lin, L., Zhao, H., & Guo, L. (2021). Rapid in-season mapping of corn and soybeans using machine-learned trusted pixels from Cropland Data Layer. *International Journal of Applied Earth Observation and Geoinformation, 102*, 102374. https://doi.org/10.1016/j.jag.2021.102374

4 ARTIGO 3 - MAPPING CROP ROTATION BY SATELLITE-BASED DATA FUSION IN SOUTHERN BRAZIL

Artigo originalmente publicado e seguindo as normas da Revista Computers and Electronics in Agriculture.

Pott, L. P., Amado, T. J. C., Schwalbert, R. A., Corassa, G. M., & Ciampitti, I. A. (2023). Mapping crop rotation by satellite-based data fusion in Southern Brazil. *Computers and Electronics in Agriculture*, 211, 107958.

ABSTRACT

Crop monitoring is a key process for agricultural management and policy making linked to food security and sustainability. The spatio-temporal assessment of crop-specific management is essential for mapping crop rotation at field-scale. The aims of this study were to generate a satellite-based data fusion approach for mapping crop rotation at field-scale and generating analyses of crop rotation patterns among different mesoregions, examining the impact of environmental factors on crop yields, within the Rio Grande do Sul, southern Brazil. Within this geographical region, soybeans (*Glycine max* (L.) Merr.), corn (*Zea mays* L.), and rice (*Oryza sativa* L.) are the major grain crops of the state. We have utilized a satellite-based data fusion crop type classification and mapping to extract spatial crop features during four growing seasons (2017-2021), associated to field boundary delineation to generate the crop rotation database. This database showed current crop rotation practices within the study region, highlighting continuous soybean (monocrop), and three soybeans with one corn year rotation as the most frequent in the state. Midwestern and Northwest mesoregions presented the highest mean municipality values of continuous crop with 76% and 61%, respectively. Crop rotation patterns in Porto Alegre metropolitan, Southwest, and Southern mesoregions showed soybean-rice rotation, but with a trend of increases in soybean area in detriment to rice in lowlands. Crop rotation effects on yield for soybeans varied from 20 up to 65% depending on the regions and years of crop rotation. Other crop rotation effects in the crop yields were not significant at the municipality level. Crop rotation showed more significant benefits for soybean yields in warm and wet climate, with higher bulk density and lower soil

organic carbon. Lastly, this study presents the first crop rotation map layer for the entire state of Rio Grande do Sul, southern Brazil, serving as a foundation for the creation of similar database for other states in the country and around the globe.
Keywords: Crop rotation map, Crop rotation pattern, Continuous crop, Crop rotation effect, Crop classification.

INTRODUCTION

Food security requires pursuing sustainable management practices for increasing crop yields over time in order to follow the pace of the population growth. Crop rotation can be defined as a practice of growing different crops in the same area across seasons following a rational plan in order to get synergism among crops. This management has been practiced for thousands of years, reporting benefits for soil health, soil biological activity, sustainability, pest and weed control, and crop yields gains (Kassam et al., 2018; Barbieri et al., 2019; Garbelini et al., 2022).

Remote sensing-based satellite are useful tool to monitor cropland (Souza et al., 2020; Brown et al., 2022), allowing track tillage system (Azzari et al., 2019), cover crop adoption (Deines et al., 2022), crop type classification (Cai et al., 2018; Wang et al., 2020; Pott et al., 2021; Pott et al., 2022), and crop yield forecasting (Lobell et al., 2015; Schwalbert et al., 2020; Deines et al., 2021), among other applications. Mapping crop rotation requires a precise crop type classification to capture temporal dynamics at field-scale. Waldhoff et al. (2017) have mapped crop rotation utilizing multitemporal remote sensing and geospatial based crop mapping in Germany. Liu et al. (2021) proposed remote sensing and deep learning to capture both crop types and temporal changes for crop rotation in China. However, reliable and timely crop rotation maps at field-scale around the globe are still rarely available.

The crop rotations plans adopted by the farmers may be motivated by regional characteristics, crop practices, cultural, social and economic aspects (Steinmann and Dobers, 2013; Sahajpal et al., 2014; Stein and Steinmann, 2018). Regional aspects are related to climate as temperature, precipitation and solar radiation, crop varieties well adapted, specific soil characteristics as nutrients and water availability besides topography and drainage, such as demanded for flooded rice (*Oryza sativa* L.). Crop management practices can limit crop rotation due cultural aspects as the experience in growing specific crops, need of specific

machinery, time and labor availability, and growing knowledge that might not be available. In addition, sanitary disadvantageous like considering the entire growing season (summer and winter) in Rio Grande do Sul, the crop rotation with canola (*Brassica napus* L.) and soybean (*Glycine max* (L.) Merr.) can promote *Sclerotinia sclerotiorum* affecting soybean yields in without disease control fields in the summer (Tomm, 2008; Passinato et al., 2021). Favorable economic return has led to a recent increase in rice-soybean rotation in lowland areas that demand specific crop management as fast drainage during soybean season and water retention during rice season (IRGA, 2021; CONAB, 2022).

Crop rotation can benefit soil health by promote changes in physical as soil aggregation, root penetration and aeration and/or chemical characteristics as availability of nitrogen, phosphorous, potassium, and other soil nutrients that are cycled more efficiently decreasing their loss or by increasing bioavailability (Sarwar et al., 2008; Cohen et al., 2019; Yu et al., 2022). Furthermore, crop rotation can restore the soil organic carbon (SOC) content that enhance overall soil health by promoting biological activity and mitigate plant effects of drought spells (Paustian et al., 2016; Degani et al., 2019). Seifert et al. (2017) reported a greater crop rotation benefit for corn (*Zea mays* L.) in dry areas, while for soybean the greater effect was reported in cooler and wetter areas. Kluger et al. (2022) showed lower benefits of crop rotation on corn yield in seasons and locations with high temperatures, differently of observed for soybean that had higher crop yield benefits in high temperatures.

Crop rotation effects on grain yield have been conducted at different scale, based on a few field trials (Edwards et al., 1988; Nafziger, 2007; Erickson, 2008; Ribas et al., 2021), carried out on large number of farmer field trails (Seifert et al., 2017), with remote sensing (Cohen et al., 2019; Kluger et al. 2022), and globally through systematic review (Zhao et al., 2022). However, the complex interaction of genotype x environment × management (GxExM) creates a growing demand of crop rotation databases with high spatial resolution for guiding farm management and facilitating implementation of new policies towards regenerative agriculture.

The research question of this study is: Can crop rotation mapping be generated by a satellite-based data fusion approach at field-based within a relevant crop region in southern Brazil? Thus, the objectives of this study were to: i) generate a satellite-based data fusion approach for mapping crop rotation at field-scale, ii) quantify and identify patterns of crop rotation within and between different mesoregions, iii) describe the major environmental

factors linked to patterns in practices of crop rotations, and iv) examine crop rotation effects on yield, all focused on Rio Grande do Sul state, southern Brazil.

MATERIAL AND METHODS

Study area

The study area comprised the entire Rio Grande do Sul state, extended from 27º to 34º S and from 50º to 58º W, in southern Brazil (Fig 1A). Rio Grande do Sul consists of approximately 281.7km², with 497 municipalities distributed in seven mesoregions (IBGE, 2021). The environmental conditions in the state are wide variety in terms of soils with predominance of Oxisols, Alfisols, Mollisols, Ultisols, Entisols and Inceptisols (Santos et al., 2018; Soil Survey Staff, 2014), and climate types with majority region of Humid subtropical without dry season with hot summer (Cfa), and minor region with temperate summer (Cfb) (Alvares et al., 2013). The three major summer crops in Rio Grande do Sul are soybean and corn, besides flooded rice in lowlands, representing totally ~95% of summer crop area, with ~83% of the total crop area for the last years in the state (IBGE 2022). Soybean cropping area in the last four years of Rio Grande do Sul was ~5.7 Mha (~16% national area), rice area was ~1 Mha (~56% of national area), and corn area was 0.76 Mha (~4% national area) (IBGE 2022).

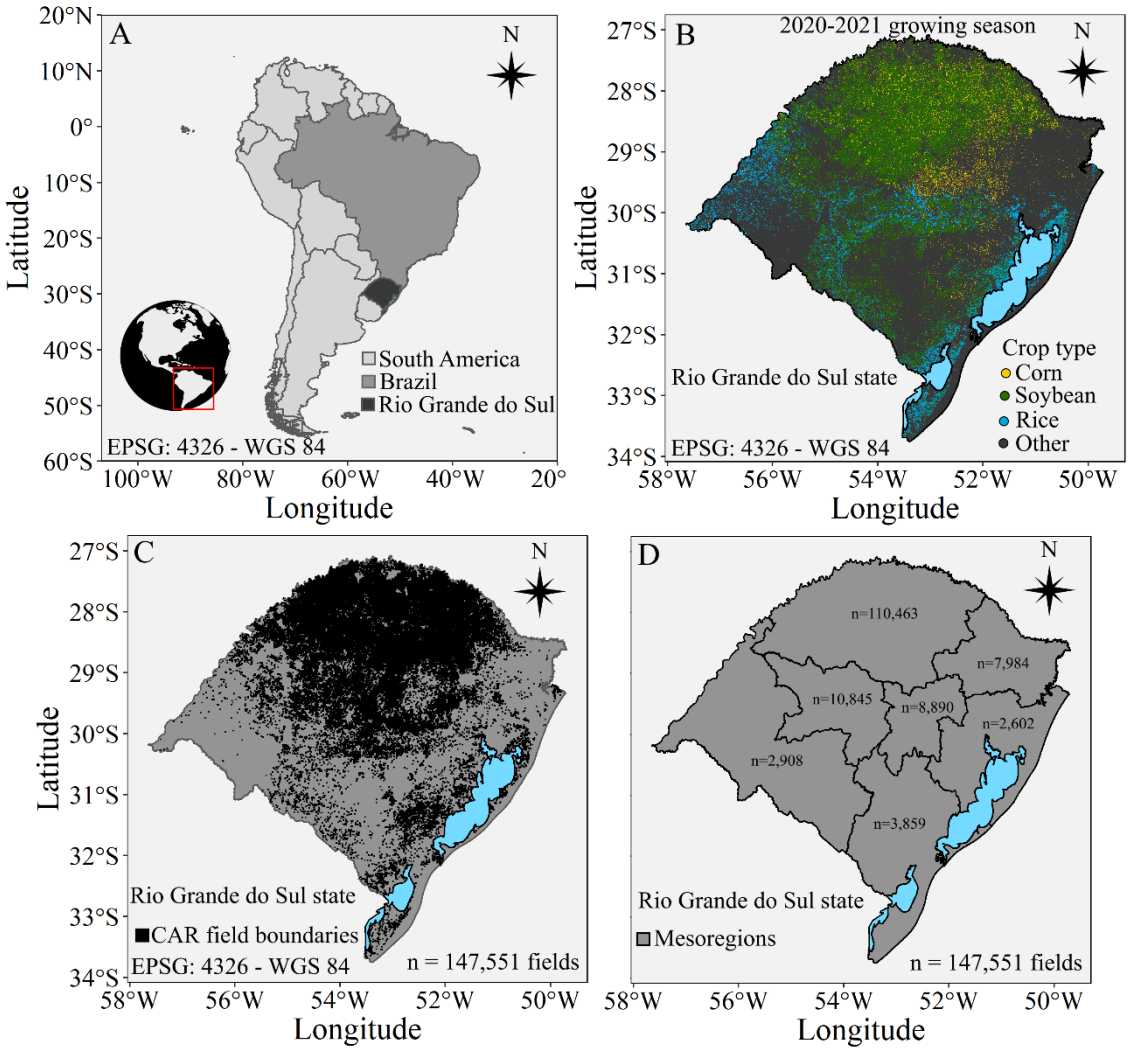


Fig. 1. (A) South America map, Brazil (shaded in grey), and Rio Grande do Sul state (shaded in black). (B) Crop type map of Rio Grande do Sul in 2020-2021 growing season (Pott et al., 2021). (C) Field boundaries of consolidated agricultural areas extracted from Cadastro Ambiental Rural (CAR) in all municipalities of Rio Grande do Sul. (D) Mesoregions of the Rio Grande do Sul state with the number of field boundaries extracted from CAR for each region.

Crop type classification in Rio Grande do Sul state, Brazil

To generate a field-scale crop rotation on a region scale requires large amount of crop rotation labels. As an alternative method, a satellite-based crop classification model may be utilized as crop type information in the growing seasons of interest. For some regions in the world a yearly crop classification is publicly available such as United States (US) where the

Cropland Data Layer (CDL) is released since 1997 by the United Sates Department of Agriculture (USDA, 2022) and Argentine where a 30 m resolution layer is available since the 2018/2019 growing season (Abelleyra and Verón, 2020). Other regions like the European Union also have available such layers but a less frequent release, every 3 years since 2006 (d'Andrimont et al. 2020). Brazil however did not have an official cropland data layer with the three summer major crops in Rio Grande do Sul state with high-resolution to provide field-level information. We have utilized the crop type classification map from the previous study of Pott et al. (2021) that performed a satellite-based data fusion model for mapping the crop type in the entire Rio Grande do Sul state (Fig 1B). Pott et al. (2021) had utilized ground truth data to access Sentinel-2, Sentinel-1 and Shuttle Radar Topographic Mission (SRTM) digital elevation data in a random forest machine learning to perform crop classification spatially coordinated with in-season predictions. The crop type classification from Pott et al. (2021) achieved an overall accuracy of 0.95 with a data fusion approach utilizing the growing seasons time-series to extract harmonic regressions coefficients of bands and vegetation indices from Sentinel-2, mean monthly composition of VV, VH and VV/VH from Sentinel-1. In addition, terrain data such as elevation, slope, and aspect from SRTM were retrieved to include these inputs in the machine learning model. Further details about the classification method to compose the crop rotation database can be found at Pott et al. (2021).

Field crop data

The model developed in this paper considered field-level to generate the crop rotation database. The crop mapping generated utilizing model from Pott et al. (2021) have other class as a general class unifying urban, forest, water, native vegetation and other minor crops, however field delineation is not achieved since the other class is rasterized englobing many different boundary objects. Land use and land cover (LULC) for entire Brazil country is generated by the MapBiomas project that make available LULC products every year (MapBiomas, 2022). Other globally LULC such as Dynamic World (Brown et al., 2022) have potential to contribute with agricultural masks around the globe. Nevertheless, automated agricultural field boundary detection is a challenging task due to the local characteristics, irregular shape and mixed-cropping systems (Taravat et al., 2021). The Environmental Rural Register (*Cadastro Ambiental Rural* – CAR - http://www.car.gov.br) is a Brazilian nationwide electronic open database, with the purpose of integrating environmental

information from rural areas for environmental planning (Sicar, 2022). The CAR database includes geoinformation related to field boundary delineation of agricultural areas (Fig 1C) that were utilized with the crop type classification developed by Pott et al. (2021) to build the crop rotation database.

Crop rotation data

The crop rotation data has been generated for each field-level boundary delineation from CAR database. Rio Grande do Sul state has 7 geographical mesoregions: Northwest, Northeast, Midwestern, Central Eastern, Porto Alegre metropolitan, Southwest and Southeast. For each mesoregion, the amount of field boundaries retrieved from CAR database is presented in Fig 1D, totalizing 147,551 fields. We have extracted the field centroid of each field boundary (Fig 2-I) to relate the geographical coordinates with the crop type mapping from Pott et al. (2021) model for the 2017-2018, 2018-2019, 2019-2020, and 2020-2021 growing seasons (Fig 2-II). The analyses were carried utilizing *sf* (Pebesma, 2018) and *raster* (Hijmans, 2022) packages in R (R Core Team, 2022). Although the study analyzed the centroid of field boundaries, to validate our approach we have estimated the crop areas (crop type for each centroid with the polygon area of the field boundary) in comparison to actual area IBGE statistics for soybean, corn and rice, achieving high precision for soybean and rice, while for corn crop area obtained the lowest accuracy, as already reported in Pott et al. (2021) in the pixel-based crop area approach (Fig 3). Crop rotation types may differ from site to site with many different possibilities. Considering that, we have grouped the number of possibilities of crop rotation patterns in the four growing seasons of the study in the following four categories: continuous crop, one year rotation, two years rotation, and intensive rotation. Continuous crop is characterized with only one of the three crops in the same field in the four growing seasons (4-yr in a row). One year rotation is the crop rotation pattern of two crops, with only one year rotation. Two years rotation is two crops, with two years rotation. Intensive rotation is the category of two or three crops in the same field without repeating the same crop for the next growing season. The four categories were presented in percentage of crop rotation pattern for each municipality and mesoregion (Fig 2-III).

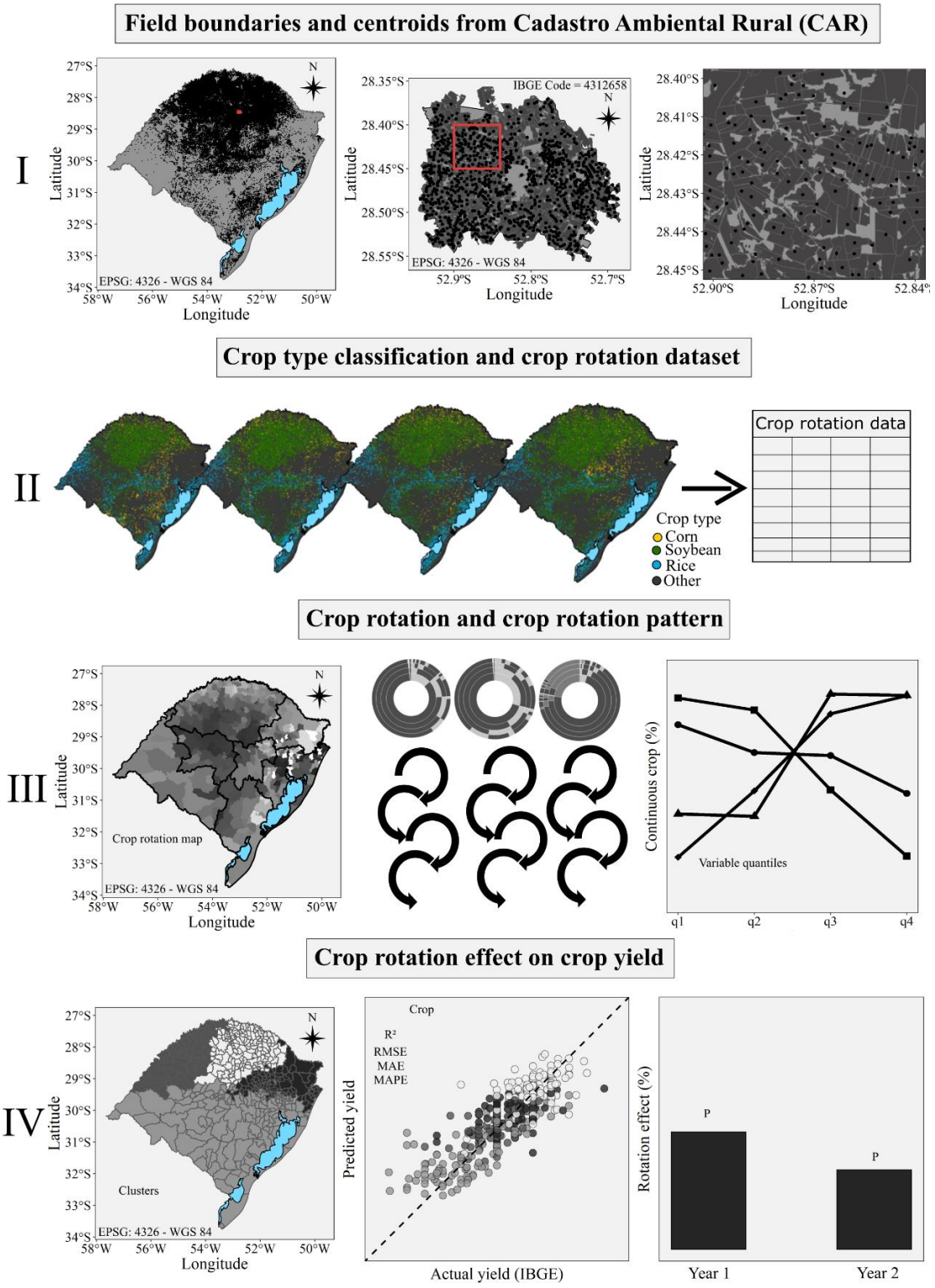


Fig. 2. Framework of the research study. I - Field boundaries and centroids of the polygons from Cadastro Ambiental Rural (CAR). II - Crop type classification layers (Pott et al., 2021) were assessed to compose the crop rotation dataset. III - Crop rotation and crop rotation pattern analyses. IV – Evaluation of crop rotation effect on crop yield.

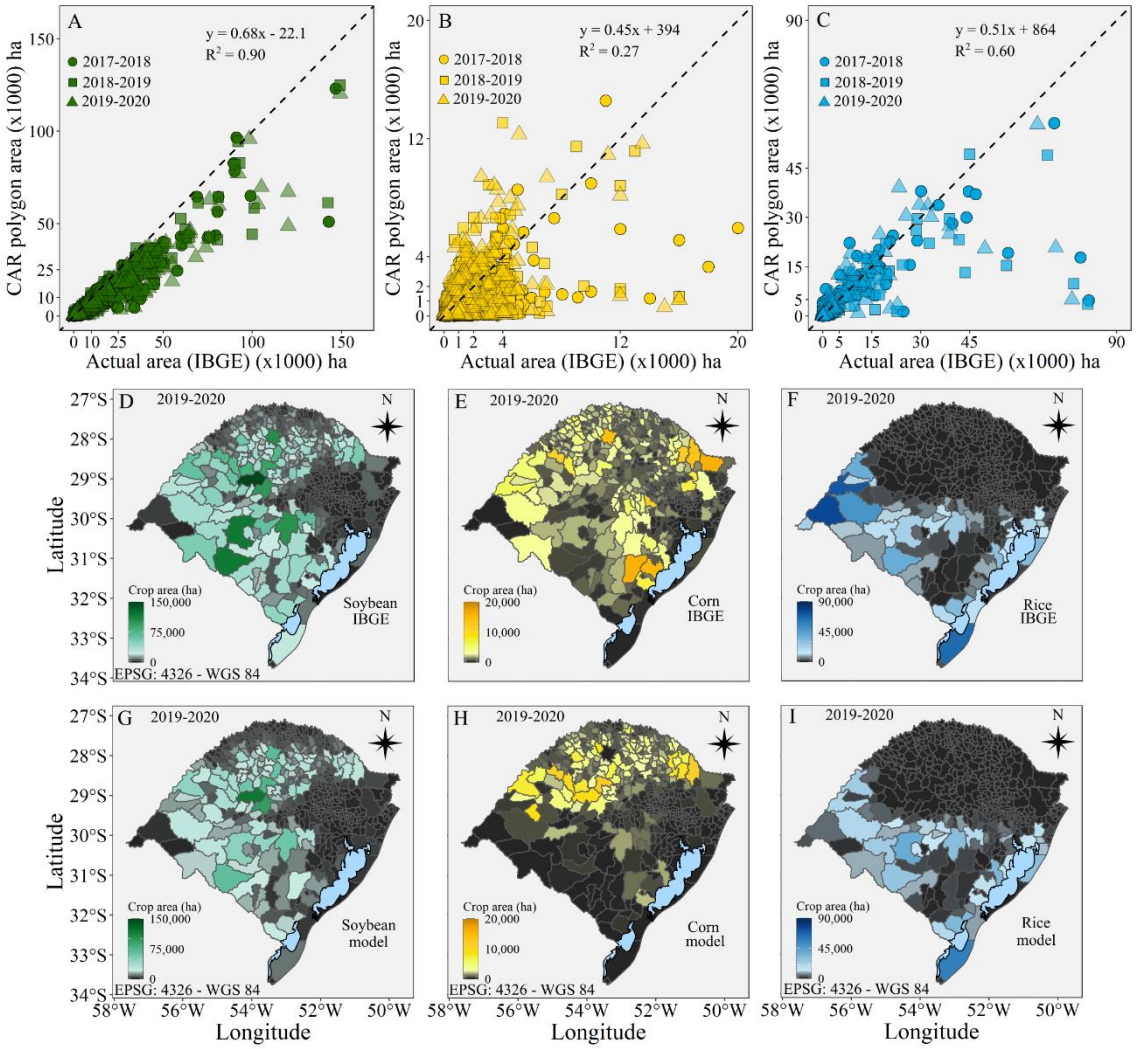


Fig. 3. (A) Relation 1:1 area of the field boundaries from Cadastro Ambiental Rural (CAR) related to crop map layers (Pott et al., 2021) and actual area IBGE statistics for soybean, (B) corn, and (C) rice. (D) Crop municipality area map of IBGE statistics from 2019-2020 growing season for soybean, (E) corn, and (F) rice. (G) Crop municipality area map of the field boundaries from Cadastro Ambiental Rural (CAR) related to crop map layers (Pott et al., 2021) from 2019-2020 growing season for soybean, (H) corn, and (I) rice.

Environmental data

Environmental data were retrieved to relate the interactions between the crop rotation pattern and quartiles of climate and soil variables at municipality level. We have utilized SoilGrids (250 m resolution) for the soil properties: texture (clay, sand, and silt), water

availability [drained upper limit (DUL), saturation (SAT), lower limit 15 (LL15), and air dry (AD)], bulk density, SOC, pH, and cation exchange capacity. For the climate, daytime land surface temperature (LST), from the Aqua satellite´s MODIS sensor (MYD11A2 product), that is strongly related to air temperatures (Mildrexler et al., 2011) were accessed and treated as mean month temperature. The MYD11A2 product provides an average 8-day land surface temperature (LST) with spatial resolution of ~1km. Daily precipitation data was retrieved by the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) treated as accumulated month rainfall with spatial resolution of ~5.5 km.

In order to analyze interactions between crop rotation pattern and environmental features, quartiles of environmental data features were utilized to compare with % continuous crop at municipality level (Fig 2-III).

Crop rotation effect

We have utilized the environmental data to cluster regions in Rio Grande do Sul state and characterize the regions for climate and soil features. Furthermore, we used multiple linear regression to examine relationships between yield and a set of climate, soil, and crop rotation covariates as follow Eq. (1).

$$yield = C_c + S_c + R_c + e \quad (1)$$

where C_c = climate covariates, S_c = soil covariates, R_c = Rotation covariates, and e = error. The linear regression approach to examine correlations between yield and a set of weather, soil, management practices covariates, in addition to rotation history have been used by Seifert et al. (2017) and Cohen et al. (2019) for crop rotation purposes.

Crop yields were retrieved from IBGE statistics at municipality level (IBGE, 2022). Climate covariates were mean month temperature, and accumulated month rainfall from July to April. Soil covariates were clay, sand, and silt, drained upper limit, saturation, lower limit 15, air dry, bulk density, SOC, pH, and cation exchange capacity. Rotation covariates were percentage of continuous crop, one year rotation, two years rotation, and intensive rotation at municipality level. We have treated correlated features to perform the multiple linear regression. We have generated the regression for 2018-2019, 2019-2020, and 2020-2021 growing seasons, once in 2017-2018 we did not have information about crop rotation (only one year). We have extracted the coefficient of the rotation covariates from the regressions for two years rotation (2018-2019), three years rotation (2019-2020), and four years rotation

(2020-2021) data as crop rotation effects for the entire Rio Grande do Sul state and for each cluster's regions (Fig 2-IV).

RESULTS

Crop summer rotation in Rio Grande do Sul state

Crop rotation mapping in Rio Grande do Sul presented mesoregions specific spatial patterns. High proportion of continuous crop were reported in Midwestern and Northwest mesoregions, with municipality mean of 76% and 61%, respectively (Fig 4A). In addition, these two mesoregions showed the lowest relative values for other categories: one year-, two years- rotation, and intensive rotation (Fig 4B, C, and D). Northeast, Central Eastern, Porto Alegre Metropolitan, Southwest, and Southeast had different patterns of crop rotation for each mesoregion (Fig 4). Fig A1 presents examples of crop rotation maps for municipalities in the study area.

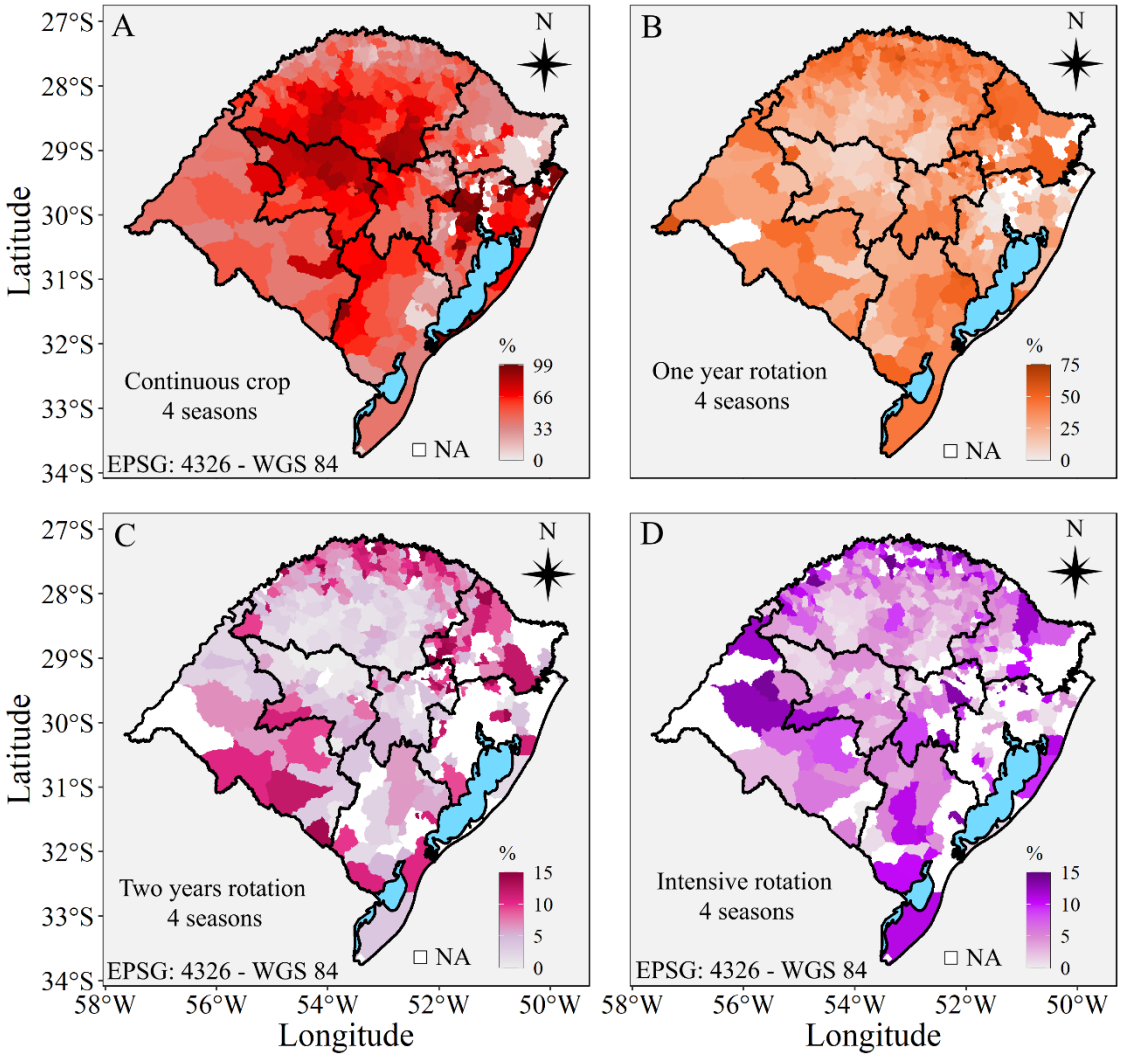


Fig. 4. (A) Crop rotation map of continuous crop, (B) one year rotation, (C) two years rotation, and (D) intensive crop rotation in the Rio Grande do Sul. The data is related to the percentage of the field boundaries area from Cadastro Ambiental Rural (CAR) of each municipality in Rio Grande do Sul, Brazil. The crop rotation data was observed for four continuous years. NA – not applicable.

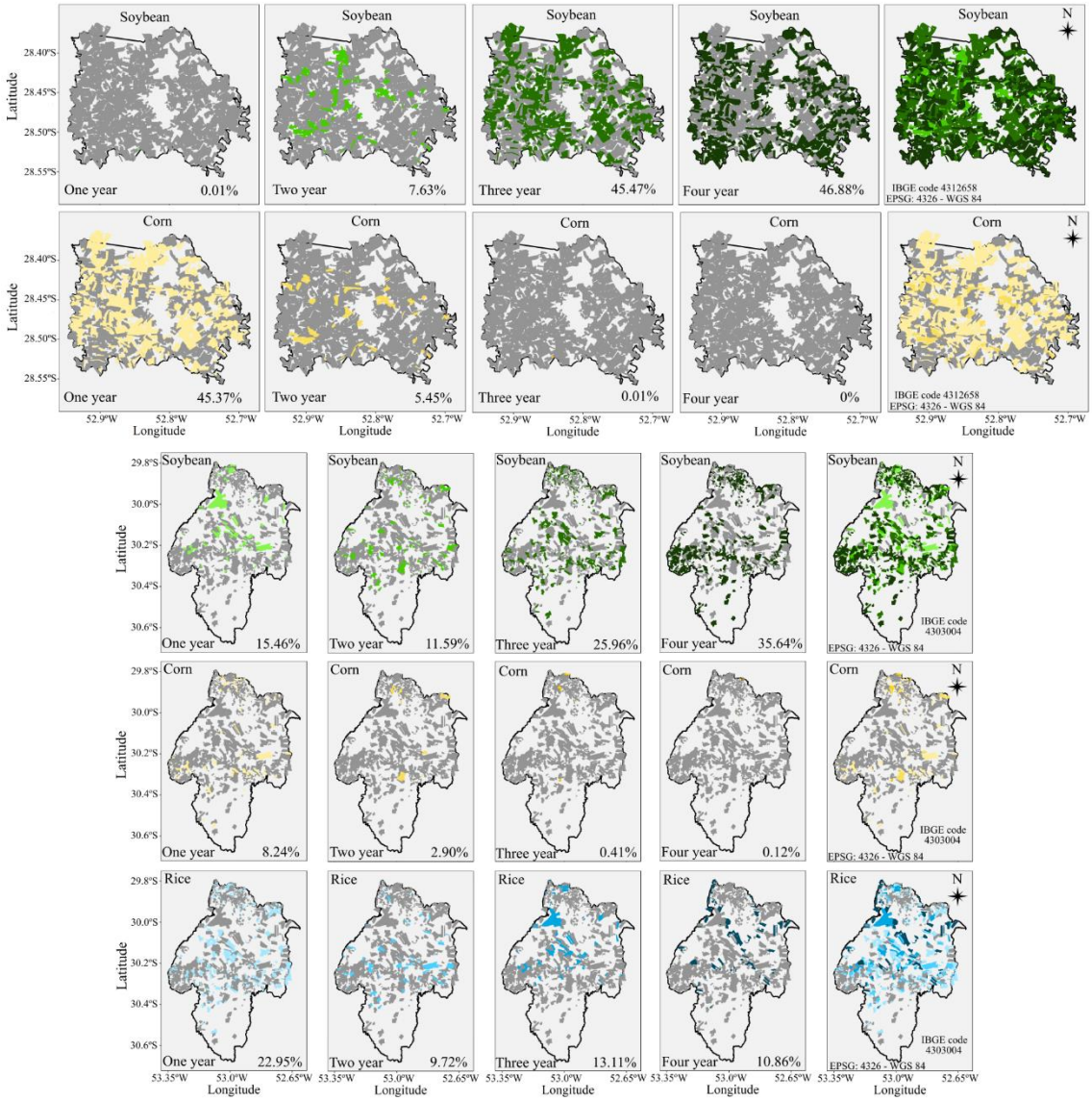


Fig. A1. Crop rotation maps of Não-Me-Toque (IBGE code 4312658), and Cachoeira do Sul in Rio Grande do Sul (IBGE code 4303004). The figure represents the crop rotation of soybean and corn for the first municipality, and soybean, corn and rice for the second one. The "n" year of the crop represent the quantity of each crop were grown in the field in total of 4 growing seasons. The last column of plots shows the maps of crops grown in one, two, three and four years. We have evaluated four years (2017-2018, 2018-2019, 2019-2020, 2020-2021).

Crop rotation patterns in the Rio Grande do Sul mesoregions

The mesoregions in Rio Grande do Sul showed different crop rotation patterns. Northwest mesoregion represents ~50% of the total crop areas of soybean and corn grown in

the state. In the Northwest, continuous crop and one year rotation sum up more than 90% of the crop rotation categories. The most utilized crop rotation pattern is continuous soybean (4yr- in a row), followed by one year rotation (3 years soybean and 1 year corn) (Fig 5A). Northeast mesoregion, mainly with corn and soybeans, presented the highest percentage of one year- and two years rotation related to the other mesoregions. One year rotation is the most common crop rotation practice (46%) with 3 years soybean - 1 corn, followed by continuous soybean (36%), highlighting also the 2 soybean - 2 corn years as two years rotation (10%) (Fig 5B). Midwestern mesoregion growth mostly soybean and rice, with low area of corn. Midwestern is the mesoregion with the highest percentage of continuous crops – soybean and rice (4yr- in a row) (76%) (Fig 5C). Central Eastern mesoregion is a more equality crop rotation with soybean, rice and corn. The most common practice is continuous crops (51%) followed by one year rotation (34.5%) (Fig 5D). Porto Alegre metropolitan mesoregion is characterized with the highest percentage of continuous rice (4yr- in a row) with 52.6% of the crop practice. One year rotation, two years rotation and intensive rotation had rice and soybean as crop rotation practices, showing decrease in rice area (~10%) and increase of soybean area (~34%) along the study growing seasons (Fig 5E). Southwest mesoregion showed continuous crop as the most common crop practice with 47%, followed by the one year rotation with soybean and corn crop rotations. Nevertheless, soybean crop area demonstrated an increase by ~27% with a decrease in rice by ~ 14% (Fig 5F). Southeast mesoregion presented a high proportion of continuous crop (44%) followed by one year rotation (38%). In addition, there were an increment in soybean area (~21%) with a reduction in rice area (~7%). In addition, in Southeast mesoregion had the highest intensive rotation practice relative to other regions (Fig 5G).

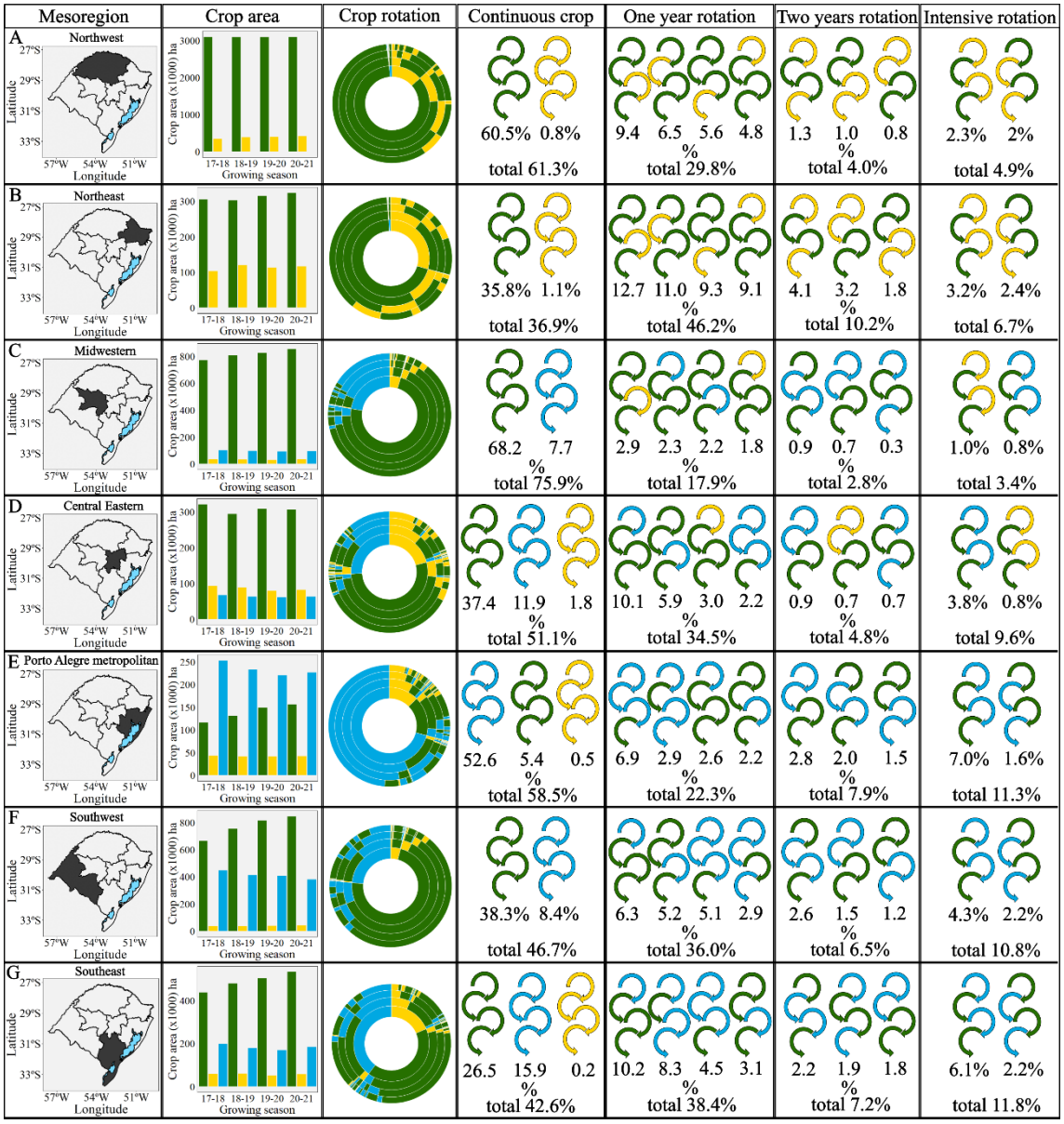


Fig. 5. Crop rotation patterns in the different mesoregions in Rio Grande do Sul, Brazil. Soybean (green), corn (yellow), and rice (blue). The mesoregions in the rows (A-G) are shaded in black. The bars plot refers to the sum of crop area of each mesoregion in the four growing seasons. The sunburst plots refer to the rotation line, 2017-2018, 2018-2019, 2019-2020, 2020-2021 growing seasons from into out circles, respectively. The rotation line, 2017-2018, 2018-2019, 2019-2020, 2020-2021 growing seasons, is up to bottom on continuous crop, one year rotation, two years rotation and intensive rotation columns.

Interactions of crop rotation with environmental data

Quartiles of climate and soil data covariates were related to continuous crop showing some degree of interactions. Related to climate covariates, higher temperatures from October to April generally are related to a large proportion of continuous summer crops. In addition, higher accumulated precipitation presented generally higher values of continuous crop (Fig 6A, and B). The relationships between continuous crop and soil covariates showed some patterns for soil physics attributes as texture, with high values of clay and silt content being related to low values of continuous crop, while high values for sand are linked to high proportion of continuous crop (Fig 6C). Water availability features such as DUL, SAT, LL15, and AD had negative relationship with continuous crop (Fig 6D). Nonetheless, bulk density showed positive relationship with the percentage of continuous crop (Fig 6E). SOC as the water availability covariates showed negative relationship with continuous crop (Fig 6F). The pH and cation exchange capacity did not show significant relationship with continuous crop rotation (Fig 6G, and H).

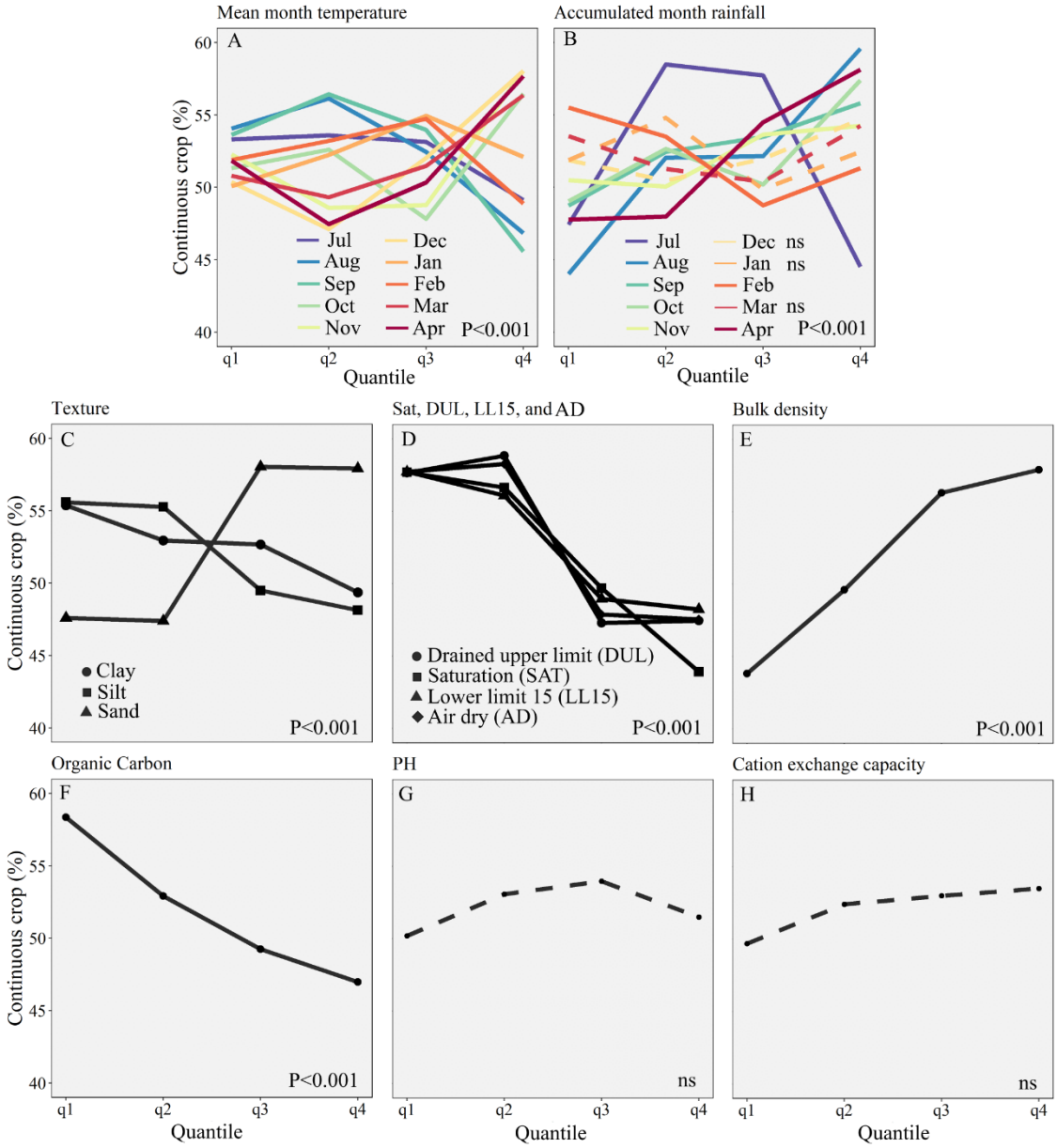


Fig. 6. Interactions between the continuous crop and quartiles of climate and soil variables. Solid lines indicate significant at P <0.001, dashed lines indicate no significant interaction. ns = not significant.

Cluster regions and crop rotation effect on crop yield

Rio Grande do Sul state presented geographical clustered regions according to soil and climate features showing four spatial regions (Fig 7A). Climate characterization for the different regions demonstrated warm and wet, warm and dry, cool and dry, and cool and wet scenarios. In general, Rio Grande do Sul state, Northwest region was warm and wet, North

region was warm and dry, and neutral in some year, South region was cool and dry, and Northeast region was cool and wet, compared to the mean of the four years study (Fig 7A, and B). Soil characterization showed variability with greater bulk density for Northwest and North regions, higher SAT for North and Northeast regions, greater SOC for Northeast region, and lower values of clay for South region (Fig 7A, and C).

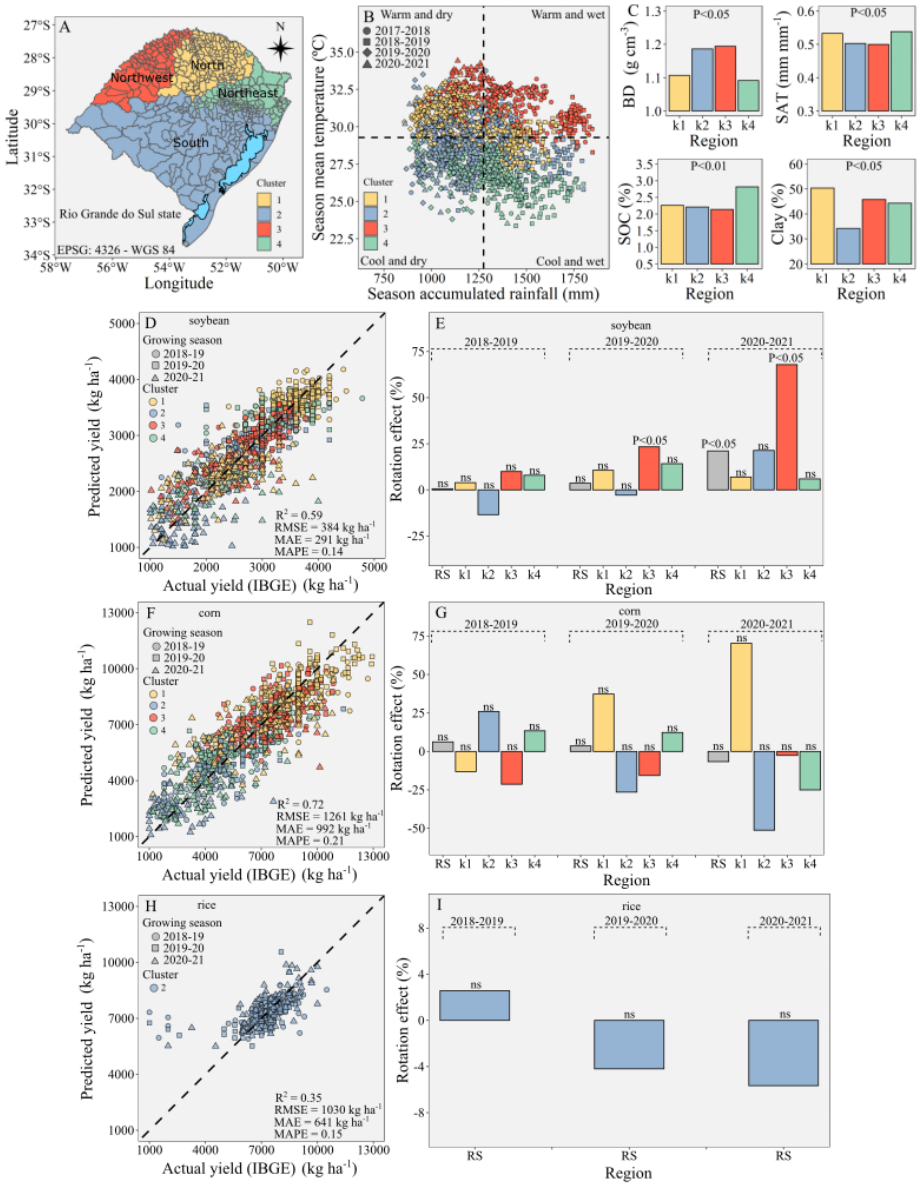


Fig. 7. Clusters of Rio Grande do Sul, Brazil according soil variables and climate variables from 2017-2018, 2018-2019, 2019-2020, 2020-2021 growing seasons (A). Meteorological condition – temperature and precipitation, of the clusters in Rio Grande do Sul (B). Soil variables – bulk density (BD), Saturation (SAT), soil organic carbon (SOC), and clay, of the

clusters in Rio Grande do Sul (C). Relation 1:1 predicted yield and actual yield IBGE for soybean (D), corn (F), and rice (H) at municipality level. Crop rotation effect of Rio Grande do Sul (bars in gray), with the cluster's regions for two growing season rotation data (2018-2019), three years rotation data (2019-2020), and four years rotation data (2020-2021) for soybean (E), corn (G), and rice (I).

The multiple linear regression with climate, soil, and rotation covariates for 2018-2019, 2019-2020, and 2020-2021 growing seasons obtained comparable results with the crop yields from IBGE statistics comparison, with R^2 of 0.59 for soybean, 0.72 for corn and 0.35 for rice, and with mean absolute percentage error (MAPE) of 0.14 for soybean, 0.21 for corn and 0.15 for rice (Fig 7D, F, and H).

The rotation covariate effects extracted from the regressions did not have significant results for corn and rice for 2018-2019, 2019-2020, and 2020-2021 growing seasons (2-3-4-yr in a row) (Fig 7E, G, and I). Soybean showed yield penalties up to ~23% for Northwest region in 2019-2020 growing season for continuous soybean (3-yr). Furthermore, for 4-yr continuous soybean, crop rotation effect showed ~20% in the state, while for the Northwest region showed up to 65% linked to low soybean yields in 2020-2021 growing season (Fig 7D). Northwest region has warm and wet climate with higher bulk density and lower SOC compared to the other regions, showing greater benefits on crop yields utilizing more intense crop rotation practices. Other regions did not show significant crop rotation effect for the short-term yield penalty evaluation (Fig 7E, G, and I).

DISCUSSION

Crop rotation mapping

This study proposed a novel method for mapping crop rotations for Rio Grande do Sul State. We showed that integrating temporal information is essential to represent field level crop rotation patterns. Waldhoff et al. (2017) and Liu et al. (2022) reported the importance of the temporal data into the classification scheme to represent complex crop rotation systems. Globally, crop rotation maps have been recently reported in Germany (Waldhoff et al., 2017), Argentina (Abelleyra and Verón, 2020), China (Liu et al., 2021; Liu et al., 2022), and the United States (Kluger et al., 2022), utilizing diverse remote sensing approaches. From our study, continuous crops were the most frequent sequence adopted, followed by the one year

rotation. Mapping crop rotation is critical to characterize crop production areas for regional agro-ecosystem modeling (Waldhoff et al., 2017), conservation and regenerative agriculture adoption (Kassam et al., 2018), and agriculture intensification planning (Kleinman et al., 2018; Lin and Huang, 2019).

Crop rotation patterns

Crop rotation patterns were very specific for each mesoregion, with an increase of soybean area in rice in lowlands of the South Rio Grande do Sul. Typically, the traditional flooded rice cropping system is continuous crop alternate with a fallow or annual ryegrass (*Lolium multiflorum* Lam.) (Dominschek et al., 2022). This cropping systems is being replacing by more intensive cropping system with rotation mainly with soybean and in some cases corn crop since the drainage system is improved and soil acidity and fertility is enhanced (Emygdio et al., 2017). Liu et al. (2022) reported high intensity and low diversity for crop rotations concentrated in Southern China. Ribas et al. (2021) reported an increase of both weed pressure and production costs as a motivation for rice farmers introducing soybean to the traditional monocrop rice systems.

Environmental interactions

The climate and physical soil attributes have an interaction that affect crop yield mainly in drought seasons, specifically the relationship between soil texture, bulk density, water availability, and SOC with continuous crop farming system can drive the magnitude of the yield penalty of crop rotation lack. Crop rotation is a suitable practice in conservation agriculture that can mitigate early-stage challenges of no-till as compaction by alleviating top-soil bulk density compared with continuous crops (Kassam et al., 2018; De Moura et al., 2021). For the region of study, both Northwest and Midwestern mesoregions presented soybean as a high continuous crop system, even when the literature is comprehensive in reporting yield penalties of this continuous crop system (Wingeyer et al., 2015; Seifert et al., 2017; Cohen et al., 2019; Kluger et al., 2022). Intensified crop rotation has been reported as an efficient practice for improving plant water availability mainly under drought years (Rosenzweig et al., 2018; Degani et al., 2019). Currently, crop rotation has been highlighted due its benefits for soil health, playing a crucial role for soil health and for contributing to the

carbon removal from the atmosphere (Sá et al., 2013; Paustian et al., 2016; De Oliveira Ferreira et al., 2018).

Crop rotation effects

From a crop productivity perspective, crop rotation effects have been widely reported as benefits for soybean, corn, and rice crops. For soybean, a yield benefit for rotation varied across studies such as 4% (Nafzinger, 2007), 8-17% (Crookston et al., 1991), and 25% (Edwards et al., 1988). For corn, the rotation benefits in yield showed generally lower value with 7.8% (Erickson, 2008). Seifert et al. (2017) reported soybean and corn yield penalties of 10.3% and 4.3%, respectively in US utilizing field data related to lack of crop rotation. Cohen et al. (2019) in a satellite data aggregated with field data study presented yield benefits of crop rotation in average of 10.8% and 1% for soybean and corn, respectively. Ribas et al. (2021) reported yield benefits of 26% for rice in soybean-rice crop rotation compared to continuous rice for Rio Grande do Sul state. In addition, environmental factors linked to yield benefit of rotation may mainly derive from contribution of nitrogen (N) availability to the following crop, for example for corn (Sindelar et al., 2016; Puntel et al., 2016; Novelli et al., 2023), reduction in pest pressure from soybean as cyst nematode (Hu et al., 2018; Strom et al., 2020), and mitigation of weed infestation for rice (Ribas et al., 2021). Furthermore, diversified crop rotations increase soil health, sustainability, and yield of major field crop production systems considering regional and economic aspects (Kassam et al., 2018; Spiegal et al., 2018; Chamberlain et al., 2021; Passinato et al., 2021; Garbelini et al., 2022).

Rio Grande do Sul State has a trend of increase in soybean area that is expanding even to lowland areas that were traditionally grown only with continuous rice or traditional explore with native pasture that were under cattle farming (Emygdio et al., 2017; Ribas et al., 2021). The predominant state cropping system is continuous crop and one year rotation during the four years study period. Although, some mesoregions did not have yield penalty by lack of crop rotation, the crop rotation mapping create a critical alert regarding to lack of crop diversification in the state and mainly in Northwest and Midwest mesoregions. In a season with drought spells (2019-2020), continuous soybean in that mesoregions showed a yield penalty ~20%, showing a low climate resilience of current cropping system adopted. Another concern is that Northwest and Midwestern, where the soils with higher agricultural potential and that were the pioneer areas (Silva et al., 2020) and hold currently the largest crop area

(IBGE, 2022) are the ones with higher lack of crop rotation. The new areas and with lower yield potential, as Central and Southern mesoregions (Cattelan and Dall'Agnol, 2018), have a trend to adopt crop rotation as a frequent strategy to grain production. Probably, the high yield gap of these mesoregions and temporal yield instability makes the farmers more caution about adopt monocropping.

Limitations and future studies

A potential limitation of this study is the lack of large field datasets as an independent validation, a common problem of many satellite-based approaches. Future studies should investigate the integration of field experiment with remote sensing aiming to improve identification of different crop rotation in the region and around the globe. Access to yield data at field-scale will permit to downscale this approach of crop rotation from municipality to field, making this forecast more relevant to farmers in the region. Furthermore, other crop type and systems such as winter crops, relay or intercropping, and double cropping options should be integrated to enhance the crop rotation dataset.

CONCLUSIONS

The presented remote sensing-based crop mapping approach was critical to demonstrate that Midwestern and Northwest mesoregions presented large areas with continuous crop (>60%), mostly with 4-yr continuous soybean. Porto Alegre metropolitan, Southwest, and Southern mesoregions presented cropland use patterns resulting on increases in soybean area in detriment to the crop rice region.

The climate presented interactions with continuous summer type crops linked to higher temperatures and greater precipitation from October to April. Soil attributes, mainly soil texture, bulk density, water availability, and SOC was linked to monocrop systems, which can substantially increase yield penalty due to the lack of crop rotation.

The model presented significant crop rotation effect on yield for soybeans with yield penalty of ~23% with continuous crop (3-yr in a row) for Northeast cluster in 2019-2020 season, and 20% for entire state, highlighting up to 65% for Northwest cluster region with continuous crop (4-yr in a row) in 2020-2021 growing season.

In summary, we have generated an end-to-end approach to build crop rotation maps using available remote sensing-based crop type classification and mapping with field

boundary delineation ancillary data. This novel approach is suitable to be applied to other regions or countries around the globe.

CRediT authorship contribution statement

Luan Pierre Pott: Conceptualization of this study, Methodology, Software, Formal analysis, Investigation, Data Curation Visualization, Validation, Writing - Original Draft. Telmo Jorge Carneiro Amado: Investigation, Validation, Writing - Review & Editing. Raí Augusto Schwalbert: Investigation, Writing - Review & Editing. Geomar Mateus Corassa: Investigation, Writing - Review & Editing. Ignacio Antonio Ciampitti: Conceptualization of this study, Investigation, Methodology, Supervision, Writing - Review & Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The first author thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The second author thanks to CNPq for research scholarship n 312769/2021-0. This is a contribution number 23-xwy for the Kansas State Research Extension.

REFERENCES

Abelleyra, D., & Verón, S. (2020). Crop rotations in the Rolling Pampas: Characterization, spatial pattern and its potential controls. *Remote Sensing Applications: Society and Environment, 18*, 100320. https://doi.org/10.1016/j.rsase.2020.100320

Alvares, C. A., Stape, J. L., Sentelhas, P. C., de Moraes Gonçalves, J. L., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift, 22*(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Azzari, G., Grassini, P., Edreira, J. I. R., Conley, S., Mourtzinis, S., & Lobell, D. B. (2019). Satellite mapping of tillage practices in the North Central US region from 2005 to 2016. *Remote Sensing of Environment, 221*, 417–429. https://doi.org/10.1016/j.rse.2018.11.010

Barbieri, P., Pellerin, S., Seufert, V., & Nesme, T. (2019). Changes in crop rotations would impact food production in an organically farmed world. *Nature Sustainability*, *2*(5), 378–385. https://doi.org/10.1038/s41893-019-0259-5

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., … Tait, A. M. (2022). Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, *9*(1). https://doi.org/10.1038/s41597-022-01307-4

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., & Li, Z. (2018). A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, *210*, 35–47. https://doi.org/10.1016/j.rse.2018.02.045

Cattelan, A. J., & Dall'Agnol, A. (2018). The rapid soybean growth in Brazil. *OCL*, *25*(1), D102. https://doi.org/10.1051/ocl/2017058

Chamberlain, L. A., Whitman, T., Ané, J.-M., Diallo, T., Gaska, J. M., Lauer, J. G., … Conley, S. P. (2021). Corn-soybean rotation, tillage, and foliar fungicides: Impacts on yield and soil fungi. *Field Crops Research*, *262*, 108030. https://doi.org/10.1016/j.fcr.2020.108030

Cohen, A. A.B., Seifert, C. A., Azzari, G., & Lobell, D. B. (2019). Rotation Effects on Corn and Soybean Yield Inferred from Satellite and Field-level Data. *Agronomy Journal*, *111*(6), 2940–2948. https://doi.org/10.2134/agronj2019.03.0157

CONAB, Companhia Nacional de Abastecimento. (2022). Séries históricas das safras. Retrieved October 7, 2022, from Conab.gov.br website: https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras#gr%C3%A3os-2.

Crookston, R. K., Kurle, J. E., Copeland, P. J., Ford, J. H., & Lueschen, W. E. (1991). Rotational Cropping Sequence Affects Yield of Corn and Soybean. *Agronomy Journal*, *83*(1), 108–113. https://doi.org/10.2134/agronj1991.00021962008300010026x

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H. I., Joebges, C., Lemoine, G., & van der Velde, M. (2020). Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. *Scientific Data*, *7*(1). https://doi.org/10.1038/s41597-020-00675-z

Degani, E., Leigh, S. G., Barber, H. M., Jones, H. E., Lukac, M., Sutton, P., & Potts, S. G. (2019). Crop rotations in a climate change scenario: short-term effects of crop diversity

on resilience and ecosystem service provision under drought. *Agriculture, Ecosystems & Environment*, *285*, 106625. https://doi.org/10.1016/j.agee.2019.106625

De Moura, M. S., Silva, B. M., Mota, P. K., Borghi, E., Resende, A. V. de, Acuña-Guzman, S. F., … Curi, N. (2021). Soil management and diverse crop rotation can mitigate early-stage no-till compaction and improve least limiting water range in a Ferralsol. *Agricultural Water Management*, *243*, 106523. https://doi.org/10.1016/j.agwat.2020.106523

De Oliveira Ferreira, A., Amado, T. J. C., Rice, C. W., Ruiz Diaz, D. A., Briedis, C., Inagaki, T. M., & Gonçalves, D. R. P. (2018). Driving factors of soil carbon accumulation in Oxisols in long-term no-till systems of South Brazil. *Science of the Total Environment*, *622-623*, 735–742. https://doi.org/10.1016/j.scitotenv.2017.12.019

Deines, J. M., Patel, R., Liang, S.-Z., Dado, W., & Lobell, D. B. (2021). A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sensing of Environment*, *253*, 112174. https://doi.org/10.1016/j.rse.2020.112174

Deines, J. M., Guan, K., Lopez, B., Zhou, Q., White, C. S., Wang, S., & Lobell, D. B. (2022). Recent cover crop adoption is associated with small maize and soybean yield losses in the United States. *Global Change Biology*. https://doi.org/10.1111/gcb.16489

Edwards, J. H., Thurlow, D. L., & Eason, J. T. (1988). Influence of Tillage and Crop Rotation on Yields of Corn, Soybean, and Wheat. *Agronomy Journal*, *80*(1), 76–80. https://doi.org/10.2134/agronj1988.00021962008000010018x

Emygdio, B. M., Rosa, A. P., & Oliveira, A. D. (2017). Cultivo de soja e milho em terras baixas do Rio Grande do Sul. Brasília, DF: Embrapa.

Erickson B. (2008). Corn/soybean rotation literature summary. Purdue University, West Lafayette, IN.

Garbelini, L. G., Debiasi, H., Junior, A. A. B., Franchini, J. C., Coelho, A. E., & Telles, T. S. (2022). Diversified crop rotations increase the yield and economic efficiency of grain production systems. *European Journal of Agronomy*, *137*, 126528. https://doi.org/10.1016/j.eja.2022.126528

Hijmans, R. J. (2022). raster: Geographic Data Analysis and Modeling. R package version 3.5-15. https://CRAN.R-project.org/package=raster

Hu, W., Strom, N., Haarith, D., Chen, S., & Bushley, K. E. (2018). Mycobiome of Cysts of the Soybean Cyst Nematode Under Long Term Crop Rotation. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.00386

IBGE, Instituto Brasileiro de Geografia e Estatística. (2021). Panorama Estados. Retrieved August 31, 2022, from Ibge.gov.br website: https://cidades.ibge.gov.br/brasil/rs/panorama.

IBGE, Instituto Brasileiro de Geografia e Estatística. (2022). Área plantada, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras temporárias. Retrieved August 31, 2022, from Ibge.gov.br website: https://sidra.ibge.gov.br/tabela/1612.

IRGA, Instituto Rio Grandense do Arroz. (2021). Boletim de resultados da safra 2020/21 em terras baixas: arroz irrigado e soja. Retrieved October 6, 2022, from Irga.rs.gov.br website: https://irga.rs.gov.br/upload/arquivos/202109/27151231-boletim-de-resultados-da-safra-2020-2021-compressed.pdf.

Kassam, A., Friedrich, T. & Derpsch, R. (2018). Global spread of Conservation Agriculture. *International Journal of Environmental Studies*, *76*, 1-23. https://doi.org/10.1080/00207233.2018.1494927

Kleinman, P. J. A., Spiegal, S., Rigby, J. R., Goslee, S. C., Baker, J. M., Bestelmeyer, B. T., … Pierson, F. B. (2018). Advancing the Sustainability of US Agriculture through Long-Term Research. *Journal of Environmental Quality*, *47*(6), 1412–1425. https://doi.org/10.2134/jeq2018.05.0171

Kluger, D. M., Owen, A. B., & Lobell, D. B. (2022). Combining randomized field experiments with observational satellite data to assess the benefits of crop rotations on yields. *Environmental Research Letters*, *17*(4), 044066. https://doi.org/10.1088/1748-9326/ac6083

Lin, M., & Huang, Q. (2019). Exploring the relationship between agricultural intensification and changes in cropland areas in the US. *Agriculture, Ecosystems & Environment*, *274*, 33–40. https://doi.org/10.1016/j.agee.2018.12.019

Liu, Y., Zhao, W., Chen, S., & Ye, T. (2021). Mapping Crop Rotation by Using Deeply Synergistic Optical and SAR Time Series. *Remote Sensing*, *13*(20), 4160. https://doi.org/10.3390/rs13204160

Liu, Y., Yu, Q., Zhou, Q., Wang, C., Bellingrath-Kimura, S. D., & Wu, W. (2022). Mapping the Complex Crop Rotation Systems in Southern China Considering Cropping Intensity, Crop Diversity, and Their Seasonal Dynamics. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 9584–9598. https://doi.org/10.1109/jstars.2022.3218881

Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, *164*, 324–333. https://doi.org/10.1016/j.rse.2015.04.021

MapBiomas Brazil. (2022). Retrieved September 13, 2022, from Mapbiomas.org website: https://mapbiomas.org/

Mildrexler, D. J., Zhao, M., & Running, S. W. (2011). A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests. *Journal of Geophysical Research*, *116*(G3). https://doi.org/10.1029/2010jg001486

Nafziger E. (2007). What will replace the corn-soybean rotation? 2007 Indiana CCA Conference Proceedings, Indianapolis, IN. 18– 19 Dec. 2007. Purdue University. https://www.agry.purdue.edu/cca/2007/2007/Proceedings/Emerson%20Nafziger-CCA%20proceedings_KLS.pdf (accessed 30 September 2022).

Novelli, L. E., Caviglia, O. P., Jobbágy, E. G., & Sadras, V. O. (2023). Diversified crop sequences to reduce soil nitrogen mining in agroecosystems. *Agriculture, Ecosystems & Environment*, *341*, 108208. https://doi.org/10.1016/j.agee.2022.108208

Passinato, J. H., Amado, T. J. C., Kassam, A., Acosta, J. A. A., & Amaral, L. de P. (2021). Soil health check-up of conservation agriculture farming systems in Brazil. *Agronomy*, *11*(12), 2410. https://doi.org/10.3390/agronomy11122410

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G. P., & Smith, P. (2016). Climate-smart soils. *Nature*, *532*(7597), 49–57. https://doi.org/10.1038/nature17174

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal 10* (1), 439-446, https://doi.org/10.32614/RJ-2018-009

Pott, L. P., Amado, T. J. C., Schwalbert, R. A., Corassa, G. M., & Ciampitti, I. A. (2021). Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing*, *176*, 196–210. https://doi.org/10.1016/j.isprsjprs.2021.04.015

Pott, L.P., Amado, T.J.C, Schwalbert, R.A., Corassa, G.M., & Ciampitti, I.A. (2022). Crop type classification in Southern Brazil: Integrating remote sensing, crop modeling and machine learning. *Computers and Electronics in Agriculture, 201*, 107320. https://doi.org/10.1016/j.compag.2022.107320

Puntel, L. A., Sawyer, J. E., Barker, D. W., Dietzel, R., Poffenbarger, H., Castellano, M. J., … Archontoulis, S. V. (2016). Modeling Long-Term Corn Yield Response to Nitrogen Rate and Crop Rotation. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.01630

Ribas, G. G., Zanon, A. J., Streck, N. A., Pilecco, I. B., de Souza, P. M., Heinemann, A. B., & Grassini, P. (2021). Assessing yield and economic impact of introducing soybean to the lowland rice system in southern Brazil. *Agricultural Systems*, *188*, 103036. https://doi.org/10.1016/j.agsy.2020.103036

Rosenzweig, S. T., Stromberger, M. E., & Schipanski, M. E. (2018). Intensified dryland crop rotations support greater grain production with fewer inputs. *Agriculture, Ecosystems & Environment*, *264*, 63–72. https://doi.org/10.1016/j.agee.2018.05.017

Sá, J. C. M., Séguy, L., Tivet, F., Lal, R., Bouzinac, S., Borszowskei, P. R., … Friedrich, T. (2013). Carbon Depletion by Plowing and its Restoration by No-Till Cropping Systems in Oxisols of Subtropical and Tropical Agro-Ecoregions in Brazil. *Land Degradation & Development*, *26*(6), 531–543. https://doi.org/10.1002/ldr.2218

Sahajpal, R., Zhang, X., Izaurralde, R. C., Gelfand, I., & Hurtt, G. C. (2014). Identifying representative crop rotation patterns and grassland loss in the US Western Corn Belt. *Computers and Electronics in Agriculture*, *108*, 173–182. https://doi.org/10.1016/j.compag.2014.08.005

Santos, H.G., Jacomine, P.K.T., Dos Anjos, L.H.C., De Oliveira, V.A., Lumbreras, J.F., Coelho, M.R., Cunha, T.J.F., (2018). Sistema Brasileiro de Classificação de Solos. Brasília, DF: Embrapa, 2018.

Sarwar, G., Schmeisky, H., Hussain, N., Muhammad, S., Ibrahim, M., & Safdar, E. (2008). Improvement of soil physical and chemical properties with compost application in rice-wheat cropping system. *Pakistan Journal of Botany*, 40(1), 275-282.

Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. V. Vara., & Ciampitti, I. A. (2020). Satellite-based soybean yield forecast: Integrating machine learning and

weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*, *284*, 107886. https://doi.org/10.1016/j.agrformet.2019.107886

Seifert, C. A., Roberts, M. J., & Lobell, D. B. (2017). Continuous Corn and Soybean Yield Penalties across Hundreds of Thousands of Fields. *Agronomy Journal*, *109*(2), 541–548. https://doi.org/10.2134/agronj2016.03.0134

Sicar, Sistema Nacional de Cadastro Ambiental Rural. (2022). Retrieved September 13, 2022, from Car.gov.br website: https://www.car.gov.br/#/

Silva, R. F. B., Batistella, M., Moran, E., Celidonio, O. L. D. M., & Millington, J. D. A. (2020). The Soybean Trap: Challenges and Risks for Brazilian Producers. *Frontiers in Sustainable Food Systems*, *4*. https://doi.org/10.3389/fsufs.2020.00012

Sindelar, A. J., Schmer, M. R., Jin, V. L., Wienhold, B. J., & Varvel, G. E. (2016). Crop Rotation Affects Corn, Grain Sorghum, and Soybean Yields and Nitrogen Recovery. *Agronomy Journal*, *108*(4), 1592–1602. https://doi.org/10.2134/agronj2016.01.0005

Soil Survey Staff. (2014). Keys to soil taxonomy, 12th ed. USDA-Natural Resources Conservation Service, Washington, DC.

Souza, C. M., Z. Shimbo, J., Rosa, M. R., Parente, L. L., A. Alencar, A., Rudorff, B. F. T., … Weber, E. J. (2020). Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sensing*, *12*(17), 2735. https://doi.org/10.3390/rs12172735

Spiegal, S., Bestelmeyer, B. T., Archer, D. W., Augustine, D. J., Boughton, E. H., Boughton, R. K., … McCarty, G. W. (2018). Evaluating strategies for sustainable intensification of US agriculture through the Long-Term Agroecosystem Research network. *Environmental Research Letters*, *13*(3), 034031. https://doi.org/10.1088/1748-9326/aaa779

Stein, S., & Steinmann, H.-H. (2018). Identifying crop rotation practice by the typification of crop sequence patterns for arable farming systems – A case study from Central Europe. *European Journal of Agronomy*, *92*, 30–40. https://doi.org/10.1016/j.eja.2017.09.010

Steinmann, H.-H., & Dobers, E. S. (2013). Spatio-temporal analysis of crop rotations and crop sequence patterns in Northern Germany: potential implications on plant health and

crop protection. *Journal of Plant Diseases and Protection*, *120*(2), 85–94. https://doi.org/10.1007/bf03356458

Strom, N., Hu, W., Haarith, D., Chen, S., & Bushley, K. (2020). Interactions between soil properties, fungal communities, the soybean cyst nematode, and crop yield under continuous corn and soybean monoculture. *Applied Soil Ecology*, *147*, 103388. https://doi.org/10.1016/j.apsoil.2019.103388

Taravat, A., Wagner, M. P., Bonifacio, R., & Petit, D. (2021). Advanced Fully Convolutional Networks for Agricultural Field Boundary Detection. *Remote Sensing*, *13*(4), 722. https://doi.org/10.3390/rs13040722

Tomm, G.O. (2008). Uma nova fase do cultivo no Brasil: produção com seguro e todo o suporte ao produtor (In portuguese). *Revista Plantio Direto, may-jun*, 4-6.

USDA, United States Department of Agriculture. (2022). National Agricultural Statistics Service. Retrieved September 1, 2022, from nass.usda.gov website: USDA - National Agricultural Statistics Service - Research and Science - CropScape and Cropland Data Layer - Announcements.

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. https://doi.org/10.1016/j.jag.2017.04.009

Wang, S., Di Tommaso, S., Deines, J. M., & Lobell, D. B. (2020). Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. *Scientific Data*, *7*(1). https://doi.org/10.1038/s41597-020-00646-4

Wingeyer, A., Amado, T., Pérez-Bidegain, M., Studdert, G., Varela, C., Garcia, F., & Karlen, D. (2015). Soil Quality Impacts of Current South American Agricultural Practices. *Sustainability*, *7*(2), 2213–2242. https://doi.org/10.3390/su7022213

Yu, T., Mahe, L., Li, Y., Wei, X., Deng, X., & Zhang, D. (2022). Benefits of Crop Rotation on Climate Resilience and Its Prospects in China. *Agronomy*, *12*(2), 436. https://doi.org/10.3390/agronomy12020436

Zhao, J., Chen, J., Beillouin, D., Lambers, H., Yang, Y., Smith, P., … Zang, H. (2022). Global systematic review with meta-analysis reveals yield advantage of legume-based rotations and its drivers. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-32464-0

5 DISCUSSÃO

Os trabalhos dispostos nos capítulos que compõem a tese trazem i) modelos de classificação de culturas através da fusão de dados de satélite de diferentes características como óptico, de radar, e modelos digitais de elevação, ii) utilização de dados de modelos de crescimento de culturas agrícolas como APSIM para gerar dados para utilização em modelos de classificação de culturas de forma isolada, ou com intuito de agregar em dados de campo obtendo maiores acurácias nos modelos de *machine learning*, iii) mapeamento da rotação de culturas agrícolas de verão para o estado do Rio Grande do Sul, bem como *pipeline* para geração dos bancos de dados, avaliação de interações ambientais e efeito de rotação de culturas na produtividade de grãos.

O monitoramento da agricultura com dados de satélites vem crescendo nos últimos anos devido à crescente disponibilidade de imagens de diversos satélites tanto de distribuição pública, bem como pagos, sendo muitas vezes vinculados com plataformas digitais impulsionado pela agricultura digital em todo o mundo (LOBELL et al. 2019). Sensoriamento remoto com a utilização de imagens de satélite com intuito de classificação de culturas agrícolas vêm sendo utilizadas globalmente com diversos grupos de culturas agrícolas, diversos tipos de satélites e resolução espacial (BELGIU et al., 2018; JIN et al., 2019; HAO et al., 2020; WANG et al., 2020). Mais recentemente a incorporação de diferentes satélites e dados espaciais como de elevação para unir *features* para utilização de modelos de *machine learning* ganharam espaço demonstrando acréscimo em acurácias dos modelos (JIN et al., 2019; ORYNBAIKYZY et al., 2020). *Feature engineering* quando se trata de imagens de satélites é um tema bastante importante, uma vez que a confecção das *features* para utilização dos modelos de *machine learning* afetam diretamente a acurácia dos modelos, assim a utilização de regressões harmônicas vem sendo bastante utilizada em estudos de classificação e previsão de culturas agrícolas (WANG et al., 2019; WANG et al., 2020; DEINES et al., 2020; DADO et al., 2020). Número de amostras necessárias, bem como a espacialização dos dados ao longo da área de interesse para classificação de culturas tem impacto no desempenho dos modelos (CAI et al., 2018; WANG et al., 2019; FOWLER, J.; WALDNER, F.; HOCHMAN, 2020). Ainda, modelos de classificação que predizem a cultura na safra agrícola de maneira antecipada (~60 dias após semeadura) possui maior apelo por partes governamentais sendo factíveis com acurácia aceitáveis (CAI et al., 2018). Assim, o primeiro estudo teve como contribuições a geração de classificação e mapeamento das culturas de soja,

milho e arroz para todo estado do Rio Grande do Sul com a utilização de fusão de dados de Sentinel-2, Sentinel-1 e dados de elevação de SRTM obtendo acurácia de 0.95. Ademais, demonstrou-se a necessidade de levar em consideração a variabilidade espacial existente no estado com intuito de formar um *dataset* representativo para geração do modelo de classificação principalmente para dados de campos de milho e soja. Número de amostras de cultivos agrícolas demonstrou-se que com maior utilização de dados e número de safras agrícolas, bem como dados da safra agrícola corrente para confeccionar o modelo apresentaram maiores acurácias, obtendo valores satisfatórios acima de 250 amostras por cultura agrícola. Predições de áreas plantadas de forma antecipada, antes do final do ciclo, demostrou-se eficiente para classificação e mapeamento com dados de imagens de satélite até 1º março, obtendo maior acurácia, no entanto classificação com dados até 1º de janeiro obtiveram acurácias aceitáveis. Por fim, o primeiro modelo de classificação e mapeamento de culturas de soja, milho e arroz para todo o estado do Rio Grande do Sul foram apresentados, com capacidade de classificações de futuras safras do estado, bem como adaptações do modelo para outras regiões nacionais e mundiais.

Modelagem de culturas agrícolas como o *framework* APSIM-NG com intuito de gerar simulações de crescimento e desenvolvimento das culturas vem sendo utilizados em diversos estudos com intuito de *data augmentation*, com aumento de dados bem como da variabilidade dos dados proporcionando essas características ao modelo de *machine learning* (LOBELL et al., 2015; JIN et al., 2018). Estudos com simulações de modelos de crescimento almejando conjuntar banco de dados para classificação de culturas agrícolas ainda não está bem reportado pela literatura, no entanto as mesmas simulações vêm sendo utilizados com muito êxito na melhoria de modelos de previsão de produtividades (LOBELL et al., 2015; JIN; AZZARI; LOBELL, 2017; DADO et al., 2020; DEINES et al., 2021). *Feature engineering* utilizando regressões harmônicas vem sendo amplamente utilizados como estratégias de extração de *features* para *input* nos modelos de *machine learning* (WANG et al., 2019; WANG et al., 2020; DEINES et al., 2020; DADO et al., 2020) com dados de sensoriamento remoto. Uma vez que é necessário combinar modelagem de culturas agrícolas e sensoriamento remoto é necessário a utilização das mesmas técnicas de *feature engineering* objetivando *data augmentation* na confecção dos modelos de *machine learning*. Classificação do tipo de culturas agrícolas necessitam prévia ou simultaneamente classificação das áreas de culturas agrícolas. Modelos de máscaras de áreas agricultáveis para o Brasil são

disponibilizadas com os produtos anuais do MapBiomas (MAPBIOMAS, 2023), através do recurso público de georreferenciamento de áreas agrícolas do CAR (SICAR, 2023), e por fim a geração de modelos específicos para regiões, anos, bem como para o mundo inteiro (BROWN et al., 2022). Combinações de *dataset*, avaliação de *transfer learning* dos modelos para novos anos agrícolas, número de amostras e da espacialização das amostras precisam ser avaliadas uma vez que influenciam a qualidade do modelo (CAI et al., 2018; WANG et al., 2019; FOWLER, J.; WALDNER, F.; HOCHMAN, 2020). Assim, o segundo estudo teve como contribuições a utilização de dados simulados de *frameworks* de modelos de crescimento de culturas agrícolas relacionados com sensoriamento remoto para a geração de classificação e mapeamento de culturas agrícolas isoladamente ou em conjunto com dados de campo com intuito de aumentar a acurácia dos modelos, alcançando no estudo acurácias de 0.94 para classificação de soja e milho para a região do estudo. Dados das simulações com APSIM-NG relacionados com a confecção das *features* das regressões harmônicas demonstraram similaridades com as *features* gerados com dados de campo adquirindo possibilidades de expansão das utilizações de simulações para maiores gerações de dados para outras áreas e anos agrícolas. As máscaras de culturas agrícolas utilizadas no estudo como geração de modelos independentes de *random forest*, produtos do MapBiomas e CAR apresentaram boas delimitações para compor a classificação de culturas agrícolas, apresentando melhores resultados os produtos na ordem apresentadas acima. Por fim, os métodos deste estudo podem ser aplicados em outras regiões, estados e países, bem como para outras culturas agrícolas gerando o mapeamento de culturas agrícolas.

Muitos estudos na literatura demonstram capacidade de distinção de culturas agrícolas com sensoriamento remoto com intuito do mapeamento dos tipos de culturas agrícolas na safra agrícola (BELGIU et al., 2018; JIN et al., 2019; HAO et al., 2020; WANG et al., 2020). No entanto a utilização do mapeamento de culturas agrícolas para compor dados temporais ao longo das safras agrícolas objetivando a confecção de mapeamento de rotação de culturas, bem como análises de padrões e mudanças de rotação de culturas e suas análises tem sido pouco explorado ainda globalmente. Mapeamento de rotação de culturas tem sido recentemente reportado na Alemanha (WALDHOFF et al., 2017), Argentina (ABELLEYRA & VERÓN, 2020), China (LIU et al., 2021), e Estados Unidos (KLUGER; OWEN; LOBELL, 2022), sendo reconhecidamente crítica para caracterizar modelagem de culturas agrícolas em escala regional (WALDHOFF et al., 2017), para conservação e regeneração da agricultura

(KASSAM; FRIEDRICH; DERPSCH, 2018), e planificação de intensificação dos sistemas agrícolas (LIN & HUANG, 2019). Rotações de culturas podem ser motivadas por características regionais, práticas de manejos, cultura do produtor, aspectos econômicos e socias (STEIN & STEINMANN, 2018). A prática de rotação de culturas tem sido reportada variando seus benefícios em função de condições ambientais para as culturas agrícolas de soja (SEIFERT; ROBERTS; LOBELL, 2017; COHEN et al., 2019, SIMÃO et al., 2023), milho (SEIFERT; ROBERTS; LOBELL, 2017; COHEN et al., 2019) e arroz (RIBAS et al., 2021). Assim, o terceiro estudo teve como contribuições geração de *framework* para mapeamento de rotação de culturas com a utilização de modelo de mapeamento de culturas agrícolas, dados auxiliares de delimitação de campos agrícolas, dados ambientais para análises de padrões de rotação de culturas relacionados regionalmente com as variáveis ambientais, e por fim análise de benefício da prática de manejo de rotação de culturas. A presente metodologia do estudo foi crucial para mostrar que as mesoregiões Noroeste e Centro Ocidental do Rio Grande do Sul apresentaram monocultivo de principalmente soja em >60% dos campos dessas regiões, e as Região Metropolitana de Porto Alegre, Sudeste e Sudoeste Rio-Grandense demonstraram padrões de crescimento da área com a cultura da soja e decréscimo da área com cultura do arroz. Os efeitos de penalização de produtividade de grãos das culturas em função da ausência de rotação de culturas foram reportados em ~23% com 3 anos consecutivos de soja para região Noroeste em 2019-2020, e ~20% para todo estado, sendo destacado a região Noroeste para a safra 2020-2021 com 4 anos consecutivos de soja a perca de produtividade da soja em ~65%. Finalmente, o estudo traz uma metodologia de mapeamento de rotação de culturas utilizando modelos de classificação de culturas baseados em sensoriamento remoto e dados auxiliares os quais podem ser utilizados em outras regiões e países ao redor do globo.

6 CONCLUSÃO

Modelos de classificação de culturas baseados em sensoriamento remoto com fusão de dados auxiliam no monitoramento da agricultura podendo ser destinado à diversos setores com intuito de políticas públicas, segurança alimentar e sustentabilidade. O modelo de classificação de culturas desenvolvido com a fusão de dados de Sentinel-2, Sentinel-1 e dados de relevo de SRTM obteve acurácia de 0.95 para classificações de soja, milho e arroz para o estado do Rio Grande do Sul. Constatou-se que as amostras de identificação de cultura agrícola dos campos requeridas para formar o banco de dados para confecção do modelo necessita a consideração da variabilidade espacial das mesmas, uma vez que se tem diferentes condições de clima, solo e manejo nas diferentes regiões. Essa variabilidade espacial demonstrou-se mais importante para as culturas de milho e soja, sendo menos importante a variabilidade espacial da cultura do arroz, podendo ser que a fusão de dados de sensoriamento remoto para a cultura do arroz é capaz de identificá-la mais facilmente. Em relação ao número de amostras o modelo demonstrou que com ~250 amostras para cada cultura têm-se acurácia similar comparado com modelos utilizando maiores números de amostras. Também, o modelo com a atuação de *transfer learning* para próximas safras agrícolas demonstrou-se satisfatório, aumentando sua capacidade preditiva quando poucas amostras, (~50), da atual safra agrícola foram adicionadas ao modelo. Classificações antecipadas obteve resultados satisfatórios para 1º de janeiro, sendo 1º de março maiores eficácias do modelo. Por fim, modelo de classificação de culturas de soja, milho e arroz para o estado do Rio Grande do Sul foi desenvolvido com capacidade de geração de classificações e mapeamento de culturas agrícolas com fusão de dados de sensoriamento remoto para as próximas safras agrícolas.

A incorporação de dados na agricultura vem sendo cada vez mais gerados, com diversos intuitos de monitoramento e tomadas de decisões. Modelos de crescimento de culturas com diversos cenários de solo e clima ao longo dos anos agrícolas vem sendo utilizados para geração de simulações de desenvolvimento de culturas agrícolas e assim ser utilizados nos mais diversos propósitos. O modelo de classificação de culturas utilizando dados de simulações de crescimento e desenvolvimento do *framework* APSIM-NG apresentou acurácia similar com o modelo utilizando dados de campo, uma vez que as *features* de regressões harmônicas resultantes das relações do modelo de crescimento e dos dados de campo obtiveram similaridades. As máscaras de culturas agrícolas proporcionaram melhores resultados na ordem de: modelo independente de *random forest*, MapBiomas e CAR, porém

todos com resultados satisfatórios. As predições de área cultivadas de milho não obtiveram resultados satisfatórios, o qual pode estar relacionado com as informações estatísticas de comparações do modelo e/ou estatísticas da cultura para silagem/grão da cultura. A utilização das simulações do APSIM-NG com diferentes climas, solos, e anos agrícolas proporciona aumento da representatividade de dados com *input* no modelo, o que foi demonstrado com maiores acurácias no modelo, 0.94, ao utilizar a técnica de *data augmentation*, neste caso quando dados simulados e dados de campo foram utilizados de forma agregada no modelo. Por fim, simulações de modelos de desenvolvimento de culturas agrícolas podem ser utilizados em conjunto com sensoriamento remoto com intuito de melhorar classificação e mapeamento de culturas agrícolas para o Noroeste do Rio Grande do Sul, sendo possível a utilização dos métodos para outras regiões, estados e países do mundo.

O mapeamento da rotação das culturas agrícolas no mesmo campo apresenta-se como estratégia de diversificação e intensificação dos sistemas de cultivos agrícolas, bem como a sustentabilidade da agricultura. O modelo de mapeamento de rotação de culturas utilizando *layers* de mapeamento de culturas agrícolas nas diferentes safras agrícolas associados com a delimitação de campo bem como incorporação de dados climáticos e auxiliares para análise da rotação de culturas em diferentes regiões bem como efeito da rotação de culturas na produtividade das culturas acessados com dados de satélite proporcionou resultados panorâmicos da agricultura do Rio Grande do Sul. As mesoregiões Noroeste e apresentaram maior preocupações em relação à grande percentual de área, >60%, de monocultivo de soja em 4 anos de avaliação. Regiões típicas de cultivo de arroz como Região Metropolitana de Porto Alegre, Sudeste e Sudoeste Rio-Grandense vêm perdendo espaço para o cultivo da soja. Ademais, a penalização de produtividade de grãos para a cultura da soja variou de ~23% para região Noroeste, ~20% para todo o estado do Rio Grande do Sul quando monocultivo de soja com três anos, e destaca-se a penalização de ~65% para a região Noroeste com quatro anos de cultivo de soja consecutivo para a safra de 2020-2021. Finalmente, o modelo e análises utilizados neste estudo podem ser utilizados em outros estados e países do mundo promovendo visão sistemática dos sistemas de cultivos na agricultura.

REFERÊNCIAS

ABELLEYRA, D., & VERÓN, S. Crop rotations in the Rolling Pampas: Characterization, spatial pattern and its potential controls. **Remote Sensing Applications: Society and Environment**, v. 18, p. 100320, 2020.

ASTER Mount Gariwang image by the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC) at the USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. 2018. Disponível em: lpdaac.usgs.gov website: https://lpdaac.usgs.gov/resources/data-action/aster-ultimate-2018-winter-olympics-observer/. Acesso em: 31 Jan. 2023.

BASTOS, L. M. et al. Winter wheat yield response to plant density as a function of yield environment and tillering potential: A review and field studies. **Frontiers in Plant Science**, v. 11, 2020.

BATTISTI, R., SENTELHAS, P. C. New agroclimatic approach for soybean sowing dates recommendation: A case study. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 18, n. 11, p. 1149–1156, 2014.

BATTISTI, R.; SENTELHAS, P. C. Drought tolerance of Brazilian soybean cultivars simulated by a simple agrometeorological yield model. **Experimental Agriculture**, v. 51, n. 2, p. 285–298, 2015.

BELGIU, M.; CSILLIK, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis**. Remote Sensing of Environment**, v. 204, p. 509–523, 2018.

BELTRAN-PEÑA, A.; ROSA, L.; D'ODORICO, P. Global food self-sufficiency in the 21st century under sustainable intensification of agriculture. **Environmental Research Letters**, v. 15, n. 9, p. 095004, 2020.

BISONG, E. (2019). **Google Colaboratory**. Building machine learning and deep learning models on Google Cloud Platform, p. 59–64, 2019.

BOOTE, K. J. et al. Simulation of crop growth: CROPGRO model. In: Peart, R.M., Curry, R.B. (Eds.), **Agricultural Systems Modeling and Simulation**, v. 18. Marcel Dekker, New York, pp. 651–692, 1998.

BROWN, C. F. et al. Dynamic World, Near real-time global 10 m land use land cover mapping. **Scientific Data**, 9(1), 2022.

CAI, Y. et al. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. **Remote Sensing of Environment**, v. 210, p. 35–47, 2018.

CHI, M. et al. Big data for remote sensing: challenges and opportunities. **Proceedings of the IEEE**, v. 104, n. 11, p. 2207–2219, 2016.

COHEN, A. A. B. et al. Rotation effects on corn and soybean yield inferred from satellite and field-level data. **Agronomy Journal**, v. 111, n. 6, p. 2940–2948, 2019.

CONAB, Companhia Nacional de Abastecimento. 2020. Calendário de Plantio e Colheita de Grãos no Brasil. Disponível em: Conab.gov.br website: https://www.conab.gov.br/institucional/publicacoes/outras-publicacoes/item/download/36427_9534db174ba2bcddb8bad4be22818839. Acesso em: 24 Jan. 2023.

CORASSA, G.M. et al. Optimum soybean seeding rates by yield environment in southern Brazil. **Agronomy Journal**, v. 110, p. 2430– 2438, 2018.

DADO, W. T. et al. High-resolution soybean yield mapping across the US Midwest using subfield harvester data. **Remote Sensing**, v. 12, n. 21, p. 3471, 2020.

DEINES, J. M. et al. A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. **Remote Sensing of Environment**, v. 253, p. 112174, 2020.

DOORENBOS, J.; KASSAM, A.M. Efeito da água no rendimento das culturas. **Estudos FAO − Irrigação e Drenagem 33**, 306 (Traduzido por Gheyi, H.R. et al. −UFPB), 1994.

FARR, T. G. et al. The Shuttle Radar Topography Mission, **Reviews of Geophysics**, v. 45, 2007.

FOWLER, J.; WALDNER, F.; HOCHMAN, Z. All pixels are useful, but some are more useful: Efficient in situ data collection for crop-type mapping using sequential exploration methods. **International Journal of Applied Earth Observation and Geoinformation**, v. 91, p. 102114, 2020.

FUNK, C., et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. **Scientific Data**, v. 2, n. 1, 2015.

GORELICK, N. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, p. 18–27, 2017.

HAO, P. et al. Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. **Science of The Total Environment**, v. 733, p. 138869, 2020.

HASHEM, I. A. T. et al. The rise of “big data” on cloud computing: Review and open research issues. **Information Systems**, v. 47, p. 98–115, 2015.

HENGL, T., et al. SoilGrids250m: Global gridded soil information based on machine learning. **PLOS ONE**, v.12, n 2, e0169748, 2017.

HOLZWORTH, D., et al. APSIM Next Generation: Overcoming challenges in modernising a farming systems model. **Environmental Modelling & Software**, v. 103, p. 43–51, 2018.

IBGE, Instituto Brasileiro de Geografia e Estatística. 2022. Área plantada, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras temporárias. Disponível em: Ibge.gov.br website: https://sidra.ibge.gov.br/tabela/1612. Acesso em: 31 Jan. 2023.

JIN, Z.; AZZARI, G.; LOBELL, D. B. Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. **Agricultural and Forest Meteorology**, v. 247, p. 207–220, 2017.

JIN, X. et al. A review of data assimilation of remote sensing and crop models. **European Journal of Agronomy**, v. 92, p. 141–152, 2018.

JIN, Z. et al. Smallholder maize area and yield mapping at national scales with Google Earth Engine. **Remote Sensing of Environment**, v. 228, p. 115–128, 2019.

KAMILARIS, A.; KARTAKOULLIS, A.; PRENAFETA-BOLDÚ, F. X. A review on the practice of big data analysis in agriculture. **Computers and Electronics in Agriculture**, v. 143, p. 23–37, 2017.

KARMAS, A.; TZOTSOS, A.; KARANTZALOS, K. Geospatial big data for environmental and agricultural applications. **In: s.l.: Springer International Publishing**, pp. 353–390, 2016.

KASAMPALIS, D. et al. Contribution of remote sensing on crop models: A review. **Journal of Imaging**, v. 4, n. 4, p. 52, 2018.

KASSAM, A.; FRIEDRICH, T.; DERPSCH, R. Global spread of Conservation Agriculture. **International Journal of Environmental Studies**, v. 76, p. 1-23, 2018.

KLUGER, D. M.; WANG, S.; LOBELL, D. B. Two shifts for crop mapping: Leveraging aggregate crop statistics to improve satellite-based maps in new regions. **Remote Sensing of Environment**, v. 262, p. 112488, 2021.

KLUGER, D. M.; OWEN, A. B.; LOBELL, D. B. Combining randomized field experiments with observational satellite data to assess the benefits of crop rotations on yields. **Environmental Research Letters**, v. 17, n. 4, p. 044066, 2022.

KUNISCH, M. Big Data in agriculture–perspectives for a service organization. **Landtechnik**, v. 71, n. 1, p. 1-3, 2016.

LANEY, D. 3D Data management: controlling data volume, velocity and variety. **META Group Research Note**, v. 6, 2001.

LIAKOS, K. et al. Machine learning in agriculture: A review. **Sensors**, v. 18, n. 8, p. 2674, 2018.

LIN, M. & HUANG, Q. Exploring the relationship between agricultural intensification and changes in cropland areas in the US. **Agriculture, Ecosystems & Environment**, v. 274, p. 33–40, 2019.

LIU, Y. et al. Mapping crop rotation by using deeply synergistic optical and SAR time series. **Remote Sensing**, v. 13, n. 20, p. 4160, 2021.

LOBELL, D. B. et al. A scalable satellite-based crop yield mapper. **Remote Sensing of Environment**, v. 164, p. 324–333, 2015.

LOBELL, D. B. et al. Eyes in the sky, boots on the ground: Assessing satellite- and ground-based approaches to crop yield measurement and analysis. **American Journal of Agricultural Economics**, v. 102, n. 1, p. 202–219, 2019.

LOKERS, R. et al. Analysis of big data technologies for use in agro-environmental science. **Environmental Modelling & Software**, 84, 494–504, 2016.

MAPBIOMAS, BRASIL. 2023. Disponível em: Mapbiomas.org website: https://mapbiomas.org/en. Acesso em: 24 Jan. 2023.

MASSRUHÁ, S, S. F. M.; LEITE, A. A. M. M. Agro 4.0 – Rumo à agricultura digital. **JC na Escola Ciência, Tecnologia e Sociedade: Mobilizar o Conhecimento para Alimentar o Brasil**, 2017.

MYKLEVY, M.; DOHERTY, P.; MAKOWER, J. The New Grand Strategy; St. **Martin's Press: New York**, NY, USA, p. 271, 2016.

NENDEL, C. et al. The MONICA model: Testing predictability for crop growth, soil moisture and nitrogen dynamics. **Ecological Modelling**, v. 222, n. 9, p. 1614–1625, 2011.

NIETO, L. et al. An integrated approach of field, weather, and satellite data for monitoring maize phenology. **Scientific Reports**, v. 11, n. 1, p. 15711, 2021.

ORYNBAIKYZY, A. et al. Crop Type Classification Using Fusion of Sentinel-1 and Sentinel-2 Data: Assessing the Impact of Feature Selection, Optical Data Availability, and Parcel Sizes on the Accuracies. **Remote Sensing**, v. 12, n. 17, p. 2779, 2020.

POTT, L. P. et al. Mitigation of soil compaction for boosting crop productivity at varying yield environments in southern Brazil. **European Journal of Soil Science**, v. 71, n. 6, 1157-1172, 2019.

POTT, L. P. et al. Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 176, p. 196–210, 2021.

POTT, L. P. et al. Crop type classification in Southern Brazil: Integrating remote sensing, crop modeling and machine learning. **Computers and Electronics in Agriculture**, v. 201, 107320, 2022.

POTT, L. P. et al. Mapping crop rotation by satellite-based data fusion in Southern Brazil. **Computers and Electronics in Agriculture**, v. 211, 107958, 2023.

R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**. Vienna, Austria, 2023.

RIBAS, G. G. et al. Assessing yield and economic impact of introducing soybean to the lowland rice system in southern Brazil. **Agricultural Systems**, v. 188, p. 103036, 2021.

RODRIGUEZ, D. et al. To mulch or to munch? Big modelling of big data. **Agricultural Systems**, v. 153, p. 32–42, 2017.

SAIZ-RUBIO, V.; ROVIRA-MÁS, F. From smart farming towards agriculture 5.0: A review on crop data management. **Agronomy**, v. 10, n. 2, p. 207, 2020.

SCHWALBERT, R. A. et al. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. **Agricultural and Forest Meteorology**, v. 284, p. 107886, 2020.

SEIFERT, C. A.; ROBERTS, M. J.; LOBELL, D. B. Continuous corn and soybean yield penalties across hundreds of thousands of fields. **Agronomy Journal**, v.109, n. 2, p. 541–548, 2017.

SICAR, Sistema Nacional de Cadastro Ambiental Rural. 2023. Disponível em: Car.gov.br website: https://www.car.gov.br/publico/imoveis/index. Acesso em: 24 Jan. 2023.

SIMÃO, L. M., et al. Crop rotation and tillage impact yield performance of soybean, sorghum, and wheat. **Agronomy Journal**, 2023.

SINCLAIR, T. R. et al. Assessment across the United States of the benefits of altered soybean drought traits. **Agronomy Journal**, v. 102, n. 2, p. 475–482, 2010.

SPARKS, A. nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. **Journal of Open Source Software**, v. 3, n. 30, p. 1035, 2018.

STEDUTO, P. et al. AquaCrop-The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. **Agronomy Journal**, v. 101, n. 3, p. 426–437, 2009.

STEIN, S., & STEINMANN, H.-H. Identifying crop rotation practice by the typification of crop sequence patterns for arable farming systems – A case study from Central Europe. **European Journal of Agronomy**, 92, 30–40, 2018.

TZOUNIS, A. et al. Internet of Things in agriculture, recent advances and future challenges. **Biosystems Engineering**, 164, 31–48, 2017.

WALDHOFF, G.; LUSSEM, U.; BARETH, G. 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**, v. 61, p. 55–69, 2017.

WANG, S. et al. Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. **Scientific Data**, v.7, n.1, 2020.

WANG, S.; AZZARI, G.; LOBELL, D. B. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. **Remote Sensing of Environment**, v. 222, p. 303–317, 2019.

WEBER, R.H.; WEBER, R. Internet of Things. **Springer, New York.** NY, 2010.

WEISS, M.; JACOB, F.; DUVEILLER, G. Remote sensing for agricultural applications: A meta-review. **Remote Sensing of Environment**, v. 236, p. 111402, 2020.

WHITE, J. W. Methodologies for simulating impacts of climate change on crop production. **Field Crops Research**, v. 124, n. 3, p. 357–368, 2011.

WOODARD, J. D. et al. The power of agricultural data. **Science**, v. 362, n. 6413, p. 410–411, 2018.

ZHANG, Y. The Role of Precision Agriculture. Resource, 2019.

ZHAO, C. et al. A SIMPLE crop model. **European Journal of Agronomy**, v. 104, p. 97–106, 2019.